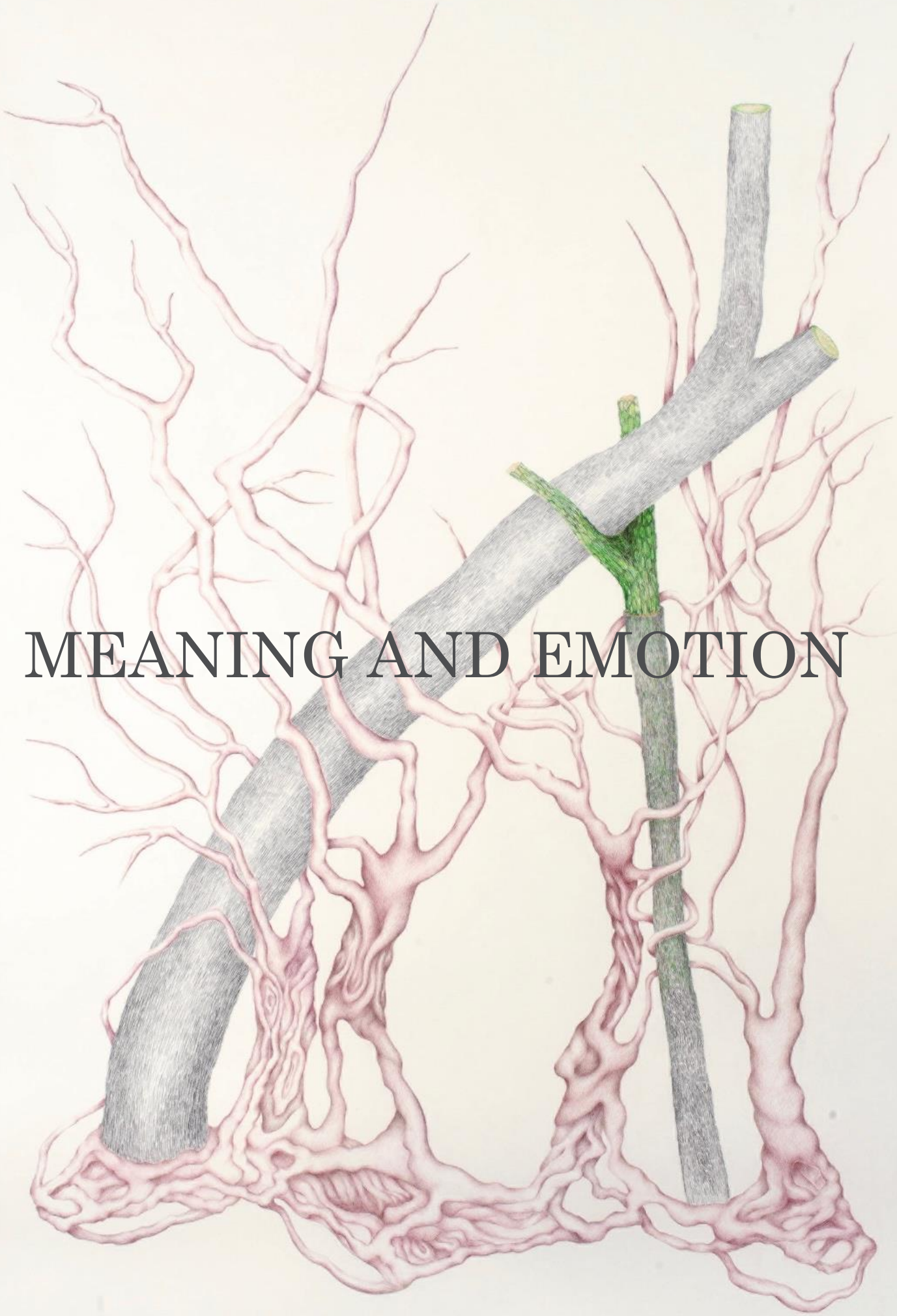

MEANING AND EMOTION

Constant Bonard

Meaning and Emotion

Constant Bonard

A dissertation jointly submitted to the Department of Philosophy, University of Geneva and to the Department of Philosophy, University of Antwerp in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Philosophy.

Supervised by Julien Deonna (University of Geneva) and Bence Nanay (University of Antwerp).

Jury members: Dorit Bar-On (University of Connecticut), Elisabeth Camp (Rutgers University), Manuel García-Carpintero (University of Barcelona), Fabrice Teroni (president of the jury, University of Geneva).

Cover picture: 'Enchevêtrement' by Pierre Bonard and Sylvie Mermoud.

A Joe

TABLE OF CONTENT

0. GENERAL INTRODUCTION

> « Thus we have a sort of canonical pattern that some creature X nonvoluntarily produces a certain piece of behavior α, the production of which means, or has the consequence, or evidences, that X is in pain. That is the initial natural case. »
> – Paul Grice, 'Meaning Revisited'

After a preamble (§0.1), I give an overview of the dissertation (§0.2) and express my acknowledgments (§0.3).

0.1. PREAMBLE

Most people I have met in the last number of years have asked what my dissertation was about. Despite the repeated practice, I am afraid I have never found the optimal answer. If I say, in an effort to be jargon-free, that it is about the communication of emotions, the next question often is whether I work on a specific emotion or a specific case study. Nope. 'On a particular communicative medium then?' Neither. 'Oh ok. But then, what is it really about?'. I have tried explaining that I work on models of communication which were developed to account for the transmission of neutral, affect-less, information and that my goal is to modify them so that they apply satisfyingly to affect-loaded information transmission. But this did not speak to many of my interlocutors. In the end, the most efficient trick that I found was to begin with what my PhD *used* to be about. And so this is how I will introduce my subject here as well.

What used to be the central question of my PhD is the following: how is it possible that we can communicate our emotions through music? After all, music is made of abstract patterns of sounds. We cannot explain this by appealing to an agreed-upon lexicon, like with languages and road signs. How then can a melody be meaningful and express sadness or joy?

The answer I developed was that musical communication employs the same mechanisms that are used to communicate emotions in more common, less mysterious cases. Roughly, musical expression piggybacks on both nonverbal and verbal expressive capacities. We can understand music in part because of how we understand that someone is angry from her abrupt, unpredictable gestures or that she is sad from the sound of her voice, because music mimics these nonmusical expressive features. Sad melodies, like sad voices, tend to be slow, soft, unsteady, with descending pitch contours. We also understand music in part because of linguistic-like

abilities, for instance the pragmatics-like ability to better grasp what musicians want to express in light of their personal life, just like we may better get what is expressed in a sentence in light of the speaker's personal life (e.g. an expression of genuine admiration vs. mockery).[1]

So, to explain our intriguing capacity to communicate emotions through music, just look at how we usually communicate emotions in garden-variety cases and see if you can apply the explanations from these better known, less mysterious situations.

To flesh out this answer, I dug into the literature on verbal and nonverbal emotional expression both within and outside philosophy. During this process, I came to realize that the less mysterious, garden-variety cases of emotional expressions actually were ill-understood and insufficiently elucidated. They themselves asked for more satisfying explanations before they could be used as explanans, before they could really illuminate what happens in musical communication.

Take, for instance, laughter. How do we know when laughter indicates embarrassment or mirth? Well, good luck finding a solid answer in the existing literature! To the best of my knowledge, there are no good theories to help us understand this apparently trivial phenomenon. Similarly, I found that pragmatics – both in the Gricean tradition and in speech act theory – failed to give satisfying accounts of affect-loaded language, especially in light of what I was learning about emotions at the Swiss Center for Affective Sciences, where I have done most of my PhD.

So the plan to explain musical expression by using explanations from garden-variety cases ended up not being entirely satisfying – at least until the theories developed for garden-variety cases were improved. This is one of the reasons that led me to the actual subject of my dissertation – or rather to its multiple subjects. At some point, the plan was to have a first part on nonmusical emotional expression, a part where I would do my best to improve the existing theories, and then a second part on music when I would apply the updated theories of the first part. But as the first part

[1] Another important aspect of the answer I developed, but one which bears no links with the present work, is that we also understand music in part because we have learned to parse sounds through phonology-like, morphology-like, and syntax-like structures. This hypothesis, by the way, would explain why many musical features are universal among human cultures and why nonhuman animals don't share our musical capacities: they lack the relevant linguistic capacities. This would also explain why we sometimes fail to properly understand foreign music: because we haven't got used to parsing such music as we did for the music we are familiar with. Compare this with how we fail to parse words correctly in languages we don't know well (irrespective of our knowledge of the lexicon). For a cross-cultural study on this hypothesis, see Bonard (2018).

increased in size, I let go of the idea to work on music at all in this dissertation.[2]

I ended up setting aside what used to be my theme of predilection. I hope to come back to it in the future and I am sure that the work achieved in this dissertation will allow me to answer it much better than if I had not delved into the labyrinth of nonmusical emotional expression in the first instance.

Let me now turn to the actual subjects of my dissertation, whose unity hopefully will be made apparent against the background presented in this preamble.

0.2. Overview of the dissertation

This dissertation may be divided into two parts. The first is about the Extended Gricean Model of information transmission, a new model which I here introduce. The second is about what emotional signs mean, in various senses of the term 'mean'.

Part 1. The Extended Gricean Model

Part 1 is constituted of four chapters: the first one sets a problem that needs to be solved, the second one presents a solution – the Extended Gricean model – while the third and the fourth are applications of the model.

In Chapter 1 – 'A Blind Spot in the Standard Picture of Information Transmission' – I ask the question I mentioned above: How do we know when laughter indicates embarrassment or mirth? I explain why there are no satisfying answers to be found in the relevant literature. My diagnostic is that the standard picture of information transmission presupposes that there are two ways in which we may communicate or otherwise exchange information, each respectively being accountable for by the code models and the existing Gricean models. However, neither of them adequately applies to the cases I present. The meaning of laughter resists both kinds of models. It resists the code model because what is transmitted by a laugh often goes beyond what is encoded in it. We usually understand more from a laugh than what could be predicted based merely on a code, and by 'code' I mean a pre-established pairing between kinds of laughter and what information they carry. This is because the same sounds, the same

[2] I have nevertheless written two papers about how language sciences may help us understand musical expression (Bonard, 2018, in preparation).

laughter, may mean that the person is embarrassed, mirthful, afraid, joyful, and many other things. The same conclusion applies to many other emotional expressions: a smile may mean happiness, compassion, and aggressiveness; a frown may indicate anger, incomprehension, and concentration; a sigh may signal relief, fatigue, and disappointment; etc.

The cases I present also resist the prevailing Gricean models because the latter only applies to so-called speaker-meaning, i.e. what sign-producers intend their signs to mean. The problem is that we often laugh spontaneously, without intending the laughter to mean what it nevertheless means.

In Chapter 2 – which is the central chapter of Part 1 and, in fact, of the entire dissertation – I present the Extended Gricean Model of information transmission. This model is supposed to apply to cases, such as the case of laughter from Chapter 1, that can be accounted for by neither the prevailing Gricean models nor the code models of information transfer. This model preserves much from its antecedents, the prevailing Gricean models, but contrary to them it is not restricted to what people intend to mean with the signs they produce. Instead, it extends to what they *allow* the signs they produce to mean. The central notion is not anymore that of speaker-meaning, but that of *allower-meaning*.

While Chapter 2 presents the Extended Gricean Model quite abstractly, Chapter 3 is dedicated to illustrating the model, thereby exploring its breadth as well as its boundaries. It begins with the examples of laughter presented in Chapter 1, showing how it can explain what information is carried by such stimuli, and then discusses other kinds of stimuli: nonverbal affective signs, some behavioral signs, clothing, but also what one allows one's speech to mean beyond what one intends it to mean.

In Chapter 4, I show how the Extended Gricean Model is an interesting tool to interpret the meaning of narrative artworks. The central idea here is that the meaning of a novel or a movie may be found in what the authors *allow* their work to mean even though it is not (and we know it is not) what they intended it to mean.

The four chapters of Part 1 thus constitute a presentation of the need, the nature, and the use of the Extended Gricean Model and its central concept: allower-meaning. This kind of meaning corresponds to a non-negligible portion of the information transmitted in everyday life but for which, to the best of my knowledge, there was no theory – at least in analytic philosophy and in linguistics.

PART 2. WHAT EMOTIONAL SIGNS MEAN

In Part 2, I turn to existing theories of meaning and see how they apply to emotional signs, i.e. signs which give us information about the affective state of the sign producer.

In Chapter 5, I discuss how to distinguish expressives – utterances whose (illocutionary) goal is to express affect – from descriptives – utterances whose goal is to describe the world truthfully. Expressives include, for instance, insults, encouragements, and interjections (ouch, wow, yuk, etc.) while descriptives include assertions, conjectures, or suppositions. I spell out three features that importantly distinguish these types of utterances. Drawing on recent insights from the philosophy of emotion and value, I then show how the three features derive from the nature of emotions, understood as felt, bodily, value-tracking attitudes. I also indicate how speech act theory helps us clarify this claim.

Chapter 6 discusses three possible accounts of what understanding expressives amount to. The first account, doxasticism, claims that the audience must only take the utterer to be in a certain doxastic state (to believe, judge, suppose, doubt, …). The second view, moderate affectivism, claims that the audience must believe that the utterer undergoes, or is disposed to undergo, emotions. The third view, radical affectivism, claims that it is not sufficient that the audience *believes* that the utterer expresses an emotion, the audience must *resonate affectively* with the expresser in order to properly understand the expressive utterance. I discuss some advantages and disadvantages of these three views, arguing that moderate and, especially, radical affectivism are in a better position to explain the distinctive features of expressives discussed in Chapter 5.

In Chapter 7, I turn to how affect may be 'naturally' encoded in stimuli, i.e. without the stimuli being intentionally designed to convey affective states. For instance, how can we explain that red cheeks can mean embarrassment or that vervet monkey alarm calls can indicate fear of a predator? I discuss two main accounts proposed in the literature: *natural meaning* and *probabilistic meaning*. I evaluate how useful they are when it comes to analyzing what emotional signs mean. I argue that natural meaning is too strict for this purpose. The notion of probabilistic meaning seems adequate to analyzing non-communicative emotional signs (e.g. pupil dilatation, perspiration, blushing), but it faces several difficulties when it comes to analyzing communicative signs (e.g. vocal, facial, or gestural emotional expressions).

In Chapter 8, to fill the gap left by the notion of probabilistic meaning, I present and develop the notion of *teleocoded meaning*, which is largely based on previous so-called teleosemantic theories.[3] The idea is that certain signals *encode* certain information – i.e. these stimuli are somehow associated with certain information by communicators, as explained in Chapter 1 – and that this encoding is best explained through an evolutionary process, as opposed to an intentional design. In other words, it is the evolutionary function (hence 'teleo') of these signals to encode certain information (hence teleo*coded* meaning). I argue that this notion can overcome the difficulties that we saw probabilistic meaning was facing in the last chapter while preserving its advantages over natural meaning.

In the final chapter, Chapter 9, I turn to what emotions mean in and of themselves. I ask whether emotions are supposed to indicate something to the organism having them about the situation in which the organism is. I argue that they do: one of the functions of emotions is to give us information about evaluative properties, i.e. what is good or bad for us. More specifically, I argue that, if we accept widespread views of emotions, representation, evaluative properties, and consciousness, then emotions involve a component – the appraisal process – that represents evaluative properties unconsciously. From this conclusion, we may further infer that emotions represent evaluative properties *tout court*. This chapter also serves as a reference for many undefended claims I make about emotions in the other chapters. It captures much of what I have learned about emotions during my time at the Swiss Center for Affective Sciences.

By the end of the dissertation, to the best of my knowledge, I will have discussed and explored all the philosophical accounts of meaning that are relevant to answer the question 'What do emotional signs mean?'. In fact, trying to answer this question will even have led me to define a new kind of meaning: allower-meaning.

I have added as an Appendix a (long) discussion of four different definitions of speaker-meaning, i.e. different ways in which the locution 'S means p by X' may be captured. These are Grice's (1968), Neale's (1992), Green's (2007, Chapter 3), and one based on Sperber and Wilson's (1986) definition of ostensive-inferential communication. In the conclusion of this Appendix, I

[3] Unlike the most famous teleosemantic notions (Dretske, 1986, 1988; Millikan, 1984; Papineau, 1984) it is, as Sterelny (1990, sec. 6.6) puts it, a modest account. It is modest insofar as its scope is not supposed to include Gricean meanings. To come back to the distinction of Chapter 1, it is restricted to what is *encoded* in signals. As such, it is akin to existing 'modest teleosemantic' proposals such as Green's *organic meaning* (2019) or Shea, Godfrey-Smith, and Cao's *functional content* (2018), but I show how they nevertheless differ.

offer my favored definition by synthetizing what we have learned from the discussion.

0.3. ACKNOWLEDGEMENTS

I once read that it takes a village to write a book. This is certainly true of my dissertation. Here, I want to thank some important villagers. I present my sincere apologies to the inhabitants that I may have forgotten.

First of all, my thanks go to my supervisors, Julien Deonna and Bence Nanay, as well as to Fabrice Teroni who supported me as if he were a third supervisor.

Julien had already supervised my master's thesis. After that, I was not sure about doing a PhD. I wanted to become a music journalist. I had just been hired by the Paléo Festival to do a one-year internship in their press department when Julien called me. He proposed that I apply to become his and Fabrice's teaching assistant. This meant doing a PhD. I was very pleasantly surprised by this proposal. I didn't hesitate long; I cancelled my internship, and that's how it all started. Since then, Julien hasn't stopped providing all that a PhD student could ask: lots of liberty but with continuous support – this combination is as great as it is rare – countless careful, attentive, considerate discussions on any kind of philosophical or non-philosophical questions I would have, timely reminders to achieve this or that task without ever unduly worrying me about anything, thoughtful telephone calls, especially during the end of the process, to check on how things were going, and a plethora of other wonderfully friendly gestures. I feel very lucky to have had Julien as a supervisor for all these years.

I met Bence later in my dissertation process, when I visited him and his group in Antwerp for six months in 2018. The discussions I had with him during this period helped me immensely to have a different, more global point of view on many issues that I was working on. His clear vision of what is best for a PhD student also has redefined my goals and the way I now think about my work in extremely beneficial ways. But the strongest contribution that he made to this dissertation was the comments he rapidly made every time I sent him a chapter. Not only were these comments numerous, detailed, and full of great references, but they sometimes deeply challenged my way of thinking and, I hope, greatly improved the present work. Some of them would certainly have merited deeper changes in the dissertation and I'm sorry not to have been able to take all of them into account.

Fabrice had no obligation whatsoever to provide guidance for my dissertation, but he gave me lots of help anyway. He read and commented on tons of drafty pages and his perspicacity was immensely helpful. He also contributed to shaping my way of thinking on numerous subjects in the periphery of my dissertation, for instance while I was helping him supervise exams on John Locke or Thomas Reid and on how these classic thinkers relate to contemporary philosophy of mind.

Julien, Bence, and Fabrice should also be thanked for the environment they provided, especially the Thumos group in Geneva and the Center for Philosophical Psychology in Antwerp. During my 5 years at Thumos, I feel privileged to have met numerous amazing philosophers working on emotions who have changed my way of thinking about affects and beyond. In Antwerp, although my time there was briefer, I received very precious feedback at a critical time, when I was developing my Extended Gricean Model and had great discussions will some people who have become close friends. Thank you very much to all of you Thumosians and Philosophical Psychologists! In Geneva, thanks also go to non-Thumosians for very helpful discussions: Jacques Moeschler, Benjamin Neeser, Anna Piata, Pierre Saint-Germier, and David Sander.

I am also very grateful to the people who accepted to read and discuss some of my work and for their very valuable comments: Mitch Green, Louis de Saussure, Tim Wharton, and Deirdre Wilson. Thanks also go to all the people who gave me feedback at the talks I gave during my PhD (if I count correctly, this amounts to 51 talks, so too numerous to list them here).

Huge thanks go to the people who so kindly accepted to proofread a chapter of this dissertation: Mary Carman, Richard Dub, Roberto Keller, Chris McCarroll, Tris Oliver-Skuse, Edgar Phillips, Laura Silva, Gerardo Viera, and Nick Wiltsher. Some of them not only kindly corrected my English as discussed but additionally gave precious philosophical comments. Thanks a lot also to Pauline Perret who helped me with the cover, the layout, and the graphics. The great generosity you all offered – des quatre coins du monde – has really touched me!

Special thanks also go to Greg Currie who supervised my work for 6 months while I was a visiting graduate student at the University of York. The subject on which I was working at the time did not end up in the dissertation either (yes, it was a third topic[4]), but the half dozen meetings

[4] The question was: 'What is so special about human art that makes it universal among human cultures, but inaccessible to other species?'. It was an extension of the music topic discussed above and the answer I was exploring was somewhat similar: what is really

we had and the challenging questions he asked made me realize that I needed to change the topic of my dissertation. This certainly was of great help, although it does not show in the present work.

I also want to thank the jury members of my PhD defense for having accepted to play this time-consuming role: Dorit Bar-On, Elisabeth Camp, Manuel Garcìa-Carpintero, and Fabrice Teroni.

During the writing of this dissertation, I benefited from the financial support of the République de Genève, to which I owe my position (*assistant-doctorant*), of the Swiss National Science Foundation, and to which I gratefully acknowledge a 13-month grant spent in York and Antwerp, and in addition the Swiss Center for Affective Sciences and the Department of Philosophy at UNIGE for travel grants. Many thanks to all these institutions and to the Swiss taxpayers who financed my work. Although this dissertation won't directly profit my co-citizens, I hope that the process of writing it constituted a formation which will allow me to indirectly give back at least some of the benefit I gratefully enjoyed during these PhD years.

Finally, I want to express my gratitude to my family and friends. My parents in particular have always been extremely supportive in whatever enterprise I have undertaken and I couldn't have wished for a better background for my studies. Furthermore, their concern for aesthetics – see the cover picture – actually is the source of my first philosophical interest. Lastly, to my friends, I will say this: whatever your contribution to this dissertation is, thanks to the conversations, the dreams, the laughter, the music, the support, and all the other ingredients necessary to the great friendship we have built. I owe you what I cherish most in my life. This is why I dedicate this work to Joe the Sailor.

Lausanne
August 2020

special is to be found in our mindreading, pragmatic capacities. Art was thought of as the ostensive display of value-loaded perceptual qualities, and by 'ostensive' I mean what pragmaticists such as Sperber and Wilson mean. What I realized thanks to Greg was that defining art properly is so hard that my project was somewhat doomed. But this experience fortunately translated into an encyclopedia article on the definitions of art (Bonard & Humbert-Droz, 2020).

PART 1 – THE EXTENDED GRICEAN MODEL

1. A Blind Spot in the Standard Picture of Information Transmission

« 'Be our sister,' everyone's smiles seemed to be asking. »
– Robert Walser, *The Tanners*

Abstract. Within philosophy and linguistics (among other fields), it is usual to distinguish between two broad kinds of meaning which we can call 'non-Gricean' and 'Gricean'. They are respectively accounted for by two kinds of models of information transmission: the code models and the Gricean models (in which I also include post- and neo-Gricean models) (§1.1). The code models (§1.2) explain information transmission based on pre-established pairings between information and the stimuli that carry it. They successfully apply to diverse kinds of information transmission, from bacteria signaling to the semantics of natural languages. Grice and his heirs have nevertheless insisted on the insufficiencies of the code models concerning many aspects of human communication (§1.3). Gricean models were designed to account for such cases and in particular cases where linguistic codes are unable to account for what is meant by speakers. To do this, the prevailing Gricean models postulate that the linguistic encoding–decoding process must be supplemented by a process involving the overt display of communicative intentions and their mindreading, based on pragmatic principles (this is also called ostensive-inferential communication) (§1.4).

This is roughly what I call the standard picture of information transmission: information may be accounted for by either the code models or the prevailing Gricean models. In this chapter, I argue that this picture fails to account for certain cases, and thus has a blind spot (§1.5). These are cases where information is transmitted through stimuli that are not overtly intended for communication – and so cannot be accounted for by the prevailing Gricean models – but where the relevant codes are insufficient to account for all the information transmitted – and so where the code models also fail. I focus on cases of laughter, but the conclusion applies more widely, as we will see especially in Chapters 2 and 3.

1.1. INTRODUCTION: THE STANDARD PICTURE OF INFORMATION TRANSMISSION

According to what I will call 'the standard picture of information transmission' – which I take to be the generally received view in the philosophy of language and of communication, in linguistics, and beyond – we may distinguish two broad kinds of processes through which information may be transmitted. The two processes may be accounted for by two broad kinds of models of information transmission: the code models and the Gricean models.[5] Accordingly, they correspond to two broad kinds of meaning, which we may call 'Gricean' and 'non-Gricean meaning'.[6] I will detail what these models are in §§1.2–1.4, but let me briefly introduce them.

The code models account for information transmission through a simple coding–decoding process, based on a pre-established pairing between information and the stimuli that carry it. Code models have been successfully developed to account for information transmission through conventional and non-conventional meaning. Concerning conventional meaning, a notable example is formal semantics (for a textbook account, see Heim & Kratzer, 1998), where information transmission is explained by a conventional pairing between messages and lexical entries together with compositional rules. Concerning non-conventional meaning, codes between information and stimuli have been proposed based on strict correlations (Dretske, 1981; Grice, 1957), probabilistic correlations (Hauser, 1996; Millikan, 2004; Scarantino & Piccinini, 2010; Shannon, 1948; Shea, 2007; Skyrms, 2010; Stegmann, 2015), or biological functions (Dretske, 1986; Godfrey-Smith, 1991; Green, 2019b; Millikan, 1984; Papineau, 1984).

[5] Note that I use the term 'Gricean' very broadly to include not only Grice's and neo-Gricean theories, but also post-Gricean theories (e.g. D. Blakemore, 1987; Borg, 2004; Carston, 2002; Recanati, 2004; Sperber & Wilson, 1986, 2015; Wharton, 2009). So what I call the 'Gricean model' corresponds to what has been called the 'inferential model' (Sperber and Wilson, 1986, Chapter 1). Neither of these expressions is entirely happy because post-Griceans usually refuse the label 'Gricean' and because the code models can function through inferential mechanisms. I have nevertheless chosen the former because post-Gricean models undeniably are continuations of Grice's work and so, in some very broad sense, are Gricean.

[6] This picture and the distinction at its core is largely based on Grice's (1957) distinction between natural and non-natural meaning, as well as on the numerous refinements of this distinction (see below as well as my Chapters 7, 8 and Appendix). Predecessors of Grice who made a similar distinction include Anton Marty, Victoria Welby, and, arguably, medieval thinkers such as Roger Bacon.

Gricean models were developed to account for more complex processes through which information may be transmitted, which are not exhausted by a pre-established ('mechanical') pairing between information and stimuli, but which require postulating pragmatic competences, mindreading processes, and pragmatic principles (see §1.4 below). The scope of all the prevailing Gricean models – i.e. those of the current standard picture, which also include neo- and post-Gricean models (see footnote 5) – is restricted to what Grice (1957) called *non-natural meaning* and which is now usually called *speaker-meaning* (see my Appendix for different definitions of this notion).[7] Speaker-meaning requires the production of signals that are overtly intended for communication. The scope of the prevailing Gricean models is thus restricted to such signals.

In this chapter, I will challenge an assumption of the standard picture (see §1.5): that the prevailing Gricean models can account for all the information transfers that cannot be accounted for by the code models. In other words, the standard picture has a blind spot. I will show this by focusing on cases of laughter that are not overtly intended for communication – and which thus fall outside the scope of the prevailing Gricean models – but which nevertheless carry information that the code models cannot account for. Even though I will focus on laughter here, cases which fall outside the scope of both the code and the prevailing Gricean models are diverse and widespread (see especially Chapter 3).

Let me by the way note that I am not the first to suggest that the code and the prevailing Gricean models fail to account for all cases of information transmission (see e.g., Dorit Bar-On, 2013; Schlenker et al., 2016). However, the blind spot that I will point to – roughly, implicatures made by stimuli that are not intended for communication – has not, as far as I know, been highlighted as such. Schlenker et al. (2016), for instance, talk about pragmatics without mindreading, while the cases I present require mindreading abilities. Bar-On (2013) is interested in cases where the receivers 'do not *rationally infer* what they are supposed to be informed about' (2013, p. 361), while the cases I present require some kind of rational inferences – at least this is what I will argue. Relatedly, the conclusions to which my examples will lead – roughly, that we need to extend the Gricean model of communication – has not been pursued either. Bar-On and Schlenker et al., for instance, don't propose to 'go Gricean', unlike I. Finally,

[7] Sperber and Wilson (1986) and other relevance theorists do not talk about speaker-meaning, but instead focus on *ostensive-inferential communication,* a notion that nevertheless is defined in a very similar way (see Chapter 2 and the Appendix). I will here ignore the difference between speaker-meaning and what is successfully ostensively-inferentially communicated.

the cases I present cannot, as far as I know, be dealt with the non-code-based but non-Gricean proposals that I know of.[8]

The assumption that I will challenge in this chapter is related to another assumption of the standard picture according to which all Gricean models must deal exclusively with overtly intentional signals. I will reject this second assumption in the next chapter, where I will present a new Gricean model: the Extended Gricean model. The latter is meant to deal with the blind spot revealed in the present chapter. The scope of the Extended Gricean model is not restricted to speaker-meaning, contrary to the prevailing Gricean models. It extends to what I will call allower-meaning.

Fig. 1.1 is an illustration of the typology of meaning with which we will be working. Code-based meaning is the kind of meaning that can be accounted for by the code models. It includes, for instance, semantic meaning as well as types of meaning which need not be based on conventions and that we will be analyzing in Chapters 7 and 8 (natural meaning, probabilistic meaning, and teleocoded meaning). Gricean models apply to Gricean meaning. The prevailing Gricean models are restricted to speaker-meaning (or cognate notions such as ostensive-inferential communication, see the Appendix). I will argue that the standard picture has a blind spot insofar as it cannot account for certain implicatures*[9] that are not speaker-meant. I plan to fill this gap with my Extended Gricean model and the notion of allower-meaning.

[8] Schlenker et al.'s Informativity Principle and Urgency Principles (Schlenker et al., 2016) cannot explain the cases that I will present because they don't yield the relevant implicatures. Schlenker's 'presupposition algorithm' (Schlenker, manuscript) cannot either since this algorithm is based on what the context probabilistically means, and so, as we will see, belongs to the code model, which cannot predict the relevant implicatures.
[9] I explain below why I put an asterisk.

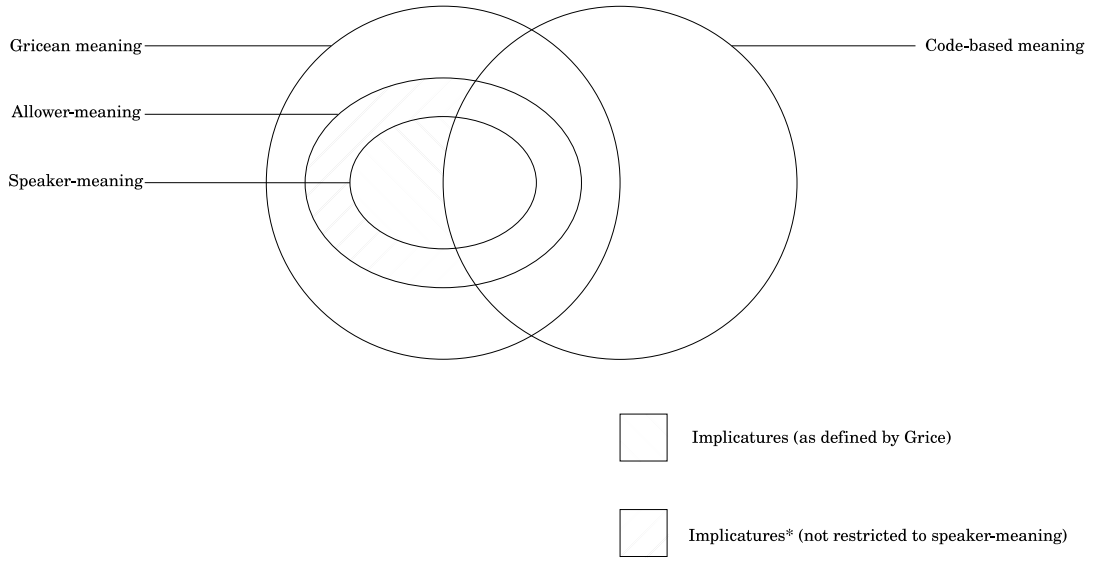

Fig. 1.1. Our typology of meaning.

We may summarize the general argument that I aim to defend in this and the next chapter as follows:

(A) Two assumptions of the standard picture are that (A1) if the information is transmitted through a stimulus that is not produced with overt communicative intentions, a code model can account for it, and (A2) if the information transmitted must be accounted for by a Gricean model, it must be communicated through a stimulus produced with overt communicative intentions (an overtly intentional signal).

(B) Contrary to (A1), information which cannot be accounted for by a code model can be transmitted through stimuli that are not produced with overt communicative intentions (this chapter) and, contrary to (A2), these cases must be accounted for by a Gricean model (the Extended Gricean model, see next chapter).

(C) Therefore, we should revise these assumptions of the standard picture to allow the explanatory scope of the Gricean model to extend beyond overtly intentional signals.

Putting aside this general argument, the claim defended here, that neither the code models nor the existing Gricean models can account for certain information transfers, has, if correct, some noteworthy consequences. One of them concerns the scope of pragmatics. The code and Gricean models used in the philosophy of language and linguistics are usually thought to

correspond respectively to semantics and pragmatics (Korta & Perry, 2020, sec. 3). The 'border disputes' between semantic and pragmatic meaning can be formulated by the relative scope of these models:

> « Contemporary philosophical approaches to pragmatics are often classified by their view of the two models [i.e. the code and Gricean models]. 'Literalists' think that semantics [i.e. codes model] is basically autonomous, with little 'pragmatic intrusion'; 'contextualists' adopt the basic outlines of the Relevance Theory view of the importance of pragmatics [i.e. Gricean models] at every level » (Korta & Perry, 2020, sec. 3.4)

And here is how Schlenker puts it:

> « The informational content conveyed by utterances has two sources: meaning as it is *encoded* in words and rules of semantic composition (often called 'literal' or '*semantic meaning*'); and further inferences that may be obtained by reasoning on the speaker's motives (the conjunction of these inferences with the literal meaning is often called the 'strengthened' or *'pragmatic meaning'* of the sentence). » (Schlenker, 2016, my italics)

If the argument presented here is correct, one consequence is that the scope of pragmatics is broader than usually thought. We may even talk of 'super pragmatics' echoing what Schlenker (2018) calls 'super semantics'.

Consequences may extend beyond philosophy and linguistics as the standard picture and the distinction between the code and Gricean models is also prevalent in the study of language evolution (Dorit Bar-On, 2017; Moore, 2018; Reboul, 2017; Scott-Phillips, 2015; Sterelny, 2017; Tomasello, 2008, Chapter 5), developmental psychology (Csibra, 2010; Csibra & Gergely, 2009; Gergely & Király, 2019; Tomasello, 2008, Chapter 4), or primatology (Sievers & Gruber, 2016; Tomasello, 2008, Chapter 2; Townsend et al., 2017). In all these fields, it is often assumed that if information cannot be accounted for by a code model, it would require postulating speaker-meaning.

The conclusion of this chapter can also be significant for the affective sciences, where code models seem to be the only ones used to analyze emotional expression and recognition (usually the models of Shannon (1948) or Brunswik (1956), but see also Owren and Bachorowski (2003)). The argument defended here shows that there is a need for a new model in this domain.

Let us now see in more detail what the code models are.

1.2. THE CODE MODELS: A GOOD OLD ACCOUNT OF INFORMATION TRANSMISSION

The code models include a wide range of theories of meaning from that given in Aristotle's *De Interpretatione* to formal semantics and biological signaling theory (Dawkins & Krebs, 1978; Skyrms, 2010; Zahavi, 1975), Shannon's mathematical model of communication (1948), or Saussure's semiology (1916). The code models analyze the transfer of information in terms of a system consisting of a sender (or source), that encodes information (or a message) into a stimulus (or a signal), which travels to a receiver who gets the information by decoding the stimulus.[10]

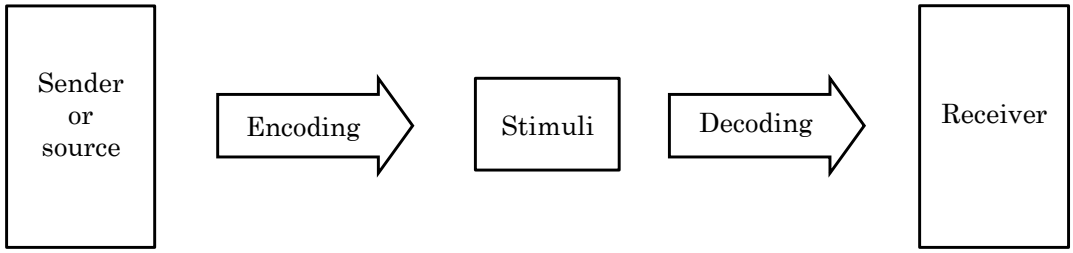

Fig. 1.2. A typical representation of the code models

Encoding is the process where the information is converted by the sender or the source (using the sender's coding rules) into the cues which will constitute the stimuli detected by the receiver. *Decoding* is the reverse process, where the sender converts the stimuli back into the information (using the receiver's decoding rules). For the encoding and decoding process to work and the information to be transferred, they must be based on the same *code* (the sender's and receiver's rules must somehow correspond to each other). The code is a set of pre-established pairings between stimuli and pieces of information (which may be expressed as a set of sender's and receiver's rules in a sender-receiver signaling game, see Chapter 8).

[10] To be more precise, the sender or the source produces *distal cues* (relative to the receiver) which travel through a channel to the receiver, who decodes *proximal cues* or *stimuli*, a cue being an entity conveying information or misinformation (Green, 2007, Chapter 5; Scarantino, 2015). If the channel is 'noisy', the receiver will only get some of the cues produced by the sender, and they might be distorted. Here I will ignore the noisiness of channels and will assume that the distal cues and the proximal cues (or stimuli) are identical. For this reason, and because the term 'cue' is defined in contradictory manners by different relevant authors (Compare (Hauser, 1996; Scott-Phillips, 2008) and (Green, 2007; Juslin & Laukka, 2003; Scherer, 2003)), I will only use the term 'stimulus'.

Code models are usually used to explain communicative phenomena[11] – the term 'code' comes, I believe, from Shannon (1948)'s model which was developed for telecommunication through machines such as telegraphs or telephones. It is indeed easier to imagine how it applies to a sender and a receiver who are communicating. However, they may also elucidate non-communicative information transmission. The term 'code' in such cases will be used in a very abstract way, far from its everyday use. As an example, we can think of an animal hunting which has detected the presence of a prey. This, of course, is not communication, but we may say that the prey has somehow sent certain information to the predator by having encoded certain information – say, 'A deer has walked by here recently' – into stimuli – say, such snow tracks. The code in this case is a pairing between properties of the snow and the relevant information. The stimuli are detected by the predator and the latter will be able to decode the relevant pieces of information as long as it masters the appropriate code. It will indeed need to possess the capacities allowing it to extract the relevant pieces of information from the detection of the snow tracks. In other words, the predator will need to master the code whereby the relevant information was encoded into these stimuli.[12] We will come back to similar cases in Chapters 7 and 8.

In this chapter, I will mostly concentrate on communicative phenomena, i.e. cases where both the sending and the receiving were designed for the transfer of the transmitted information. Let us keep in mind however that code models may also be applied to cases where information is transferred without being communicated.

Communicative codes can be widely different. They can be quite simple, like the vervet monkey alarm call system, or extremely complex, like the codes underlying human languages. Indeed, languages' lexicon and grammar (in the sense of Chomsky (1957)) can be seen as a code consisting of a pairing between stimuli (phonemes for spoken languages) and messages (semantic representations). This code, being combinatorial, allows the pairing of an infinite number of different messages with a finite number of stimuli (something like 40+ phonemes in English). An important

[11] I use 'communication' in a broad sense as the transfer of information between a sender and a receiver whereby both the sending and the receiving were designed to transfer this information (Green, 2007; Hauser, 1996; Maynard Smith & Harper, 2003; Scott-Phillips, 2008; Skyrms, 2010). I follow Scarantino (2013) in considering that this definition need not be opposed to the manipulation-based definition of communication (Dawkins & Krebs, 1978).

[12] Note that one may resist applying a code model to this scenario because one would rather want to account for the predator's behavior without postulating that it manipulates information, and instead through non-representational, mechanistic processes.

aim of generative linguistics is to help to spell out what these codes are. In particular, formal semantics makes hypotheses on what the pairings (i.e. the rules) are between, on the one hand, sentences (typically, from a fragment of English) and, on the other, their literal meaning (Coppock & Champollion, 2020).

The code models are economical, intuitive, and apply to a wide range of cases, but they do have one critical constraint: the code *must pre-exist the information transfer for it to take place successfully*. During the encoding process, the sender or the source must make use of a pairing between information and stimuli that the receiver must already be able to use before the decoding process. This constraint limits the explanatory scope of the code models, as we will now see.[13]

1.3. Trouble is brewing: Limitations of linguistic codes

Grice (1989) and his heirs have forcefully argued that, even though the information transferred through the grammar and lexicon of languages can be accounted for by the code models, linguistic communication doesn't boil down to codes. As the consecrated phrase puts it: *linguistic codes underdetermine meaning*. This is what led to the development of the Gricean model and, indeed, of contemporary pragmatics.

To see what this means, take the following dialogue:

(1) – Sam: 'Where is Joe?' – Maria: 'There is a little red Corvette in front of Maggie's house.'

Through the lens of the code models, Maria's answer doesn't make much sense. Here is a typical code model account: Maria has encoded a message (to be decoded) into a signal (the sounds of her utterance) by using a pre-existing code (the lexicon and grammar of spoken English) which pairs stimuli (phonemes) and messages (semantic representations) through

[13] In *Convention,* Lewis (1969) has given an account of how senders and receivers can achieve successful communication through a code model without starting their interactions with a pre-established code. Skyrms (1996, 2010) has further developed this account in a naturalist framework where senders and receivers need not possess the sophisticated cognitive abilities postulated by Lewis. Such accounts are not counter-examples to my claim that codes must pre-exist information transfer for it to take place successfully. Rather, they are complementary accounts of how to build up a code starting from none. Importantly, for both Lewis and Skyrms, this process requires multiple signals to be sent, multiple responses, and that the responses are more or less mutually beneficial or harmful. For this reason, they cannot explain what goes on in the cases that will interest us below.

generative rules. Sam, to get Maria's message, decodes the signal using the same code, a pre-established, generative pairing of sounds and semantic representations.

The problem is that this code can only lead Sam to decipher a message detailing the location of a car, although this is obviously not all that Maria meant. Instead, what we very naturally understand is that Joe probably is at Maggie's. A code model thus yields inaccurate predictions of the communicative phenomenon created by this sentence, i.e. of all the messages carried to an ordinary receiver by this signal.

Examples where the codes underdetermine the meaning, where grammar and lexicon are insufficient to predict what messages are sent, are easy to multiply and are well known and widely studied, especially since Grice's William James Lectures. There are important debates as to what the limit of linguistic codes is exactly. As mentioned above, some – known in the debate as 'contextualists' – defend that barely anything is linguistically encoded (Carston, 2002; Korta & Perry, 2007; Neale, 2004; Recanati, 2004; Sperber & Wilson, 1986). On the opposite side, some – known as 'minimalists' and 'literalists' – argue that much of what is meant is encoded (Cappelen & Lepore, 2005). But even the most radical literalist should nevertheless agree that the messages we send when we speak, the information that is meant to be communicated through speech, far exceeds what is linguistically encoded. Typical cases include what Grice (1975) called 'conversational implicatures'. Here are some more examples:

(2) – Sam: 'Are you going to Joe's party?' – Maria: 'I have to work.'
(3) – Sam: 'How much longer will you be?' – Maria: 'Mix yourself a drink'.
(4) Sam drops the tray with the dishes. – Maria: 'That is beautiful, Sam!'

Clearly, besides the meaning that is encoded through English grammar and lexicon, Maria's utterances send other messages; something like 'No' for (2), 'A while longer' for (3), and 'I blame you for breaking the dishes' for (4). The fact that codes underdetermine meaning is what led Grice and his followers to develop a new model of communication.

1.4. The prevailing Gricean models and their application to linguistic cases

What I call 'the prevailing Gricean models' are the models elaborated by Paul Grice and his heirs, including both post-Griceans and neo-Griceans.[14]

[14] My presentation of the prevailing Gricean models focuses on the similarities between various rather established versions (Bach & Harnish, 1979; Grice, 1989; Horn, 1984;

Just like the code models, they postulate a sender who sends some information through a set of stimuli and a receiver who gets the information after having interpreted the signal.[15] They need not be understood as rivals to the code model: firstly, they usually are taken to apply to different phenomena (although that is not true concerning the 'border dispute' between semantics and pragmatics) and, secondly, Gricean communication can involve the encoding/decoding process of the code model.

An important difference between the code and the prevailing Gricean models is that if there is a code used in a Gricean communication, it must always be supplemented by *the expression and recognition of communicative intentions*. This expression and recognition can be considered as essential to the prevailing Gricean model – but not to Gricean models in general, as we will see in the next chapter. According to the prevailing Gricean models, for a Gricean communication to take place successfully, the sender must produce, as part of the signal, some stimuli which could allow the receiver to figure out that she wanted to communicate something with this signal. Such stimuli are sometimes called 'ostensive stimuli' or 'overtly intentional signals'. In turn, the receiver must recognize that the sender has overtly displayed communicative intentions and try to figure out what these are.

The mindreading processes of displaying communicative intentions and figuring them out work together with a *common background* consisting, roughly, in background information or representations that are mutually cognized (believed, presupposed, desired, attended, feared, …).[16]

As part of the common background, a key element of Gricean models is what I will call *pragmatic principles*. In their most general form, pragmatic principles stem from the mutual assumption that the communicators are (imperfectly) rational agents and that they thus try to maximize their goals

Levinson, 2000; Lewis, 1969; Neale, 1992; Schiffer, 1972; Searle, 1969; Sperber & Wilson, 1986; Stalnaker, 1978, 2014) as well as more recent ones (Carston, 2002; Green, 2007; Moore, 2017; Neale, 2016; Sperber & Wilson, 2015; Stalnaker, 2014; Tomasello, 2008; Wharton, 2009).

[15] Contrary to most code models, Gricean models may not postulate that the information sent and received needs to be identical for communication to take place successfully, see Sperber and Wilson, 1986/95, Chapter 1.

[16] I use the terms 'common background' and 'cognized' as broad, theory-neutral terms in order to avoid committing to any particular Gricean model (more on this in the next chapters). We can here put aside the differences between 'common ground' (Grice, 1989, p. 65; Stalnaker, 2002), 'mutual knowledge' (Lewis, 1969), 'Background' (Searle, 1969), 'contextual assumptions' (Sperber & Wilson, 1986), or other cognate notions. The way they differ shouldn't affect the argument of this chapter.

in an intelligent way (Kasher, 1982). This allows them to suppose that they are mutually respecting something like Grice's *Cooperative Principle* (Grice, 1975). This consists of a mutual assumption of the communicators that they are trying to maximize the common goal that they have in their communicative interaction.[17] Grice further characterizes this principle through his famous maxims and submaxims. Most contemporary models have replaced the original Cooperative Principle with their preferred pragmatic principles. The most popular ones today may be Horn (1984)'s Q-based and R-based implicatures, Sperber and Wilson's (1986) Communicative Principle of Relevance, and Levinson's (2000) Q, I, and M heuristics.

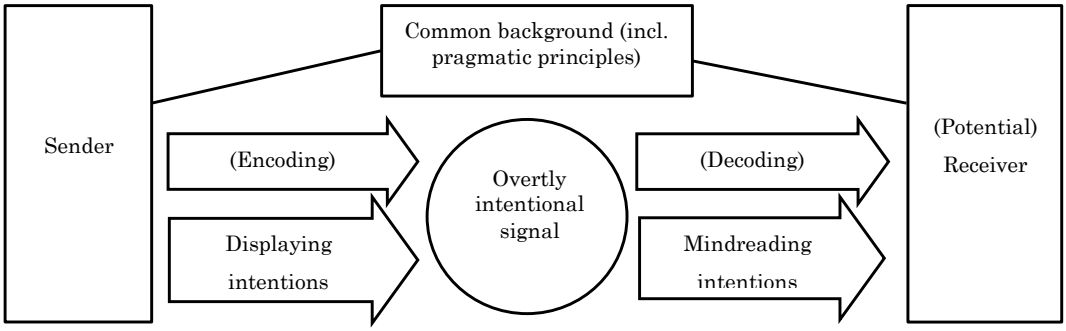

Fig. 1.3. The Gricean model of communication

Let me roughly sketch how this allows accounting for cases where 'linguistic codes underdetermine speaker-meaning' and what this phrase means. Let us take example (1):

(1) Sam: 'Where is Joe?' Maria: 'There is a little red Corvette in front of Maggie's house.'

According to the prevailing Gricean models, Sam not only needs to decode the signal using pre-established pairings between the sounds of Maria's utterance and its semantic representation. As we saw, using only the relevant code (English lexicon and grammar) is not sufficient to account for the messages Maria is sending. This is where the process of mindreading Maria's communicative intentions kicks in (represented by an arrow in Fig. 1.3). Since Maria, at least prima facie, is being cooperative, the messages she encoded in the signal must be relevant to Sam's question (pragmatic principle). Since Sam asked about Joe and she answered, presumably relevantly, about a car, Sam can suppose that Maria thinks he is aware

[17] Even when communicators are insulting each other, if they do so in a mutually intelligible language, this is already an indication that they are respecting something like the Cooperative Principle (see Horn 2006: 6-8).

that the little red Corvette is Joe's car. How else could she try to maximize the purpose of their interaction with this utterance? From these assumptions, the simplest and most efficient explanation of Maria's production of this signal is that she meant that Joe is probably at Maggie's.[18]

Importantly for us, according to the prevailing Gricean models, such explanations rely on the fact that Maria has produced a special kind of signal with her utterance: a signal that overtly displays her intentions to communicate, or an *overtly intentional signal* (a.k.a. an *ostensive stimulus*). Only overtly intentional signals are accounted for by the prevailing Gricean models. We will see that this assumption is problematic – this is why prevailing Gricean models leave a blind spot in the standard picture of information transmission. Let me thus present in more detail what overtly intentional signals are and why they are essential to prevailing Gricean models.

That the explanatory scope of the prevailing Gricean models is delimited by overtly intentional signals stems from the notion of *speaker-meaning*, originally called 'non-natural meaning' (Grice, 1957).[19] The process of defining speaker-meaning, first initiated by Grice, continues to this day. It has yielded many different versions of the Gricean model, some of which have replaced 'speaker-meaning' by a cognate notion (such as ostensive-inferential communication, see my Appendix for discussions of these notions). In any case, the explanatory scope of all prevailing Gricean models is defined by the two following necessary conditions:[20]

(i) The sender has a first intention (let us call it *Intention 1*) to make something manifest or to generate a particular effect e in the (potential) receiver with the signal x (such as generating a belief in the receiver, evoking an emotion, inducing a behavior, etc.), and

[18] A more complete explanation requires further assumptions from the common background. See e.g. Grice (1975), Sperber & Wilson (1995), or Horn (2013).

[19] Both the expressions 'speaker-meaning' and 'non-natural meaning' are rather unfortunate. First, speaker-meaning is not restricted to speech. Secondly, non-natural meaning is in a sense natural since it can be defined by concepts used in natural sciences (if we include psychology). Thirdly, there are cases of meaning which are neither natural nor non-natural (Denkel, 1992), a fact that this expression conceals.

[20] These conditions might not be sufficient. Extra conditions have been proposed for a complete definition of speaker-meaning, notably by (Bach & Harnish, 1979; Grice, 1989; Neale, 1992; Schiffer, 1972; Searle, 1969; Strawson, 1964a). To the best of my knowledge, clauses (i) and (ii) are nevertheless necessary in the different definitions proposed by these authors. A dissident voice is Davis' (1992, 2003), but, as I remark below, his account is not Gricean in the sense used here.

(ii) The sender has a second intention (*Intention 2*) that, through x, Intention 1 is made manifest (publicly discernable) or mutually manifest to both sender and (potential) receiver.

Many versions of the Gricean model have been proposed over the past 50 years or so, but all of them require something like Intentions 1 and 2. For instance, even though Sperber and Wilson (1995, 2015) don't use the notion of speaker-meaning, they instead use that of ostensive-inferential communication and define the latter through an 'informative intention' and a 'communicative intention', which are subsets of Intentions 1 and 2. Another example: Green (2007, p. 66) defines speaker-meaning through a single intention which has two parts, but these two parts can also be interpreted to correspond to our Intentions 1 and 2. Furthermore, neither Sperber and Wilson nor Green appeal to effects on a receiver in their definitions, but rather talk about making something manifest *tout court* (see my Appendix for my reasons to think that this implicitly requires a mention to a potential receiver). In any case, the variations between versions of the prevailing Gricean models shouldn't affect our argument.

Let us note by the way that some authors do not define speaker-meaning through anything like Intentions 1 and 2. For instance, Davis (2003) advocates a return to a Lockean definition of speaker-meaning and an abandonment of a large part of the Gricean program.[21] However, his non-Gricean model is subject to the same kind of counterexamples that we will discuss below.

What is essential for our purpose is that, within all versions of the prevailing Gricean models, for communication to be successful, *the sender must produce a signal that makes manifest* (i.e. publicly discernable) *an Intention 1* (an intention to inform, express, order, make something manifest, etc.). In other words, the sender must produce a signal which can fulfill Intentions 1 and 2. It is important to emphasize that if the sender produces stimuli which can only fulfill the Intention 1, this won't be sufficient to qualify as a case of speaker-meaning and the prevailing Gricean models would not apply.

To understand why the prevailing Gricean models appeal to two different intentions (or a two-part intention as in Green, 2007), consider the following case (from Grice 1957, p. 380). A put B's handkerchief near the scene of a murder to induce the detective to believe that B was the

[21] For critics of Davis's definition of speaker-meaning, see e.g. Green (2007, p. 78–9), Buchanan (2012), a reply to the latter is Davis (2013), and a reply to the reply is Zeman (2014).

murderer. Here A had an Intention 1 to generate a certain effect in the receiver (the detective) and might very well succeed in fulfilling it. But A didn't have an Intention 2 that her Intention 1 is made manifest. Indeed, it is quite the opposite: the handkerchief is supposed *not* to reveal A's intention to induce a belief in the detective. The prevailing Gricean models don't apply to this case. Note that this doesn't depend on whether the handkerchief possesses stimuli (such as A's DNA) that can lead the detective to infer A's Intention 1. The prevailing Gricean models not only require the recognition of Intention 1, but the intentional public display of Intention 1, the Intention 2 to make the Intention 1 manifest. This is what is required by clause (ii).

The rationale for this clause is that, if the signal doesn't possess the stimuli necessary to fulfill clause (ii), the signal wouldn't give any reason to the receiver to go through the process of figuring out what the sender is communicating beside what is encoded in the signal. Below, we will see that this is not quite right, but for now, let us assume with the prevailing Gricean models that this is so.

To fulfill clause (ii), the sender must thus produce an *overtly intentional* signal, also called 'wholly overt' or 'ostensive' (Bach, 2004; Dorit Bar-On, 2017; Csibra, 2010; Green, 2007, Chapter 3; Moore, 2018; Recanati, 2008; Sperber & Wilson, 1986; Strawson, 1964a). It is important to be clear about what this means because this notion is central to the argument of this chapter. An overtly intentional signal is a stimulus that makes manifest (publicly discernable) the fact that it was produced to fulfill an Intention 1. In other words, it is a signal that overtly shows one's intention to communicate. Typical examples of signals overtly intended for communication include pointing, winking, waving, creating prolonged eye contact, or speaking. In all these cases, the sender makes it clear (manifest, publicly discernable) that she intends to produce a certain effect in the audience (cases where there is no audience are not typical, see my discussion in the Appendix). I will detail what is meant by 'intention' in this context below.

If a signal is overtly intentional and recognized as such by the receiver, this will allow the sender to satisfy her Intention 2, i.e. the intention to make an Intention 1 manifest. Overtly intentional signals thus are what allow a successful communication within the prevailing Gricean models. They are the signals carrying the type of information that is required for senders to

speaker-mean something and for receivers to infer what is speaker-meant.[22]

1.5. The blind spot: Limitations of both models

Contra one of the current assumptions of the standard picture, I believe that there are cases of communication, and of information transfer more generally, that can be accounted for by neither of the two models I have presented. To show this, I will focus on cases of laughter where the laughter is not overtly intentional.

Before I do so, let me remark that, although my argument in the present chapter is negative, my general claim is not to be interpreted as anti-Gricean – on the contrary! I believe that accounting for the cases which resist the prevailing Gricean models actually leads us to extend the scope of Gricean models, as I will explain in the next chapter. My general goal thus is to amplify, and not to thwart, the type of explanation given by Griceans.

The argument in this section will go as follows. First, I will make explicit the distinction between laughter that is overtly intentional and laughter that is not. Then, I will expose how the code models could be used to explain some of what is communicated by laughter that is not overtly intentional. After that, I will give examples of laughter that is not overtly intentional. These cases are comparable to (1)–(4) in that the code models fail to account for all the information transmitted. However, contrary to cases (1)–(4), because they are not overtly intended for communication and thus cannot respect the conditions for speaker-meaning, these examples won't be accounted for by the prevailing Gricean models.

I choose laughter as an example for several reasons. First, codes for laughter are empirically well studied. Second, cases of laughter where the laughter is not overtly intentional are common. Third, laughter is a non-conventional form of emotional expression that is culturally universal (Sauter et al., 2015) and even shared with other species (Panksepp & Burgdorf, 2003; Preuschoft, 1992). This distinguishes laughter from many emotional expressions such as interjections ('Wow!', 'Yuk!', etc.) or emotional gestures (thumbs up, middle finger, etc.) which are at least partially conventional. This is relevant because non-conventional emotion

[22] Once again, certain authors, such as Sperber and Wilson (1986) and Green (2007), define overtly intentional signals slightly differently, but in ways that shouldn't affect our argument. Green describes it as the product of an overt action, which is 'an action done intending that (a) something be publicly discernable, and (b) this intention itself be publicly discernable as well'.

expressions have been thought to be neatly accounted for within code models of communication (Dezecache et al., 2013; Moore, 2018).[23] As Bar-On (2013) puts it:

> « In the philosophical literature, [expressive] behaviors are regularly portrayed as mere reliable indicators of the internal states that regularly cause them—as signs with natural meaning. »

But we will see that what laughter means is not only irreducible to natural meaning, but cannot be accounted for by any sort of code model.[24]

To explain when laughter is overtly intentional or not, it is useful to distinguish between four different cases of laughter, based on Green's typology of signals (2007, p. 12):

(a) *Uncontrollable laughter*. The impulse to laugh is too strong to be repressed and I thus cannot help but laugh, against my will.
(b) *Not willed, not uncontrollable, but not suppressed.* I have an impulse to laugh, but the impulse is not so powerful that I couldn't suppress it. Furthermore, I neither have the will to produce the laughter, nor the will not to produce it (because, say, there is no audience). The mild impulse results in my not suppressing the laughter, perhaps without thinking about it.
(c) *Willed but not overt.* I have an impulse to laugh, but this time I also want or intend to produce it, because, say, I have an Intention 1 to generate a belief in the audience that I am mirthful. However, I don't have an Intention 2 to make my Intention 1 manifest.
(d) *Willed and overtly intended for communication*. I not only have an intention to laugh as in the preceding case, but I also intend to make manifest my intention to produce an effect with the signal (I have an Intention 2).[25] To fulfill my Intention 2, I have to do something other than merely produce a regular laugh. Since, in this case, I

[23] This last point doesn't mean that non-conventional emotion expressions cannot be used as overtly intentional signals (Wharton, 2009, Chapter 5). Rather, that they typically are not, and that is how they are generally considered in the literature. This contrasts them with overtly intentional emotional expressions such as thumbs up, whistling with admiration, or the use of expressive language

[24] As said above, Bar-On agrees with me that neither the prevailing Gricean models nor the code models can account for certain expressive behaviors. But unlike hers, the cases that I present must be accounted for by a Gricean model, because they involve some kind of implicatures.

[25] Green (2007: 66) doesn't characterize overtness in terms of Intention 1 and Intention 2, but in terms of a self-referring intention with two components. As I noted above, he defines an overt action as 'an action done intending that (a) something be publicly discernible, and (b) this intention itself be publicly discernible as well.' However, as I noted, the differences shouldn't affect my argument.

want my Intention 1 to be made manifest to the audience, I have to produce stimuli overtly intended for communication while laughing (see also Csibra, 2010; Scott-Phillips, 2015, p. 66; Sperber, 2000).

Importantly, for laughter to fall in category (d), it must be produced in a way that makes it *discernibly different* from the laughter of categories (a), (b), and (c). The laughter must be produced with some recognizable, manifest, publicly detectable evidence that overtly shows it was *intended* for communication.

For instance, a bout of laughter can overtly show it was intended for communication by being accompanied with a pointing finger, sustained eye contact, a particular posture, etc. (see Csibra, 2010 for an empirical investigation on different ways to display overtly intentional signals). In the following, we will concentrate on the acoustic stimuli that laughter is made of. In this respect, a bout of laughter can be overtly intentional by, e.g., being exaggerated or stylized in some way. You can probably recall or imagine laughter which, merely through its sounds (or its silences), makes manifest an Intention 1 to produce an effect in the audience. Imagine for instance laughter which sounds particularly ironic or Machiavellian and which thus overtly display communicative intent. Such laughter could be exaggeratedly long, or loud, or low-pitched, or slow, particularly articulated, being produced with a noticeably unnatural timing, or stylized in any other way that would allow making manifest (i.e. publicly discernable, mutually recognizable) the fact that it was produced to fulfill an Intention 1, and would thus be distinguished from a laugh that is not overtly intentional.

A laugh that is not overtly intentionally communicative can be (a) uncontrollable, (b) not willed, not uncontrollable, but not suppressed, or (c) willed but not overtly intended for communication.[26] In each of these three cases, the laughter isn't intended to make an Intention 1 manifest. Most cases of laughter, I believe, fall in either of these three categories.

We see something funny or embarrassing and laugh, without any intention to publicly display any communicative intent. We may argue that we lack these intentions on several grounds, coming from different philosophical perspectives. One is that, in such circumstances, it is as automatic to laugh

[26] Green (2007, p. 96) notes that if emotion expressions 'are not inhibited [i.e. are allowed] but, (a) at the time they are manifested, could have been, and (b) we refrain from inhibiting them for a reason, then they merit treatment as intentional [i.e. willed].' Depending on how demanding is the criterion for knowing one's reasons (to allow an emotional expression), the limit between (c) and (b) might thus be very thin. What is important for our argument however is that such cases are distinguished from (d).

as it is, in normal circumstances, to breathe, to sneeze, or to blink: it just happens without us intending anything, we let it happen although we might be able to control it to a certain degree – often we could have refrained from breathing, sneezing, or blinking if we had wanted to. This is why O'Shaughnessy (2008, p. 359ff) contrasts 'the *semi-helpless inclinatory phenomenon* of laughter' with (sub-)intentional actions and classify this type of laughter with 'the merely bodily event of the twitching of an eyelid' and 'the autonomic phenomenon of breathing' (359). Note that this so even though O'Shaughnessy has a very wide notion of what counts as an (sub-)intentional action, e.g. the latter includes 'idly and unaware moving one's tongue in one's mouth as one drives' (352).

Another indication is that, in these cases, if one tries to figure out one's *reasons* for laughing, one doesn't find that the laughter was a signal produced with the intention to make manifest another intention. In such cases, this is not the reason for one's laughter and, according to a common view according to which intentional actions are actions performed for a reason (e.g. Davidson, 2001), this indicates that the laughter is not overtly intended for communication. If we asked the person 'Why did you laugh?' she wouldn't say that it was to show one's intention to communicate something (see Anscombe (1957) and her discussion of 'Why?' questions in her analysis of what intentional actions are).

Relatedly, the means-end format typical of, and perhaps essential to, intentions is absent: in such cases, it is not true that we plan to fulfill the Intentions 1 and 2 and that we laugh as a *means* to fulfill them (see Bratman (1987, 2018) as well as Mele and Moser (1994) for the roles that intentions play in means-ends practical rationality and the importance of action plans to analyze intentional actions).

At this point, one may ask: these arguments are only valid if the Intentions 1 and 2 must somehow be accessible to consciousness, but couldn't we have these intentions and not be able to access them? Well, first, intentions are usually defined as mental states that *are* accessible to consciousness (Pacherie & Haggard, 2010) and, secondly, most importantly, the Intention 2 defining speaker-meaning requires that the Intention 1 is made mutually recognizable, or mutually manifest, to both senders and receivers. Thus, at least the Intention 1 must be consciously accessible if one is to produce a signal which can be accounted for by the prevailing Gricean models.

Furthermore, on a causal account of intention such as that favored by Davidson (2001) or Searle (1983), we have once again reasons to consider that laughter is often not overtly intended for communication because the

etiology of laughter in many cases is *emotional*. Indeed, it is widely accepted that emotions can cause expressions (facial, vocal, etc.) without intentions (Scherer & Moors, 2019, sec. C). In fact, according to laughter specialist Robert Provine (Provine, 2001, 2017), thinking that laughter always is intentional is committing to the 'error of intentionality'. Moreover, thinking that emotional expressions such as laughter must always be produced with an Intention 2 to make an Intention 1 manifest seems to be a serious over-intellectualization of emotional reactions.

Even though there certainly are cases where laughter is produced with the Intentions 1 and 2 and where such communicative intentions are recognized, for my argument to work, we just need to accept that there are cases of laughter that is not and that the cases I discuss below can be construed as part of this category. Given what I have said so far, these requirements should be uncontroversial. And indeed, in line with what I have said so far, the relevant Gricean literature considers laughter as a typical type of stimulus which normally is produced without the relevant communicative intentions (Dezecache et al., 2013), and, in general, spontaneous emotional expressions are considered as such (Bar-On, 2013; Green, 2007; Grice, 1957; Moore, 2018).

What we will now see is that both the code models and the prevailing Gricean models are unable to account for the meaning of cases of laughter that is not overtly intentionally communicative. To do so, let us begin by looking at what a code model can tell us about what laughter means.

A first, rather naïve, attempt at designing a code for laughter could be the following: senders undergo positive emotions such as mirth or amusement and non-consciously, automatically encode this emotional state (the information) into the set of acoustic stimuli making up laughter (the stimuli) according to a pre-established association (the code) between these positive emotions and the acoustic stimuli. A receiver then unconsciously, automatically uses the same code to decipher the laughter (the stimuli) and thus recognizes that the sender is in a state of positive emotion (the information).

However, what laughter communicates isn't restricted to positive emotions. Based on ethnographic and literary evidence, Poyatos (2002, pp. 71–76) lists ten different communicative functions that laughter can play, including the expression of negative affects such as embarrassment or aggressiveness.

In response to such claims, some researchers have attempted to create codes pairing different emotions with different types of laughter (Gervais

& Wilson, 2005; Lavan et al., 2017; McGettigan et al., 2015; Provine, 2004; Szameitat et al., 2009; Tanaka et al., 2011; Tanaka & Campbell, 2014; Todt & Vettin, 2005; Wild et al., 2003).[27] The result of these empirical investigations is that we can devise a code constituted of two, but only two, laughter–emotion pairs. In other words, if we are interested in what psychological states laughter can express, we find that we can distinguish empirically between two categories of laughter. Note that these two categories form a continuum, with intermediary cases failing to fall neatly in one of the two categories. I will follow Gervais and Wilson (2005) and call the two categories *Duchenne laughter* and *non-Duchenne laughter*.[28] Importantly for us, only Duchenne laughter can be reliably paired with positive emotions (Gervais and Wilson, 2005; Tanaka and Campbell, 2014). Non-Duchenne laughter cannot be paired with a specific affective state. Non-Duchenne laughter isn't fake laughter and can sincerely express a wide range of psychological states. Very often in social contexts, we emit such brief, soft, medium-pitch laughter, without paying attention, and certainly without the intention to deceive or fake anything. 'Chuckle', 'titter', or 'snigger' usually refer to non-Duchenne laughter.

In light of these results, it appears that the best that a code model for the sounds of laughter could do would be to offer a pairing between, first, the acoustic stimuli of Duchenne laughter and positive emotions, and second, between the acoustic stimuli of non-Duchenne laughter and different types of psychological states that it might express (see Table 1.1).[29] Let me underline that the empirical literature thus shows that we cannot acoustically distinguish between cases of, say, nervous laughter, sardonic laughter, embarrassed laughter, etc.

[27] Space constraints prevent a review of the literature. For an overview, one can read the introductory remarks of (Curran et al., 2018; McGettigan et al., 2015) as well as the more complete, but less recent, paper by Gervais and Wilson (2005).

[28] This terminology is rather unfortunate because many problems have recently been highlighted concerning the original distinction first made by Duchenne in 1862 for smiles. The *Duchenne smile* is supposedly recognizable through certain muscle activations around the eyes and reliably correlated with positive emotions, unlike non-Duchenne smile (Messinger et al., 2001). For criticisms see e.g. Krumhuber and Manstead (2009). Non-Duchenne laughter is also called 'controllable', 'voluntary', 'polite', 'emitted', and 'soft' while Duchenne laughter can be called 'uncontrollable', 'involuntary', 'sincere', 'mirthful', and 'loud', but I find these terminologies ever more problematic. Again: contrary to some researchers, I am not excluding the possibility that there is a continuum of cases between Duchenne and non-Duchenne laughter (or smiles), and that the distinction is clear only with paradigmatic expressions.

[29] I discuss below attempts at devising a more sophisticated code.

Information encoded	Stimuli
Positive emotion (mostly mirth, but also joy, relief, or playfulness)	Acoustic stimuli of Duchenne laughter (louder, higher-pitched, lasts longer, more calls per bouts, …)
Amusement, contempt, fear, incredulity, joy, sadness, Schadenfreude, social anxiety, urge to affiliate, urge to aggress, ticklishness.	Acoustic stimuli of non-Duchenne laughter (softer, lower-pitched, briefer, fewer calls per bouts, …)

Table 1.1. The code for the sound of laughter, based on empirical investigations (Gervais & Wilson, 2005; Lavan et al., 2017; McGettigan et al., 2015; Provine, 2004; Szameitat et al., 2009; Tanaka et al., 2011; Tanaka & Campbell, 2014; Todt & Vettin, 2005; Wild et al., 2003).

I do not doubt that this code can successfully account for many cases of information transmission. There surely are many instances of Duchenne laughter where all that is understood by the audience is that the person laughing is undergoing a positive emotion and there surely are many cases where the audience only understands that the person producing non-Duchenne laughter is undergoing one or the other affective states, as per Table 1.1.

However, in many other cases, this code fails to give a satisfying account of the information transmitted by laughter, even though the laughter is not overtly intended for communication. Take the following example and construe it as one of those typical cases where the laughter is not produced so that it overtly displays an intention to communicate. You can construe it as appearing to be either (a) uncontrollable, (b) not willed, not uncontrollable, but not suppressed, or (c) willed but not overtly intentionally communicative:

(5) – Emily: 'Where did your wife go?' – Frank: 'She is actually calling the doctor to see if she can meet him about her gastroenteritis. Huhuh. Heh. Huh. (low pitched, soft)' – Emily: 'I will keep that to myself.'[30]

The prediction of a code model would be that, since Frank's laughter is more like non-Duchenne than Duchenne laughter (being low pitched and

[30] The example is adapted from a corpus example from (Ginzburg et al., 2015).

soft), he sends a stimulus which carries the information that he is either undergoing amusement, contempt, fear, incredulity, joy, sadness, Schadenfreude, social anxiety, an urge to affiliate, an urge to aggress, or feeling ticklish.

But this is not a satisfying account of the information we understand Frank's laughter to carry. For instance, we[31] very naturally understand from the laughter that (p) Frank's wife would rather avoid that Emily or other people be informed of her gastroenteritis, (q) that Frank is embarrassed to reveal this private information (he laughs out of embarrassment), but (r) that the situation is not too worrisome (otherwise he wouldn't have laughed). Furthermore, these pieces of information are not only transmitted to Emily, but they even update the common background between Emily and Frank. This is why it is perfectly natural for Emily to reply to Frank's laughter with 'I will keep that to myself'. This reply, I take it, presupposes that Frank has sent something like (p) with his laughter.[32] Although Frank doesn't literally *mean* p, q, and r by his laughter, I take this information to be part of what he *allows* his laughter to mean (see next chapter for this notion).

Case (5) can thus be compared to the linguistic cases (1)–(4) where the messages encoded in the signals weren't the only ones transmitted to a normal receiver. A normal receiver can understand more of what is communicated by the laughter than what is made available by the relevant code. The best prediction available to a code model, if we trust the empirical investigations behind Table 1.1, cannot begin to explain this fact.

At this point, a reader might think: 'Couldn't a code model make use of other codes that the one presented in Table 1.1, such as English grammar and lexicon and use these codes in conjunction with Table 1.1 to predict what information is carried by the laughter?'. The problem is that, as far as I can tell, even the combination of the sentences uttered by Frank and

[31] To be cautious, I should restrict 'we' to Ginzburg et al. (2015) (from whom this interpretation stems) the colleagues I have discussed this example with, and myself. I have not met anyone who disagreed with this interpretation (especially once we imagine the sound of a soft, brief, low-pitched non-Duchenne laughter), but there surely is room for disagreement and so 'we' may not refer to anybody.

[32] If Frank had not laughed, Emily may, somewhat surprisingly, have replied in the same way, but it would then have been her who would have updated the common background with the presupposition that she thinks that either Frank or his wife wouldn't want the gastroenteritis news to spread. It would have been an informative presupposition. Speakers can indeed presuppose information that is not already part of the common ground when they expect that the audience will just accept it without objections. if Frank had not laughed and Emily had given the same reply, that is how I would understand her: as performing the informative presupposing that p.

the laughter won't constitute a signal which can be paired, through pre-established pairing, with the information we understand Frank's laughter to carry: this information is too idiosyncratic. And if that were possible, the burden of the proof rests on those who claim that a code could predict that Frank's laughter is carrying (p), (q), and (r). Until one has built such a code, a code model is not an available explanation.

The code models, unlike the Gricean models, cannot appeal to the mindreading abilities based on pragmatic principles, nor to any form of inference based on contextual information that is not predicted by pre-established codes. If one appeals to something other than the pre-established pairing between information and stimuli, one is outside the scope of the code models.

Another attempt to defend the code models here would be the following. We may make use of a code for laughter that is more sophisticated than the ones discovered by scientists: a code more fine-grained than the one presented in Table 1.1. We can imagine that a sophisticated code would pair information not only with acoustic stimuli, but also with contextual stimuli. For instance, one could attempt to list contextual stimuli typically associated with negative emotions (death, sickness, fights, loud noises, rapid and unpredictable movements, great heights, …) and then associate these stimuli with laughter to obtain the following pairing: 'typical negative stimuli + non-Duchenne laughter (stimuli) => expression of negative emotion (information encoded)'.

First problem: I cannot imagine a list of stimuli typically associated with negative emotions where the stimuli are also never associated with positive emotions. Second problem: the typical stimuli would not include idiosyncratic stimuli (e.g. someone being afraid of clowns). Third, and worst, problem: even if we could build this code, it will not be enough to account for all the information that we understand laughter to transmit. For instance, such a code wouldn't yield the associations needed between Frank's laughter and the information we understand him to suggest that his wife wouldn't want Emily to be informed. The latter, I take it, is too specific to be coded in any pre-existing pairing. In order to understand Frank's laughter to carry pieces of information such as (p), (q), and (r), the receiver needs to perform inferences based on a common background and a few other assumptions similar to the ones found in the prevailing Gricean models. We will see that in detail in the next chapter.

Code models are not sufficient. However, since the laughter in (5) is not an overtly intentionally communicative stimulus, and it does not overtly

display an Intention 1, it is also out of the scope of the prevailing Gricean models.

To avoid this last conclusion and the ensuing consequence that the standard picture has a blind spot, one might refuse the claim that Frank's laughter is not overtly intended for communication and that it does display an Intention 2 to make it mutually recognizable that Frank has an Intention 1 to generate in Emily the beliefs that (p), (q), and (r) (or to make these manifest).

In response to this rescue attempt, let me emphasize that, for my argument to go through, I don't need it to be the case that no laughter is ever overtly intended for communication, but only that there exist some cases where the laughter is not overtly intentional, that is, where the laughter is not produced to fulfill an Intention 2 that an Intention 1 is made (mutually) manifest and that, in these cases, we understand the laughter to mean more than what a code model can predict. I postulated that the example I gave in (5) is of that type. I don't thereby suggest that the words I use in (5) cannot be used to describe another example where the person laughing has the Intentions 1 and 2 and succeeds in displaying them through the laughter (through, for instance, a special kind of gaze). In other words, I am not saying that the description given in (5) makes it necessary that the laughter is not overtly intended for communication. The reader only needs to allow that (5) is a *possible* description of laughter that is not overtly intended to communicate (p), (q), and (r), but that it nevertheless carries these pieces of information. Remember, by the way, that non-intentional laughter is, according to much empirical research, the most common type of laughter (see Provine 2000, 2004, 2017). Remember also that, if we take influential views about intentions such as Anscombe's (1957), Davidson's (1980), Searle's (1983), Bratman's (1987, 2018), or O'Shaughnessy's (2008: 359ff), we find that (5) can very well be a case where their criteria for intentions are absent (see above).

Another way to put my point is the following: the laughter in (5) is creating something like conversational implicatures. But they are not conversational implicatures because those are a subset of speaker-meaning and (p), (q), and (r) are not speaker-meant, since the laughter is not produced with the Intentions 1 and 2. [33]

[33] Note that we wouldn't say that, by laughing, Frank *meant* that (p), (q), and (r). This fits well with the fact that Frank's laughter doesn't fulfill the conditions for speaker-meaning. The definition that Grice gave respects the intuitions guiding how to use the phrase 'By this, the sender meant that …'.

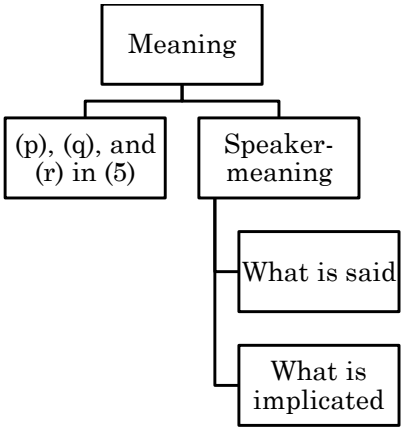

Fig. 1.4. Pieces of information (p), (q), and (r), even though they are similar to conversational implicatures, do not belong to the prevailing Gricean model, because they are not part of what is speaker-meant.

Here is another example of laughter that resists both the code and the prevailing Gricean models. Once again, let us construe this example so that the laughter does not display an Intention 2 to make an Intention 1 mutually recognizable:

(6) (David and Chuck are good friends, who share progressive, politically left-wing, values) David: 'You know, I was thinking: maybe Sarah Palin is the future of the Republican party…' Chuck : 'Hh hh, heh heh heh, huhuh, hahahahaha (high pitched, loud)' David: '…I even think she's got her chances for the next election.' [34]

The context and the conversation in (6) make it plain that two pieces of information transmitted by Chuck's laughter, and which update Chuck and David's common background, are the following:

(s) Chuck doesn't take his friend's prediction very seriously.
(t) Sarah Palin being the future of the Republican Party is a risible, absurd, or ridiculous idea.

Once again, a code model doesn't come close to being able to account for this. Even if we can categorize the laughter as rather Duchenne-like and that the code can thus tell us that Chuck is probably undergoing a positive emotion (such as mirth, amusement, or relief), this doesn't suffice to account for the fact that his laughter carries s and t. That Chuck undergoes a positive emotion is coherent and complementary with s and t and can be considered as further information that is carried by the laughter:

[34] Adapted from Ginzburg et al (2015, p. 137).

(u) Chuck undergoes a positive emotion.

A code model can account for (u) and thus for some of the information that Chuck's laughter is transmitting, but it cannot account for all the information we naturally understand his laughter to carry.

Some readers might worry about the conclusion I have reached, namely that a code model cannot account for the pieces of information carried by Chuck's or Frank's laughter, on the grounds that I have caricatured and oversimplified the way code models are used in the relevant literature. This worry is unfounded. The models that are used to account for the communication of emotions in affective sciences really are similar to the one I have discussed (see e.g. Ekman (1993) for facial expression, Scherer (2003) for vocal expression, Juslin & Laukka (2003) for musical expression, Dael, Mortillaro and Scherer (2012) for bodily movements). The code in Table 1.1 does not caricature such studies. More broadly, all of the many researchers who base their work on Shannon's (1948) mathematical model of communication or Brunswick's (1956) lens model use models along these lines. In philosophy, models of information transmission which are built on notions such as probabilistic meaning (Dretske, 1981; Millikan, 2004; Scarantino, 2015; Scarantino & Piccinini, 2010; Shea, 2007; Skyrms, 2010; Stegmann, 2015) or teleosemantic meaning (Dretske, 1986; Godfrey-Smith, 1991; Millikan, 1984, 1989; Papineau, 1984, 1993) are also sufficiently similar to the model I have discussed to be subjected to my argumentation, since they are essentially based on pre-established correlations and that the latter won't be sufficient to account for all the information that is transmitted by Chuck or Frank's laughter.[35]

Finally, once again, if some readers are worried that I have unjustifiably stipulated that the laughter in (6) is not overtly intended for communication, i.e. that it is not produced with the Intention 2 that an Intention 1 is made (mutually) manifest, remember that I don't need (6) to *necessarily* be a description of laughter produced without the Intentions 1 and 2. To go back to Green's distinction, if (6) truly describes a possible case where Chuck's laughter is either (a) uncontrollable, (b) not uncontrollable, not suppressed, but not willed, or (c) willed but not overtly intended for communication, then my stipulation that (6) is not a case of

[35] I will discuss these notions in more detail in Chapters 7 and 8. For a detailed argumentation of why Millikan's model is a code model, which can be applied mutatis mutandis to many other teleosemantic accounts, see Origgi & Sperber (2000) or Reboul (2017: 30ff).

speaker-meaning is unproblematic. I don't see any reason to doubt that this is a reasonable requirement.

But if the code models and the prevailing Gricean models are unable to account for cases (5) and (6), how should we explain the information that is transmitted with such laughter? I believe that the answer is very similar to the one provided by the prevailing Gricean models in so far as it appeals to something like a common ground between senders and receivers and to mindreading inferences based on some kind of pragmatic principles. In other words, I believe that such theoretical constructs can be modified to apply to cases that do not involve overtly intentional signals and so that the Gricean model can extend beyond speaker-meaning or ostensive–inferential communication. This is not trivial but I think it can be made to work. I develop this answer in the next chapter.

1.6. IS THAT COMMUNICATION ANYWAY?

Before I move to the conclusion, let me briefly address an objection.[36] Is the information sent by the laughter in (5) and (6) *communicated* or is it merely inferred by the relevant audience? If they are not communicated, then it is normal that the code and the Gricean models don't apply to it, because they were designed for communicative phenomena. So, it is not true that the standard picture of information transmission has a blind spot: it is not its job to account for such cases.

Let us first ask whether the cases discussed are communicative phenomena or not. By 'communication', I refer to the transfer of information between a sender and a receiver whereby both the sending and the receiving were designed for this information transfer (Green, 2007; Hauser, 1996; Maynard Smith & Harper, 2003; Scott-Phillips, 2008; Skyrms, 2010). The design in question can come from the intentional behavior of an intelligent creature as well as from natural and cultural selection.

Now, in (5) and (6), the information is not transferred through an intentional design, but it might well be communication nevertheless because it may be that the sending and the receiving have been designed by natural selection for the transfer of such information. This is, ultimately, an empirical, evolutionary question. As far as we can tell, it is at least plausible that it is the case (for a development of this line of reasoning, see Green, 2007, Chapter 1). Indeed, not only have all kinds of

[36] Thanks to Mitch Green, Kevin Lande, and Deirdre Wilson for discussing this question with me.

laughter most probably been selected by natural selection for communicative purposes (see e.g. Gervais and Wilson, 2005), but the mindreading processes which are involved in our understanding of cases (5) and (6) were probably also selected to fulfill a communicative function, a point on which otherwise conflicting evolutionary theory of language seem to converge (R. I. Dunbar & Shultz, 2007; Reboul, 2017; Scott-Phillips, 2015; Tomasello, 2008; Zuberbühler, 2018). Thus, the empirical evidence so far seems to suggest that the sending and receiving involved in cases (5) and (6) were plausibly naturally selected for the transfer of non-coded information.

That being said, if it turned out that it was not the case, e.g. that laughter actually wasn't selected as a signal but rather is merely a cue we use to infer people's psychological states, the analysis given in this chapter wouldn't be affected and my claim would still hold: neither the code models nor the prevailing Gricean models would be able to account for the information transferred by the laughter in cases (5) and (6). The standard picture of information transmission would still be unable to account for such cases. The only difference would be that it would have the following excuse: cases (5) and (6) aren't communication proper and the models of the standard picture were designed for communication. But then, if it so turned out, what kind of theory should we turn to for an account of the information transferred in such cases? We would still be left with the same questions. We would still wonder what kinds of information is carried by our laughter in this or that situation. We would still understand the laughter in (5) and (6) to mean things that cannot be accounted for by the prevailing Gricean or code models. Furthermore, even if cases (5) and (6) aren't communication, I believe that we can devise an extended version of the Gricean model to explain them, one that is not restricted to overtly intentional signals, and so use a model that was first designed for communication to understand the meaning carried by these stimuli. We will see how in the next chapter.

Finally, let me emphasize that even if the meaning of the laughter in (5) and (6) shouldn't be called 'communicative', it cannot be accounted for by a theory of non-communicative natural information either, since natural information just is a type of coded information, i.e. information which is paired, through a pre-established correlation, with a certain set of stimuli. The natural information carried by laughter is what codes such as the one presented in Table 1.1. try to capture. And we have seen why this would not be sufficient.

1.7. CONCLUSION

If the argument of this chapter is on the right track, it is false that any information transferred through stimuli that are not overtly intentional signals must be accounted for by a code model, since there are cases where *codes underdetermine the meaning of stimuli that are not overtly intended for communication.*

Let me observe that my argumentation in §1.5, where I discussed the two laughter cases, was essentially the same as in §1.3, where I presented the limits of the code models for linguistic communication. In other words, the reasons I gave for thinking that the code models cannot account for cases of laughter where the laughter is not overtly intentional are of the same kind as the ones which led Grice and his followers to think that the code models cannot account for cases of linguistic communication, such as cases (1)–(4). Thus, if somebody comes up with a new code model which can account for the information transmitted by the laughter in (5)–(6), it would be very surprising if this model weren't also able to account for linguistic cases (1)–(4). Whether or not such a new code model can be devised, one which would basically reduce conversational implicatures to semantics, I believe to have shown that the code models fail in the same way for all cases presented here. In other words, if one agrees that the code models fail for linguistic cases, one should also agree that it fails for laughter that is not overtly intended for communication.

We are now faced with a problem. We produce stimuli which we understand to carry information whose analysis escapes the existing models of information transmission. How, then, can we account for this? In other words, how can we account for the cases where codes underdetermine the meaning of stimuli that are not overtly intended for communication? The answer, I believe, is to be found in an extended Gricean model, one that is not limited to speaker-meaning, as I will argue in the next chapter.

2. THE EXTENDED GRICEAN MODEL

« We take a painting to be a product of a rational agent. Accordingly, we assume that no major detail of a painting is superfluous. »

– Asa Kasher, *Gricean Inferences Reconsidered*

Abstract. In this chapter, I introduce the Extended Gricean Model and the kind of meaning that it is supposed to account for: allower-meaning. Its goal is to account for the transmission of information that is accountable neither by the code models nor by the prevailing post- or neo-Gricean models. Such information is conveyed by stimuli that do not overtly display communicative intentions – and so which fall outside the scope of the prevailing Gricean models – but where a pre-established pairing between information and stimuli cannot account for all the information that is transmitted, and so which fall outside the scope of the code models. The Extended Gricean Model can account for such information transmission by preserving features of the prevailing Gricean models – in particular, mindreading processes based on pragmatic principles – but without being restricted to speaker-meaning or ostensive communication.

2.1. INTRODUCTION

As we saw in the last chapter, an assumption of what I have called the standard picture of information transmission is the latter phenomenon can be accounted for by either the code models or the prevailing Gricean models. I have argued that there are counterexamples to this assumption and that the standard picture thus has a blind spot. I have focused on two examples where we understand laughter to convey pieces of information that go beyond the predictions of both kinds of models. The reason was that the information which we could infer from laughter was underdetermined by the codes available – and so out of reach for the code models – but the laughter was not produced with the overt communicative intentions required by the prevailing Gricean models, and so out of reach for the latter. Instead, as we will see in detail in this chapter, the person allowed her laughter to mean certain information which (a) was not overtly

intended to be communicated and (b) was not encoded in the laughter. These are cases of implicature* without speaker-meaning.[37]

At the end of the chapter, we were thus faced with a puzzle: how to account for such cases if the standard picture of information transmission cannot? In the present chapter, I offer a solution by presenting the Extended Gricean Model, a Gricean model whose scope is not restricted to signals produced with overt communicative intentions. Before I do so, let me remind you of what are the code and the Gricean models.

Roughly, the code models – which includes traditional models of communication such Aristotle's, de Saussure's, Pierce's, or Shannon's – is a model where a sender, or information source, encodes information into a stimulus which is sent to a receiver who accesses the information by decoding the stimulus thanks to the code used for the encoding. A code just is a pre-established pairing between information and types of stimuli. A code can be conceived as a sort of dictionary: for each type of stimulus belonging to the code, it gives you one or several pieces of information, just like each lexical entry of a dictionary gives you a definition.

Unlike the content of dictionaries though, codes need not be conventional. Communication between trees, bacteria, or bees can be explained with the code models of communication, at least if preeminent researchers working on this question are on the right track (Gorzelak et al., 2015; Menzel & Giurfa, 2001; Millikan, 1989; Rescorla, 2012; Skyrms, 2010).[38] The alarm call system of vervet monkeys is also a typical example.[39]

Codes can also be much more sophisticated. Formal semantics (compositional, truth-conditional, model-theoretic semantics), if considered as a model of information transmission, falls within the code model. It does so because it analyzes literal linguistic meaning through pre-established associations between stimuli and the pieces of information they carry: the literal meaning of a sentence is accounted for through a generative code which pairs through syntactic and semantic rules pieces of information

[37] As mentioned in the last chapter, I add an asterisk here because Grice defined implicatures as being a subspecies of speaker-meaning.
[38] These researchers do not use the phrase 'code model' but their accounts nevertheless fit the description (for more on why they are code models, see Origgi & Sperber, 2000; Reboul, 2017, sec. 2.4).
[39] It is based on a code pairing three types of alarm calls (different barks) with three types of messages, each corresponding to a predator (snake, eagle, or leopard). The idea is that vervet monkeys master a code that is largely innately determined (Price et al., 2014) and which allows them to encode and decode these predator-related messages.

(usually: what is denoted) and stimuli (usually: words of the fragment analyzed) (Coppock & Champollion, 2020; Heim & Kratzer, 1998).

By the 'prevailing Gricean models' I refer to aspects of Grice's work on meaning (1957, 1968, 1969, 1975, 1989) as well as to both neo-Gricean and post-Gricean theories.[40] The prevailing Gricean models include influential and established theories of various sorts (Bach & Harnish, 1979; Grice, 1989; Horn, 1984; Levinson, 2000; Lewis, 1969; Neale, 1992, 2016; Schiffer, 1972; Searle, 1969; Sperber & Wilson, 1986, 2015; Stalnaker, 1978, 2014; Strawson, 1964a) as well as more recent ones (Carston, 2002; Green, 2007; Moore, 2017; Recanati, 2010; Tomasello, 2008; Wharton, 2009).

All the prevailing Gricean models share the following hypotheses: (a) successful linguistic communication is a matter of what intentions speakers have. (b) These intentions need to be inferred by the relevant audience thanks to assumptions about the pragmatic competence of participants to the conversation, i.e. assumptions that they respect certain pragmatic principles. (c) Doing so allows inferring what is meant by the speaker beyond what is encoded in her utterances; it allows understanding pragmatic meaning beyond semantic meaning.

Contrary to the code model, the *explananda* of the prevailing Gricean models include the transmission of pieces of information that is not based on a pre-established pairing between information and types of stimuli.[41] As we saw in the last chapter, according to a widespread view on the distinction between semantics and pragmatics, code models account for

[40] Post-Griceans usually refuse the label 'Gricean', which they associate with Grice and neo-Griceans. Even if post-Griceans went in another, more psychologically-oriented, direction than neo-Griceans (who have followed Grice more closely), both branches stem out of Grice's pioneering work. So, despite the important differences between the two kinds of theories, I group both of them under the label 'Gricean models'. I recognize that this is not optimal, but I could not find a better expression. Sperber and Wilson (1986) use the general label 'inferential models' instead. However, this label is confusing since communicative phenomena accountable by the code models can be inferential as well, because inferences are so pervasive in animal cognition, as has been acknowledged by Sperber (Mercier & Sperber, 2017, Chapter 2). More recently, Sperber (on his blog) used the label 'mentalistic models' instead of 'inferential models', which could have been an option for us instead of 'Gricean models'. However, it is not clear that explanations based on the code models never are 'mentalistic'. This term also possesses a non-naturalistic flavor which I would rather avoid.

[41] A typical example is conversational implicature: when the infamous professor writes a recommendation letter about a pupil which only states 'Dear sir, Mr. X's command of English is excellent, and his attendance at tutorials has been regular. Yours, etc.' (Grice, 1989, p. 33), the messages sent go beyond those accountable by semantics, because it uses words in a way that is not predicted by the semantic code. To account for such messages, i.e. those that resist code models, Gricean models postulate pragmatic competencies, which include mindreading abilities based on pragmatic principles, such as Grice's Cooperative Principle and its four maxims (1975) (see Chapter 1 and §2.4 below).

semantic phenomena and the prevailing Gricean models account for pragmatic phenomena (Korta & Perry, 2020, sec. 3; Schlenker, 2016).

You will remember from the last chapter that the prevailing Gricean models are designed to account for speaker-meaning or ostensive-inferential communication and that the latter is defined with the following intentions, or variants thereof (see the Appendix for a discussion of these variants):

(i) Intention 1: to produce a stimulus which generates an effect in the receiver or makes something manifest, and
(ii) Intention 2: that the stimulus makes Intention 1 mutually manifest (or, as I prefer to put it, *mutually recognizable*) to sender and receiver.

Stimuli which are produced without the Intentions 1 and 2 – stimuli which are not overtly intended for communication – fall outside of the prevailing Gricean models' explanatory scope.

In the last chapter, I have argued that the standard picture of information transmission has a blind spot because the prevailing Gricean models cannot apply to all the cases left out by the code models. I illustrated this claim by presenting two cases where, on the one hand, a piece of laughter is not produced with the Intentions 1 and 2 (or their variants), but where, on the other hand, information which cannot be analyzed through a code is nevertheless transmitted. This blind spot – information transmitted which is accountable neither by the prevailing Gricean nor by the code models – is not restricted to laughter, far from it, as we will see especially in the next chapter.

In this chapter, to solve this problem and shed light on this blind spot, I will lay the first stones of the Extended Gricean model (EGM for short). In the next chapter, we will apply the EGM to the laughter examples of the last chapter as well as to other types of stimuli (other affective signs, non-affective non-verbal behaviors, clothing, and more). In all these cases and more, the EGM can give an analysis of what information is transmitted, and why it is rational to infer that this information is transmitted: an analysis that is available to neither the code nor the prevailing Gricean models. The EGM nevertheless deals with these cases in a way very similar to that of the prevailing Gricean models: by postulating mindreading processes based on a common background and pragmatic principles. But the mindreading processes and the principles in question are larger in their scope: they apply beyond speaker-meaning to what I call *allower-meaning*, a notion that I will present in §2.2. Correspondingly, it applies to stimuli

beyond those overtly intended for communication to stimuli with Effects Mutually Recognizable As Controllable (stimuli with EMRAC) a notion I introduce in §2.3.

Two of the main goals of the EGM thus are the following.

> Goal one: to account for information transfer that cannot be explained by the code models – this is the wished-for-virtue of all Gricean models.
>
> Goal two: to do so without requiring that stimuli are overtly intended for communication: contrary to the prevailing Gricean models, the sender shouldn't need to produce a signal with the Intentions 1 and 2 (or their variants), and in fact need not have any communicative intentions.

Here is another way to formulate the two goals: the EGM is to account for implicatures* that are not subspecies of speaker-meaning nor of ostensive-inferential communication (OIC for short). I add a '*' to 'implicature' because, for Grice (1989, pt. 1) and his heirs (see e.g. Horn, 1984; Levinson, 2000; Neale, 1992), implicatures are a subspecies of speaker-meaning and, for Relevance theorists, it is a subspecies of OIC.[42]

Fig. 2.1 illustrates the typology of meaning relevant to this chapter. The code models can only account for messages that are encoded. The prevailing Gricean model can only account for speaker-meaning (or OIC). Speaker-meaning includes both 'what is said', which is encoded, and 'what is implicated', which is not encoded (Grice, 1968; Levinson, 2000, p. 13; Neale, 1992). The blind spot of the standard picture of information transmission are the implicated* messages that are not speaker-meant (in bold). The EGM aims to account for them, and in fact for all types of allower-meaning, a notion to which we will now turn.

[42] Not everyone agrees on this point, however. For instance, Green (2019a) argues that some information transmission which has been classified as implicatures actually do not belong to speaker-meaning, fall outside the scope of the prevailing Gricean models, and should instead be accounted for through biological models of communication. See also Schlenker et al. (2016) for the claim that monkeys which lack the cognitive abilities for speaker-meaning may nevertheless communicate non-coded information and so make implicatures*. Both of these research programs are very much compatible with mine and are complementary. Nevertheless, as far as I can tell, neither would be able to account for the cases accounted for by the EGM that I present below and in the next chapter.

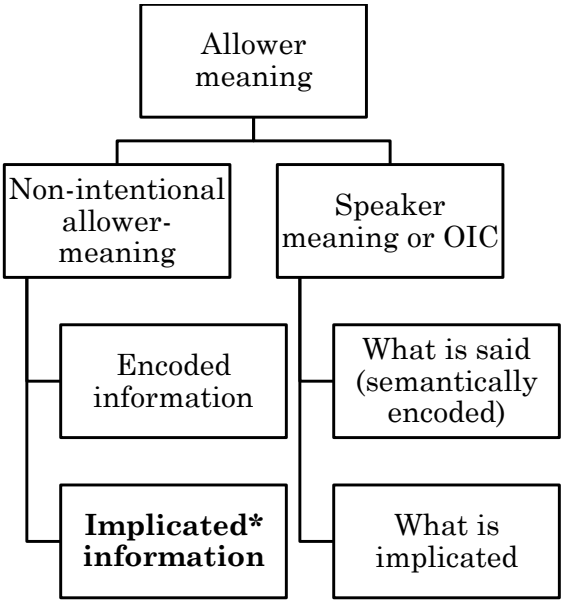

Fig. 2.1. The Extended Gricean Model (EGM) aims to account for all cases of allower-meaning, including the messages that are implicated* without being speaker-meant (in bold). These are cases that cannot be accounted for by the code models and the prevailing Gricean models. See below for definitions of 'allower-meaning' and 'non-intentional allower-meaning'. 'OIC' stands for ostensive-inferential communication and is treated as equivalent to speaker-meaning here (see the Appendix for a discussion of these notions).

2.2. ALLOWER-MEANING

The EGM preserves all elements of prevailing Gricean models besides the ones that derive from the definition of speaker-meaning or of OIC because the latter is replaced with the notion of *allower-meaning*, of which speaker-meaning and OIC are species.

I will define allower-meaning based on my favored way of defining speaker-meaning, but will also give alternative definitions based on other definitions of speaker-meaning and OIC (see The Appendix for their respective strengths and weaknesses of the original definitions). I will consider all of them as equivalent here and will also put aside the difference between speaker-meaning and OIC, to which I will both refer to as 'speaker-meaning'.

Allower meaning – definition:
A sender S allows x to mean something to the receiver R (or the appropriate conditional receiver R) if, and only if,
S produces x while:
(i) S allows x to generate effects e in R, and
(ii) S allows x to make (i) mutually recognizable for R and S.

Variant 1, Grice style (1957, 1989, p. 99):
S allows x to mean something to R if, and only if,
S produces x allowing:
(i') x to generate in R some effects e,
(ii') x to recognize that S allows (i'), and
(iii') R's recognition that S allows (i') to function, in part, as a reason for (i').

Variant 2, Neale style (1992), based on Grice (1982), see also Moore (2018, p. 4):
S allows x to mean something to R if, and only if,
S produces x while
(i") allowing x to generate in R some effect e, and
(ii") R to recognize that S allows (i"),
(iii") and not intending that R should be deceived about (i") and (ii").

Variant 3, Sperber and Wilson style (1986, 2015):
S allows x to mean something to R if, and only if,
S produces x while
(i'") S allows x to make manifest or more manifest to R an array of propositions I, and
(ii'") S allows x to make it mutually manifest to R and S that (i'").

Variant 4, Green style (2007, p. 66ff):
S allows x to mean something to R if, and only if,
S produces x allowing
(i"") x to make something manifest, and
(ii"") to make it manifest that S allows (i"").

In all of them, roughly, 'intending' in the original definition is replaced by 'allowing'. This might sound like an insignificant modification, but the consequences are vast and far-reaching. This small change enables the EGM to account for many more phenomena than the prevailing Gricean models while being restricted to a domain where central Gricean insights still apply. Before I can show how, I need to present the tools I will use. First, let us turn to the term 'allowing'.

2.2.1. Allowing requires control

Allowing, as I use the term, implies a form of control over what is allowed. If S allows x to have the effects e, then S possesses control over the

production of e. S is, in some sense, *free* to produce e. This condition is important for what follows, so we need to discuss it in some detail.

Following Fischer and Ravizza (1998), we may distinguish two kinds of control: reactive-control and guidance-control. S possesses *reactive-control* over her course of action A insofar as S could have done otherwise, in the sense that 'there was a time at which [S's] doing otherwise was in [S's] power'. (Chisholm, 1967, p. 417). In other words, if S has reactive-control over A, then S may freely select alternative possibilities to A. By contrast, S possesses *guidance-control* over A, if the process leading to A is S's own, *reasons-responsive* mechanism. Guidance-control does not require a possibility to do otherwise, it does not require reactive control, but reactive-control requires guidance-control. Reactive-control thus involves more conditions than guidance-control, but we will only need to rely on the later, less demanding, notion.[43]

Let me illustrate the distinction with an example from Fischer and Ravizza (1998, p. 29) inspired by the famous Frankfurt-type cases (Frankfurt, 1969), i.e. the type of cases which show that reactive-control is not necessary for (moral) responsibility.

Sam is bad and Jack is no better. Sam has planned to shoot the mayor and has told his plan to Jack, who is very pleased with it. But Jack worries that Sam will waver and so has placed in Sam's brain a chip which gives Jack the ability to monitor Sam's behavior. If Sam were to give up on his plan, Jack would activate the chip and force Sam to keep up with it. Sam however does not waver and methodically acts as he planned. Jack and his device thus play absolutely no role. Sam did not possess a reactive-control over the course of his actions – he could not have acted otherwise – but he freely and masterfully guided them through his own choices and deliberations: he acted on the basis of mental mechanisms that are perfectly reasons-responsive. He possessed guidance-control although he had no reactive-control. Note by the way that Sam is to be held responsible for his actions even though he could not have acted otherwise.

Here are, by contrast, some situations where, according to Fischer and Ravizza, we lack guidance-control (1998, p. 40ff): Joe has been hypnotized and thus forced to punch the first person he meets. An evil person has

[43] This is good news because defining reactive-control – the possibility to do otherwise – is a notoriously difficult task (for a concise review of the debate, see O'Connor & Franklin, 2020, sec. 2). And, of course, it is widely debated whether it is metaphysically possible at all, since it may be incompatible with determinism (the idea that the future is dependent upon the present such that, given the present, only one possible future exists). By contrast, the compatibility of guidance-control and determinism is unproblematic.

wired Smith's TV set to subject him to powerful subliminal advertising which causes Smith to murder his neighbor. Both Joe and Smith lack control over their actions in a way in which Sam didn't. And note that neither Joe nor Smith should be held responsible for their actions. Other situations where one lacks guidance-control involve powerful forms of coercion, potent drugs, manipulation of the brain, brain lesions, or mental disorders.

According to Fischer and Ravizza, what unites the cases where one possesses guidance-control is that the kind of mental mechanism that leads to the action is *reasons-responsive* and is the agent's own mechanism. Typically, Sam deliberately chooses to do what he does. By contrast, the kind of mechanisms that operate in cases where one lacks guidance-control are not reasons-responsive or are not the agent's own. Whatever reasons Joe and Smith would be given, they would not have responded to them: the mental mechanisms that led to their actions were not reasons-responsive because of the hypnosis and the subliminal manipulation.

More precisely, here is how they define the reasons-responsiveness which defines guidance-control:

> Reasons-responsiveness[44]
>
> « A mechanism of kind K is … responsive to reasons to the extent that, holding fixed the operation of a K-type mechanism, the agent would recognize reasons … in such a way as to give rise to an understandable pattern (from the viewpoint of a third party who understands the agent's values and beliefs), and would react to at least one sufficient reason to do otherwise (in some possible scenario). » (Fischer & Ravizza, 2000, p. 444)

By 'mechanism' they mean the process that leads to the action, the way the action comes about and they give examples such as: doing A after having deliberately chosen to do A, doing A as a result of a stroke, as a result of an irresistible urge to take a drug, or as a result of a brain manipulation (e.g. with Sam's chip). The mechanism kind which needs to be held fixed across

[44] Actually, this is the definition of *moderate* reasons-responsiveness, which they contrast to strong and weak kinds, but I will ignore the distinction. Also, I have deleted the part of their definition which requires that some of the reasons must be *moral* reasons. They add this condition because they are interested in defining moral responsibility. By contrast, I am interested in the responsibility we have over what information we transmit, which may or may not be moral.

scenarios is the one that appears as 'most relevant to our ascription of responsibility'.[45]

Let us illustrate this with Sam's case. The action we assess is his killing the mayor. The kind of mechanism that issues in this action is his own deliberate decision. There is a possible world where, holding fixed this kind of mechanism, Sam would recognize a reason as sufficient for acting otherwise. Let us say: a world where Sam learns that if he kills the mayor, his entire family will be tortured to death and so where he acts otherwise on the basis of this reason (in this possible world, Jack does not implement a chip in Sam's brain so that the alternative mechanism won't stop Sam from acting otherwise). This reason to act otherwise is understandable to someone acquainted with Sam's values – in particular the value his family has for him – and his beliefs – in particular, the belief that his killing the mayor will result in his family being tortured. So the mechanism kind that issues in the action is reasons-responsive.

By contrast, the kind of mechanism that would have led Sam to kill the mayor in the alternative Frankfurtian scenario where Jack would have used the brain-chip is not Sam's own, reasons-responsive mechanism. The kind of mechanism relevant to the ascription of responsibility in this alternative scenario is something like 'a device implemented in a third person's brain to manipulate her'. There is no possible world where, holding fixed this kind of mechanism, Sam would recognize a reason as sufficient for acting otherwise. The chip manipulates Sam's brain in such a way that, even if he were told that his family would be tortured, he would still kill the mayor. This is true for any sort of reason given to Sam, however strong it is.

[45] If this is the most fundamental individuation criterion, then their theory of responsibility is circular, but, even in that case, I would consider the circle to be sufficiently wide and informative for the circularity to be unproblematic. While admitting that their theory cannot specify in a general way how to determine what mechanism is 'the' mechanism to consider, they draw an illuminating comparison with ethical theories: « It is simply a presupposition of this theory as presented here that for each act, there is an intuitively natural mechanism that is appropriately selected as the mechanism that issues in action, for the purposes of assessing guidance control and moral responsibility. The problem here is, of course, similar to that of 'generalization' theories in ethics. On such an approach, an act is (say) wrong if there would be (for example) certain bad consequences of actions of *that type* generally being done … On these approaches, it is assumed that there is some natural, unproblematic way of selecting the relevant general 'type' by reference to which the act is to be assessed. A similar assumption lies behind [our notion of mechanism]. » (1998, 47)
What exactly they mean by a 'natural, unproblematic' way of selecting the kind of mechanism is vague and is certainly a weakness of their theory. However, the concept of guidance-control is sufficiently powerful and useful for us to considerably benefit from it despite this weakness.

This is why, according to Fischer and Ravizza, if Sam acts on the basis of his deliberation, he possesses guidance-control and that, if he acts on the basis of the chip in his brain controlled by Jack, Sam lacks guidance-control.

To capture the condition according to which reasons-responsiveness must 'give rise to an understandable pattern', we can contrast Sam with Sam*. Sam* is a strange person for whom the *only* reason to act otherwise would be that the mayor's wife is in the same room as a red-nosed clown. Let us suppose that, even from the viewpoint of a third party who understands Sam*'s values and beliefs, the fact that he may act otherwise *solely* on the basis of this specific reason does not give rise to an understandable pattern. Sam*'s deliberation is reasons-responsive, but not in the appropriate way, and so lacks guidance-control (we may imagine that he suffers from a strange mental disorder).

We now have the tools needed to capture the kind of control that is implied by 'allowing', as I use the term.

> Allowing and guidance-control
>
> If S allows x to generate effects e in the audience R, then S exhibits guidance-control over the production of e, which means that the kind of mechanism that actually issues in S's allowing x to generate e in R is S's own and is reasons-responsive.

Let us illustrate. My friend makes a remark and I have an irresistible impulse to laugh. As a result, my friend thinks 'Constant thought my remark was funny'. Because I know my friend well, I am capable of knowing that my laughter would generate this belief unless I tell him that, despite my laughter, I don't find the remark funny. However, I don't think it is problematic that he forms this belief. For me, it is perfectly okay that he thinks that I find his remark funny. So I don't produce any such excuse and, in fact, it doesn't even cross my mind to produce such an excuse. I just candidly laugh and don't add anything, remaining silent until my friend makes another remark.

Now, we can only say that I *allow* the laughter and the silence afterward to generate the belief in my friend if I have guidance-control over my allowing this belief. I have this control if there is a possible scenario where, holding fixed the kind of mechanism that issues in the production of my friend's belief, I would have acted otherwise on the basis of a reason and in a way that can be understandable from a third-person perspective.

Let us say that the following is such a possible scenario: if I had considered that my friend would be deeply hurt by the belief that I find his remark funny, this would have constituted a sufficient reason for me to tell him that, despite my laughter, I didn't find the remark funny, so as to avoid producing the belief in question.

The mechanism kind that is held fixed between the two scenarios is the one that leads me not to produce an excuse in the actual scenario and that leads me to produce one in the alternative scenario. It is the mechanism that is most relevant to our ascription of responsibility. It certainly includes my capacity to infer my friend's belief and my capacity to produce an excuse that I know would change what my friend thinks. In both scenarios, I possess these capacities and the same kind of mechanism is in place. And such a mechanism is reasons-responsive because, in the alternative scenario, I acted otherwise based on a reason. Thus, in the actual scenario, I possess the guidance-control necessary for *allowing* my behavior to produce the belief in my friend that I find his remark funny.

Let us observe that in this example, I allowed my laughter *and the silence afterward* to generate my friend's belief. Stimuli which we allow to have effects thus sometimes are sets of stimuli ordered in a temporal sequence.

Let us also observe that the silence afterward constitutes an omission: the failure of having produced an excuse. Accordingly, the relevant mechanism is the one that led to whatever I did at the time where I *could have instead* given the excuse, if I had a sufficient reason to do so.

Below, I will sometimes say that something was done 'freely' or 'with control' but let us keep in mind that by that I mean 'with guidance-control'.

2.2.2. The control must be manifest

Here is a further condition on allowing: If S allows x to have the effects e by producing a set of stimuli <x1, …, xn> between t0 and t1, then, given S's background knowledge and mental capacities between t0 and t1, S's control over e must be manifest to her. This means that, holding fixed S's background knowledge and mental capacities, there is a possible scenario where S finds out between t0 and t1 that her producing x may have the effects in question, but that she possesses control over them. Thus, allowing is a *de se* attitude (Lewis, 1979a) in the sense that S must be able to consider *her own* behavior as having the relevant consequences. This creates an *opaque context* for what is allowed (Quine, 1960, sec. 30).

By 'manifest', I mean the following (adapted from Sperber & Wilson, 1986, Chapter 1):[46] a mental content is *manifest* to a subject S at time t if, given S's mental capacities and knowledge at t, we can reasonably expect S to be capable of having a conscious mental state about that content, whether it is through perception, inference, emotion, imagination, or another type of psychological mode (by 'at t', I mean the laps of time relevant to determine whether S has the mental capacities and knowledge to perceive, infer, imagine, etc. the content). For instance, the thought that Snoop Dogg never smoked with Jesus surely was manifest to me yesterday at 5 p.m., even though I have never consciously entertained this thought before just now. It was manifest because, given the mental capacities and knowledge I had yesterday at 5 p.m., I was surely capable of inferring that Snoop Dogg never smoked with Jesus and so you could have reasonably expected me to be capable of consciously entertaining this thought at this moment. By contrast, that Snoop Dogg's real name is Cordozar Calvin Broadus Jr was not manifest to me yesterday at 5 p.m. I looked it up just now and didn't know it before. So, given my knowledge yesterday at 5, one couldn't have reasonably expected me to have a conscious state about this content.[47]

Instead of 'being manifest', we may say that something *is apprehensible* by S where 'to apprehend' is understood as an umbrella term regrouping all conscious mental states with an informational content (more on this below). A content may be strongly or weakly manifest to S at t depending on whether it is respectively easy or difficult for S to apprehend it at t. Note that something may be manifest to S without S having apprehended it: it must only be apprehens*ible*.[48]

Let us take an example of how control can be manifest or not. Joe does not know that you have an imaginary friend. If Joe's behavior makes you think about your imaginary friend, even if Joe displays guidance-control, and even if Joe knows his behavior is under control, Joe does *not* allow his behavior to generate this thought in you. This is because, given his

[46] See the Appendix for a more detailed discussion of manifestness.
[47] There is some possible world where we hold fixed the mental capacities and knowledge I had yesterday at 5 and where I have a random thought that Snoop Dogg's real name is 'Cordozar Calvin Broadus Jr', but we couldn't reasonably expect me to have this random thought. My definition is more constrained than Sperber and Wilson's in this sense, although it is also less constrained in another sense because, according to their definition, something can only be manifest if it is represented as true or probably true (1986, p. 39).
[48] What is the difference between 'manifest' and 'recognizable'? The latter is cognitively more demanding: you may see or hear something without recognizing it (c.f. Dretske, 1995, Chapter 1) and so something may be manifest (apprehensible) without being recognizable. The way I use 'recognizable' is supposed to match that of the relevant Gricean literature (see e.g. the definitions of speaker-meaning above and in the Appendix).

background knowledge, it is not manifest to him that his behavior would produce this thought and that he possessed control over it.

By contrast, if Joe knows about your imaginary friend and can know that his behavior is reminiscent of it, given that he possesses control over the production of your thought, the control in question is more or less strongly manifest. The strength of the manifestness depends on how easily Joe can apprehend that his behavior produces your thought at this moment. If you have not talked about your imaginary friend for 20 years, the relevant control will be very weakly manifest. But if you have told Joe the other week that some of his ways of behaving remind you of your imaginary friend, then manifestness will be much stronger.

For the control to be manifest to S, it is not necessary that S masters, nor even can master, the concept 'guidance-control'. Rather, S's control may be apprehended by S in any kind of way. For instance, the control may be felt by S, S may possess a know-how of her control (e.g. by mastering a skill), S may have the intuition that she can prevent the effects from happening, S may hope that she controls them, S may be afraid not to have the control, S may have the intention to e (and thus think she has control over e since we only intend what we think is within our control), etc.

The conditions A and B, i.e. control and manifestness, are together sufficient to define 'allowing' as I use the term (I hope!). Let me nevertheless make a few more observations before I give a definition.

2.2.3. ALLOWING DOES NOT IMPLY INTENDING

Although allowing implies a form of control and that this control is manifest to S, allowing does not imply intending: if S allows x to F, it does not follow that S intends for s to F. We have already reviewed in the last chapter some of the features of intentions that explain why not all controllable behavior is intentional. Let me go back to some of these features to see how allowing differs from intending.

Firstly, following influential accounts by Bratman (1987) and Mele (1992), it is usually agreed that the content of intentions is action plans. However, if S allows x to F, it does not follow that it was S's plan to F. This is obvious in certain omissions. If I am supposed to water my neighbors' plants while they are on holidays, sincerely intend to do so, am free to do so in the sense of having the guidance-control necessary for doing so, but that I let them die due to my negligence, I have allowed my behavior to kill these plants even if that was not my plan. On the contrary: my plan was that they do not die. It seems clear in this case that I may allow something to happen

without intending it to happen because it was not part of my plans. Note that this is true even if I have previously considered the possibility that I may, unfortunately, let the plants die. The fact that I have considered the possibility of my omission does not imply that my omission was intentional.

Relatedly, it is also usually held that F-ing intentionally requires having a reason to F (Anscombe, 1957; Bratman, 1987; Mele, 1992; O'Shaughnessy, 2008; Searle, 1983). But it is not the case that allowing x to F requires having a reason to F. For instance, by humming a tune, I may generate in you the belief that I know the first notes of Beethoven's 14th quartet and thus allow my humming to generate this thought in you, but I may do so without having any reason to generate this belief.

Thirdly, even for philosophers who do not think that every type of intention requires reasons or plans, intentions nevertheless are considered as having the function of guiding one's behavior (Pacherie, 2006; Pacherie & Haggard, 2010). Pacherie, for instance, argues that our motor-intentions, those which guide the movements of our body (e.g. guide my fingers as I grab something) do not require reasons, or plans, because reasons and plans are cognitively too demanding for motor-intentions. Nevertheless, motor-intentions must serve as guides to our behavior. That is not true for allowing: I may allow my humming to generate a belief in you without having any mental states guiding the generation of your belief. That my behavior may produce your belief may be weakly manifest to me in the sense that it is inferable even if I have not thought about it at all. This consequence is potentially inferable but not actually inferred. None of my mental states have this consequence as content and so none of my mental states guide my behavior toward this consequence.

Furthermore, for Pacherie, motor-intentions (like all intentions) must phenomenologically appear to me as guiding my behavior but, clearly, the phenomenology of guidance is entirely absent from certain cases of allowing such as my allowing (e.g. in the humming and the plants examples).

Let me give a last example which, I believe, intuitively illustrates why allowing does not imply intending: we normally sneeze not because we form the intention to sneeze (is it even possible to intentionally sneeze?), but because we have a reflex-like reaction to dust or light or something. Now, we usually have control over our sneeze in the sense that, were we to have a reason strong enough not to sneeze, we would not sneeze. This control, furthermore, may well be manifest. Now, let us say if I sneeze, this will produce a sound that my dog hears. This is knowable to me. I neither

intend nor want to avoid this consequence: I just don't care about it and don't even think about it. In such circumstances, because the effects of my sneeze are manifestly controllable, I *allow* my sneeze to produce a sound that my dog will hear. But I don't intend my sneeze to have this consequence. Instead of 'allowing my sneeze to produce an effect', we could also say in such cases that I unintentionally *let* my sneeze produce an effect.

Although allowing does not imply intending, as I use the term, successful intending implies allowing: if S F-s intentionally with x, then S thereby allows x to F. If I intentionally move the bottle with my gesture, then I allow my gesture to move the bottle. This is so because doing something intentionally requires the guidance-control as well as the manifestness of the control that define allowing.

2.2.4. WHAT IS ALLOWED ARE THE EFFECTS

If it was not clear already, what is allowed in our definition of allower-meaning are the *effects* of the stimuli, not the stimuli themselves. So, for instance, if we want to know what S has allowed her laughter to mean, the most relevant question is not whether S could have refrained from laughing, but rather whether S possessed guidance-control over the effects of her laughter that are manifest to her. Now, of course, if I know that stimuli x can have certain effects and that I can refrain from producing x, then I can thereby refrain from producing these effects. So, for instance, if I laugh knowing that I could have prevented myself from laughing and thereby prevented you from thinking that I laughed, I both allow myself to produce the stimulus (the laughter) and, by the same token, allow the stimulus to produce the effects (your belief that I laughed). It is the effects that interest us.

2.2.5. ALLOWING IS NOT COGNITIVELY DEMANDING

Producing stimuli while allowing potential effects doesn't require that the sender actually can entertain the thought 'x might generate effects e in R' in the sense that S doesn't need to be capable of entertaining the concepts making up this thought (GENERATE, EFFECTS, or RECEIVERS). Instead, what is required is a *capacity* to use the stimuli so as to allow or not to allow them to produce effects.

Take the following anecdote as an example (Perry & Manson, 2009, p. 47). A capuchin monkey, chased by a group of aggressive conspecifics, emits a snake alarm call, knowing perfectly well that there is no snake around, but so that the other monkeys stop threatening him and instead focus their

attention on the snake (which doesn't exist).[49] One may reasonably agree, on the one hand, that this monkey can intend his call to produce effects on his conspecifics while, on the other hand, disagreeing that the monkey master the concepts EFFECTS or RECEIVERS because one considers that such concepts cannot be mastered by capuchin monkeys. The monkey can be understood as having a capacity, perhaps a know-how, which makes him able to induce fear in the other monkeys with his call and to intend it to do so, even if he is not able to entertain the structured proposition that 'If I produce a snake alarm call, this will scare the other monkeys'.

Now, if we agree that the monkey can intend his alarm call to produce effects on others, then, because intending implies allowing, we must agree that the monkey also *can allow* his alarm call to produce these effects.

Because allowing requires guidance-control and that the latter implies reasons-responsiveness, it follows that the monkey's intention to scare the other monkeys must be based on a mechanism that is reasons-responsive. Many philosophers (e.g. Davidson, McDowell, Brandom) would not agree that capuchin monkeys can recognize and act on the basis of reasons, in a certain sense of 'reasons'.[50] However, the way I use the term 'reason' is not as demanding as that of these authors. It rather should be understood as belonging to the framework which Dretske calls 'minimal rationality' (1988, 2006). As such, many non-human animals act on the basis of reasons. We may postulate for instance that the monkey acts on a belief or belief-like state that he will avoid harm by emitting the snake alarm call (and on his desire or desire-like state to avoid harm). This belief (or the belief-desire pair) constitutes a reason for him to act as he does. Another reasons-ascribing explanation would be to say that he acts on the basis of his fear and that this fear is based on a reason: he may very well be harmed if he does not act swiftly to avoid the other monkeys, and this constitutes his reason to be afraid.

So, the monkey's behavior was, I take it, reasons-responsive. Furthermore, he certainly can be taken to know (to possess the know-how relevant to) what effects his behavior would generate in his audience. There is no

[49] Similar behaviors have been reported with baboons (R. Dunbar, 1996, pp. 23–24), vervet monkeys (R. Dunbar, 1996, p. 101), and chimpanzees (Sievers & Gruber, 2016, p. 765).
[50] Note, as I already mentioned in a preceding footnote, that Fischer and Ravizza's definition of 'moderate reasons-responsiveness', unlike my adaptation, is restricted to morality: they require that at least some of the reasons defining guidance-control are *moral* reasons. I do not want to be so restricted. Consequently, it is normal that my capuchin monkey example would not be considered as possessing guidance-control for Fischer and Ravizza although it does for me.

problem in ascribing allowing to capuchin monkeys. This illustrates that allowing is not cognitively demanding.

2.2.6. DEFINING ALLOWING

Taking these remarks into account, here is my proposed definition:

S allows x to F – definition

A sender S allows the stimuli x – made of individual stimulus <x1, x2, …, xn>, produced by S between t0 and t1 – to generate the effect e (doxastic, affective, evaluative, behavioral, …) on the actual or conditional audience R if, and only if,

(a) S had guidance-control over the production of e between t0 and t1,[51] and
(b) It was manifest to S between t0 and t1 that S may generate e in R with x.

We may add: Furthermore, S didn't need to intend x to generate e in R, or to entertain the thought 'x might generate effects in R', or to possess the concepts GENERATE, EFFECTS, or RECEIVERS, but merely to have a capacity, a know-how, for (a) and (b).

Let us also observe that, as the examples show, the 'x' comprise a wide range of stimuli: intentional actions, reflex-like reactions (e.g. sneezes), omissions, emotional expressions (frowns, sighs, laughter, growls, shouts, etc.), etc.

The definition of allower-meaning is roughly a definition of speaker-meaning where 'intending' is replaced with 'allowing'. This implies that, contrary to the stimuli carrying speaker-meaning, those carrying allower-meaning don't need to be produced with the Intentions 1 and 2. Thus, allower-meaning doesn't require the production of stimulus overtly

[51] For those who believe that we can make the notion of reactive-control work better than the notion of guidance-control (e.g. libertarians about free will), we could replace the first condition with the following: S had the power within herself between t0 and t1 to have intentionally prevented x from generating e in R (either by refraining from producing x or by producing another set of stimuli y between t0 and t1 which would have canceled e), but S did not behave in this way. Under this interpretation, allower-meaning requires that the sender could have had certain intentions that she did not have. Actually, this was my first attempt at a definition and in a previous version of this chapter, I had the following quote as epigraph: '… what was really required in a full account of speaker-meaning was *the absence* of a certain kind of intention.' (Grice, 1989: 303). (However, Deirdre Wilson told me that I thus allowed this epigraph to mean that Grice was engaging in the same project as mine at the end of his life (the quote is from the 'Meaning revisited') although that is not what he speaker-meant with this utterance.)

intended for communication. This is one of the two goals we set for the EGM, the other being that it can nevertheless account for implicatures* (see §2.1 for the notion of implicature*). We have reached the first goal by stipulating what defines allower-meaning. What will be important of course is that, thanks to this notion, we can reach the second goal. But before I can do that, I need to make a few more points. Let us now turn to the kind of stimuli defining the scope of the EGM.

2.3. Stimuli with EMRAC

Although allower-meaning doesn't require stimuli overtly intended for communication, not all stimuli can carry allower-meaning.

As a reminder, here are the two clauses defining allower-meaning:

> (i) S allows x to generate effects e in R (often: to make something manifest to R), and
>
> (ii) S allows x to make (i) mutually recognizable for R and S.

These clauses require S to produce a specific type of stimuli: stimuli with *Effects* on the (potential) receivers that are *Mutually Recognizable As Controllable*. I will call them *stimuli with EMRAC*, for short.

Crucially, as we will see, if S produces a stimulus with EMRAC, even though it is not an overtly intentional stimulus, R can nevertheless take S to be subject to certain pragmatic principles resembling Grice's Cooperative Principle. This is what gives R reasons to interpret the stimulus beyond what it encodes and to consider that this information is now part of S and R's common background. This is in turn what allows R to be rational in going through Gricean Derivations (Dänzer, 2020) and to form hypotheses about messages that are not encoded, that are implicated* even though they are not speaker-meant. As we will see below and in the next chapter, these consequences of the production of stimuli with EMRAC and how they are accounted for within the EGM is what allows this model to fulfill our second goal: to account for the transfer of information beyond what is encoded in the stimuli and what is speaker meant.

First, let me detail why clauses (i) and (ii) imply that stimuli carrying allower-meaning are restricted to stimuli whose effects on the (potential) receivers are mutually recognizable as controllable and what this means exactly.

2.3.1. MUTUALLY RECOGNIZABLE AS CONTROLLABLE

The effects must be controllable because clause (i) requires that S allows x to generate e in R and, as we have seen, allowing implies guidance-control. So, the relevant kind of mechanism leading to the production of e must be such that there is a possible scenario where, holding fixed the mechanism kind, S acts so as to not produce e in R with x because of an understandable reason. In this scenario, S could either refrain from producing the stimulus (e.g. not laughing) or create a further stimulus (e.g. an excuse) which would cancel the relevant effects that the first stimulus would otherwise have.

Let us now see what *mutually recognizable* in (ii) means. This phrase refers to a rather intuitive phenomenon, but one which is particularly hard to define. Philosophers have referred to it with the following labels: *overtness* (Green, 2007; Strawson, 1964a), *common knowledge* (Lewis, 1969), *mutual knowledge* (Schiffer, 1972), *mutual manifestness* (Sperber & Wilson, 1986), *collective intentionality* (Searle, 1995), *joint attention* (Campbell, 2005; Peacocke, 2005; Seemann, 2011), *participatory sense-making* (De Jaegher & Di Paolo, 2007), *shared intentionality* (Tomasello, 2008), and more. I prefer to use the fresh expression 'mutually recognizable' to avoid being committed to the definitions of these terms.[52]

Although no consensus exists on how to define what I call 'mutually recognizable', the examples are largely agreed upon and shall suffice for present purposes. Here is one: If you and I are watching television together, what happens on the screen becomes mutually recognizable. This is different than if we just happen to watch the same channel from our own individual houses. In the latter case, we both have the same experience of what happens on the screen, but without a shared recognition, without a mutual awareness of, or a joint responsiveness to these stimuli. Note also that to be mutually recognizable, a stimulus need not be mutually recognized. So, even if we never think about the fact that we both know that we both know that we are watching TV together, this is mutually recognizable, because we can mutually recognize it.

Here are some other examples: If we are sitting at a dinner table together, even though we don't look at the vase on the table in open view, the fact that it is there is mutually recognizable. If we are speaking English together, that we can speak English is mutually recognizable. If we are

[52] I give my reasons in the Appendix. Let me note however that, among this list, I consider that the most successful candidates at defining the phenomena in question are Lewis's common knowledge (1969), Sperber and Wilson's 'mutual manifestness' (1986: 39-43), and Green's 'overtness' (2007: 66ff).

taking a walk together, it is mutually recognizable that we are doing so, but it wouldn't be the case if we happened to perform the same movements by chance, without taking a walk *together*.[53]

Note that something may be mutually recognizable by A and B without being recognizable by C. Being mutually recognizable does not mean that anybody can recognize it; it does not imply anything like universal recognizability.

We can contrast the production of a stimulus that makes something mutually recognizable with a case where S has produced a stimulus which she doesn't know can have effects on a certain audience. For instance, S might think that all her work colleagues are now gone from the office and so that nobody can hear her sing. Now, even if one of her colleagues is still in the office, S's singing cannot be said to have mutually recognizable effects on her and her colleague, because she thinks she is alone.

We can also contrast a stimulus with mutually recognizable effects with *covert* stimuli, i.e. stimuli whose production is intended to be hidden. Take for instance Grice's (1957) famous case where A leaves B's handkerchief on the crime scene to induce the belief in the detective that B committed the crime. Although A respects clause (i) by allowing the handkerchief to generate certain effects in the detective's (namely: the belief that B committed the crime), A doesn't respect condition (ii) as A doesn't allow the handkerchief to make it mutually recognized that (i). Indeed, A precisely wants to hide from the detective the fact that she aims to produce the belief in question. Thus, putting the handkerchief on the crime scene is not a stimulus with EMRAC and is not a case of allower-meaning.

[53] The question of where in human ontogeny and phylogeny the emergence of this cognitive capacity lies is disputed. However, I have not seen any strong reason to refuse to attribute it to other social species, for instance to certain (or perhaps all) primates and canines. On the contrary, I think it would explain many communicative phenomena among nonhuman species. I take it that there is plenty of evidence concerning great apes, baboons, vervet monkeys, capuchin monkeys, and more (R. Dunbar, 1996; Perry & Manson, 2009; Sievers & Gruber, 2016). Closer to us, when I play with a dog, or when a dog begs me to open a door or fill up a plate, I certainly have the impression that the toy, the door, or the plate is mutually recognizable, and that this explains the behavior of the dog, such as gazing at me, scratching the door, or making special vocalizations. A skeptic may say this is an anthropocentric illusion, but I am skeptical of this skeptical interpretation. It sounds to me to be rooted in speciest biases.
Concerning ontogeny, there is evidence that at least by 9 months we can detect that something is jointly attended to, and I would say is therefore mutually recognizable (Cleveland & Striano, 2007). Perhaps this capacity emerges in even younger infants and we are as of yet unable to measure it. So, I take the cognitive capacity which allows something to be mutually recognizable to be relatively easy to come by.

Now, in the definition of allower-meaning, what is mutually recognizable is (i). In other words, the fact that the effects in (i) are allowed should itself be mutually recognizable. Because allowing e implies a form of control over e, this means that, to fulfill the definition of allower-meaning, S must produce a stimulus with effects that are *EMRAC*, a *stimulus with EMRAC* for short.[54]

Contrasting stimuli with EMRAC with the following example may be helpful. Say I very convincingly pretend to have a broken leg. My entire leg is in a (fake but realistic) plaster cast and I sit in a wheelchair. We arrive next to staircases and my behavior thus induces in you the belief that I can't walk up the stairs. Let us say that, given your background knowledge and mental capacities at this moment, you can't infer or otherwise apprehend the fact that I can control your belief. You consider your belief to be caused by stimuli (the cast, the wheelchair) over which I had no control; you don't (and, given the circumstances, can't) think I was free to generate this belief in you. Consequently, although this belief is an effect I can in fact control (I could take off the cast and just walk up), its controllability isn't mutually recognizable insofar as its controllability isn't manifest to you and so is not recognizable. So, your belief is not an EMRAC. If we focus on this belief, my behavior thus doesn't respect clause (ii). Even though I intendedly produced this belief in you, I don't allow my behavior to mean that I can't walk up the stairs to you. I had an intention corresponding to (i), but not an intention corresponding to (ii). (This is similar to Grice's example (1957) where B displays A's handkerchief on the crime scene to induce in the detective the belief that it was A who committed the murder).

2.3.2. SAM'S THROWING OF A POUND

Let me end the introduction of these concepts by highlighting how stimuli with EMRAC are different to overtly intentional signals. I will do so by using an example from Grice (1957: 385).

Sam has a very greedy man in his room – let us call him Scrooge. Sam wants Scrooge to go. To do so, he throws a pound out of the window. Sam doesn't intend Scrooge to recognize his intention to make him go, he just thinks that Scrooge's greediness will make him leave.[55] Does that

[54] Remember though that neither the sender nor the receiver needs to think about clause (i) through the concepts corresponding to the terms I use to discuss (i). That (i) is mutually recognizable may depend on some kind of ability or know-how.

[55] Admittedly, this is a bizarre example. Sam seems to perform a nakedly hostile action and it is strange that he lacks any intention to overtly communicate with Scrooge. Nevertheless, this is how Grice presented the example and I will follow him in construing

constitute a case of speaker-meaning? No, because, as Grice points out (using different terms), the throwing of a pound is not overtly intended for communication. It is not a signal produced with Intentions 1 and 2. The intention to generate an effect in Scrooge is not itself intended to be made mutually recognizable. However, we should construe this as a case of allower-meaning, where Sam's throwing of a pound is a stimulus with EMRAC. Let us see why.

Sam's behavior generates effects in Scrooge that are undoubtedly *controllable*. Indeed, Sam deliberately and freely chooses to throw the pound out of the window, and deliberation is a paradigmatic reasons-responsive kind of mechanism.

Are these controllable effects of his behavior mutually recognizable as such? Well, certainly, if Sam and Scrooge consider each other to possess normal cognitive abilities (e.g. mindreading). If Sam considers Scrooge to be able to understand that his behavior is deliberate, then Sam must consider that Scrooge can recognize that the throwing of a pound was a stimulus that produces controllable effects. If we assume that Scrooge also considers Sam as possessing normal cognitive capacities, then Sam's behavior produces effects that are mutually recognizable as controllable. So Sam's behavior may be interpreted as a stimulus with EMRAC, although it is not interpreted as a stimulus overtly intended for communication.[56] As such, even though Sam does not speaker-mean anything, he nevertheless allows his behavior to mean something to Scrooge. What would that 'something' be? To answer this question, we have to figure out what effects are mutually recognizable as controllable.

One candidate for such effects is that, by behaving as he does, Sam creates the following belief in Scrooge (let us call it 'p'): Sam thinks that there now is a good reason for Scrooge to get out of the room. We may say that Sam allows his behavior to mean that p to Scrooge if the following conditions hold: (i-p) he allows his behavior to make Scrooge believe that p, and (ii-p)

it as such. We can imagine that Sam doesn't intend to be overtly hostile because he is a careless and/or heartless person and does not even think about the hostility of his action and the effects it could have on Sam besides making him go.

[56] Sam's behavior would not be a stimulus with EMRAC if Sam considered Scrooge as cognitively unable to mindread his behavior. In this case, Sam would act toward Scrooge as we act toward, say, an insect (e.g. I open up the window to let a fly out): without thinking that Scrooge can interpret his behavior as intending or as allowing anything. In this case, despite its having controllable effects, the throwing of a pound wouldn't be a stimulus with mutually recognizable effects, since Sam wouldn't consider Scrooge as able to interpret the stimulus as having controllable effects. Under this (strange) interpretation, Sam's behavior wouldn't be a stimulus with EMRAC, Sam wouldn't respect (ii), and so Sam wouldn't allow his behavior to mean anything to Scrooge.

he also allows his behavior to make (i-p) mutually recognizable to him and Scrooge.

We can reasonably take condition (i-p) to be fulfilled for the following reasons. Sam rather obviously intends Scrooge to think that there now is a good reason for him to get out of the room and Scrooge can thus infer from Sam's behavior that p. Sam can know that Scrooge can make this inference but he nevertheless freely acts in such a way that generates the belief that p in Scrooge, although he doesn't intend to do so. He merely allows his behavior to generate this belief, i.e. (i-p) holds.

For condition (ii-p) to be fulfilled, Sam must have a guidance-control over his making (i-p) mutually recognizable and this control must be manifest to him. He certainly has guidance-control over his making (i-p) mutually recognizable because of what we said above about the deliberate nature of his behavior. And there is no reason to think that this control is not manifest to him: he is certainly able to know that the relevant consequences of his behavior are under his control. Thus, Sam allows his behavior to mean to Scrooge that p, i.e. that he thinks there is now a good reason for Scrooge to get out of the room.

Sam allows his behavior to mean other things to Scrooge. For instance, that he doesn't think very highly of Scrooge (let us call this 'q'). Indeed, (i-q) Sam certainly allows his behavior to make Scrooge believe that q – after all, Sam could very well have politely asked Scrooge to get out of the room instead of throwing a pound as one would throw a stick to a dog. Furthermore, (ii-q) Sam certainly allows his behavior to make (i-q) mutually recognizable for him and Scrooge. This is so because, first, the fact that he may generate the belief that q in Scrooge is easily recognizable for both of them and they both know that. Second, there is no reason to think that his making (i-q) mutually recognizable was not free nor that he could not know it.

Therefore, even though Sam cannot be considered to be speaker-meaning anything with his behavior, he can certainly be considered to be allower-meaning that p and q (and more). Below, I will explain in more detail how the interpretation of stimuli with EMRAC works and why it is rational to consider it as updating the common background between senders and receivers. First, let me make a few remarks about what I mean by 'common background'.

2.3.3. COMMON BACKGROUND

By 'common background', I mean the information that senders and receivers are warranted to take as mutually recognizable. This notion is thus very close to Stalnaker's common ground (1978), Lewis' conversational score (1979b), and Sperber and Wilson's mutual cognitive environment (1986). However, I believe that the EGM requires a slightly different notion.

This is because 'common background', as I use the expression here, is a normative rather than a psychological notion. It has to do with what is *warranted* and *reasonable* to assume, not necessarily with what actually happens in the senders' and receivers' head. This is apparent in Sam and Scrooge's example. As I have construed the example, it may well be the case that Sam does not actually assume that his behavior transmits to Scrooge the pieces of information I have discussed. Sam may not have thought about the fact that Scrooge can now assume that q, i.e. that he doesn't think very highly of Scrooge. Nevertheless, I believe that it is now warranted for Scrooge to take this information to be part of their common background and to act accordingly. It would be reasonable for Scrooge to presuppose that Sam and him now both have access to q. For instance, Scrooge would now be warranted to say to Sam something like 'You know, Sam, I have never held you in high esteem either.' or to make other kinds of presuppositions based on the assumption that q is now in their common background. There is no need for Sam to have actually updated his beliefs and for the corresponding neurons to have actually fired. Scrooge would nevertheless be warranted to take their common background to be updated by the information which Sam allows his behavior to mean.

'Common background' then is not identical to what Lewis calls the conversational score (1979b), nor what Stalnaker calls the common ground (2002), nor is it nor is it what Sperber and Wilson call the cognitive environment (1986) because these three notions are defined in terms of what happens in the mind of the participants to the conversation. They are defined through psychological variables only. Indeed, conversational score is 'by definition, whatever the *mental* scoreboards say it is' (Lewis, 1979b, p. 346 my italics), common ground is defined in terms of what is shared among what speakers actually *accept* (and believe they accept, and believe they believe they accept, etc., see Stalnaker, 2002, p. 716), while mutual cognitive environment is defined in terms of what is manifest to the

relevant people[57], which is, as we saw above, a mental notion, even though it is very broad. What is manifest is what can be apprehended given the person's actual cognitive state. But the notion I am after rather concerns what we are *warranted* to take as mutually recognizable and so it is even broader than mutual cognitive environment.

Since 'being warranted' is a normative notion, 'common background' is defined normatively. As such, it seems to me that the notion of common background which I want to employ may be closer to the normative notions of contexts discussed by Brandom (1983). MacFarlane (2011), or García-Carpintero (2015).[58]

On the other hand, what I mean by common ground must also be defined in psychological terms because I see no reason to think that what Lewis, Stalknaker, and Sperber and Wilson talk about should not belong to what I call common background. Indeed, we *are* warranted to take as mutually recognizable what is in our common ground, in our conversational score, or in our mutual cognitive environment. This means that, if we follow Camp (2018c) in considering that Stalnaker's common ground includes and is broader than Lewis' conversational record (which we will here not distinguish from the conversational score), and if we agree that 'being manifest' is broader than 'being accepted' so that mutual cognitive environment is broader than common ground, then here is a scheme which represents the relation between these notions.

[57] 'In a mutual cognitive environment, for every manifest assumption, the fact that it is manifest to the people who share this environment is itself manifest.' (Sperber & Wilson, 1986, p. 42).

[58] The latter in particular – called 'shared commitments' – seems very close to what I mean by common background, because it makes room for many different kinds of attitudes, including e.g. emotions.

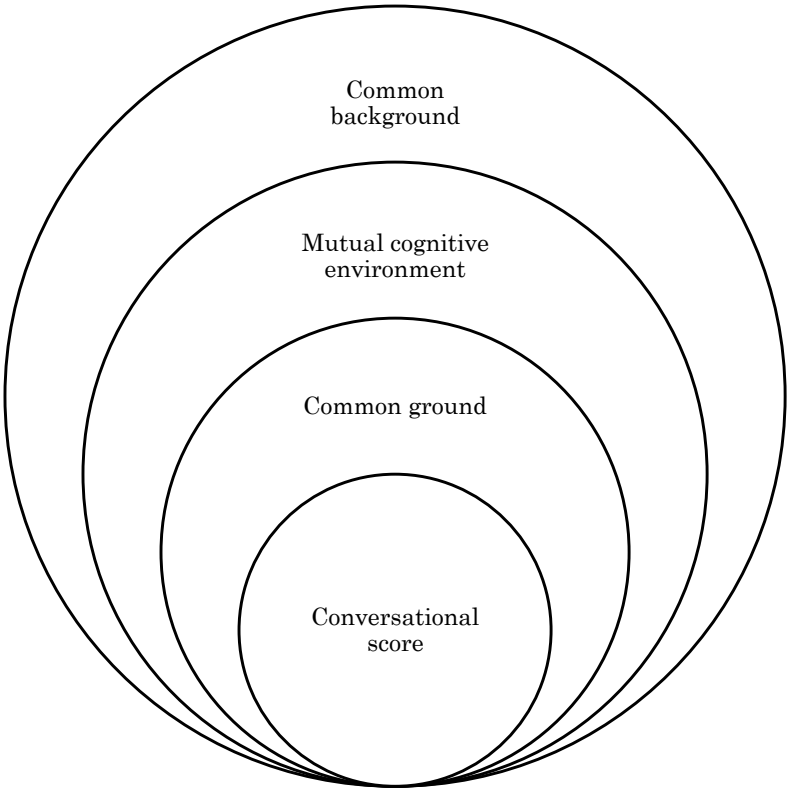

Fig. 2.2. The relation between Lewis's (1979) conversational score, Stalnaker's (2002) common ground, Sperber and Wilson's (1986) cognitive environment, and what I call 'common background'.

2.3.4. On the relation between stimuli with EMRAC and other types of stimuli

Before I turn to another subject, let me make a few points about the typology stimuli that is emerging from what has been said above.

Overtly intentional signals – the type of stimuli that define the scope of the prevailing Gricean models – are a proper subset of stimuli with EMRAC. Indeed, overtly displaying intentions to communicate through a signal implies producing a stimulus with EMRAC. In other words, ostension is not necessary, but it is sufficient to fulfill clauses (i) and (ii) and so for allower-meaning. This is because 'intending x to F' implies 'allowing x to F'. Speaker-meaning is a species of allower-meaning. I call the other species of allower-meaning 'non-intentional allower meaning'. Sam throwing a pound (adapted from Grice, 1957) illustrates such non-intentional allower-meaning. Let me by the way note that another example given by Grice in the same paper, the 'spontaneous frown' (1957: 381), a case which he also excluded from speaker-meaning because it is produced without the Intentions 1 and 2, can nevertheless count as non-intentional

allower-meaning for similar reasons than Sam's throwing of a pound. I will come back to similar cases in the next chapter where I will discuss affective signs.

Fig. 2.3 shows the picture of stimuli at which we have arrived and, in parenthesis, different kinds of meanings whose notions were developed to analyze how these stimuli can transmit information. While the prevailing Gricean models and the notions of speaker-meaning (or OIC) were designed for *overtly intentional signals* (box 1), the EGM and the notion of allower-meaning was designed for all *stimuli with EMRAC* (boxes 1 and 2). Finally, *stimuli without EMRAC* (box 3) should be accountable by a version of the code model, and notions such as natural meaning, statistical meaning, or teleosemantics were developed to account for the transmission of information through this kind of stimuli. We will study how the latter may carry affective meaning in Chapters 7 and 8.

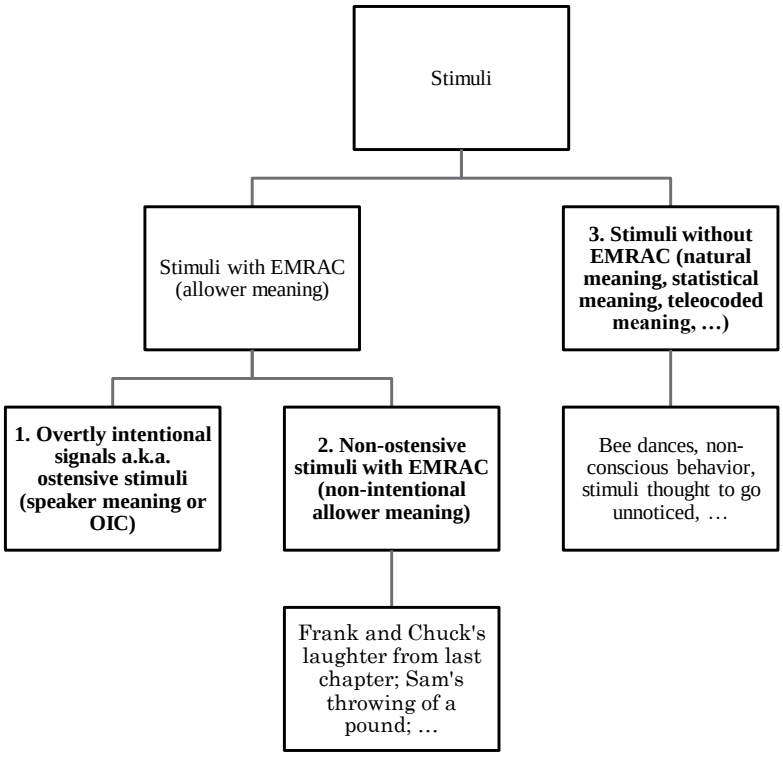

Fig. 2.3. The Extended Gricean Model's typology of stimuli (and the corresponding notion of meaning used to analyze them). The code models account for box 3[59], the prevailing Gricean models for 1 and 3, and the Extended Gricean model for 1, 2, and 3.

[59] We will discuss these cases in Chapters 7 and 8 and see what natural meaning, statistical meaning, and teleocoded meaning refer to.

Let me also mention already that, although overtly intentional signals are subsumed under stimuli with EMRAC, I am definitely *not* recommending that we dispense with the prevailing Gricean models and just replace them with the EGM. The reason is that the prevailing Gricean models are specifically designed for speaker-meaning (or OIC) in ways that make them optimal for this restricted scope, unlike the more general EGM. I will come back to this in the conclusion

Let us now discuss a last feature where the prevailing Gricean models and the EGM differ.

2.4. Pragmatic principles and the Extended Gricean Model

The EGM is broader than what Grice has focused on, but it nevertheless preserves much of the Gricean spirit. It does so especially because of the way it appeals to pragmatic, rationality-based, principles. These are analogous to Grice's Cooperative Principle and they are similarly used to explain why it is warranted and rational to interpret stimuli with EMRAC as carrying the information beyond what they encode, beyond their semantics. 'Rational' and 'rationality' here are understood as *instrumental* rationality, as the (bounded, limited) capacity to adopt suitable means to one's ends (Kolodny & Brunero, 2020).

Here is the important Gricean connection: If someone produces a stimulus with EMRAC, it justifies a receiver in interpreting the signal in a way similar to the interpretation of overtly intentional communicative stimuli. This is because if S produces x while (i) allowing x to generate a particular effect e in a receiver R and while (ii) allowing x to make (i) mutually recognizable, assuming that R wants to understand S, it gives R reasons to mindread what I will call S's *informative dispositions* and do so under the assumption that both senders and receivers are respecting pragmatic, rational principles.

What I mean by 'mindreading S's informative dispositions' is illustrated by the following questions:

- Why didn't S refrain from producing the stimulus x since we both know that x normally carries information I and that S publicizing I is in tension with S's goals?
- Why didn't S produce stimulus y since displaying the information associated with y *prima facie* appears to be conducive to S's goal?

- Did she think about the fact that I could interpret x as sending this information or not?
- If she did, why didn't she prevent me from interpreting her as sending this information with x?
- Did she know that it would have these effects on me, e.g. that I am asking myself these questions?
- If she did, why didn't she prevent x from generating such effects? Why did she allow these effects?
- And why did she allow me to recognize that she allowed these effects?
- Could x be used by S to achieve some other goals of hers?

Assuming that S and R are (imperfectly, bounded) rational agents, i.e. agents who try to maximize their goals as effectively as possible given normal human cognitive abilities, answering these questions – and more generally mindreading S's informative dispositions – should be carried out in light of assumptions similar to Grice's Cooperative Principle (1975) or cognate notions such as Horn's Q- and R-Principles (1984), Sperber and Wilson's Relevance Principles (1986), and Levinson's I-, M-, and Q-heuristics (2000). I call such assumptions 'pragmatic principles'.

The EGM however cannot use these exact pragmatic principles. They were designed for speaker meaning (or OIC) and for this reason they are not suitable for allower-meaning. To see why, take for instance Grice's Cooperative Principle (1975), which he splits into four maxims:

> Cooperative Principle: Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged
>
> 1. Maxim of quality: Try to make your contribution one that is true.
>
> Submaxims: Do not say what you believe is false. Do not say that for which you lack adequate evidence.
>
> 2. Maxim of quantity: Make your contribution as informative as is required (for the current purposes of the exchange). Do not make your contribution more informative than is required.
>
> 3. Maxim of relation: Be relevant.
>
> 4. Maxim of manner: Be perspicuous.
>
> Submaxims: Avoid obscurity of expression. Avoid ambiguity. Be brief (avoid unnecessary prolixity). Be orderly.

The Cooperative Principle and its maxims clearly are meant to apply to speaker meaning and to be restricted to overtly intentional communication. More specifically, they work best for assertions, since the maxims of quality and quantity do not apply for questions or orders. Similar remarks can be made for the pragmatic principles of Horn, Levinson, or Sperber and Wilson.[60] None of them apply well to non-intentional stimuli with EMRAC, such as spontaneous emotional expressions. These pragmatic principles are unhelpful to explain the interpretation of such cases. We need pragmatic principles with a broader scope, which go beyond overtly intentional communication.

This project has been illuminatingly pursued by Kasher's (1982) who argues that Grice's Cooperative Principle and its four maxims can be grounded in the following general principle of rationality: [61]

> Principle of Effective Means
>
> « Given a desired end, one is to choose that action which most effectively, and at least cost, attains that end, *ceteris paribus*. » (Kasher, 1982, p. 32)

Kasher argues that when this principle is put to use in the specific domain of actions that are speech acts, we can derive the following, more precise, principle:

[60] Horn's (1984) pragmatic principles are the following. R-Principle: « Say no more than you must (given Q). » Q-Principle: « Say as much as you can (given R). » Levinson's (2000) pragmatic principles are the following. Q-heuristics: « What isn't said, isn't. » (31) I-heuristics: « What is simply described is stereotypically exemplified. » (32) M-heuristics: « What's said in an abnormal way, isn't normal. » (33) Again we see that these principles are restricted to signals carrying speaker-meaning, and more specifically to linguistic speaker-meaning. Although we could easily adapt them to other types of signals carrying speaker-meaning (e.g. to road signs), they are not fit for all stimuli with EMRAC. And here is Sperber and Wilson's Communicative Principle of Relevance: « Every *ostensive* stimulus conveys a presumption of its own optimal relevance. » (Wilson & Sperber, 2006, p. 612, my italics) Even though the latter has the broadest scope when compared to the pragmatic principles of Grice, Levinson, and Horn, it is still restricted to ostensive-inferential communication, i.e. to signals that are produced with overtly communicative intentions. These prevailing Gricean models do not give us the pragmatic principles we need for stimuli with EMRAC. They won't help us explain all cases of allower meaning.

[61] This project was not pursued by Grice himself, but he nevertheless clearly indicated that such an extension would align with his project. This is apparent in the following passage: « As one of my avowed aims is to see talking as a special case or variety of purposive, indeed rational behavior, it may be worth noting that the specific expectations or presumptions connected with at least some of the foregoing maxims have their analogues in the sphere of transactions that are not talk exchanges. » (Grice 1989: 28)

Rationalization Principle

> « Where there is no reason to assume the contrary, take the speaker to be a rational agent. His ends and beliefs, in a context of utterance, should be assumed to supply a complete justification of his behaviour unless there is evidence to the contrary. » (Kasher, 1982, p. 33)

He then goes on to show how to ground Grice's four maxims in the latter principle, and how each maxim is justified by this general principle.[62] Toward the end of the article, Kasher ventures on a comparison between interpreting language and artistic productions such as paintings (see the epithet of this chapter). In such a context, the Rationalization Principle applies if we merely replace 'speaker' with 'painter' and 'utterance' with 'painting'.

Despite the extent of the scope of Kasher's Principle of Effective Means, it is nevertheless limited to *intentional actions* and more specifically to actions that are chosen as a means to attain an end. However, as we have seen in both this chapter and the last, the blind spot of the standard picture of information transmission – that which we aim to elucidate with the EGM – includes non-intentional actions. The effects generated by the stimuli carrying allower-meaning – the EMRAC – need not be means that are intentionally chosen to achieve an end. I may laugh without having chosen to laugh in order to achieve an end, including communicative ends. Kasher's Principle won't apply to this stimulus. Thus, Kasher's Principle is not broad enough.

Instead, I propose the following principle:

Goal Principle 1: Maximization

> A rational agent will act and react to events that are appraised as relevant to her goals in ways that are apprehended as maximizing the probability of attaining her goals while respecting her goal-hierarchy, *ceteris paribus*.

Let me clarify some terms. By 'appraised' I refer to a basic, non-demanding evaluation, a mental process that is displayed by all creatures that have

[62] Kasher's principles can be considered as a grounding and generalization not only of Grice's Cooperative Principle, but also of the pragmatic principles given by Grice's most influential heirs in this area. As Horn (2006: 24) puts it: « It will be noted that Kasher's principle incorporates the minimax of effort and cost that also underlies models as diverse as the apparently monoprincipled relevance theory (Sperber and Wilson 1986), the dual Q- and R-based approach of Horn (1984, 1993), and the tri-heuristic Q/I/M theory of Levinson (2000). »

goals. 'Goal' in this context is understood as a very broad notion which encompasses the most basic kind of world-to-mind or imperative representations (for the notion of imperative content of a representation, see e.g. Martínez, 2015; Millikan, 1995) – such as those that subtend the desire to eat, to sleep, or to avoid pain – as well as the most sophisticated ones – such as one's reasons to deconstruct post-colonialism or one's plan to compose a symphony. 'Cognized' also refers to a basic, cognitively non-demanding mental mechanism displayed by all creatures that are capable of having preferences among different goals. By contrast, it has been observed that some insects (locusts, aphids, flies) continue trying to feed even as they are being eaten alive (Tye, 2016, p. 140). Their feeding behavior seems to be reflex-driven and not controlled by a goal hierarchy. This is unlike what we observe in some fish, for instance (Tye, 2016, p. 98) (see also Dretske (2006) for a discussion of 'minimal rationality' and why some birds act on the basis of the representations of goals). These processes and representations need not be accessible to consciousness. We will come back to these notions – goals, imperative representations, appraisals, accessibility to consciousness – in the second part of this dissertation (Chapters 5–9).

The Goal Principle is broader than the Principle of Effective Means because it is not restricted to actions that are chosen as *a* means to achieve an end, nor to intentional actions (their superset). Let me illustrate by focusing on emotional reactions. If we follow what has become the consensus view of emotions in affective sciences, emotional reactions should be considered as non-intentional reactions that are nevertheless subjected to the Goal Principle. Indeed, according to this view, emotions are reactions to events appraised as goal-relevant, reactions which are supposed to maximize the agent's goals (Scarantino & De Sousa, 2018, secs. 6–8; Scherer & Moors, 2019). A fear episode is a reaction to an event appraised as relevant to the goal of preserving one's safety, a reaction which is suppose to maximize one's goal by increasing the chance of being safe (and so prepare the organism for fleeing, freezing, fighting, etc.). Again: this process – appraising of an event as goal-relevant and reacting in what is apprehended as a goal-maximizing way – may be, and arguably always is, entirely inaccessible to consciousness.[63]

Thus, a nervous laugh, a spontaneous frown, or an involuntary grim tone of voice, even though they are not intentional actions and fall outside the scope of Kasher's principle, are nevertheless subjected to the Goal

[63] In Chapter 9, I present in some details the view of emotions at which I hint in this paragraph.

Principle, because they are components of an emotional reaction and so are constitutive of a goal-driven reaction. If, for instance, I am embarrassed by the subject of our conversation and I laugh without intending to laugh, without having chosen to laugh as a means to achieve a further end, I nevertheless react to an event appraised as goal-relevant – the embarrassing conversation – with an attitude – embarrassment – whose purpose is to react optimally to this goal-relevant event, to maximize the probability of attaining my goals while respecting my goal-hierarchy, *ceteris paribus* (or just 'to maximize my goals' for short).

You may wonder: how is the triggering of an embarrassed laugh supposed to be a reaction that is apprehended unconsciously as maximizing my goals? Well, by displaying my nervousness, I may make my audience understand that not everything is completely okay from my point of view, perhaps draw some sympathy, and even get help from them about what has caused my embarrassment (e.g. my audience may henceforth avoid coming back to this subject of conversation, or stop expecting me to deal with the present subject with assurance and detachment). Even if the laughter may not be the best reaction to have, what affective sciences seem to teach us about emotions indicate that it nevertheless is the kind of reaction that is caused by an appraisal of the situation and a very fast, rough, unconscious cognitive processes which selects this reaction as that which will maximize the organisms' goals (see in particular (Moors, 2017)).

Now, just like Kasher derived the more specific Rationalization Principle from the general Principle of Effective Means to focus on speech acts, we may derive a more specific principle from the very general Goal-Conducive Principle to focus on the production of stimuli with EMRAC. Here is a proposition:

> Goal Principle 2: Rationalization of allowing
>
> If S allows x to F, then F-ing is more conducive than not to S's goal, *ceteris paribus*.

More specifically, we are concerned with the following case:

> Goal Principle 2: Rationalization of allower-meaning
>
> If S allows the stimuli x to generate effects e in R and to make this mutually recognizable, then assume that these mutually recognizable effects are conducive to S's goal, including her goals in interacting with R, *ceteris paribus*. Accordingly, assume that S would not have allowed x to have e if allowing e was more obstructive to S's goals than not allowing e, *ceteris paribus*.

So, in the following, I will explore how stimuli with EMRAC can be interpreted under a mutual assumption of senders and receivers that they are (imperfectly) rational agents and so subjected to the Goal Principles. The latter are the pragmatic principles of the EGM.

Let us take stock. I said above that the essence of the prevailing Gricean models was that the coding-decoding process is always supplemented by the expression and recognition of *communicative intentions*. In the EGM, a very similar process is happening, but instead of communicative intentions being overtly displayed, the stimuli with EMRAC carry S's *informative dispositions*: the disposition of S to share this or that information, the intentions that S might or might not have had, that S could or could not have had, and how these are coherent or not to her apparent goals. As we will see, even when they don't involve intentions, the way in which these informative dispositions are interpreted by the receiver is very similar to the way in which communicative intentions are interpreted in the prevailing Gricean models, because in both cases the interpretation is done through mindreading processes based on a mutual assumption of rationality and so a mutual assumption that some pragmatic principles are respected by both senders and receivers.

If this is on the right track, a stimulus with EMRAC, similarly to an overtly intentional signal, would carry not only a set of encoded messages, but also a reason to interpret the stimulus in light of some pragmatic principles, and from this assumption, a reason to infer that the stimulus is carrying *information beyond the information which a pre-established code can predict*.

This is why, I believe, the EGM can reach the second goal that we set above and attain what I called the wished-for-virtue of the Gricean models: the ability to account for the transmission of information that goes beyond the prediction of the code model. We will see how in the next section and, in the next chapter, we will illustrate how the EGM can help us analyze different kinds of examples, including the laughter presented in Chapter 1.

2.5. MINDREADING ALLOWER-MEANING

I now have completed my presentation of the EGM. We thus have all the tools necessary to see the machinery at work and, in the next chapter, we will see how the EGM applies to diverse kinds of stimuli. For now, I will only give a short preview of how the EGM explains the derivation of

information beyond that encoded in stimuli. To do so, I will draw a few comparisons between the EGM and the prevailing Gricean models.

Putting all the ingredients presented above together, here is the scheme illustrating the EGM:

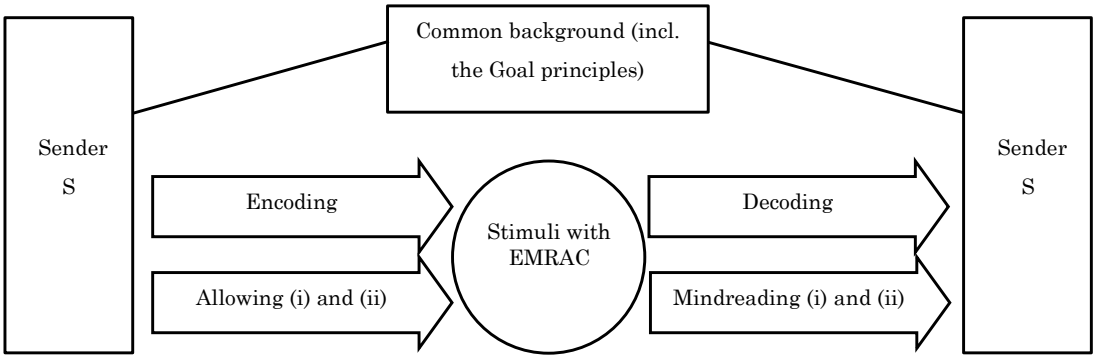

Fig. 2.4. The Extended Gricean Model (EGM).

This scheme looks a lot like the one I used to illustrate the prevailing Gricean models in the last chapter (Chapter 1, Fig. 1.2, §1.4), but there are three important differences, which are all quite tightly related.

First, the display of the communicative Intentions 1 and 2 is replaced by the display of allowing (i) and (ii). Accordingly, it is what S allows x to mean, i.e. (i) and (ii), and not what S speaker-means with the Intentions 1 and 2, that are mindread by the receiver. Once R has inferred (i) and (ii), S and R's common background is updated with the information transmitted by S. R is warranted to take what S allows x to mean as part of what is mutually recognizable between S and R.

The second difference is that instead of an overtly intentional signal we have stimuli with EMRAC, stimuli whose effects on the (potential) receivers are mutually recognizable as controllable.

Third, the pragmatic principles that are mutually assumed by senders and receivers are the Goal principles rather than the more specific pragmatic principles found in Grice, Horn, Levinson, Sperber and Wilson, or even in Kasher.

The other elements of the scheme are shared with the prevailing Gricean models: the sender and the (potential) receiver, the encoding and decoding processes, and the common background (modulo differences between common background and cognate notions of these models).

As illustrated in Fig. 2.4 above, there are two mental processes that the model postulates:

(A) Decoding the information paired with the stimuli
- Is there a conventional meaning associated with such stimuli? What are they lawfully (statistically) correlated with? Do these stimuli have the function of conveying some information? (For more on these questions, see Chapters 7 and 8 on natural information.)

(B) Mindreading the senders' informative dispositions
- May the stimuli have effects (cognitive, affective, …) on me that are mutually recognizable as controllable?
- If so, in light of our common background and the Goal Principles, is the decoding process a satisfying explanation that S produced this stimulus?
- If not, what reasonable hypothesis may I make to rationalize her behavior? What does S allow these stimuli to mean?

The letters A and B are not supposed to refer to the order of the processing: these two processes may run in parallel or in sequence, feedback into one another, etc. However, I find it natural to hypothesize that there is, first, a simple decoding process taking place and then, if the decoding process is deemed insufficient – i.e. unsatisfying in light of the Goal Principles – the mindreading process kicks in. Once the mindreading process has yielded new information and hypotheses, another decoding process may take place, based on the new information obtained from the mindreading process. For instance, a first decoding process does not make sense but, thanks to the mindreading process, I may form the hypothesis that the speaker is being ironic and mocking something and, on this basis, find a new pre-established association between what she said and what she mocks, a decoding process which was not available before the mindreading, since the first decoding didn't take into account the irony (for different views on decoding vs. mindreading processes, see e.g. Cappelen & Lepore, 2005; Recanati, 2004; Wilson & Sperber, 2006).

Let me now give a brief comparison of how implicatures and implicatures* are derived according to the prevailing Gricean models and the EGM. I have highlighted in bold where the two models differ.

Derivation of implicatures in the prevailing Gricean models: (a) Sender S produces an overtly intentional communicative stimulus x. (b) Receiver R decodes x, but, based on their common background, a mere decoding of x is not a satisfying rationalization of S's behavior according to Grice's

Cooperative Principle or another pragmatic principle from the prevailing Gricean models. (c) Because S has produced an overtly intentional communicative stimulus and because R has no reason to think that S does not respect such pragmatics principle, R is led to make hypotheses about other pieces of information that S is sending beside, or instead of, what is encoded in x. (d) Hypothesizing that S is sending pieces of information p in addition to, or instead of, what x encodes permits the best available rationalization of R's behavior. (e) S can know that R can make this hypothesis and S has done nothing to prevent R from making this hypothesis. R can thus reasonably conclude that S means that p with x and that this may now be added to their common background.

Derivation of implicatures*[64] in the prevailing Gricean models: (a) Sender S produces a stimulus which may generate in receiver R effects that are mutually recognizable as controllable. (b) R decodes x, but, based on their common background, a mere decoding of x is not a satisfying rationalization of S's behavior according to the Goal Principles. (c) Because S has allowed x's effects in R and that this is mutually recognizable as controllable and because R has no reason to think that S does not respect the Goal Principles, R is led to make hypotheses about other pieces of information that x carries beside, or instead of, what is encoded in x. (d) Hypothesizing that x carries information p in addition to, or instead of, what x encodes permits the best available rationalization of R's behavior. (e) S can know that R can make this hypothesis and S has done nothing to prevent R from making this hypothesis. R can thus reasonably conclude that S has allowed x to mean that p and that this may now be added to their common background.[65]

If what I have said in this section is correct, I have completed my defense of the general argument given in the introduction of the last chapter and we can now come back to it.

[64] I put a * here because the word 'implicatures' normally is used as a subset of speaker meaning. See §2.1.

[65] Let me note that, in both kinds of derivation processes, the second step involves a tension between what the code says and the pragmatic principles: the code alone is deemed unsatisfying. By this, I mean that, if the receiver only understood the sender to transmit the information encoded in the stimulus, it would be more difficult to rationalize her behavior in light of the pragmatic principles than if the sender hypothesized that more information is transmitted by the stimuli than that which is encoded. If no such tension exists, if the hypothesis that what is transmitted by the sender is optimally rationalized by what is encoded in the stimulus, then there is no impetus to make a hypothesis about what other information could be transmitted by the sender.

2.6. CONCLUSION

In this chapter and the preceding, I have defended the following argument, an argument that will be given more flesh in the next chapter, where the EGM will be put to work.

(A) Two assumptions of the standard picture of information transmission are that (A1) if a piece of information is transferred through a stimulus that is not overtly intended for communication, this information transfer is accountable within the code model, and (A2) if what is communicated requires to be accounted for by a version of the Gricean model, it is communicated through an overtly intentional signal.

(B) But, contrary to (A1) there are cases of information transfer through stimuli that are not overtly intended for communication but that cannot be accounted for by the code models (last chapter) and contrary to (A2) these cases require to be accounted for by a version of the Gricean model (this chapter and the next).

(C) Therefore, we should revise these assumptions of the standard picture in order to allow the explanatory scope of the Gricean model to extend beyond overtly intentional signals.

I have claimed that the EGM can do the job and this will need to be substantiated in the next chapter. In the EGM, the notion of speaker-meaning (or OIC) is replaced with the more generic notion of allower-meaning. Accordingly, the stimuli need not be overtly intended for communication but can be any stimuli with EMRAC, i.e. with Effects (in the audience, potential or actual) that are Mutually recognizable as Controllable. What stimuli with EMRAC are allowed to mean, how they may tentatively be taken to update the common background, can be inferred based on pragmatic principles which I have called the 'Goal Principles', which generalize Grice's Cooperative Principle. Besides these three modifications (allower-meaning, stimuli with EMRAC, Goal Principles) the EGM preserves the components of the prevailing Gricean models and, in particular, how the coding-decoding process is supplemented by mindreading processes based on a common background.

Since stimuli with EMRAC form a much larger set of stimuli than overtly intentional ones, the conclusion reached in this chapter leads us to the following questions: how far does the EGM extend? If the scope of the EGM is the set of stimuli with EMRAC, then what are the limits of the latter? And if the prediction is that this model can account for the information carried by all stimuli with EMRAC, can it live up to this expectation? I will

address these questions in the next chapter by looking at diverse kinds of stimuli.

Let me finish this chapter by highlighting that, even though overtly intentional signals are a subset of stimuli with EMRAC and that speaker-meaning is a subset of allower-meaning, I am definitely *not* recommending that we dispense with the prevailing Gricean models and just replace them with the EGM. The scope of the prevailing Gricean models is that of speaker-meaning, and the latter possesses special kinds of properties that non-intentional allower-meaning does not possess.[66] We benefit from distinguishing them and preserving models dedicated to speaker-meaning and specialized for its analysis.

Let me draw a comparison: the prevailing Gricean Models are like Earth observation satellites, while the EGM is more like a space probe. Earth observation satellites are designed to track what goes on our planet (for weather reports, maps, spying, etc.). And so they just orbit around it. Space probes, on the other hand, are designed to explore extraterrestrial objects, for instance to make observations on the weather of other planets. In principle, a space probe can be used to make observations on the Earth, but because it is designed to travel further away and not to stay in Earth's orbit, it is not as optimal as the Earth observation satellites for this job. The prevailing Gricean models were designed specifically for speaker meaning (or OIC), while the EGM is designed to explore other territories. So, even though speaker meaning also falls within its observation scope, just like Earth falls within the observation scope of space probes, the design of the Extended model does not make it as efficient as the prevailing

[66] To see how the scope of the prevailing Gricean models is special, note for instance that a good definition of speaker-meaning corresponds to what is denoted by the phrase 'S meant p by x', but that is not true of the definition of allower-meaning. So, for instance, if we go back to some of our examples, we won't say that Frank *meant* with his laughter that he is embarrassed. And we won't say that Sam *meant* by his behavior that he did not think highly of Scrooge. Even though this information is part of what Frank and Sam allowed their behavior to mean, it is not what they speaker-meant: speaker meaning corresponds to an important class of phenomena, those for which we use the ordinary expression 'S means p by x', while allower-meaning correspond to a broader class of phenomena. The latter may correspond to the ordinary expression 'S sends messages p with x' as in 'Do you know what messages you are sending with this behavior?'.
Besides the fact that speaker-meaning correspond to an important expression in ordinary English (or in French: *vouloir dire*), the class of phenomena which correspond to the scope of the prevailing Gricean model possesses many other special features which makes it an object of investigation that is worth pursuing on its own, separated from the broader scope of the Extended Gricean Model. This is why I am opposed to a *replacement* of the prevailing Gricean models with the Extended one. It does not make sense to use the Extended Gricean Model to study speaker-meaning, because we already have models that are designed for this particular phenomenon, and are as such better suited to this end.

Gricean model for this specific case, this specific planet. We need both Earth observation satellites and space probes.

3. APPLYING THE EXTENDED GRICEAN MODEL

> « … l'adolescent qui suit point par point la mode actuelle des adolescents, la 'mode militaire', communique par là même à tous ceux qui l'entourent une information, à savoir qu'il entend être reconnu comme appartenant à un certain groupe, avec sa mentalité et ses valeurs. »
> – Roland Barthes, Le grain de la voix

Abstract. In the last two chapters, I have defended two related hypotheses: (a) We may send certain information and update the common background in ways that can be accounted for by neither the prevailing Gricean models nor by the code models. (b) Such cases may be accounted for by the Extended Gricean model (EGM). In this chapter, I will illustrate these claims with several examples. I will begin with the two examples of laughter presented in Chapter 1 and then discuss other kinds of stimuli (nonverbal affective signs (§3.1), the sound of one's voice (§3.2), clothing (§3.3), and speech acts (§3.4). This will allow me to illustrate the working of the EGM and explore some of its boundaries.

3.1. NONVERBAL AFFECTIVE SIGNS

3.1.1. FRANK'S LAUGHTER

Let us go back to Frank's laughter, an example given in Chapter 1.

(1) (At a restaurant) – Emily: 'Where did your wife go?' – Frank: 'She is actually calling the doctor to see if she can meet him about her gastroenteritis. Huhu. He. Hu. [low pitched, soft]' – Emily: 'Oh! I will keep that to myself.'[67]

As we saw, it is natural to understand Frank to transmit the following pieces of information with his laugher:

- p: Frank is embarrassed to reveal a piece of private information about his wife.
- q: Frank's wife would rather avoid that Emily be informed of her gastroenteritis.
- r: The situation is not too serious or worrisome.

[67] The example is adapted from a corpus example (Ginzburg et al., 2015).

And we saw that neither the prevailing Gricean models nor the code models can account for this fact. How can the EGM do so?

First of all, observe that Emily can reasonably make the following hypotheses:

- Frank knows that his laughter may produce a certain set of effects e in Emily (formation of beliefs, modification of her affective state, …).
- Frank can control at least some of the effects produced by his laughter, in the sense that the mechanism leading Frank to generate effects in Emily (beliefs, affects, …) is reasons-responsive. In other words, by holding fixed this kind of mechanism, there is a possible scenario where Frank would have acted otherwise because he recognized a reason as sufficient for acting otherwise (in a way that is understandable to a third-person perspective). For instance, there is a possible scenario where he does not want Emily to think that he is embarrassed and so where he either refrains from laughing (if the laughter is not uncontrollable) or produces further stimuli (an explanation, a confident smile, taking a self-assured posture, etc.) to make Emily think that he is not embarrassed.
- Frank did not refrain from laughing nor did he produce further stimuli to prevent the effects e.
- It is mutually recognizable for both Frank and Emily that they can make these hypotheses.

So, it is reasonable to believe that Frank has produced a stimulus with effects in Emily that are mutually recognizable by both Frank and Emily as controllable, Effects Mutually Recognizable As Controllable, or EMRAC. This means that Frank allows his laughter to mean something, according to how I have defined allower-meaning in Chapter 2.

These notions were defined in Chapter 2, but, maybe it will be easier to follow if I reproduce some definitions here. Here is that of allower-meaning:

> Allower meaning – definition:
>
> A sender S allows x to mean something to the receiver R (or the appropriate conditional receiver R) if, and only if,
>
> S produces x while:
>
> (i) S allows x to generate effects e in R, and

(c) (ii) S allows x to make (i) mutually recognizable for R and S.

And here is that for 'to allow':

S allows x to F – definition

A sender S allows the stimuli x – made of individual stimulus <x1, x2, …, xn>, produced by S between t0 and t1 – to generate the effect e (doxastic, affective, evaluative, behavioral, …) on the actual or conditional audience R if, and only if,

(d) S had guidance-control over the production of e between t0 and t1, and
(e) It was manifest to S between t0 and t1 that S may generate e in R with x.

A mental content is manifest to S at t when that mental content is consciously perceptible, inferable, imaginable, or could be the content of another conscious mental state of S at t holding fixed S's mental capacities and memories at t. And by 'guidance-control' (a notion from Fischer and Ravizza (1998)), I mean that the kind of mechanism that actually issues in S's allowing x to generate e in R is S's own and is reasons-responsive. The mechanism is reasons-responsive if, holding fixed the mechanism kind, the agent would react to at least one sufficient reason to do otherwise. See Chapter 2 (§2.2) for more on these notions.

Because his laughter and the omission to produce further stimuli – I will just say 'the laughter' for short – are stimuli with EMRAC, which means that Frank allows his laughter to mean something, and since there are no reasons to believe that Frank is not a rational agent, Emily can assume that he is subjected to the Goal Principles. In particular, she can assume that here is subjected to this particular application:

Goal Principle 2: Rationalization of allower-meaning

If S allows the stimuli x to generate effects e in R and to make this mutually recognizable, then assume that these mutually recognizable effects are conducive to S's goal, including her goals in interacting with R, *ceteris paribus*. Accordingly, assume that S would not have allowed x to have e if allowing e was more obstructive to S's goals than not allowing e, *ceteris paribus*.

Now, to know what Frank allows his laughter to mean, we need to figure out what are the EMRAC of his laughter (in light of Goal Principle 2). A first set of EMRAC is to be found in what the laughter encodes according

to the codes shared by Frank and Emily, i.e. according to the pre-established associations between laughter and messages. As we saw in Chapter 1, as far as we know from empirical studies, what laughter encodes is best predicted by Table 3.1 (see below and Chapter 1). Since Frank's laughter is soft, low pitched, and brief, it is more of a non-Duchenne than a Duchenne kind. Assuming that Frank and Emily implicitly master the code in Table 3.1 thanks to their past exposure to laughter (like we master the syntactic rules of our language implicitly after sufficient exposure), here is a belief that Frank allows his laughter to produce:

- (B1) Frank expresses amusement, contempt, fear, incredulity, joy, sadness, Schadenfreude, social anxiety (including embarrassment), an urge to affiliate, an urge to act aggressively, or ticklishness.

Information encoded	**Stimuli**
Positive emotion (mostly mirth, but also joy, relief, or playfulness)	Acoustic stimuli of Duchenne laughter (louder, higher-pitched, lasts longer, more calls per bouts, …)
Amusement, contempt, fear, incredulity, joy, sadness, Schadenfreude, social anxiety, urge to affiliate, urge to aggress, ticklishness.	Acoustic stimuli of non-Duchenne laughter (softer, lower-pitched, briefer, fewer calls per bouts, …)

Table 3.1. An acoustic code for laughter (reproduced from Chapter 1).

As we saw in chapter 1, (B1) would not satisfy an engaged receiver, one who wants to answer the question, 'Why did Frank laugh?'. This is because the code in Table 3.1 does not provide enough information to make sense of Frank's laughter by rationalizing his behavior. After all, it is far too vague. Its vagueness may also be unsatisfying for a receiver who cares about how the common background is updated through Frank's laughter.

Frank's laughter is a stimulus that is relevant to the conversation. As a rule of thumb, all emotional reactions that are mutually recognizable are relevant to a conversation. But merely allowing the production of (B1) is not particularly conducive to Frank's goal of interacting with Emily. Interpreting Frank's laughter to merely allow his laughter to generate (B1)

seems to be in tension with the following goal, which is reasonably attributable to Frank:[68]

- (G1) When engaged in a conversation, try to avoid producing stimuli that appear to be relevant to the conversation but are too vague to allow a rationalization of why the stimuli were produced.

So, at this point, what the code models can tell us is in tension with Frank's goal (G1). This would give Emily reasons to make hypotheses about Frank's informative dispositions that are not based on the code model. But before going into this (Gricean) direction, let me look at another potential (semantic) way to resolve the tension between (B1) and (G1) by taking into account statistical regularities.

Among all the emotions that non-Duchenne laughter can express, overall, it is most often perceived as expressing positive affects rather than negative ones (McGettigan et al., 2015, p. 248). Plausibly, this reflects the fact that non-Duchenne laughter may be statistically skewed toward positive emotions, even though it is not statistically significantly correlated with positive emotions like Duchenne laughter is. Additionally, the belief that laughter normally expresses amusement or another positive emotion is widespread, whether or not it is a true belief (Provine, 2001). For these reasons, the association between laughter and positive emotions may be more salient than associations of laughter with other emotions. Following this line of reasoning, it is at least reasonable to suppose that Frank allows his laughter to generate the following EMRAC:

- (B2) It is more probable that Frank is undergoing a positive emotion than a negative one about what he has just said (i.e. that his wife is calling the doctor because of her gastroenteritis).

(B2) is much more precise than (B1) and, as such, it may be considered as avoiding the unwanted vagueness and so resolving the tension with (G1). This may appear as an economic way to rationalize Frank's behavior as it does not appeal to pragmatic principles or mindreading abilities. However, allowing his laughter to generate (B2) seems incompatible with other goals of Frank, goals that can also reasonably be taken to be part of the common background:

[68] It is reasonably attributable to Frank because this goal can plausibly be derived from basic principles of rationality that apply to all rational agents in a way similar to how Grice's maxims of quantity and of relation are derived (Kasher, 1982) since these two maxims are very similar to (G1).

- (G2) Frank wants to be caring toward his wife.
- (G3) Frank's wife's condition is at least bad enough to go to see the doctor and so Frank does not want to appear as undergoing a positive emotion about her condition (because of G2).
- (G4) People generally prefer not to publicize their gastric issues and Frank would rather respect his wife's preferences (because of G2).

I don't see how the tensions between, on the one hand (B1) and (G1) and, on the other hand, between (B2) and (G2–4), can be resolved merely based on pre-established pairings, merely with the tools of the code model, or merely on semantic grounds. This leads us down the Gricean way.

An engaged receiver may suppose that Frank's laughter carries more than the information that it encodes. This supposition may allow resolving the tension between what laughter encodes (its semantic meaning) and the goals and beliefs in the common background.

Here are some hypotheses that Emily may form about beliefs that Frank allows his laughter to generate:

- p: Frank is embarrassed to reveal a piece of private information about his wife.
- q: Frank's wife would rather avoid that Emily or other people be informed of her gastroenteritis.
- r: The situation is not too serious or worrisome.

More specifically, Emily may reasonably suppose that:

> (i) Frank allows his laughter (and the absence of further stimuli) to generate her beliefs that p, q, and r (or: to make p, q, and r manifest), and
>
> (ii) He allows his laughter to make this mutually recognizable to both of them.

So, following our definition, Emily may reasonably suppose that Frank allows his laughter to mean that p, q, and r. Making this hypothesis permits Emily (and us) to rationalize Frank's behavior as we will now see.

Supposing that Frank allows his laughter to mean that p, i.e. that he is embarrassed, is coherent with Frank having goal (G4), i.e. that Frank would rather respect his wife's preferences not to publicize her gastric issue. Embarrassment is an emotion whose purpose is to make us react to situations such as this one: a situation where we don't want to reveal a

piece of information, but where we have no choice to do so, or where it would be worse if we did not reveal it (e.g. because a more important goal is to answer to our friends' questions truthfully and with the appropriate amount of information). Supposing that Frank allows his laughter to mean that he is embarrassed is also conducive to Frank's goal (G3), i.e. to not appear as undergoing a positive emotion about her wife's issue, since embarrassment is not a positive emotion. It is also coherent with (G2), i.e. that Frank wants to be caring toward his wife since being embarrassed about this situation shows that Frank cares about his wife's privacy.[69]

Supposing that Frank allows his laughter to mean that q, i.e. that his wife would rather avoid that Emily be informed of her gastroenteritis, is, of course, conducive to (G4), i.e. that Frank would rather respect his wife's preference that the issue remain private. It also gives further support to the hypothesis that p, i.e. that Frank is embarrassed, and as such supposing that Frank allows his laughter to mean that q reinforces the cohesion with goals (G3) and (G2).

Supposing that Frank allows his laughter to mean that r, i.e. that the situation is not too serious or worrisome, is coherent with a further goal that can also reasonably be taken to be part of the common background:

> (G5) Try not to laugh at topics that are too serious or worrisome for you or your audience. If you cannot help or did not know that the issue was too serious or worrisome, present your apologies.

Finally, supposing that Frank allows his laughter to mean that p, q, and r makes it conducive to (G1), i.e. to not allow stimuli that appear to be relevant but are not informative enough.

So, we see how supposing that Frank is sending pieces of information that go beyond what is encoded in his laughter permits a more optimal rationalization of his behavior. It gives a satisfying explanation for why Frank laughed.

Since there is no reason to think that Frank would be unable to think that Emily can make these hypotheses and since Frank possessed guidance-control over the ensuing beliefs, Emily can reasonably conclude that Frank allows his laughter to mean that p, q, and r.

[69] Observe also that p is consistent with what the code models would predict, i.e. with (B1) and the fact that laughter expresses several emotions including embarrassment, and with (B2), i.e. that laughter is most often associated with positive emotions. The laughter in question is just not part of the statistically most frequent type of laughter.

Furthermore, since Frank has not only allowed the beliefs that p, q, and r to be generated in Emily, but has also allowed this to be mutually recognizable, Emily's beliefs that p, q, and r may *tentatively* be added to their common background.

I say 'tentatively' because there may be alternative, divergent hypotheses which would rationalize Frank's behavior just as well as the present one, or better. In order to be sure that we can add these pieces of information to the common background, we need to be sure that the information in question is the best rationalization available and that it is so for all the participants. In this case, the description of the scenario does not give enough information to be certain about what would be the best rationalization of Frank's behavior. The more we know about the common background between Emily and Frank, the more chance we have to know how exactly their common background should be updated by stimuli with EMRAC.[70]

The fact that p, q, and r can tentatively be added to the common background is an important fact about Frank and Emily's conversation because tentative additions to the common background can be used to reorient the conversation, or even can be used to make presuppositions which, if they remain unchallenged, will reinforce the tentative additions and make them definite additions. In our example (1), after Frank's laughter, Emily utters 'Oh! I will keep that to myself.' We can reasonably take this response to make the presupposition that q, i.e. that Frank's wife would rather avoid that Emily or other people be informed of her gastroenteritis. The laughter has primed this presupposition; with her response, Emily confirms that she now takes q to be part of the common background. If Frank does not challenge this presupposition, this piece of information will be added to their common background. Similar reasoning applies for p and r.

In conclusion, it seems to me that the EGM allows us to give a satisfying account of the information we naturally understand Frank to convey with his laughter, of what he allows his laughter to mean beyond what is encoded in it. This model allows us to explain how the laughter updates the common background through pragmatic explanations (implicatures*) even though these pieces of information were not speaker-meant.

[70] In Relevance theory's terminology, p, q, and r are akin to weak implicatures (Sperber & Wilson, 1986, Chapter 4). We could say they are weak-implicatures* since implicatures *stricto sensu* belong to what is ostensively communicated, which is not true of p, q, and r.

3.1.2. CHUCK'S LAUGHTER

Let us now see, more briefly, how the EGM may apply to Chuck's laughter, the other example that I presented in the first chapter:

> (2) (David and Chuck are good friends, politically left-wing, who share progressive values) David, on a serious tone: 'You know, I was thinking: maybe Sarah Palin is the future of the Republican party…' Chuck: 'hh hh, heh heh heh, huhu, hahaHAHAHAHA' (laughs while David is continuing his sentence, his laughter begins rather softly and middle pitched, raises in pitch and ends up pretty loud) David continues: '… seriously I even think she's got her chances for the next elections.'

According to the code in Table 3.1, Chuck is emitting a burst of Duchenne-like laughter (long, high-pitched, loud) and this can tell us that he is most probably undergoing a positive emotion. This, however, is in apparent tension with the meaning of David's statement (which is pronounced on a serious tone, not joking around), since it is part of their common background that, because they are left-wing progressives, Sarah Palin being the future of the Republican party is not good news. Because of this tension, there is a reason for an engaged receiver such as David to figure out Chuck's informative dispositions: what is he disposed to let his audience infer from his behavior? Chuck doesn't provide an excuse or an explanation after his laughter or any other behavior that is meant to cancel some of the effects that his laughter could have on David (e.g. by saying 'I'm sorry, I'm tired, it was a nervous laughter', or by showing nonverbally that he wanted to suppress his laughter, or otherwise). So, for reasons analogous to the one discussed with Frank's laughter above, we can suppose that the laughter is a stimulus with EMRAC.

Another way to describe the situation is to say that David would be justified in engaging in the following reasoning: 'Chuck could have suppressed his laughter or changed some of its communicative effects by explaining why he laughed. This is mutually recognizable. Thus, either he doesn't mind that the affective state usually associated with this type of laughter is available to me or he pretends to make it available. In either case, he is open to the idea that I interpret him as willing to share this affective state. He thus is disposed to generate such effects in me with this laughter. Would these effects help explain his behavior? What are the ways in which they may be taken to update our common background?'

So, David would be justified in thinking that:

Chuck laughed (and did not produce any further relevant stimuli) while

(i) Chuck allowed his laughter (and the absence of further stimuli) to generate certain effects (beliefs, affects, etc.) in me, and

(ii) Chuck allowed the laughter (and the absence of further stimuli) to make (i) mutually recognizable for both him and me.

Assuming that there is no reason to doubt that Chuck is a rational agent, David can take Chuck to respect the Goal Principles and in particular their application to stimuli with EMRAC. Furthermore, we can infer from their common background that these friends are engaged in a cooperative interaction and thus assume that his laughter should, *prima facie*, be taken as optimally contributing to the goals of their interaction and their communicative exchange.

Here is some relevant information available from the common background:

- (B3) Duchenne laughter is most often the signal of an attitude of amusement toward the stimulus and thus of not taking the stimulus seriously, but
- (G6) Chuck possesses left-wing, progressive values that are threatened by the possibility of Sarah Palin being the future of the Republican party and so he probably is not, and does not want to appear as, merely amused by this possibility.

(B3) and (G6) are in tension and as such these pieces of information do not permit a satisfying rationalization of Chuck's behavior in light of the Goal Principles.

However, a plausible hypothesis that David can make is that Chuck allows his laughter to mean the following:

- s: Although people could think that Sarah Palin is the future of the Republican Party and, although her influential political opinions could thus be considered a threat to my/our values, she actually is not a serious threat, and so your remark merely amuses me.

Supposing that Chuck allows his laughter to mean that s would be consistent and help explain the following assumptions:

- There is no reason to suppose that Chuck is not respecting the Goal Principles. If by laughing, he allows his laughter to mean that s, his laughter would satisfyingly contribute to David and Chuck's interaction, and thus respects the Principles.

- Chuck can know that David could form the hypothesis that he allows his laughter to mean (something like) s, but Chuck didn't do anything to stop David from thinking so. So, he probably is not against David making this hypothesis.
- Chuck would expect David not to be satisfied with stimuli carrying information incoherent or in tension with their common background, *ceteris paribus*.
- (B3) and (G6) *prima facie* are in tension, but if Chuck allows his laughter to mean that s, David can update their common background by taking this hypothesis as a plausible rationalization of Chuck's behavior, a hypothesis to be reinforced by their future interaction if it is not sufficiently secured by what is already in their common background (e.g. does Chuck usually laughs for reasons comparable to s?). If their common background is updated with s, the tension between (B3) and (G6) is resolved: Chuck is not amused by Sarah Palin's being the future of the Republican Party, but by (something like) the ridicule of taking her as a serious threat.

These make the hypothesis that Chuck allowed his laughter to mean that s reasonable. David and Chuck may thus tentatively update their common background accordingly. As with Frank and Emily's case, this tentative update may reorient the conversation by priming it toward new topics, and it permits making some presuppositions that would not otherwise be warranted. For instance, David may reply, 'No, but seriously, she's a bigger threat than you'd think. It's scary…'. This reply presupposes that one may think that Sarah Palin is not a threat, which is part of what Chuck allowed his laughter to mean. If Chuck had not laughed, David could not have made this presupposition (or at least not so naturally). In turn, depending on how Chuck replies to this remark, what he allowed his laughter to mean would be more or less anchored into their common background.

3.1.3. OTHER EMOTIONAL EXPRESSIONS (FACIAL, POSTURAL, VOCAL, MUSICAL)

It is not only laughter that may indicate various kinds of affective states whose disambiguation requires an explanation along the lines given by the Extended Gricean model. Most, and perhaps all, affective signs carry context-dependent information in the sense of encoding only vague, ambiguous information. This information needs to be supplemented by knowledge of the context, where the context often is Gricean despite the absence of speaker-meaning, as it must include the Goal Principles and a common background if someone is to make sense of what information is transmitted within that context. Think, for one, about smiling: we may

smile when happy, but also when we are embarrassed, slightly disgusted, revengeful, polite, etc. Think also about frowning: we may frown because we are angry, but also because we are concentrating hard, or because we don't understand something. This vagueness of facial expressions indicates that, in many cases, the code models won't be sufficient. But because facial expressions often are not voluntary and rarely are intended for communication, the prevailing Gricean models frequently won't be satisfying either.

Take, for instance, the following case:

> (3) Sam tells an ethnic joke. Maria starts frowning. She doesn't intend to communicate anything. She doesn't intend to make anything manifest. Her frown is just an unintended affective reaction.

Arguably, Maria allows her frown to mean the following:

> (p3) Sam's joke is not funny.
>
> (q3) Sam's joke is offensive.

Here, once again, I believe that we need the EGM to account for the fact that the common ground may be updated with p3 and q3.

To see why the code models have little chance to successfully account for this, take another frowning example:

> (4) Maria and Sam are playing chess. Maria is not far from winning. It is now her turn. She starts frowning. She doesn't intend to communicate anything with Sam. She doesn't intend to make anything manifest to Sam. Her frown just is an unintended affective reaction.

Arguably, in this case, Maria allows her frown to the following:

> (p4) She is thinking hard about her next move.

I won't discuss cases (3) and (4) any further. I just wanted to mention how what is encoded in a frown, just like in laughter, often underdetermines what the expression means, even if the frown is not an overtly intentional signal.

The same applies to many other types of emotional stimuli. It is easy to multiply the examples besides laughing, smiling, or frowning. Take, for instance, sighing: it has been shown that sighs are correlated with mental states as diverse as pain, panic, relaxation, relief, sadness, stress, and the

will to give up (Teigen, 2008; Vlemincx et al., 2009). But in certain situations, we understand much more from a sigh than that it may express any one of these psychological states. For instance, we usually understand that a sigh expresses disappointment and not relief, even though this information is not encoded in the sigh, and that the sigh does not signal an overt intention to communicate. If we just look at the (super-)semantics[71] of sighing, we won't find what we are looking for. Whether we look at what sigh is conventionally associated with (conventional meaning), what is it statistically correlated with (statistical meaning, see Chapter 7), what it has the function of expressing (teleosemantic, organic meaning, see Chapter 8), none of these will be sufficient to explain what we may understand a sigh to mean. And because it is not overtly intended for communication, the prevailing Gricean models don't apply. Let us turn to another example.

Contrary to what Darwin, Duchenne, or Ekman thought, even prototypical facial expressions can be interpreted as expressing different emotions depending on the context. Despite the enormous success of the code proposed by Friesen and Ekman (Ekman & Friesen, 1971; Friesen & Ekman, 1976) – a code which is supposed to pair so-called basic emotions (the 'message') with patterns of facial muscles activation (the 'signal'), (see Table 3.2 and Fig. 3.1) – this code is not always a good guide to what is expressed. Just like we need pragmatics in addition to semantics to make sense of what sentences mean, we need something in addition to a code model to make sense of what information facial expressions convey. In both these nonverbal and verbal cases, the code underdetermines the meaning.

Emotions (messages)	**Facial stimuli (signals)**

[71] Super-semantics is a framework where semantic tools are applied beyond traditional semantic objects of studies, e.g. to primate calls, music, pictures, and, we could imagine, to emotional expressions (Schlenker, 2018).

Anger	Brow lowerer + Upper lid raiser + Lid tightener
Fear	Inner brow raiser + Brow lowerer + Upper lid raiser + Lip stretcher + Jaw drop
Disgust	Nose wrinkler + Lip corner depressor + Lower lip depressor
Surprise	Inner brow raiser + Upper lid raiser + Jaw drop
Joy	Cheek raiser + Lip corner puller
Sadness	Inner brow raiser + Brow lowerer + Lip corner depressor

Table 3.2. The code for facial emotional expression proposed by Friesen and Ekman (1976), which is supposed to pair facial muscle activations with six basic emotions.

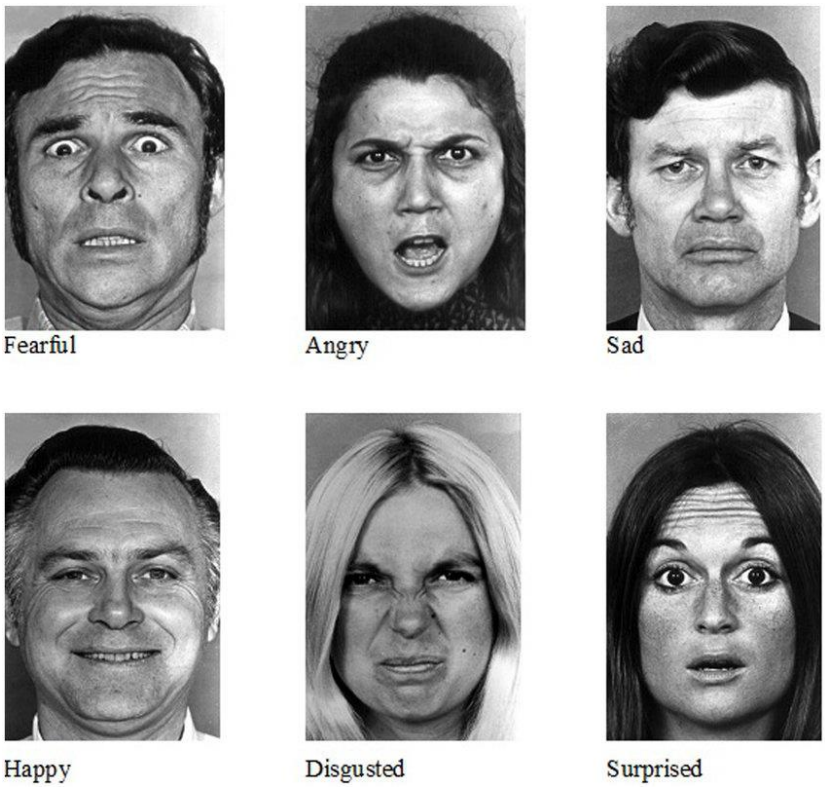

Fig. 3.1. Illustration of Friesen and Ekman (1976)'s code.

An increasing number of counterexamples to Friesen and Ekman's code can be found in the psychology literature (as samples, see Aviezer et al., 2008; Barrett et al., 2011, 2019; Carroll & Russell, 1996). An intuitive illustration is displayed in Fig. 3.2 below. Here, an 'angry-face' pulled by Paul Ekman is Photoshopped onto a body in different emotion-loaded

contexts. As a result, the 'angry-face' is not understood as expressing anger as it would if only the face were shown. Contrary to what Friesen and Ekman's code would predict, it is interpreted as disgust and sadness by a significant number of participants. The body postures and the context (a diaper and a gravestone) make them interpret Ekman's 'angry-face' as expressing other emotions. Note that this is very similar to the Kuleshov effect, a phenomenon known for nearly a century and whose experimental tests have been replicated several times (Barratt et al., 2016).

Fig. 3.2. Ekman's disgust-face' is understood as expressing disgust and sadness when put into different contexts. From Avezier et al. (2008).

One way to respond to this counterexample would be to try to devise a more sophisticated code that would take into account bodily postures and contextual stimuli as well as facial muscle activation patterns. So, we would have something like:

- Brow lowerer + Upper lid raiser + Lid tightener + bodily postures $BP_{n–m}$ and contextual cue $CC_{o–p}$ => anger
- Brow lowerer + Upper lid raiser + Lid tightener + bodily postures $BP_{q–r}$ and contextual cue $CC_{s–t}$ => disgust
- Brow lowerer + Upper lid raiser + Lid tightener + bodily postures $BP_{u–v}$ and contextual cue $CC_{w–y}$ => sadness

There are (at least) two big problems with these codes. First, devising a code pairing bodily postures and emotional expressions seems not to be realizable (Dael et al., 2012). Despite the remarkable efforts that Dael et al. have put into trying to start establishing such a code, they concluded that 'an emotion can be encoded by a variety of behavior patterns' (Dael et al., 2012, p. 1099), which made it impossible to formulate a pre-established pairing between bodily postures and emotion expressed. The second, more important, problem concerns the contextual stimuli. Listing the contextual

stimuli that would trigger one or another interpretation in audience A or B (is it anger, disgust, sadness, etc.) seems even less attainable because of the infinite number of stimuli that can make one interpret an 'angry-face' as expressing disgust (diaper, rotten food, certain insects, …). This is especially troubling once we take into account cultural and individual variability. For instance, if you know that a person comes from a culture where a certain dish is considered disgusting, you may interpret her 'angry-face' as meaning disgust because of this culturally variable stimulus. And if you know that a person is easily disgusted by, say, rubber ducks, then you may interpret an 'angry-face' as expressing disgust in a very idiosyncratic context. Such factors make it unimaginable to draw an appropriate code to explain how we understand facial expressions.

Instead of trying to devise a more sophisticated code, I propose that we respond to the counterexamples to Friesen and Ekman's claims by giving up on the idea that only a pre-established code can predict what is conveyed by facial expressions. I propose that we follow instead the path taken by Griceans: language researchers have, during the 20th century but especially since Grice's *William James Lectures*, gradually abandoned the idea that linguistic utterances may be understood only through semantics (or semiotics). They have thus developed pragmatic models to fill in the explanatory gap. But, since most facial expressions are not overtly intentional signals, the prevailing Gricean models cannot be used. Ergo, our *deus ex machina* enters the stage. However, I won't detail how the EGM would be applied here, because the laughter examples (1) and (2) are sufficiently similar to let the reader fill in the details.

Before I leave the topic of emotional expression and affective stimuli, let me briefly discuss another code besides Friesen and Ekman's that is very influential in affective sciences. This code is illustrated in Table 3.3 below. Its features include the following:

- It pairs emotion (the messages) and acoustic stimuli (the signals).
- It is designed to apply to both musical and vocal expression.
- It predicts, for instance, that both angry music and angry voices tend to be faster, higher-pitched, noisier, louder, etc. than sad music and sad voices.

This code, established by Juslin and Laukka (2003) based on 100+ empirical studies, largely confirms the code proposed for vocal expression by Scherer (2003).

Summary of Cross-Modal Patterns of Acoustic Cues for Discrete Emotions

Emotion	Acoustic cues (vocal expression/music performance)
Anger	Fast speech rate/tempo, high voice intensity/sound level, much voice intensity/sound level variability, much high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and microstructural irregularity
Fear	Fast speech rate/tempo, low voice intensity/sound level (except in panic fear), much voice intensity/sound level variability, little high-frequency energy, high F0/pitch level, little F0/pitch variability, rising F0/pitch contour, and a lot of microstructural irregularity
Happiness	Fast speech rate/tempo, medium–high voice intensity/sound level, medium high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and very little microstructural regularity
Sadness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contour, slow voice onsets/tone attacks, and microstructural irregularity
Tenderness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contours, slow voice onsets/tone attacks, and microstructural regularity

Note. F0 = fundamental frequency.

Table 3.3. Code pairing emotion expressed (the messages) and musical or vocal cues (the signals), reproduced from Juslin and Laukka (2003, p. 802).

Despite its undeniable merits and predictive power, it is easy to multiply the counterexamples for this code as well. For instance, one of its unavoidable, unfortunate predictions is that the vast majority of hard rock pieces should always express anger. Indeed, the vast majority of hard rock pieces possess a fast tempo, a high sound level, much sound level variability, much high-frequency energy, fast attacks, much microstructural irregularity (non-linear timbre from distorted guitars and screaming voices), etc. These are the acoustic stimuli associated with anger, according to this code. However, it is not the case that the vast majority of hard rock pieces express anger.

Take for instance *Highway to Hell* by AC/DC.[72] Juslin and Laukka's code would predict that it is an angry song. But this is inaccurate. It is apparent judging only by the music and the lyrics confirm that the song does not express anger. If one had to choose among the five emotions from Table 3.3, the closest surely is happiness.[73] The best code for musical expression forces on us an inaccurate reading of what this song expresses.

The limitations of Juslin and Laukka's code model is, of course, not restricted to hard rock. For instance, although techno nearly always meets

[72] Song available at https://youtu.be/fa82Qpw6lyE.

[73] As an illustration, here is the first verse: « Living easy, living free/ Season ticket on a one-way ride/ Asking nothing, leave me be/ Taking everything in my stride/ Don't need reason, don't need rhyme/ Ain't nothing I would rather do/ Going down, party time/ My friends are gonna be there too. »

the criteria associated with happiness (fast tempo, medium-high sound level, medium high-level frequency, fast tone attacks, very little microstructural variability, etc.), what this genre may express cannot be reduced to this emotion. And as you can imagine, once we start exploring non-Western music, the limitations are even more apparent.

The limitations of the code are not restricted to music. We may easily find counterexamples in vocal expression. Imagine, for instance, an Italian soccer commentator commenting on her favorite team. The team is about to score a goal and the commentator is super excited. Typically, she will talk with a fast speech rate, high sound level, much sound level variability, much high-frequency energy, fast attacks, microstructural irregularity (from screaming). Although it is obvious that she expresses joy and excitement, these are the cues paired with anger in Juslin and Laukka's code.

How do we come to the right interpretation, then, if it is not with a code such as Juslin and Laukka's? Well, once again an important factor that this code model, like the others, does not seem to deal easily with is context and more specifically common background. To come back to AC/DC, I would venture to hypothesize that it is because this band and their target audience share a common background that it is not anger that is expressed in *Highway to Hell*, where the common background includes what is generic to hard rock and what deviates from the norm that the audience can easily understand.

I am not going to analyze how exactly the common background plays a role here or why I think that the prevailing Gricean models won't suffice for such a task. However, let me note that musical stimuli usually *are* overtly intentional signals – they are usually produced with the intention (Intention 2) to make it mutually recognizable that the music was produced with another intention (Intention 1) to create some effect on the audience. The reason why the prevailing Gricean models do not apply straightforwardly to such cases is thus different from the cases we have discussed above – it rather has to do with the fact that musicians generally do not have *informative* intentions but rather want to generate other kinds of effects in their audience, effects that are not encompassed by the prevailing Gricean models.[74]

[74] I initially planned to dedicate a chapter of my dissertation to this topic but decided against it because of the dissertation's length. If interested, however, see my 'Super-semantics, super-pragmatics, and musical meaning' (Bonard, ms).

My main point in this brief discussion of Juslin and Laukka's code was to bring in more evidence for the hypothesis that codes underdetermine meaning in most – perhaps all – kinds of affective signs (vocal, facial, postural, musical, etc.). In addition, because these signs often are produced in ways that fall outside the scope of the prevailing Gricean models, we are led to see how the EGM would work.

3.2. THE SOPRANO VOICE AND THE COMMON BACKGROUND

Let us further explore the boundaries of allower-meaning with an idiosyncratic example. This will help me detail what I mean by 'common background' and why it differs from cognate notions such as conversational score, common ground, or mutual cognitive environment.

Imagine that you ask a stranger, 'Do you have a soprano voice?' and that she replies, in a soprano voice, 'Leave me alone!'[75] Does she thereby update your common background with the piece of information that she has a soprano voice? My answer would be: it depends on whether the stimuli she produced are to be considered stimuli with EMRAC or not. It depends on whether she allows her reply to mean that she has a soprano voice.

In particular, the question that interests me is whether she displays stimuli whose relevant consequences can *reasonably* be taken to be mutually recognizable. Or, in other words, whether she displays stimuli with consequences you would be *reasonable* to take as mutually recognizable. I would say that, as long as it is reasonable to assume that she respects the two clauses of the definition of allower-meaning (see Chapter 2 or (i) and (ii) in §3.1.1 above), i.e. that she has manifest control over your belief that she has a soprano voice and that this is mutually recognizable, then it makes sense to assume that she does update your common background. Both of you would then be *warranted* to assume that you both know she has a soprano voice. This is so independently of whether she has thought about this consequence. As we have seen in the preceding chapter, this makes common background a normative rather than a psychological notion and so makes it different from Lewis' conversational score (1979b), Stalnaker's common ground (2002), and Sperber and Wilson's mutual cognitive environment (1986, Chapter 1).

Assuming that she understood your question, that you both know how to recognize a soprano voice, and that neither of you suffers from any relevant

[75] Thanks to Kevin Lande for helping me to come up with this example.

psychological issues, it seems reasonable to assume that she does (i) allow her reply to generate your belief that she has a soprano voice while (ii) allowing (i) to be mutually recognizable. This is so because, even if she did not realize it, it seems reasonable to assume that she possessed control over the generation of your belief and that this control was manifest to her (it was apprehensible through, e.g., inferences). So, condition (i) holds. And it also seems reasonable to assume that she had control over whether (i) was mutually recognizable or not and that this control was manifest to her, fulfilling condition (ii).[76]

According to the EGM, she thus has updated your common background. For this reason, you are warranted in assuming that you now both know the answer to your question and respond to her accordingly. You may, for instance, ask straightforwardly, 'And does your daughter also have a soprano voice?', which expresses the corresponding presupposition (without being an informative presupposition). This is the case even if she did not actually have any mental state about her soprano voice, even if she did not update your conversational score, your common ground, or your cognitive environment.

Now, imagine a variation of the scenario. You ask a woman, 'Do you have a soprano voice?' and she replies by saying 'Fous-moi le camp!', which means leave me alone (get out of here) in French. Although you can once again hear in the tone of her voice that she has a soprano voice, the situation is different. In this case, since she replied in a language different from that of the question, as far as you know, she might not have understood the question, and thus might not have been able to think about the fact that she has displayed a soprano voice. In other words, the proposition 'I have a soprano voice' may not be accessible to her mind at this moment, being not manifest enough for her, being too unrelated to her cognitive environment (to borrow again expressions from Sperber and Wilson, 1986). If you should reasonably assume that the information 'I have a soprano voice' is *not* accessible to her at this moment, then it is not the case that (i) she allowed her answer to generate the thought in you that she has a soprano voice. If, at this very moment, it is not reasonable to

[76] Holding fixed the relevant mechanism kind (i.e. the one that is relevant to the ascription of responsibility) that led her to utter the sentence in question and thereby creating your belief, there is a possible scenario where, because of an understandable reason, she prevents (i) from obtaining, e.g. by remaining silent. This scenario shows that she has the guidance-control necessitated by (i). A very similar scenario where she remains silent because she does not want (i) to be mutually recognizable shows that she also possesses the guidance-control necessary for (ii). If it is manifest to her that she could have remained silent for the relevant reasons, then the control required by both (i) and (ii) are manifest. If that is so, the conditions for allowing in (i) and (ii) obtain.

suppose that she could think about this effect on your cognition, we cannot legitimately say that she had the relevant control over the production of your belief that she has a soprano voice.[77] We thus cannot say that she allowed her response to mean that she has a soprano voice. By the same reasoning, we arrive at the conclusion that this information does *not* update your common background.

3.3. SAM'S CRUMPLED SHIRT AND CONTROL

Here is another example that illustrates the boundaries of the EGM. While the preceding helped in further presenting what the limit of the common background is, this one will help with the limit of control.

Imagine that you are the boss of a restaurant. One day you have a chat with Sam, one of the waiters, to tell him that he should put some effort into his appearance, which tends to be neglected. The week after that, Sam comes to work with a completely crumpled shirt. Sam is not intending to communicate anything with his shirt at all – and you know that. He knows you are here and that you have seen him. He has not come up to you to discuss anything, however. You go to him and, without pointing to his shirt, without staring at it, without giving any evidence about what you are going to talk about, you say: 'This is unacceptable! Next time, you are fired.' When you utter this sentence, you are presupposing that Sam will know what 'This' refers to. It is indeed natural – and it may well be rational – for you to consider that, by behaving as he did, Sam has updated your common background with the following piece of information:

- p2: Sam doesn't really care about making the efforts you have asked him to.

[77] She intuitively lacks the relevant control over the production of your belief, but can we account for this intuition through Fischer and Ravizza's guidance-control? It would mean that, holding fixed the relevant mechanism kind, there is no possible scenario where she acts on the basis of an understandable reason in such a way that she prevents her answer from creating this belief in you. However, isn't there a myriad of possible scenarios where these conditions hold? Well, it depends on what the mechanism kind is that must be held fixed. Fischer and Ravizza tell us that it is the kind that is relevant to the ascription of responsibility. Here is an example of something highly relevant to whether she should be held responsible for the generation of your belief: whether the fact that she has a soprano voice is something that she can apprehend after you have asked her the question. I take it then that the mechanism kinds that must be held fixed are the ones which determine whether or not she can apprehend the fact that she has a soprano voice at this moment (e.g. her inability to speak English must belong to the relevant mechanism kind). And, indeed, if those mechanisms are held fixed, there is no possible scenario where she would have prevented herself from creating this belief in you on the basis of an understandable reason because she would not have access to reasons that involve the fact that she has a soprano voice. She could not recognize a reason that is based on this fact since this fact is, *ex hypothesi*, not apprehensible to her at this moment.

- q2: Sam doesn't respect you as much as you expected.

More precisely, I take it that it is rational to consider that Sam updates the common background with p2 and q2 *if* he allows his shirt to mean p2 and q2. We will explore some conditions for this to be the case below.

But first, at the risk of being redundant, let me observe that, because Sam did not have any communicative intentions, the prevailing Gricean models don't apply to him. Observe that, as I have described the scenario, this is so even if Sam stopped while dressing, asking himself whether it is ok to go to work like this, and then decided to dress in this way anyway. To count as having successfully speaker-meant something with his shirt, he must not only have acted intentionally in wearing a shirt that he knows will produce in you the beliefs that p2 and q2, he must also have intended you (what I called the Intention 2 in the previous chapters) to notice an intention (the Intention 1) to produce these beliefs in you. Furthermore, he must have succeeded in achieving Intentions 1 and 2. In the description of the scenario, I mentioned that you did not take Sam to be intentionally communicating anything with his shirt, so these conditions don't hold.

Further, once again, there does not seem to be any pre-established code that could help us fully understand what is happening here. You may think: but isn't there a conventional code which pairs, on the one hand, a waiter wearing a crumpled shirt and, on the other hand, the waiter disrespecting her boss? No: think about those laid-back restaurants where there is no such dress code. You may thus want to modify the code as follows: 'a waiter wearing a crumpled shirt at a non-laid-back restaurant => the waiter disrespects her boss'. But then what about cases where the waiter is naïve in the sense that she does not know that you expect her not to wear a crumpled shirt (she hadn't been asked to take care of her dress)? You could modify the code as follows: 'a waiter wearing a crumpled shirt at a non-laid-back restaurant and who is not naïve (in the relevant sense) => the waiter disrespects her boss'. But then what about the cases where the waiter could not have worn an ironed shirt for good reasons (e.g. see the flood example below) and who apologized? You could once again modify the code and have: 'a waiter wearing a crumpled shirt at a non-laid-back restaurant and who is not naïve and who has not apologized => the waiter disrespects her boss'. But then what about cases where the waiter has not apologized yet because she has not seen that her boss is present? You could modify the code, of course. But you have to admit that modifying the code every time that one brings a counterexample is an *ad hoc* strategy. The code models do not seem fitting to explain how Sam's crumpled shirt can reasonably be interpreted to carry the information that p2 and q2.

So, once again, neither the prevailing Gricean models nor the code models seem adapted to account for this reasonable interpretation. In contrast, I believe that the EGM can. Let us see how it can by concentrating on what EMRAC Sam produces with his crumpled shirt and by thinking about how to rationalize Sam's behavior with the Goal Principles.[78]

Because you recently had a chat with Sam about his appearance, it is reasonable to hypothesize that, if Sam really cared about making the efforts you have asked him to make, he would have either worn an ironed shirt, or he would have come to you to explain why he did not. Even if, for some reason, Sam couldn't wear an ironed shirt this morning – we can imagine that his flat was flooded during the night so that he was unable to wear anything but a crumpled shirt – when he arrived that morning, he was free to tell you why he couldn't wear an ironed shirt. So, you may hypothesize that:

- p2: Sam does not really care about making the efforts you have asked him to make.

Sam could know that you may make this hypothesis since you discussed this subject just the other week, but he has done nothing to prevent you from making it, although he certainly had the control to prevent you from making it. As said, even if he could not wear anything else, *prima facie,* he was free to come up to you to explain why his wearing a crumpled shirt does not mean that p2. And by 'free to do so', I mean that he possessed guidance-control over his omitting to either wear an ironed shirt or explain why he did not. That is, the relevant mechanism kinds leading to his generating your thought that p2 are reasons-responsive: there is a possible scenario where, holding fixed the relevant mechanism kinds (those that are relevant to whether or not Sam is responsible for his generating the thought that p2), Sam would have acted otherwise because he recognized a sufficient reason to do so (e.g. if he thought that he would be fired right away for his neglected appearance). So, it seems reasonable to consider that Sam allowed his behavior to generate in you the hypothesis that p2.

[78] In the preceding chapter, I said that the impetus to start the rationalizing process comes from a certain tension between what the relevant code says and what the Goal Principles require: the information encoded is not enough to rationalize the behavior. As we saw in the preceding paragraph, what a crumpled shirt by itself encodes is extremely vague. There is, though, a pre-established association in Western culture between crumpled shirts and the absence of well-cared-for, well-to-do appearances (this also holds in laid-back restaurants). Sam's not having a well-cared-for, well-to-do appearance is sufficiently in tension with some of your expectations (e.g. that Sam does not want to lose his job) to give the impetus for the mindreading process: why on Earth didn't Sam iron his shirt?

Furthermore, this is a mutually recognizable fact. Sam knows you are here, that you have seen him, and that both of you know that. He could certainly have thought about how he allowed his behavior to generate your thought that p2. It seems also reasonable to consider that Sam allowed this to be mutually recognizable.

So, Sam (i) allows his behavior to generate your thought that p2 and (ii) allows (i) to be mutually recognizable. Thus, you may, with justification, consider p2 to be tentatively added to your common background. 'Tentatively', once again, because there may be alternative, incompatible hypotheses that do a better job at rationalizing Sam's behavior.

Similarly, because you can reasonably hypothesize that p2 is added to the common background and because if p2 is the case, then respecting what you have told Sam is probably not among Sam's top priorities, you may hypothesize that q2:

- q2: Sam doesn't respect you as much as you expected.

Sam could know that you may hypothesize that q2, but he has done nothing to prevent you from hypothesizing q2. This is mutually recognizable and Sam has done nothing to prevent it from being mutually recognizable. So, you may, with justification, consider that Sam allows his behavior to mean that q2, and also add it tentatively to the common background.

This reasoning explains why it is perfectly understandable that you go to Sam and, without giving any evidence for what you are going to talk about, say: 'This is unacceptable! Next time, you are fired.' You can expect Sam to know what you are talking about and what the reasons for your anger and your threat are. The explanation is that, by behaving as he did, Sam has allowed his crumpled shirt to mean that p2 and q2.

I have used the expressions '*prima facie*' and 'tentatively' several times. This is because, even though it is natural for you to hypothesize that p2 and q2 and to take them as part of the common background, there can be reasons to revise your beliefs, reasons you have not thought about.

Imagine, for instance, that Sam replies, 'Oh, I'm so sorry, but what are you talking about?' After a few minutes of further discussion, you learn that Sam is suffering from amnesia and has no idea that you have previously discussed the lack of effort he puts into his appearance. Because of that, Sam may not be able to think about the fact that his behavior would generate your belief that p2. The relevant mechanisms leading to the generation of your belief are thus not reasons-responsive because Sam

lacked the memory necessary for it to be manifest to him that his appearance is too neglected. Accordingly, you should revise your judgment that Sam allowed his shirt to mean that p2. Because your belief that q2 was based on the idea that he allowed the shirt to mean p2, it should also be revised.

Here is another scenario where you should revise your beliefs. Imagine that poor Sam not only had a flood in his apartment, leaving him with only a crumpled shirt to wear for work, but also that, since he started his shift in the morning, he has not had a spare moment to come to you and apologize about his appearance. He considered it more important to be on time at the restaurant than to buy a new shirt or to borrow an ironed shirt from a friend. He also considered that it was more important to first respond to the extremely demanding clients before explaining to you why he was wearing a crumpled shirt. When you came to him saying, 'This is unacceptable', he apologized profusely and explained why he could not have worn an ironed shirt nor talked to you about it beforehand.

What happens, in this case, is that you were wrong about Sam's priorities. Contrary to what you thought, his goals to be on time and to be at the service of the pressing clientele prevented him from wearing an ironed shirt and from explaining to you why he was dressed in the way he was. It was not his lack of care for his job, quite the contrary! His priorities seem to indicate that he cares about his job and, as such, respects you as a boss. You should revise your hypotheses accordingly.

In this case, what happens is that Sam allows his behavior to mean that p2 and q2 until he succeeds in canceling your thought that p2 and q2. Given this exceptional situation, the best way to rationalize his behavior before Sam's explanation may have been to consider that he allower-meant that p2 and q2. As such, it was reasonable to consider that the common background was updated with these propositions. In fact, Sam himself should also have thought that the common background was so updated. Only after the extraordinary situation is revealed, should p2 and q2 be withdrawn from the common background and Sam be considered as actually not allower-meaning them.

Before we move on, let me briefly address the following question: in the original scenario where Sam allows his shirt to mean that p2 and q2 and has no excuse, is that communication? Remember that, by communication, I mean the transfer of information where both the sending and the receiving were designed for that purpose (see Chapter 1). Sam's crumpled shirt certainly is not a signal, i.e. a sign designed for communication, and

he does not intend to communicate with it. His wearing a crumpled shirt is not designed to carry the information that p2 and q2. However, the crumpled shirt can well be conceived of as *the absence of a signal*, where the signal would be the *ironed* shirt. Because ironed shirts are conventionally associated with a certain standing in the West, they are cultural signals: stimuli which have been designed, through cultural evolution, to carry information about the ones who wear them. The information in question may be something like 'I care about, and put efforts into, my appearance' and/or 'I respect a certain (Western bourgeois/aristocratic) dress code'. Consequently, even if we don't want to call Sam's behavior 'communication' *stricto sensu*, we need to recognize that it is very close to it: the information sent by Sam, and the fact that you receive it, are both explainable because of the communicative function that ironed shirts have; it is the absence of this signal that grounds the information transfer. The information transfer piggybacks on communication, on the absence of a communicative sign.

I would not mind calling this phenomenon *communicating by omission*, to be understood on the model of homicide by omission. One who kills by omission does not intend to kill. It is the *absence* of certain behaviors that resulted in the homicide. For the person to be responsible, she must have possessed the guidance-control necessary to prevent the homicide. Similarly, one who communicates by omission would not intend to communicate but would have the guidance-control necessary for preventing the information transfer. It would be the unintended absence of certain behavior that would result in the information transfer. The way we respond to such unintended behavior is very close to the way we respond to intended crimes (e.g. through blame and punishment). Think about Sam's case, or think about the fact that people may be blamed for their offensive language although they do not intend it to be offensive. They can nevertheless be held responsible for it because they were free not to be offensive.

3.4. BOB'S INAPPROPRIATE REMARK AND UNINTENDED MEANING OF SPEECH ACTS

Before I turn to the conclusion, let me briefly present an example where it is apparent that the EGM can apply to stimuli overtly intended for communication in ways that are not available to the prevailing Gricean models. This example shows that the extension of EGM does not primarily concern the type of stimuli, but rather it concerns the effects.

Bob, one of my philosophy students, uses an inappropriate expression in my class, for instance a term deemed extremely rude. Bob doesn't have any intention to be disrespectful. The other students and I know that. We know that Bob does not mean (speaker-mean) anything disrespectful. Nevertheless, given the circumstances, Bob transmits the information that he is being disrespectful.

Does Bob transmit the information that he is being disrespectful because he produced a signal that is statistically or conventionally correlated with a disrespectful behavior? I don't think this proposal explains everything, as we will see with Bob*'s case below. A more potent explanation, it seems to me, is that we can reasonably take Bob to allow his remark to mean something disrespectful.

Here is the reasoning. I assume that we can expect university students to be aware of what terms are recognizably rude. Accordingly, it seems reasonable to consider that it is mutually recognizable for Bob and his audience that using the expression in question can generate the belief that he is being disrespectful. The possibility of such effects may not have crossed his mind when he used the expression, however. Let us assume that he did not think about it. Still, it is reasonable to expect that the effects are *recognizable* for him, even if not recognized at the moment, and that they are mutually recognizable in this university class. Bob certainly had access to the relevant information. Even if the offensiveness of Bob's remark did not go through his flow of consciousness, even if the corresponding neurons did not fire at this moment, Bob can nevertheless be considered as *able,* as having the right preconditions, to recognize the offensiveness.

If we assume that Bob doesn't suffer from any psychological issue that would prevent him from controlling what he utters, we can take the potentially disrespectful effects to be controllable. In fact, they should be mutually recognizable as controllable, since Bob's remark was not only deliberate but publicly deliberate. Thus, we can suppose that the rude expression is a stimulus with potentially disrespectful EMRAC.

Taking into account the common background and the Goal Principles, the audience could ask: why did Bob use an expression with a recognizably rude connotation instead of a neutral one with the same reference? If Bob cared enough about avoiding being disrespectful, if that was an important goal of his, Bob would not have used this term. So, Bob does not care that much about being respectful. Even if Bob did not speaker-mean it, Bob allowed his remark to mean disrespect.

To see why a merely semantic, code-based account probably is insufficient to explain the relevant effect, consider an alternative scenario where Bob* suffers from Tourette's syndrome. Because of this condition, he cannot control his uttering of inappropriate expressions. In this situation, we would not consider Bob* to transmit the information that he is being disrespectful, and the way he would update our common background would thus be different. Importantly, this is the case even though the pre-established pairings between the rude expression and disrespectful behavior is the same as in the preceding scenario. The difference in the information transmitted by Bob and Bob* thus seems not to be accountable for by a code model.

3.5. Conclusion

In this chapter, I have explored various examples where I believe that the EGM may prove useful and which involved different kinds of stimuli: nonverbal affective signs (§3.1), the sound of one's voice (§3.2), clothing (§3.3), and speech acts (§3.4). In all of these cases, neither the code models nor the prevailing Gricean models would yield the interpretation available to the EMG. This is so because (a) the pieces of information which we discussed could not be inferred merely based on pre-established codes, i.e. merely on (super-)semantic[79] grounds, and because (b) the information in question was not intended to be conveyed, at least not in the way necessitated by the prevailing Gricean models (i.e. speaker-meant).

Although I gave a rather detailed explanation for the first couple of examples, I was quick going through the details of the later ones. I thus have left many claims undefended and have not pursued a myriad of interesting threads. I must leave such tasks for future works. In any case, I hope to have convinced the reader that the EGM constitutes an interesting tool for the exploration of areas that have not yet been analyzed in this way. And by 'in this way' I mean, roughly, a Gricean way, contrasted to approaches that may have explored the same territory based on rather different theoretical assumptions, such as Barthes' semiotics (see the epithet of this chapter, which concerns clothing) or Bourdieu's sociology. I know that many details would need to be filled in to transform the intuitions presented here into precise predictions or to back the intuitions up with more solid justifications. But, as it stands, as I put it in the last chapter's conclusion, I hope that the model already has proven its capacity to serve as a space probe satellite.

[79] See footnote 2 and Schlenker (2018) for what super-semantics is.

The kind of analyses proposed in this chapter, strongly inspired by the Gricean derivation of implicatures, is not an exact science, of course, but it does not seem to me to be less precise or less predictive than that proposed by Grice (1989: 31ff). Nevertheless, there is much room for improvement – just as the work Grice first elaborated in the 1950s and 1960s has itself been much improved since. For instance, we could try to find pragmatic principles that are more precise, more predictive, and empirically more implementable than the Goal principles. This could be done by focusing on specific kinds of implicatures*, e.g. by following the lead of Horn (1984) on scalar implicatures. We could also try to find principles that apply to a more restricted domain than the immensely vast stimuli with EMRAC, for instance by drawing on Brown and Levinson's (1987) work on politeness. We could also try to rigorously formalize aspects of the EGM to connect it with dynamic compositional semantics (e.g. Portner, 2007; Roberts, 1996). Another important improvement would be to investigate the mental mechanisms at play and connect what I have said here with the (neuro-)psychological literature, or more generally with the cognitive sciences, as Relevance theorists have insisted since the 1980s.[80] However, I must postpone all these important improvements for future work and leave this first sketch of the EGM.

[80] Whether or not the Gricean derivation of implicatures is a kind of explanation that qualifies as psychologically realistic, targeting actual brain processes, is a debated issue. Grice himself considered his theory as 'rationalist' rather than 'psychological' (1989: 29). Sperber and Wilson (1986), as well as other Relevance Theorists, have forcefully argued that Grice's theory is psychologically unrealistic They propose a 'post-Gricean' theory, which is meant to be more coherent with what we know from the cognitive sciences. By contrast, Dänzer (2020) argues that Grice's and neo-Griceans' explanations *are* psychologically realistic.

4. Allowism and the Meaning of Narrative Artworks

« 'When *I* use a word,' Humpty Dumpty said, in rather a scornful tone, 'it means just what I choose it to mean — neither more nor less.' »

– Lewis Carroll, *Through the Looking Glass*

Abstract. In this chapter, I will show how the Extended Gricean Model may be an interesting tool to interpret the meaning of narrative artworks (understood broadly as including entertainment works). In the first part (§4.1), I present allowism: the claim that the meaning of literary works should be identified with what authors allow their works to mean. I argue that it should be considered at the very least as a live candidate to choose from when trying to identify what kind of meaning literary meaning is. To do so, I will concentrate on a comparison between allowism and hypothetical intentionalism – an influential theory on the meaning of literary works defended by Jerrold Levinson and which makes predictions similar to allowism. In the second part (§4.2), I argue that, whether or not the meaning of narrative artworks should be identified with what the authors allow their works to mean, this construct nevertheless helps to account for certain intuitions we may have about what messages are being unintentionally transmitted with artworks.

4.1. Literary works (Is Dumbledore gay?)

Jerrold Levinson begins the first section of his influential article 'Intention and Interpretation in Literature' as follows:

> « When we ask ourselves what literary texts mean and how they embody such meaning as they have, I think there are only four models to choose from in answer. One is that such meaning is akin to word sequence (e.g., sentence) meaning *simpliciter*. Another is that it is akin to the utterer's (author's) meaning on a given occasion. A third assimilates it rather to the utterance meaning generated on a given occasion in specific circumstances. A last model pictures it, most liberally, in terms of what might be called ludic meaning. » (Levinson, 1996, p. 176)

Word sequence meaning is semantic, encoded meaning. By 'utterer's (author's) meaning', Levinson refers to what the author intends the work to mean and that I have called speaker-meaning in the previous chapters. Utterance meaning is the meaning that the appropriate (or 'ideal') audience arrives at 'by aiming at utterer's meaning in the most comprehensive and informed manner we can muster as the utterance's intended recipients.' (Levinson, 1996, p. 178). Ludic meaning is the meaning one can arrive at through 'interpretative play constrained only by the loosest requirements of plausibility, intelligibility, or interest' (177).

Levinson then advocates that, among the four kinds of meaning, utterance meaning is the one that is the closest to what literary interpretation practices normally take literary texts to mean.[81] The first point that I want to make here is that even if we agree that his arguments exclude the other three kinds of meaning defined above, they do not exclude allower-meaning[82] as a live option.

Levinson first argues that the meaning of literary texts cannot be equated to word sequence meaning mainly because '[w]e don't treat literary texts the way we would random collections of sentences, such as might be formed in the sands of a beach or spewed out by computer programs.' (177) Instead, we presume literary texts to issue from a single mind or several minds working together. We presume them to be the product of agents with certain purposes who engage in an act of communication, widely construed (177). This argument can also be used to exclude ludic meaning because the latter is not constrained by postulating that the text is produced by an agent with certain purposes. The ludic meaning of a poem is arrived at through the same processes as the ludic meaning of sentences generated randomly by a computer, or of a grocery shopping list: ludic meaning just is the meaning that is most interesting or fun. Because we do interpret literary texts differently than random sentences or grocery shopping lists, word sequence meaning and ludic meaning are not appropriate candidates for what literary texts mean.

This argument, however, does not exclude allower-meaning. Attributing allower-meaning to a text is attributing certain purposes, certain goals, to its author(s), and these goals certainly can be communicative, in the broad sense of the term. Even though we don't take literary texts to communicate practical information, we assume that their authors have goals such as sharing something of relevance (e.g. an aesthetic impression, what is

[81] By 'normally' he excludes certain practices such as 'the excesses of deconstructionist theory' (175).
[82] See Chapter 2 for this notion and a presentation of the Extended Gricean Model.

moving in a story, an ironical thought, a sense of the absurd, etc.), provoking certain effects in the audience (e.g. enjoyment, admiration, a developed political sensibility, etc.), or both. Considering that an author has allowed a text to mean certain things by assuming that she is rational – in particular: that she is willing and able to respect the Goal Principles – is different from interpreting a text as a mere sequence of words and coheres with the assumptions we have when interpreting a literary text.

Levinson then argues that literary meaning cannot be equated with utterer's meaning because doing so would dissolve the distinction between normal linguistic activity and literary interpretation. He gives the following, telling, example:

> « When a poet vouchsafes us, in plain language, what some enigmatic poem of his might mean, we don't react by then discarding the poem in favor of the offered precis. » (177).

In contrast, in ordinary verbal communication, we *do* discard obscure, unclear signals in favor of clearer explanations. For instance, if I send you a text message that you don't understand and I then send a new text stating 'What I meant was that we should meet at the supermarket at 6 pm', you will naturally discard the first, obscure, text message and assume that whatever I meant in it is that we should meet at the supermarket at 6 pm. In contrast, even if the author of a poem vouchsafes an interpretation, we may rationally come back to the poem and try to interpret it in our own way. We may rationally form different hypotheses on what its very words could mean beyond what the poet officially declared. For instance, we may explore what kinds of connotations, absent from the poet's explanation, could be found in some particular phrasing, in its sounds, in its place in the general structure of the poem, etc. And we can dispute the poet's meaning as well. As Levinson puts it, interpretation in a literary mode, contrary to normal practical linguistic activity, involves holding the text 'to have a certain amount of autonomy, to be something we interpret, to some extent, for its own sake, and thus not jettisonable in principle if we could just get, more directly, at what the author had in mind to tell us.' (177).

This argument, if correct, forces us to distinguish the utterer's meaning (or, as I call it, speaker-meaning[83]) from the meaning of the literary text. However, allower-meaning, once again, is not excluded as a legitimate candidate. We may well search for meaning beyond what the author declares to be the intended meaning of a poem by searching for its allower-

[83] See Chapter 1 and the Appendix for this notion.

meaning. An interpretation based on what a poet allows its work to mean is much freer than one based on what the poet speaker-means. It is freer than speaker-meaning, but it is nevertheless constrained by some assumptions. These assumptions include that there is an (imperfectly) rational agent behind the creation of the poem, i.e. an agent with certain cognitive abilities, certain goals, and a capacity to use these abilities to avoid goal-obstructive events and to seek goal-conducive ones, *ceteris paribus*. It is also constrained by the common background between the authors and their potential audience(s) (see below).

So, if what precedes is correct, the arguments given by Levinson force us to reject an identification of literary meaning with either word sequence meaning (i.e. semantic, encoded, literal meaning), ludic meaning (i.e. constraints-free interpretations), or utterer's meaning (i.e. speaker-meaning). This leads Levinson to conclude that we should identify literary meaning with utterance meaning. This is what leads him to develop the theory he calls 'hypothetical intentionalism'. However, his conclusion is based on the hypothesis that only four kinds of meaning are available. Even if we accept his arguments, identifying literary meaning with allower-meaning remains a live option – let us call this option and the potential theory we could develop around it *allowism*.

What is the difference between hypothetical intentionalism and allowism? Well, roughly, hypothetical intentionalism claims that the meaning of a literary work is to be identified with the best hypothesis that an adequately informed (or ideal) audience would form about the author's semantic intentions. By 'adequately informed', Levinson means that the audience should take into account 'the intrinsic features of the text, the operative conventions and norms of the language and genre involved, and a number of author-specific though public contextual factors as well' (1996, p. 206). Given this audience, the meaning of a literary text is recovered by, and only by, aiming at the author's intents.

Allowism, on the other hand, would not aim at what authors have *intended*. It would rather ask us to focus on what authors allowed their work to mean, and so would require literary interpretation to aim beyond authorial intents. More precisely, allowism would defend that the meaning of a literary work W is to be found in (i) the effects that the authors allow W to generate in their audience given that (ii) the authors allow W to make (i) mutually recognizable between them and their audience. Because of the way I use 'to allow' (see Chapter 2), the audience in question is an audience

which the authors could reasonably be expected to be able to think about during the writing of their text.[84]

I don't want to defend here that allowism is superior to hypothetical intentionalism. Nevertheless, I believe that it is an alternative that is worth taking seriously. In other words, it may be worth aiming at allower-meaning in literary interpretation.[85]

To see more concretely what some of the differences between hypothetical intentionalism and allowism are, let us take a simple example and ask: is Dumbledore gay?[86]

In the Harry Potter book series, written by J. K. Rowling and published between 1997 and 2007, Dumbledore is an old man, a professor, and director of the wizardry school where Harry studies. Despite the centrality of this character, not much private information is given about him. In particular, we have no information whatsoever about his romantic life.

Now, here is an extract from a public lecture given by J. K. Rowling at Carnegie Hall in 2007:

> « – [A member of the audience:] Did Dumbledore, who believed in the prevailing power of love, ever fall in love himself?
>
> – [J. K. Rowling:] My truthful answer to you… I always thought of Dumbledore as gay. »[87]

Since hypothetical intentionalism claims that the meaning of a work is to be found by aiming at the author's intents and that we should do so by taking into account author-specific, public contextual factors – among which we should certainly include public lectures given by the author – it seems to me that, according to this theory, it is part of the meaning of Harry Potter that Dumbledore is gay even though there is nothing in the books to ground this interpretation. To be fair, Levinson does state that

[84] This ability would be determined just like guidance-control: by holding fixed the mechanisms relevant to the ascription of responsibility and figuring out whether there are possible scenarios where the author thinks about this audience for a reason available to her (see Chapter 2).
[85] Maybe both point to equally legitimate attitudes one can have while reading literature.
[86] Thanks to Steve Humbert-Droz for encouraging me to explore this example. Another useful example that he gave me is the following. In *Blade Runner*, whether Deckard (Harrison Ford) is a human or a replicant is ambiguous. Years after the release of the movie, director Ridley Scott declared that Deckard actually *is* a replicant. Does this declaration affect the meaning of the movie and what its correct interpretation should be?
[87] Retrieved from http://www.the-leaky-cauldron.org/2007/10/20/j-k-rowling-at-carnegie-hall-reveals-dumbledore-is-gay-neville-marries-hannah-abbott-and-scores-more on 12 July 2020.

'the author's direct pronouncements will be given *minor* weight' (208), but, despite this, he insists that there needs to be only one possible interpretation and it is reached by aiming at the real intentions of the author, given – roughly – a public context, something like a common background between authors and their intended audience. So, arguably, it is part of the meaning of *Harry Potter* that Dumbledore is gay, according to hypothetical intentionalism.

According to allowism, on the other hand, whether Dumbledore is gay is indeterminate. This is because J. K. Rowling did not allow her books to mean either that Dumbledore is gay or that he is not, as I will now explain.

Concerning the belief that Dumbledore is gay, Rowling can be considered as respecting clause (i) of the definition of allower-meaning, but not clause (ii). She can be taken to respect clause (i) because she could have – and probably has – thought that her book may generate the belief that Dumbledore is gay. Indeed, this effect was sufficiently manifest to her at the time of her writing – she declared that she has always thought of Dumbledore as gay – and she did not prevent the audience to think so in any way. However, we cannot reasonably take her to have allowed (i) to be mutually recognizable between her and her audience, and so cannot take her to respect clause (ii). This is because, since there is no evidence in the books that Dumbledore is gay, it is not reasonable to consider that her public had enough information for the thought that Dumbledore is gay to be manifest to them. So, her allowing the books to generate this thought cannot be taken as mutually recognizable by her and her audience.

Since there is no evidence in the books that Dumbledore is not gay either, Rowling also did not allow her books to mean that he is not gay. So, if we follow allowism and equate literary meaning with allower-meaning, then, as part of the semantic content of *Harry Potter*, it is indeterminate whether Dumbledore is gay.

One may reply: but wait! Since Rowling declared that Dumbledore is gay at a public lecture, shouldn't we take that into account as part of the common background between her and her (dedicated) readers, so that she *did* allow her books to mean that Dumbledore is gay? Indeed, the Extended Gricean Model claims that what is allower-meant must be inferred in light of the common background, which includes all the information that is mutually recognizable as shared between senders and receivers. Why wouldn't her declaration be taken into account?

The answer is that we are looking at what Rowling allowed the *Harry Potter book series* to mean. And remember that, by 'allowing x to F'[88], the temporality is important: x is a set of stimuli produced between t0 and t1 – in our case: between roughly 1995 and 2005 while she wrote the books. One of the necessary conditions for allowing is that the sender (Rowling) could reasonably be taken to have known *at t1* (in, say, 2005) that x may have the relevant effect (in our case: making it mutually recognizable that Rowling allowed her book to generate the belief that Dumbledore is gay). And, at t1, when she finished writing the books, she certainly could not reasonably take her audience to be able to know that she considered Dumbledore to be gay to be mutually recognizable.[89]

Now, consider the following case which further illustrates the temporal element. In 2018, the movie *Fantastic Beasts: The Crime of Grindelwald* was released. It is a spin-off and prequel of the original Harry Potter story and its script was written by Rowling. It focuses on Grindelwald, a wizard with whom Dumbledore was in love according to what Rowling declared in the same public lecture that I quoted above. In this movie, there are scenes where young Dumbledore and Grindelwald make a blood pact that prevents them from dueling each other. According to allowism, should we or should we not consider that it is part of the meaning of these scenes that Dumbledore is making a pact out of romantic love for Grindelwald? The scenes, by themselves, certainly do not allow a naïve audience to believe that the pact is made out of romantic love (it may be out of a non-romantic friendship instead), but since Rowling declared in 2007 in a public lecture that Dumbledore was in love with Grindelwald, and since the movie was released in 2018, should we consider this piece of information as being part of the common background? While Rowling did not think that each and every spectator of this movie would have access to this information, she certainly knew that some of her dedicated fans would. And, as far as these fans are concerned, Rowling certainly allowed the scenes to mean that Dumbledore makes a pact with Grindelwald out of romantic love.

[88] As a reminder, here is how I defined 'to allow' in Chapter 2:
A sender S allows the stimuli x – made of individual stimulus <x1, x2, …, xn>, produced by S between t0 and t1 – to generate the effect e (doxastic, affective, evaluative, behavioral, …) on the actual or conditional audience R if, and only if,
(a) S had guidance-control over the production of e between t0 and t1, and
(b) It was manifest to S between t0 and t1 that S may generate e in R with x.

[89] Note by the way that this argument about the time of production does not change what is predicted by hypothetical intentionalism because Rowling stated that she has always thought of Dumbledore as gay, and so presumably she has always had the intention that Dumbledore be considered as gay, which is the intention that a hypothetical intentionalist must track.

So, according to allowism, the meaning of the movie possesses several 'layers' that differ from one audience to the next, depending on how much information the author considers these different audiences have.[90] I find this consequence to be worthy of further exploration and, potentially, to lead to interesting new questions in the debate around intentionalism, hypothetical intentionalism, and anti-intentionalism (for a review, see Irvin, 2006). For such reasons, it seems to me that allowism may be profitably explored.

Furthermore, the predictions made by allowism may be more in line with certain interpretive literary practices than those made by hypothetical intentionalism. This also constitutes a reason to explore further the consequences of this theory. It would be interesting to see how it deals with several case studies – and, in particular, with some that are a little more profound than the simplistic example I have discussed.

4.2. TRIVIALIZATION IN GAME OF THRONES

Whether or not allowism is a good theory of what literary interpretation should aim at, the Extended Gricean Model can in any case help us understand certain claims we might want to make about messages conveyed by art and entertainment works.

Take the following example from an episode of the TV series *Game of Thrones* (GOT for short), which was aired on 5 May 2019: [91]

- Context: Sansa is a young woman who has been manipulated and emotionally abused by Littlefinger, and who has been tortured and raped by Ramsay. The Hound is an old, rather friendly, acquaintance of hers who knew her as a young teen. This scene depicts their reunion. « – The Hound: 'You've changed Little Bird. … None of it would have happened if you had left King's Landing with me. No Littlefinger, no Ramsay, none of it.' – Sansa: 'Without Littlefinger, Ramsay, or the rest, I would have stayed a Little Bird all my life.' »

[90] This phenomenon is similar to 'dog-whistling politics' where someone uses coded or suggestive language that can be understood only by a fraction of the audience, e.g. to attract the intended political audience without provoking anger from opposing audiences (see Saul, 2018).

[91] Another example that I could have used it the following scene from *Blade Runner*: Deckard tries to kiss Rachel. She refuses multiple times. Deckard then violently forces her to kiss him. She ends up acting like that is how she wanted him to act. The way the actors play the scene, the camera angles, and (importantly I find) the clichéd, suave saxophone music conspire to make the scene look like a normal romantic scene, and not like a sexual abuse that could have turned into rape. However, the fact that the GOT episode was aired in 2019, while *Blade Runner* was released in 1982, makes the GOT example more evident.

(Transcribed from the HBO series *Game of Thrones*, Season 8, Episode 4, 2019)

Amy Collier, a writer and journalist, commented on this scene with the following Tweets:[92]

> « Sansa didn't need to go through all that trauma to become a powerful, intelligent person, and the show implying she did is just...ugh »

> « that exchange between The Hound and Sansa was definitely written by a man. For the millionth time, rape and abuse of women isn't just character development or a way to show a female character has matured »

> « And I'm actually not saying I dislike that Sansa is a survivor. In some ways I'm glad there's a representation of a character who is. But there is a way to write it without falling back on lazy tropes. That writing was bad. »

I concur with Collier in thinking that it is reasonable to reproach the GOT writers for (unintentionally) sending the following message (or, as she says, they *imply* it, c.f. the first Tweet):

> (p) Sansa needed to go through the trauma to become a powerful, intelligent, mature person.

Furthermore, Collier also probably interpreted this scene as sending something like the following message (I read this from the 'ugh' in the first Tweet and from the second Tweet):

> (q) Being sexually abused can be turned into an empowering event, can make you stronger.

We may even argue that Collier interpreted the scene as sending the following message, an interpretation with which I would concur:

> (r) Game of Thrones writers do not greatly care about avoiding a trivialization of rape.

By 'trivialization of rape', I mean contributing to a conception of rape where its hurtfulness is minimized (Zillmann & Bryant, 1982). In this particular case, this would be done through the use of rape as an explanation of how a female character has matured, thus reinforcing the idea that being raped

[92] Tweeted on 6 May 2019, retrieved on 14 July 2020 from https://twitter.com/Amy_Corp/status/1125241818474528768.

makes you stronger and so that there is something positive to be found in rape.

The Extended Gricean Model may be used to explain why it may be reasonable to take GOT writers to send p–r with scene (7). This is so because we may reasonably consider that:

(i) GOT writers allow this scene to make certain people think that p–r.
(ii) They allow this scene to make (i) mutually recognizable.

That (i) and (ii) holds for p and q seems quite obvious since these pieces of information (or something similar) are strongly implicated by Sansa's line, 'Without Littlefinger, Ramsay, or the rest, I would have stayed a Little Bird all my life.' In fact, it is probably not only allower-meant, but speaker-meant by the GOT writers. In addition, it seems reasonable to me to consider that (i) and (ii) also hold for r because, as Collier puts it, the claim that rape and abuse of women should not be used merely as a way to show that a female character has matured is a claim that has been repeated a million times. I take it that it is thus reasonable to take accusations about the trivialization of rape to be in the common background of GOT writers. So, if GOT writers greatly cared about avoiding trivializing rape, wouldn't they have written the script differently? It seems indeed reasonable to consider that, when they wrote this scene, the GOT writers could have thought about p–r, could have thought that certain people would know this and would think that they allowed making these pieces of information manifest. And they could have thought that all of this may well be mutually recognizable between them and the relevant audience (e.g. an audience sensitized to this issue). But the writers nevertheless did not do anything about it. So, they transmitted the pieces of information p–r by allowing this scene to mean that p–r.

Let us think about a contrasting example. Imagine that the same story was written in a culture that had no awareness that the trivialization of rape was an issue, such as in Ancient Greece (Omitowoju, 2002). In this context, we may understand the scene as sending messages p and q, but certainly not r. Sending messages p and q in such a context can already be considered as blameworthy, and as participating in an unacceptable trivialization of rape, but, I take it, it is certainly not as bad as doing so in 2019 United States of America (USA). This is because in 2019 USA people have been alerted to concerns around the trivialization of rape many times. The way in which we may interpret and comment on, say, *Lysistrata* by

Aristophanes (written circa 411 BC), a play where rape is trivialized[93], is and should be different from the way in which we may interpret and comment on the GOT writers' scenario. The GOT writers have access to claims against the trivialization of rape that were not available to Aristophanes. The fact that they do not take these claims into account, although they certainly are available, means that they communicate by omission things that were not communicated by omission by Aristophanes since omission implies a failure to act in a way that can reasonably be taken as available and Aristophanes cannot be taken as having access to the relevant claims. To use the terminology from Chapter 2, Aristophanes did not have the guidance-control necessary for allowing his piece to mean that r.

Again, I am not saying that Ancient Greek writers shouldn't be blamed for perpetuating rape culture, only that GOT writers have more responsibility given the contemporary context. The difference between, on the one hand, the messages sent by GOT writers and, on the other hand, by writers of the same story in a different context is an important aspect of how we may interpret art and entertainment works. And the difference in question cannot, I take it, be accounted for as comprehensively by the prevailing Gricean models or by the code models. This is because, on the one hand, some of the messages in question were not intended[94] and, on the other hand, it is hard to see what pre-established code shared by writers and audience could predict that the GOT writers would send the messages in question (and that, in a context where trivialization of rape is not thematized, the authors would not send r).

Conclusion

In this chapter, I have tried to show how the Extended Gricean Model may prove to be an interesting tool for the interpretation of narrative artworks. In the first part, I have done so by showing that the arguments used to defend hypothetical intentionalism – one of the main contenders when it comes to identifying the meaning of literary works – can be used to argue for the validity of *allowism*, a theory according to which literary meaning should be identified with what authors allow their work to mean to the

[93] See the following academic blog post for more information about rape culture in Ancient Greece and a description of the relevant scene in Aristophanes' play *Lysistrata*: https://womeninantiquity.wordpress.com/2017/12/06/consent-and-rape-culture-in-ancient-greece/

[94] Or, more precisely, the ideal audience would not hypothesize that the writers would have intended their work to be interpreted in that way.

relevant audience. In the second part, I have tried to show how the Extended Gricean Model can make predictions that correspond to interpretations of what information is transmitted by narrative works.

Let me observe that the arguments presented in this chapter could easily be applied to other representational artworks such as pictures or statues. With the appropriate modifications, I believe that they might also be applied to the interpretation of abstract arts, such as non-representational painting or instrumental music – but I leave for future work the task of spelling out what exactly should these modifications be.

PART 2 – WHAT EMOTIONAL SIGNS MEAN

5. THE MEANING OF EXPRESSIVES

> « The aim of a lyrical poem in which occur the words 'sunshine' and 'clouds' is not to inform us of certain meteorological facts, but to express certain feelings of the poet and to excite similar feelings in us. »
> – Rudolph Carnap, The Rejection of Metaphysics

Abstract. This chapter starts by spelling out three features that importantly distinguish expressives – by which I mean utterances whose aim is to express emotions and other affects – from descriptives – utterances whose aim is to describe the world truthfully (§5.1). Drawing on recent insights from the philosophy of emotion and value (§5.2), it then shows how these three features derive from the nature of emotions, understood as felt, bodily, value-tracking attitudes (§5.3). It then clarifies how views from speech act theory allow us to claim that expressives inherit their meaning from the nature of affects (§5.4).

In the chapter following this one, I will give three accounts of what it takes to understand expressives, and thus explore further what the meaning of expressives consists in. Another main aim of the present chapter is therefore to introduce the framework that I will use in the next chapter.

5.1. EXPRESSIVES VS DESCRIPTIVES: SOME INTUITIONS

Supposing that utterances (1)–(3) and (4)–(9) respectively refer to the same phenomena, compare groups A and B.

Group A:

(1) Outrageous!
(2) Ouch!!!
(3) The frogs won it again!

Group B:

(4) What the government did was wrong.
(5) I feel outraged by what the government did.
(6) This boiling oil has burned my hand and this is bad for me.
(7) I feel great pain.
(8) The French won the world cup again and I believe that the French are contemptible.

(9) The French won the world cup again and I feel contempt towards the French.

Even if we take into account the fact that the utterances in B are about the same states of affairs as the ones in A, there still is an intuitive sense in which they do not mean the same thing: that the meaning of (1) is not exactly that of (4) nor (5), that the meaning of (2) cannot be reduced to the meaning of (6) nor (7), and that the meaning of (3) is somehow different from that of (8) or (9). The kind of meaning found in group A is usually called 'expressive meaning' and the corresponding utterances 'expressives'. The kind of meaning found in group B is called 'descriptive meaning' and the corresponding utterances 'descriptives' (see Wharton, 2016 for a review of 20th-century studies on expressives).

But wait! What exactly do we mean by 'kinds of meaning' here? Is there a theoretically cogent way of making the distinction between expressive and descriptive meaning? If so, what is it? Shouldn't we distinguish expressives and descriptives otherwise than through kinds of meanings? For instance, through syntactic particularities? Or with the tone of voice employed? Before I try to answer these questions, let us make a few preliminary remarks.

To start with, observe that the distinction between expressives and descriptives is not as sharp as we may initially think, as there are cases where the two seem to blend. Consider this sentence from a letter by María Casares to Albert Camus (March 1952):

« When I think of us it seems absurd to not believe in eternity. »

This sentence neither falls completely on the descriptive side – because it makes Casares' passion so clearly apparent – nor completely on the expressive side – because it is, after all, a description of her thoughts that presumably is literally true.

This example also indicates that sentences may possess an expressive meaning even though they have the linguistic form of descriptives, i.e. without exclamation marks, swear words, marked prosody, syntactic inversions, or any other linguistic markers of expressives. [95] So it seems

[95] Even though there are conventional ways of encoding that an utterance is an expressive. For instance, wh-exclamatives do so through syntactic forms such as « How [ADJECTIVE] ! » or « What a [NOUN] ! » (Chernilovskaya et al., 2012; Foolen, 2012). Observe that we may draw the distinction between expressives and descriptives on two levels: a semantic level, where expressives differ from descriptives because of their encoded, literal meaning, and a pragmatic level, where expressives differ from descriptives because of what they are meant to communicate. As said, this chapter is rather concerned with the second level.

impossible to make the distinction we are after based on the linguistic form, merely through syntactic, prosodic, phonological, lexical, or morphological structures. We can even imagine that any linguistic item might, given a certain context, become an expressive. A sentence as vapid as 'The boat has departed' even if pronounced on a neutral tone could be used, provided a tragic background, to mean something very close to an emphatic exclamation such as 'Alas, how regretful I feel!'. The distinction we are interested in then is not about the linguistic form, but about *what is meant by people using these utterances*. In other words, we are here interested in what speakers mean rather than in the conventional meaning of words (Grice, 1968), because there is no way to exhaustively distinguish expressives from descriptives merely based on what is conventionally encoded in words. As a consequence, although the way in which sentences in group B are to be construed here is as descriptives, the same string of words could very well be construed as expressives provided certain background conditions.[96]

With these important warnings in mind, let us review three intuitive considerations – which I shall discuss in more detail below – that we may think support the distinction. They will also serve as important benchmarks for our effort to account for the nature of expressives in the next chapter.

(a) *Hot vs cold.* Expressives always appear to convey affects (emotions, desires, moods, sentiments, pleasures, pains, whims, etc.), which is not true of descriptives. In contrast, descriptives seem to convey beliefs or other doxastic attitudes that the speaker might have (doubt, supposition, conjecture, etc.) about how the world is. As such, and to use a common metaphor, descriptives seem to communicate mental states which are 'cold' as opposed to the so-called 'hot' affective states – we will see below what grounds this metaphor. So while the meaning of (4)–(9) is of course tightly linked with affects, the type of meaning to which they belong – descriptive meaning – need not be. By contrast, there are no examples of expressive meaning which are completely detached from affects: expressives always have the function of expressing the expresser's *affects*, whether they are sincere and really do so or whether they piggyback on the sincere occurrences and are expressives because they imitate them.

This is why I won't focus on formal, conventional features which encode expressive semantic meaning, but rather on how words are used.

[96] They could also be expressives because of the way they are pronounced, e.g. 'What the government did was wrong' said in an overtly angry voice.

(b) *Appropriate vs. true*. Truth and falsity are the normative standards by which descriptives are evaluated. This however does not seem to be the case for expressives.[97] Compare for instance (2) and (7) while imagining that the person doesn't feel pain. We would say of (7) that it is literally false, but not of (2) (Kaplan, 1999). Similarly, whether we think it appropriate or not to use the word 'frog' as in (3), it seems independent of whether we take the sentence to be true or false. Or think of 'Congratulations!' vs. 'The speaker is congratulating the addressee': the latter may be true or false, but not the former. Expressives, like the affects that they convey, seem to answer normative standards of (in)appropriateness, (un)meritedness, or (un)deservedness, while descriptives, like the doxastic attitudes they communicate, appear to answer the normative standards of truth and falsity.

(c) *Direct vs indirect.* While descriptives can and do sometimes convey affects, expressives, distinctively, do so directly. If one describes one's affects, as in (4)–(9), one in fact communicates a thought one has about an affect, a thought that is typically hidden from the audience. When expressing an affect, however, as in (1)–(3), it seems that one directly shows the affect, or at least some of its components (e.g. facial, vocal, and gestural expressions, certain action tendencies, certain verbal behaviors). Expressives thus constitute a communicative path that is more direct than the one taken by descriptions about affects, which represents a doxastic attitude of the communicator. One way of putting this direct vs. indirect distinction is to say that in descriptives (4)–(9),

[97] Note however that some authors working on the semantics of slurs (Bach, 2018; Diaz-Leon, 2020; Hom, 2008; Hom & May, 2013, 2018; Lycan, 2015; Schlenker, 2007; Williamson, 2009) argue that these expressions, which are usually thought of as possessing an expressive meaning, nevertheless are entirely accounted for by regular truth conditions. This view however is rejected by what appears to be a majority of the philosophers working on slurs (Anderson & Lepore, 2013; Camp, 2013, 2018a; Cepollaro & Thommen, 2019; Copp, 2009; Croom, 2011; García-Carpintero, 2017; Hedger, 2012; Langton et al., 2012; Marques & García-Carpintero, 2020; McCready, 2010; Potts, 2007; Richard, 2008; D. J. Whiting, 2008). Furthermore, observe that what speakers mean by an utterance may be expressive even if all the terms composing it have a descriptive semantics, i.e. a literal meaning that is easily accounted for by regular truth-conditional semantics. In such cases, the truth-conditional content cannot account for all that is meant. Think for instance of Captain Haddock's insults: 'Bashi-Bazouk', 'Visigoths', 'sea gherkin', 'anacoluthon', 'pockmark', 'steam rollers', 'vegetarians', 'floundering oath', 'carpet seller', 'blundering Bazookas', 'pinheads', 'ectomorph', 'pickled herring', 'freshwater swabs', 'molecule of mildew', 'logarithm', 'orang-utans', 'cercopithecuses', 'fancy-dress freebooter', 'dizzard', 'black-beetle', or 'pyrographer'.

one is *told* about an affect, while in expressives such as (1) to (3), one is *shown* an affect.

These intuitive considerations and others have convinced many linguists and philosophers that expressives and descriptives form two distinct categories of utterances (but, of course, not the only two: imperatives and questions being two other important classes). The same scholars, however, disagree on how exactly to account for the relevant dissimilarities and on how and whether the meaning they convey is different.

The noun 'an expressive' is usually restricted to signals carrying speaker meaning (see below, Chapter 1, and the Appendix for this notion). This includes linguistic utterances as well as certain gestures (e.g. giving the middle finger), overtly communicative facial expressions (e.g. a joyful wink), but also certain pictures (e.g. the emoji used in text messages), and other artifacts (e.g. perhaps a song might count as an expressive). Many signals which are not thought to possess speaker meaning and which are therefore not called 'expressives' may nevertheless be expressive of an affect: for instance, the cry of a newborn baby or the laughter of a rat. I will call such signals pre-expressives. I will come back to these below and especially in Chapters 7 and 8.

In this chapter, I will concentrate on expressives, but keeping an eye on pre-expressives, so as to avoid an artificially strict separation between them. Such a separation may indeed be more problematic than helpful, especially when we think about the phylogenetic or ontogenetic origins of speaker meaning (Dorit Bar-On, 2013).

In the next section (§5.2), I will present relevant insights from the main current theories of emotion and value. In §5.3., I explain how these insights can help us account for the properties that distinguish expressives from descriptives. In §5.4, I draw on views in speech act theory to explain why understanding the meaning of expressives requires understanding the affects that communicators express.

5.2. Philosophical insights on emotions

Let us start by stressing something quite obvious about the three benchmarks we have just reviewed – (a) hot vs. cold, (b) appropriate vs. true, and (c) direct vs. indirect. The distinction between expressives and descriptives seems to revolve around the existence of the relation that expressives bear to affects. It is even tempting to think that the three benchmarks all derive from the nature of the affects that expressives aim to communicate. In this section, drawing on the recent philosophy of

emotions and value, I show how thinking about the nature of emotions and cognates not only makes sense of the intuitions we started with but also promises to put constraints on what it is that an audience must recover when understanding the meaning of an expressive.

Note that I use 'affect' to refer to a broad class of phenomena which includes emotions as well as psychological states such as affective dispositions (acrophobia, arachnophobia, francophilia), moods (grumpiness, elation, feeling depressed), certain kinds of desires (cravings, sexual arousal, the irresistible urge to slap someone), or affective character traits (generosity, courage, greed). Affective *dispositions* are opposed to affective *episodes*. The latter are events with a certain, relatively short, duration while the former are psychological dispositions to undergo affective episodes, dispositions which may last a lifetime. Acrophobia, for instance, is a disposition to feel more fear toward heights than is deemed normal. Francophilia is something like a disposition to appreciate French language and culture. Affective dispositions thus are defined through their tendency to manifest in affective episodes.

Affective episodes comprise a large class of psychological states – pains and pleasures, emotions, urges, moods, and more – which possess three prototypical features. (i) Affective episodes typically are *valenced*: when we undergo an affect, we apprehend something positively or negatively. This is linked to the fact that we are attracted or repulsed by it, want to preserve it or destroy it, and feel good or bad in its presence. However, some affects, like surprise, may be neither positively nor negatively valenced (even though, arguably, surprise may always be either positive or negative). (ii) Affective episodes typically possess a *salient phenomenal character t*hat includes positive or negative hedonic tone as well as various felt reverberations from changes in the body – muscular tension, sweat, heat, heartbeat changes, goosebumps, a lump in the throat or knot in the stomach, etc. However, some affective episodes may be wholly unconscious (even though they may still possess a phenomenal character that is inaccessible to consciousness at this moment in time, much as one might not notice the painfulness of a wound whilst engaged in sport). (iii) Affective episodes are typically accompanied by strong *action tendencies* – a motivation to run away, aggress, cuddle, emulate, give up, etc. However, some affects, such as contemplative awe, or depression, may not result in any actions (although contemplation or the tendency to give up might reasonably be understood as action tendencies in the relevant sense).

These three features – being valenced, phenomenologically salient, and linked to strong action-tendencies – are what grounds the metaphor of

affects as 'hot' mental states. Consider by contrast doxastic states such as suppositions, judgments, beliefs, or conjectures: if these are not linked to an affect (the belief is about what we desire, admire, hate, etc.), then they will not move us, we won't be aroused, we will remain 'cold'. The metaphor does not fit the distinction perfectly, though, since we may well talk of cold affects, such as depression or cold anger.

Expressives can relate to any affect. For instance, certain slurs may be used to express affective dispositions (such as homophobia or racism) as opposed to punctual emotional episodes undergone by the expresser. Expressives can also express moods: think of a remark whose point is to express grumpiness or gloominess. Grumpiness and gloominess are usually considered moods, because these psychological states, unlike emotions, are usually not directed at anything specific – one usually is not grumpy or gloomy *about* one thing in particular, unlike happiness, sadness, anger, fear, etc.

Although expressives can relate to any affect, I will nevertheless focus on emotions. This is for several reasons. The main one is that, when it comes to affects, the central concept is that of emotion, and, arguably, the other affective states can be understood derivatively (see Deonna & Teroni, 2012: ch. 9; Prinz, 2004: ch. 8). For instance, homophobia, although not an emotion, may be defined as a tendency to undergo certain emotions (contempt, disgust, etc.) toward homosexuality. I believe that the same applies to expressives: first, let us focus on emotions, and then we will see how to adapt what we found to other cases. Another reason to start our inquiry by focusing on emotions is that they are better studied than the other kinds of affects we have considered. Thirdly, it seems that paradigmatic expressives more often than not express emotions. For these reasons, understanding expressives by understanding emotions seems to be an indispensable starting point.

According to the main contemporary philosophical theories of emotions (see reviews by Rossi & Tappolet, 2019; Scarantino & De Sousa, 2018), emotions are psychological episodes, more specifically experiences, that present aspects of the environment as having this or that significance or, as philosophers like to say, as having this or that evaluative property. For example, in fear we experience something as relating to *dangerousness* (one evaluative property). In moral anger, we experience something as relating to *offensiveness* (another evaluative property). In amusement, we experience something as relating to *funniness* (yet another). We might indeed say that emotions are value-tracking attitudes. Psychologists use the term 'appraisal': emotions involve an appraisal of the situation which

helps us detect, and react to, the aspects of our environment that are positive or negative for us, given the various concerns or goals we have while negotiating that environment.[98] Emotions are thus value-tracking attitudes insofar as their function is to react to evaluative properties in ways that[99] are optimal for the organism. Once the danger is detected, for instance, fear triggers various kinds of protective mechanisms, it prepares our body to flee or fight (more blood flow, widened eyes, dilated nostrils, alert posture, etc.) and gives us the feeling that we must urgently react.

As we will see below and especially in Chapter 9, considering emotions as value-tracking attitudes means that they are cognitive attitudes in the sense that they are mental states (or mental events) that 'say' something about the world and that can as such constitute a source of knowledge. If emotions are value-tracking attitudes, they can preserve or fail to preserve information, correctly or incorrectly detecting or treating the information that is relevant to our concerns and goals, the information to which we are supposed to react emotionally (e.g. the information that this is probably harmful, that this is an affectionate gesture, that this is an insult). As such, value-tracking attitudes can be more or less attuned with the environment, more or less appropriately adjusted to the world. I will come back to this below.

The fact that emotions are value-tracking attitudes is a recurring idea from Plato and Aristotle onwards (e.g. *Rhetoric* 1378a20-23). More recently, this idea has received extended philosophical treatment in theories that I will discuss below.

However, not every philosopher accepts this idea. Pre-eminent figures such as Descartes, Malebranche, Leibniz, Hume, Kant, James, Frege, Ayer, or, more recently, Searle have treated emotions as entirely *non-cognitive* attitudes.[100] For non-cognitivists, emotions are considered essentially irrational (Kant, 1798) or arational (Ayer, 1936; Frege, 1956; James, 1884;

[98] I will concentrate on philosophical theories here, but see Chapter 9 for a more psychology-focused discussion of emotions and appraisals.
[99] Philosophers often say that emotions are evaluations or evaluative attitudes, but I prefer to say 'value-tracking attitude' because 'evaluation' can either mean (a) a cognitive attitude that is meant to appropriately track the value of something or (b) a non-cognitive attitude that subjectively ascribes a value to something and that cannot be right or wrong.
[100] 'Cognitive' is here understood broadly, as that which has to do with the acquisition of knowledge, the attainment of new information, and as such as that which represent more or less accurately what it is supposed to be about. Perception, for instance, is a cognitive process, as I use the term. Desiring or intending, on the other hand, is not a cognitive mental state as their function is not to acquire new information, but rather to make us act in certain ways.

Searle, 1983).[101] Such philosophers have supported and reinforced the cliché opposition between reason and passion – Kant went as far as calling affects 'an illness of the mind' because they 'shut out the sovereignty of reason' (1798).

However, the arguments given by philosophers and cognitive psychologists since the 1950s have convinced the great majority of philosophers that emotions do involve a cognitive component, i.e. a component whose purpose is to gather and process information from the world.[102] As mentioned, this component is usually called 'the appraisal process' by psychologists. It takes neutral information as input and yields a value-loaded representation – a representation of the organism's situation as conducive or obstructive to the organism's goals – as output (see reviews by Ellsworth & Scherer, 2003; Moors et al., 2013; Sander et al., 2018). In philosophy, this cognitive component is usually identified as that which helps us apprehend evaluative properties, i.e. the features of the world which are positive or negative to the organism, the properties which are more or less beneficial to its concerns and goals. In Chapter 9, I will merge philosophical and psychological discussions and argue that the appraisal process represents evaluative properties, although it does so unconsciously.

The widespread recognition of this cognitive component in emotion has made the stark opposition between passion and reason obsolete. This point was famously argued for by Damasio (1994) who took as a central example Phineas Gage. After suffering a brain legion, Gage's emotional aptitudes changed drastically, but he kept intact his language abilities and other non-emotional cognitive capacities (memory, perception, etc.). According to Damasio, the loss of his emotional aptitudes explains how Gage fell from being a respected, smart, efficient foreman to become a socially ill-adapted vagabond who started piling up self-destructive life choices. Emotions, Damasio argues, are central to our rationality, contrary to what Descartes or other non-cognitivist philosophers have argued.

Even though emotions relate us to evaluative properties through cognitive mechanisms, it seems a mistake to think of emotions as mere judgments –

[101] This is the reason why I don't discuss in this chapter what Searle calls 'expressives'. His views on emotion make his account of what he calls 'expressives' extremely problematic, especially with respect to the idea that, for him, these speech acts have no direction of fit, because he considers emotions to have no direction of fit (Searle, 1979), which by the way seems very much in tension with his account of emotions as implying beliefs and desires (1983: ch. 1): why wouldn't emotions have both the direction of fit of beliefs and of desires?

[102] Today, some think that emotions are mere subjective feelings with no cognitive content (Shargel, 2015; D. Whiting, 2011), but they constitute a very small minority among philosophers working in the field.

pace Robert Solomon (1993), Martha Nussbaum (2001), and Stoic philosophers such as Seneca. Contrary to what the latter have defended, it seems that, say, being afraid of x relates us to x's dangerousness in a way that is quite different from the way in which a cold judgment that x is dangerous relates us to x's dangerousness. I will elaborate on this point later and come back to it in Chapter 9 (for various defenses of the difference between judgments and emotions, see Deigh, 1994; Deonna & Teroni, 2012, 2014; Döring, 2007; Goldie, 2000; Scarantino, 2010; Tappolet, 2000, 2016).

But then, if emotions are not mere evaluative judgments, how do emotions enable us to access information about evaluative properties? In particular, how can we make more precise the difference in this respect between emotions, on the one hand, and, on the other, doxastic states (beliefs, judgments, doubts, conjectures, …)?

This comes out clearly if we think of the relevant evaluative experiences that emotions exemplify as forms of *felt engagement* with the relevant aspects of the environment. In other words, emotions are felt, bodily, value-tracking attitudes towards a range of contents.[103] What does that mean? Fear and anger are felt bodily attitudes subjects have towards the dangers and offenses that they encounter, attitudes that distinguish themselves notably through the specific bodily readiness they involve. At the phenomenological level – what they feel like – these various states of bodily readiness are accompanied by pleasant or unpleasant hedonic tones and subtended by the feelings of various patterns of physiological changes (e.g. more sweat, changes in heartbeats, stopping of digestion) and motor reactions (e.g. the muscle contractions underlying facial, corporal, and vocal expression). This is how in fear we come to feel our body as mobilized to neutralize a threat, and how in anger we come to feel a preparedness for a form of active hostility.[104] According to this picture, feeling our bodies prepared or mobilized in these various ways constitutes experiencing the value-tracking attitudes that the emotions are – this, and not judging that evaluative properties are present,[105] is the sense in which emotions can be said to relate our conscious experience to evaluative properties. While,

[103] I will here reserve the term 'emotion' for those affective experiences which possess a distinctive phenomenal character. More on this in Chapter 9.

[104] States of action readiness feature already in older psychology theories of emotion (Arnold, 1960; Bull, 1951; McDougall, 1923), but they have been systematically explored by Frijda (1986) and have more recently been put to use by philosophers (Deonna & Teroni, 2012, 2015; Gert, 2018; Scarantino, 2014).

[105] Note that, in light of empirical work on emotions, it appears quite clearly that we unconsciously, rapidly, automatically, and quite primitively represent something as dangerous when we are afraid (see chapter 9). This is a kind of representation that is quite different from the conceptual, logical, or linguistic representations that philosophers have in mind when they say, e.g., that beliefs represent states of affairs.

then, emotions (e.g. feeling spiteful toward someone) and evaluative judgments or beliefs (judging that someone is contemptible) share many features – both, when successful, relate us to evaluative properties – they do so in markedly different ways.

In my view, the ingredients described in the last paragraphs – emotions being attitudes that are felt, bodily, and directed towards various contents apprehended as value-loaded – must enter any satisfactory account of the emotions, and they do indeed feature in today's main philosophical theories of emotion. I won't rely on one in particular, but I will now briefly present the most popular ones. I won't discuss further theories that cannot account for these features, such as judgment theories (Solomon 1993, Nussbaum 2001) or the non-cognitive theories I have already mentioned (Whiting 2011, Shargel 2015, see also Hutto, 2012). Anyway, such theories do not appear to be among the main contenders in philosophy today. For a more in-depth review of philosophical theories on emotion see Scarantino and de Sousa (2018).

One popular view today is the perceptual theory, which holds that emotions are perceptions of evaluative properties (see Tappolet 2000, 2016, Prinz 2004, Deonna 2006, and Döring 2007 for various versions). This theory was mainly developed in opposition to the idea that emotions are evaluative judgments, i.e. judgments ascribing the relevant evaluative property to the object of the emotion.

Perceptualists reject the judgment theory for three main reasons. First, emotions, by contrast with evaluative judgments, do not necessitate a mastery of evaluative concepts. Even if sadness makes us apprehend what we are sad about as an irrevocable loss, one need not have mastered the concepts IRREVOCABLE and LOSS to be sad: we can agree that babies and many nonhuman animals can be sad while disagreeing that they possess these concepts. This is not true for the judgment that one suffers an irrevocable loss: this judgment does require the concepts in question. That emotions can be nonconceptual allows us to accept, on the one hand, that babies and cognitively unsophisticated animals can have emotions, while, on the other hand, rejecting that they have the conceptual capacities required for the relevant evaluative judgments.

A second reason to distinguish emotions from evaluative judgments is that we can, at the very same moment, both undergo a certain emotion and judge that the object of the emotion does not possess the relevant evaluative property. So for instance we can judge that a horror movie or a rollercoaster ride is *not* dangerous while being afraid of it at the same time.

If emotions were judgments, such a situation would imply that one both judge that p (e.g. x is dangerous) and judge that p is not the case (x is not dangerous) at the same time, which would be highly irrational. However, being afraid of a horror movie or a rollercoaster ride while believing we are not in danger seems entirely reasonable or, in any case, not as irrational as entertaining two contradictory beliefs at once. This might also show that the mental states in play in evaluative judgments and emotions are subtended by different mental mechanisms. This is a strong argument for the perceptual theory of emotion because the comparison with perception is made even stronger by cases of illusions such as the Müller-Lyer illusion, where two lines of the exact same length are seen as having different lengths because of the chevrons that surround the lines (you certainly have seen this illusion with figures resembling >—< and <—>). In such cases, we can be certain that the two lines are of the same length, but still, we perceive them as being of different lengths. For this reason, cases such as the horror movie and the rollercoaster mentioned above have been called 'emotional illusions'.

Third, emotions, like perceptions, have a salient phenomenal character – i.e. give rise to an intense subjective impression – which determines what it is like to be in these states. *What it is like* to perceive (e.g. to see a bright red rose, to hear the distinctive sound of a bell), or to undergo an emotion (e.g. to be disgusted by rotten meat) strongly determines what these perceptions and emotions are. By contrast, it is not clear that judgments possess a phenomenal character at all, and if they do, it is very mild compared to that of perceptions and emotions and does not strongly determine what judgments are. What it is like to judge that the Swiss are Europeans, the phenomenal character of this judgment, is not constituted by a strong subjective impression.

Close cousins to the perceptual theory include what Scarantino and de Sousa (2018) call the 'evaluative feeling theory' (Goldie, 2000; Helm, 2009; Kriegel, 2014; Ratcliffe, 2005) and the 'patterns of salience theory' (Ben-Ze'ev, 2000; De Sousa, 1987; Elgin, 2008; Evans, 2001). Like the perceptual theory, these theories focus on the non-conceptual and phenomenologically salient nature of emotions as well as on how emotions can help us navigate the world by supplying precious information or processing such information.

Even though the perceptual, evaluative feeling, and pattern of salience theories were mainly developed as a reaction against the judgment theory, they resemble the latter in several aspects. One striking resemblance is that all these theories focus on *cognitive* functions of emotions – on how

emotions gather and process information – rather than on the *conative* functions of emotions, the role that emotions play with respect to actions – on how emotions motivate us and tend to make us approach, get away from, try to destroy, or act in other ways toward their objects.[106]

This aspect constitutes the main contrast between, on the one hand, the perceptual, evaluative feeling, and salient pattern theories and, on the other hand, their two main rivals: the motivational theory (Scarantino, 2014, 2015a) and the attitudinal theory (Deonna & Teroni, 2012, 2014, 2015) (see also Gert (2018) for an action-based theory), which we may regroup under the label 'action-oriented theories'. These theories focus on how emotions relate to *action tendencies* and can be considered philosophical heirs to the psychological theory of Nico Frijda (1986). They accept the arguments given by perceptualists against judgmentalists: that emotions can be nonconceptual, that they involve different mental mechanisms than judgments, and that they possess a strong phenomenology (although this last is not necessary for Scarantino 2014). Yet, action-oriented theories insist that emotions are also very different from perceptions.

The most relevant difference for us is that emotion involves action tendencies, which is not true of perception.[107] Here are some examples: in fear, we tend to avoid what we are afraid of; in anger, we tend to be aggressive; in disgust, we tend to actively reject what is apprehended as disgusting; in admiration, we tend to emulate or support what we admire; in sadness, we tend to give up on certain things.

Emotions do not always cause actions. They allow relatively flexible responses. In this respect, they are different from automatic reflexes, like the gag or knee jerk reflexes. Despite their flexibility, emotions always *tend* to make us act in certain ways and the physiological changes that go with emotions prepare us to react in these ways. In fear, our blood circulates

[106] Note that if emotions are considered as perception of *calls for action* (as in Deonna, 2006) or perceptions of action-guiding properties, then one can have a perceptual theory that essentially links emotions to action tendencies. However, this is not what the main champions of the perceptual theory defend (Döring, 2007; Prinz, 2004; Tappolet, 2000, 2016). Nevertheless, Prinz's theory may allow for such an option, as he does draw explicitly a link between emotions and perceptions that involve action-tendencies, even if his overall theory does not focus on action-tendencies.

[107] For five further important differences between emotion and perception, see Deonna and Teroni (2014). Among the latter, what is perhaps the strongest is that emotions always necessitate a cognitive basis constituted by a different mental state: one cannot undergo an emotion unless one perceives something, remembers something, imagines something, infers something, etc. This is not the case for perception: perception does not necessitate another kind of mental state as a cognitive basis.

faster to better deploy our muscles so as to avoid what we are afraid of, and we have rushes of hormones, such as adrenaline, which have many consequences that aid an efficient response (e.g. digestion stops, which allows allocating more energy to avoiding the threat). These action tendencies and the physiological changes that subtend them make emotions very different from regular perception.[108] Indeed, action tendencies are not among the components that define seeing, hearing, touching, etc. Furthermore, the physiological changes subtending perception (e.g. firing of optical nerves, retraction of the pupil, activity in the visual cortex) are of a very different nature than those subtending the action tendencies of emotions (beside modifications in the central nervous system, emotions involve modifications in the sympathetic nervous system, in sweat, heartbeats, muscular activity, hormonal secretion, and more). The motivational and attitudinal theories, by focusing on how emotions are essentially related to action, can explain all these features distinguishing emotion from perception, contrary to the perceptual theories.

5.3. How the particularities of emotions subtend those of expressives

In this section, I will show how the insights from the philosophical theories of emotions presented in the last section shed light on the intuitions with which we started (§5.1). Grounding expressives in emotion is, I believe, the best strategy for making sense of the distinctive nature of expressives compared to descriptives, and thus of how language can express, and not only describe, emotions. Indeed, we can comment on the three benchmarks distinguishing expressives and descriptives – (a) hot vs. cold, (b) appropriate vs. true, and (c) direct vs. indirect – by remarking how these features relate to emotions, and thus how expressives inherit them from the nature of emotions.

(a) First, we can understand the 'hotness' of emotions in light of their experiential dimension and contrast it to the experiential dimension of beliefs or other doxastic states. As we have just seen, emotions typically have a rich and diverse phenomenology, from positive or negative hedonic states to various dimensions of bodily arousal and felt action tendencies, and this phenomenology appears to be part of what a speaker is trying to

[108] Observe nevertheless that if one accepts arguments to the effects that we may perceive action properties (see e.g. the pragmatic representations discussed by Nanay, 2013), then perception and emotion can be considered as much more similar than with more traditional theories of perception. Nevertheless, some of the differences discussed in Deonna and Teroni (2014) remain (e.g. emotions require a cognitive basis). See the preceding footnote.

convey when using an expressive, thus conveying the hotness of emotions (or other affects) by means of expressives.

(b) Second, the description of emotions given above, and especially the remarks made on its relation to action, promises to shed light on the specific normative standards or correctness conditions by which we assess emotions by contrast to beliefs, i.e. (in)appropriate, (un)merited, or (un)deserved rather than true (false). Anger is a specific form of felt engagement or attitude taken towards something appraised as relevant to our concerns or goals. We appraise what we are angry about as somehow obstructive to our goal (e.g. as being offensive in moral anger) and this emotion tends to make us act aggressively. We may try to destroy the object of our anger or to prevent it otherwise from continuing to be what we deem obstructive. This engagement is *appropriate* to have towards, *deserved* by, or *merited* by obstructive (e.g. offensive) things and state of affairs. If you are angry at me because you think I intentionally broke the vase, where in fact it was the wind that broke it, I won't have deserved your anger, the situation would not merit your anger. Acting aggressively toward me would not be appropriate to the situation, it would not be the functional attitude to have. To someone afraid of a dog on a leash on the other side of the street, we shall say that her emotion is inappropriate to the circumstances or not merited by them, because they are not dangerous. But in any case, we will not say 'Your emotion is false' (D'Arms & Jacobson, 2000).

Our understanding of expressives, then, should reflect the fact that part of what is recovered is not simply a way of representing truly or falsely how the world is evaluatively speaking – as in doxastic attitudes and descriptives – but an engagement with the world that we conceive of as more or less appropriate, merited, or deserved. Beliefs fulfill their function (they are correct) when they are true, emotions fulfill their function (and, as philosophers of emotion say, they are correct) when appropriate, merited, or deserved.

To be more precise, we can distinguish two kinds of ways in which an emotion is (in)appropriate. The first one has to do with a cognitive component of emotions: an emotion is appropriate vis-à-vis this component insofar as the emotion captures and manages the information well. If I have fear of heights, I may perceive being on a balcony in a totally biased way. I may appraise the situation as extremely likely to be harmful to me even though in fact it is not. The appraisal, the cognitive component of emotion, has an indicative function: it is supposed to represent the world and my relation to it appropriately. If it fails to perform this function well,

the emotion linked with this defective appraisal process will be inappropriate to the situation. I will say more about this in Chapter 9.

The other way in which emotions may be deemed inappropriate has to do with their conative aspect: how emotions make us act, what action tendencies emotions involve. If, because of my anger, I start breaking everything I have in my flat, insult everyone I see, and try to punch my neighbors because they dare ask what is happening, my anger makes me react inappropriately. Another example: I meet a wild creature that I know may attack me if I start running but, despite myself, I start running out of fear. A third example: because I am so stressed, I decide not to go to my exam, although this is more damaging to me than if I had gone and failed, however badly. In these three cases, the way I act may be deemed inappropriate to the situation not because I misrepresented some evaluative property (which may or may not be the case), but because the way the emotion makes me act is not pragmatically functional, it is not the action tendency that would be appropriate for me to have in the situation, given my goals and concerns. In these examples, I have described how the conative component rather than the cognitive component is defective.

In most cases, the cognitive and the conative components go hand in hand: our emotions involve an action tendency that is appropriate or not to the situation because the cognitive component represents the situation appropriately or not. Since it is (at least partially) the cognitive component that determines the action tendency (see Chapter 9), we can rarely disentangle the two components when considering whether an emotion is appropriate or not. They are usually considered together, as a whole. Unless specified, therefore, when I say that an emotion is appropriate or not, I will mean 'as a whole', assuming that the cognitive and conative components are functioning together.

This discussion as well as the arguments presented against judgmentalism should make it clear how emotions' (in)appropriateness is to be distinguished from truth (falsity). The latter is the standard by which beliefs and judgments are evaluated, but it cannot be applied to emotions, which is why we don't say that emotions are true (false), but rather (in)appropriate, (un)merited, or (un)deserved.

This connects emotions with expressives in the following two ways: first, we can now see what it means for the speaker to be affectively rather than doxastically attuned to how the world is. Second, the felt, bodily, action-ready engagement I have highlighted makes emotions quite different from evaluative judgments and beliefs, even though both concern evaluative

properties. We must keep this in mind when studying expressives because this difference in the way these different attitudes relate us to evaluative properties sharply distinguishes expressives from descriptives such as (4) and (5) or (8) and (9) which communicate evaluative judgments or beliefs rather than emotions.

(c) Recall our third benchmark regarding expressives: they seem to convey their meaning by directly showing rather than indirectly saying. The description of the emotions given above makes it clear why they, as opposed to beliefs for example, could be shown. If emotions are felt bodily attitudes towards aspects of the environment, then what is felt by the subject, i.e. her bodily attitude, may be something an observer might also become directly aware of, not by feeling it as the emoter does, but by perceiving it. The posture of an angry person, the action tendencies typical of sadness, or the facial or vocal expression of happiness are directly observable or hearable and these perceptible emotional expressions can be considered proper components of emotions, along with physiological changes, and appraisal processes (a point already made by Scheler, Wittgenstein, and Austin).

We can thus plainly see how the distinctive features of expressives we have highlighted – (a) hot vs. cold, (b) appropriate vs. true, and (c) direct vs. indirect – seem to derive quite directly from distinctive features of emotions – their phenomenology, their correctness conditions, and their nature as felt bodily attitudes.

In addition to these three features, let us observe that the philosophical theories that highlight the intimate relation between emotions and action tendencies also explain a further trait typical of expressives, which is that they seem not only to be about the states of the world and of the expresser, but also about how the addressee should react. As Dorit Bar-On puts it:

> « Expressive communication, in general, is in a sense Janus-faced. It points inward, to the psychological state it expresses, at the same time as it points outward, toward the object or event at which the state is directed, *as well as toward ensuing behavi*ors. » (Bar-On, 2017: 304, my italics)

If emotions not only have a cognitive function (i.e. gathering and processing information) but also a conative or action-oriented function, then the nature of emotion also nicely elucidates how expressives, by communicating action-oriented states, have the function of pointing 'toward ensuing behaviors' – by warning, condemning, asking for help, for retribution, etc.

Although we are focusing on emotions, it is important to see that the present account can be extended to make sense of the affective domain in general. Through expressives, we can convey not only emotions but also moods and affective dispositions. So what about these? Well, one way of going about it is to spell out how the various affective states can be analyzed through the concept of emotion. For instance, moods (grumpiness, anxiety, depression, elation, …) might be analyzed as emotions without any specific target, without a conscious target, or with one's entire environment as a target. In the latter case, one could say that anxiety consists in apprehending the whole world as a dangerous place, while in grumpiness, one sees everybody and everything as obstructive or offensive. Affective dispositions (e.g. xenophobia) on the other hand may be understood as dispositions to undergo certain emotional episodes (anger, contempt, envy, etc.) given certain situations (e.g. in the presence of foreigners on one's 'territory'). So both moods and affective explanations might be explained through the concept of emotions (see Prinz 2004: ch. 9 or Deonna and Teroni: ch. 8 for more on the relations between emotions and other affects).

I have tried to home in on some crucial features of affective states so as to unearth some important aspects of what it takes to understand their occurrence in other people. In doing this I have largely ignored the specific context of our question, namely that we are after an account of what it takes to understand the affect *of someone trying to communicate this affect through an expressive utterance*. The next section is dedicated to explaining how we can conceive of the notion of expressive meaning in the light of (neo- or post-)Gricean pragmatics and speech act theory.

5.4. COMMUNICATING THROUGH EXPRESSIVES

The initial claim I will present and defend in this section is that (§5.4.1) speaker-meaning is fixed by the psychological states the speaker intends to communicate. This will then allow us to argue (§5.4.2.) that expressive speaker-meaning is fixed by the affective states the speaker intends to communicate, concentrating on emotions.

5.4.1. NATURAL VS. SPEAKER MEANING

The terminology of natural and speaker meaning (a.k.a. non-natural meaning) comes from Grice (1957, 1989).[109] Here are typical cases of natural meaning (written meaning_N or means_N):

[109] A similar distinction can already be found in Marty (1875) and Welby (1903).

(10) Smoke means$_N$ fire.
(11) The number of rings on this trunk means$_N$ the tree was 123 years old.
(12) His red cheeks means$_N$ he is embarrassed.
(13) Typical cases of speaker meaning (meanings$_S$ or means$_S$) are the following:
(14) Those three rings on the bell means$_S$ that the bus is full.
(15) By saying 'And the dishes...' Joe meant$_S$ that Sam should do the dishes.
(16) By 'You are my Sun and stars', Sally meant$_S$ 'I love you'.

As Dretske (1981, Chapter 2, 1986) has argued, we can interpret Grice's natural meaning along the following lines: natural signs are indicators, what they mean$_N$ is what they indicate to be so. They can do this thanks to certain lawful relations (including probable associations) between the sign and what constitutes their meaning$_N$. For instance, the fact that there are 123 dark rings on a tree trunk can mean$_N$ the fact that the tree was 123 years old when it was cut thanks to the lawful constraint that, every year, winter is colder than summer, which affects the tree growth and creates these dark rings. In (12), the red cheeks are a natural sign of embarrassment because of lawful psycho-physiological relations between embarrassment and blushing.

Unlike natural meaning, speaker meaning doesn't depend on lawful relations between the signal and its meaning. It rather depends on the speaker's intentions to communicate and to inform their audience about something. In (13), even if the bus is not in fact full, and even if a certain bus driver actually uses the bell most often when the bus is not full (because, say, it makes her job easier), her ringing the bell will still means$_S$ that the bus is full, even though it does not mean$_N$ that it is. The bell possesses its meaning because people have started using it with that intention and others could figure this out. Similarly, the meaning of (14) can go through because Sam understands what Joe *intends* to means$_S$ and not because of a lawful relation between 'And the dishes…' and 'You should do the dishes.' Observe that this correctly implies that an expressive such as that in (15) does not need to mean$_N$ what it means$_S$: the person using the expressive may be lying. There can be expressives where the speaker has no emotional state whatsoever.

Since we focus on expressives in this chapter and since these belong to speaker meaning, I shall leave aside emotional natural meaning. This will be the focus of Chapters 7 and 8 where I will present different ways to interpret and amend Grice's notion of natural meaning as it applies to

emotions. But before we move on, let me make four remarks which point to important similarities between emotional natural meaning and expressives.

First, the fact that expressives inherit their meaning from affective states is first and foremost true of emotional natural meaning: certain physiological changes $mean_N$ that one is undergoing a certain emotion because there are lawful relations between these physiological changes and the emotion. It is the nature of the emotions in question which gives the physiological changes their meaning.

Second, expressives are typically based on, and makes use of, expressive natural meaning, as Wharton (2009) rightly emphasized. For instance, 'Ouch!' in English and 'Aïe!' in French $means_S$ that their utterer is in pain partially because they are conventionalized forms of the initial natural meanings of uncontrollable vocal expression of pain (we can imagine something like 'Aaaargh!!!'). Similarly, if you ask me 'Should we go to this restaurant?' and I reply by sticking out my tongue, frowning, and wrinkling my nose, I can thereby $means_S$ something like 'No, I really don't like the food there' because I imitate a facial expression that $means_N$ disgust in the first place.

Third, even in cases where there are no obvious links between natural meaning of affects and expressive speaker meaning – for instance when someone utters 'Outrageous!' – there still seems to be some ingredient of the non-linguistic natural meanings of affects that is preserved in the expressive signal. In this case, the fact that it is not a full-fledged sentence, but only a one-word exclamation points to the fact that, when we are highly aroused by anger, we tend to utter short exclamations as opposed to lengthy and sophisticated signals.

Fourth, there are signals whose purpose is to express affects, and where we may even ascribe an intention to communicate an affect to the expresser, but where the conditions for speaker meaning are not present. As I mentioned above, I call such cases pre-expressives. I will come back to these below.

5.4.2 Expressives and speech act theory

Let us now further analyze a central claim of this chapter: that speaker meaning is expressive (as opposed to descriptive) when the psychological state that is overtly communicated, and from which the utterance inherits its meaning, is an affect. This is the idea that linguistic meaning is inherited from mental states – the core of Grice's philosophy of language

(1989) – and in the case of expressives, the meaning in question is determined by the conveyed affective states (emotions, moods, whims, urges, phobias, pleasures, pains, etc.). Speech act theory offers the possibility of defending this claim and making it more precise.

Following Frege and the speech act tradition, we can distinguish between two components of speaker-meaning: force and content (Austin, 1962; Bach & Harnish, 1979; Frege, 1956; Searle, 1969, 1979; Strawson, 1964, and for recent work in speech act theory see Fogal et al., 2018). Consider the following sentences, which have the same content but different forces:

- Joe smokes.
- Does Joe smoke?
- Smoke, Joe!
- May Joe smoke!

In Austin's terminology, the force component is determined by the *illocutionary* act, what one intends to do in saying something – making a statement, asking a question, giving an order, expressing a wish, etc. Illocutionary acts are successful when the audience understands to what end we use language. I successfully achieve the illocutionary act of asking a question when my audience understands that I have used language to this end. Through a speech act analysis, it makes a lot of sense to think that the meaning of expressives, as opposed to that of descriptives, is determined by what kind of illocutionary act is performed, as opposed to what kind of content it refers to.

There is no theory-neutral way to define illocutionary acts. I will briefly present several options and will then briefly explain how one of them – intentionalism – seems to be the most helpful to us (the following is largely based on the taxonomy proposed by Fogal et al., 2018).[110]

A first way to define illocutionary acts is *conventionalism*, which is how Austin (1962) himself did it. According to this view, performing an illocutionary act is a matter of following conventional procedures, of behaving in accordance with several conditions (called 'felicity conditions') determined by localized social conventions.

Although conventionalism makes a lot of sense for speech acts such as baptisms or wedding declarations (in which Austin was very interested), it

[110] My presentation is slightly different: I have merged what they call 'intentationalism' and 'expressivism' and I don't discuss 'functionalism'. Neither do I discuss the conversational score approach to speech acts (Lewis, 1979b) which can be used to spell out precisely any of the following theories, as Fogal, Harris, and Moss show (2018, sec. 1.2).

struggles with the fact that speech acts can be performed without conventionalized procedures. For instance, when Joe successfully asks Sam to do the dishes by saying 'And the dishes…', this cannot be explained merely through conventions. Or consider the fact that, by saying 'You will pay for this', a speaker may either make a threat or a prediction or give an order. Or think about how María Casares declares her love to Albert Camus (§6.1.) or about Captain Haddock's creative uses of rare words as insults (see footnote 98). Conventionalism cannot account for such facts. In general, it cannot account for what Grice later called a 'conversational implicature' (1975). This phenomenon, however, is widespread, and perhaps especially among expressives. We will thus set conventionalism aside.[111]

A cousin to conventionalism is what we may call *normativism*: the view that illocutionary acts are defined by constitutive rules or norms. Normativism about assertions has been defended by Dummett (1973) and has more recently been influentially revived by Williamson (2000).[112] According to the latter, here is the constitutive rule for assertions:

(Rule for Assertion) One must: assert p only if one knows that p.

The idea is that this norm is what makes a speech act an assertion. Because of what knowledge implies, if S does not believe that p, if p is false, or if S has no evidence for p, then we would be warranted to consider S is faulty in asserting p.

For normativism, speech acts are comparable to moves in a game. Performing an assertion, asking a question, and giving an order are licit moves in a language game only because there are certain rules which define their proper performance, just like a knight's movement in chess is only possible because of certain rules guiding its performance.

A problem for normativism is that it is hard to see how it extends beyond assertions to other speech acts. As Fogal, Harris, and Moss (Fogal et al., 2018, p. 12) put it, it is far from obvious how one is supposed to fill the gaps in the following rules:

- One must: ask someone whether p only if…
- One must: request that someone Fs only if…
- One must: advise someone to F only if…

[111] For a recent defense of conventionalism which answers some of its classical objections, see Lepore and Stone (2015). I don't see how they can answer the worries raised here though.
[112] A related view is defended by Brandom (1983).

Attempts have been made in this direction. For instance, here is the norm proposed by García-Carpintero (2015, p. 5) for orders:

> (Obligation Rule) One must: order A to p only if one lays down on A as a result an obligation to p.

García-Carpintero has also proposed norms for derogatory speech acts (Marques & García-Carpintero, 2020) and presuppositions (García-Carpintero, 2020) and other philosophers have given accounts of other illocutionary acts as well (see Fogal et al., 2018, sec. 1.1.5). However, no systematic taxonomy of speech acts exists and, as far as I know, no attempt has been made to understand expressives through this theoretical framework.[113]

I find normativism to be a very promising theory – especially because it combines well with powerful dynamic semantic frameworks such as that of Portner (2007) or Roberts (1996) (see García-Carpintero, 2015). Nonetheless, more work would need to be done for it to help us understand the meaning of expressives, since it is not clear what the constitutive rule for expressives might be.

A further worry with normativism is that the norms governing speech acts can be explained by more primitive features of speech acts. The fact that assertions are subject to epistemic norms (about truth, belief, evidence) is uncontroversial. What is controversial is that this is the ineliminable ingredient that makes certain acts assertions. Bach (2004) for instance argues that the epistemic norms of assertion are explained by a more basic fact: that assertions express beliefs. This leads us to the next type of theory, the one which I find most helpful for a better understanding of expressives.

Intentionalism is the view that performing an illocutionary act is a matter of producing a signal with certain intentions. This view stems from Grice's

[113] Maybe expressive rules along the following lines could be made to work, where 'm-express' means 'perform a speech act with the illocutionary intent to express'.
(ER1) One must: m-express emotion e only if one undergoes e.
However, this norm may be too weak for the following reason. It may be warranted to blame someone for making an assertion based on a belief that the person knows to be highly irrational, e.g. 'I don't have any evidence for thinking that aliens live on Mars, but aliens live on Mars'. Similarly, it may be warranted to blame someone for performing an expressive speech act based on an emotion that the person knows to be highly irrational, e.g. 'Yuk! This soup tastes really bad! And I haven't tasted it.'. Possible alternatives are the following:
(ER2) One must: m-express emotion e only if one undergoes a justified e.
(ER3) One must: m-express emotion e only if one undergoes an appropriate e.
(ER4) One must: m-express emotion e only if one undergoes an appropriate and justified e.
(ER5) One must: m-express emotion e only if one undergoes e and knows it is appropriate.

work on speaker meaning and was notably developed by Strawson (1964a), Schiffer (1972), and Bach and Harnish (1979). An important version of intentionalism, which may be called *attitudinal intentionalism*, is based on the idea that we can distinguish types of illocutionary acts by the types of psychological attitudes that speakers intend to communicate. As Bach and Harnish (1979) put it:

> « Since illocutionary intents are fulfilled if the hearer recognizes the attitudes expressed by the speaker, types of illocutionary intents correspond to types of expressed attitudes. » (Bach and Harnish 1979: 39).

According to this view, we may say that assertions express beliefs (or knowledge), orders express desires that the audience does something, questions express desires to know something, promises express intentions to do something, thanks express gratitude toward the audience's deed, etc.

Attitudinal intentionalism naturally leads to an intuitive way of understanding the nature of expressives, one that is very much compatible with what we have seen above. Expressives would be the utterances whose illocutionary intent is to express affects. According to this view, thanks are expressives because their illocutionary intent is to express gratitude and gratitude is an affect; apologies are expressives because their illocutionary intent is to express regret and regret is an affect; etc.

According to intentionalism, illocutionary acts are defined through speaker meaning intentions (see Chapters 1 and 2 and the Appendix). Here, for instance, is how Bach and Harnish characterize the illocutionary act of apologizing:

> « In uttering e, S apologizes to H for D if S expresses:
>
> i. regret for having done D to H, and
>
> ii. the intention that H believe that S regrets having done D to H. » (Bach & Harnish, 1979, p. 51)

If we follow my favored definition of speaker meaning, and we generalize the analysis, we may want to characterize expressives as follows:

By producing x, S performs expressive E only if[114]

[114] I take these clauses to be necessary conditions and not to be sufficient because it may be possible to fulfill these conditions without performing an expressive. For instance, a piece of instrumental music arguably is not an expressive *stricto sensu* but it may be used to fulfill (i) and (ii).

(i) S intends (Intention 1) to make her affect A about content C manifest and
(ii) S intends (Intention 2) to make (i) mutually recognizable.

The illocutionary act is successful when the speaker fulfills her intentions. One understands an expressive when one recognizes that an affective state is overtly intended to be made manifest, and what affective state it is.

An important advantage of intentionalism over conventionalism is that expressing one's intentions can be done in nonconventional ways, e.g. through conversational implicatures. An advantage over normativism is that it effortlessly accounts for the main speech act categories (assertions, questions, orders, promises, etc.). Furthermore, as mentioned, it may well explain and ground the norms proposed by normativism and why the norms govern the types of mental state expressed. For instance, the fact that we expect speakers to be committed to the truth of what they assert can be explained by the fact that, by making an assertion, they overtly and intentionally express a belief (or knowledge), which is a mental state subject to a norm of truth. [115]

Attitudinal intentionalism seems to have the right tools to help us inquire further into the meaning of expressives. To see how, let us go back to our first example:

(1) Outrageous!

Here is how attitudinal intentionalism would interpret (1). The speaker has an Intention 1 to make manifest to her audience her outrage (her attitude) about what the government did (the content of the attitude). She also has the Intention 2 that Intention 1 is made mutually recognizable to her and her audience. Once Intention 1 is recognized the audience has understood what the speaker meant. This requires understanding what kind of psychological state the utterer is in and thus, arguably, what are the norms governing this psychological state and hence what kind of

[115] Let me observe that the different views of what illocutionary acts are can be combined. For instance, Searle's view (1969) combines conventionalism, normativism, and intentionalism while Green's (2007) combines intentionalism and normativism. This is apparent in Green's definition of illocutionary speaker-meaning: « S illocutionarily speaker-means that P φ'ly, where φ is an illocutionary force, iff 1. S performs an action A intending that 2. in performing A, it be manifest that S is committed to P under force φ, and that it be manifest that S intends that (2). » (Green, 2007, p. 74) What does it mean exactly to be committed to a certain content under the force of an expressive speech act? This question may be answered through the rules governing expressives such as the ones proposed in the penultimate footnote.

commitments one undertakes by producing this expressive (here are some candidates: the commitment to evaluate other actions of the same nature with the same force, the commitment not to wholly support the government in the future, the commitment to justify why one is outraged, etc.).

Here is another illustration, using (3) above, i.e. 'The frogs won it again!'. The speaker has the intention to make it publically recognizable that, by producing the word 'frogs', she intends to, say, make manifest that she is disposed to feel contempt (her attitude) toward the French (the content of the attitude).

You might have noticed that in these illustrations, I have disentangled the attitude (outrage, disposition to feel contempt) and the content (what the government did, the French). This is because expressives and descriptives can inherit their meaning from psychological states that possess the same content: they only differ in the attitude they express. As part of Intention 1, one may intend to make manifest that one is happy that it is raining or that one believes that it is raining. The meaning of expressives differs from that of descriptives (or from that of questions or orders) insofar as affects differ from doxastic states, not because of what the contents of these mental states are. Attitudinal intentionalism thus combines well with the claim explored above according to which expressives inherit their distinctive properties from what it is that makes affects distinctive attitudes.

Another advantage of attitudinal intentionalism is that it accounts well for the fact that certain sentences do not fall neatly on either the expressive or the descriptive side. Remember the sentence by Casares:

> « When I think of us it seems absurd to not believe in eternity. »

It makes sense to interpret this as expressing both an introspective belief as well as, indirectly, her love for Camus. Attitudinal intentionalism rightly predicts that the sentence possesses both a descriptive and an expressive nature.

Similarly, intentionalism can easily account for the fact that a sentence as vapid as 'The boat has departed", even if said in a neutral tone, can nevertheless be an expressive, given the right background context, i.e. a context which allows one to make it mutually recognizable that the speaker intends to make her affect manifest to her audience.

Before I conclude, let me briefly address one worry that has been raised against intentionalism. The worry is that we can successfully express

certain affects without any need for the complex intentions that define speaker meaning (Dorit Bar-On, 2013).

Signals which lack speaker meaning might still possess a meaning that can be analyzed as having both a force and a content component. For instance, an infant might not yet have the capacity to make articulate speech acts, nor to create signals with the complex intentions necessitated by speaker meaning, but she might, on the one hand, ask that we give her food, and, on the other hand, express contentment that we give her food. These communicative acts can be interpreted as having the same content but two different forces: one is imperative and the other is expressive. Similarly, nonhuman primates might emit signals that have the function of requesting something as well as signals whose function is to inform others about something (Tomasello, 2008, Chapter 6). Even if chimps don't have the mindreading capacities required for speaker meaning (Scott-Phillips, 2015), their communicative acts can nevertheless have these two different kinds of force. Similarly, we may interpret a wolf baring its teeth as a *warning* of an imminent attack, as opposed to informing of, or describing, an imminent attack (Scarantino, 2017).

We can also interpret the examples of allower meaning given in Chapters 1–4 as involving pre-illocutionary acts: for instance, the allower meaning of laughter may be analyzed as possessing both an expressive pre-illocutionary force (e.g. express embarrassment) as well as an imperative force (e.g. something like 'please, let's not talk about that'). In sum, when it comes to communication, and perhaps in all its forms, we may analyze meaning as made up of (at least) two components: the force and the content.[116]

I will follow common usage and reserve the expression 'illocutionary force/act' for speaker meaning. For the meaning of signals sent without the communicative intentions necessary for speaker meaning, but where it nevertheless makes sense to distinguish a force and a content component, I will use the expression *pre-illocutionary* force/act.[117] The suffix 'pre' is not meant to be value-loaded, but phylogenetic and ontogenetic. Pre-illocutionary acts come first from both an evolutionary and a developmental perspective.

[116] I say 'at least' because Frege, among others, thought we needed a third component: the *Färbung* (tone).
[117] The term 'pre-illocutionary' comes from a conversation with Mitch Green. Thanks to him for suggesting this terminology.

Pre-illocutionary forces/acts may not come from the intentions of the sender, but from an evolutionary function of the signal. Instead of explaining their forces as intentionalism does, we may use the evolutionary notion of *teleosemantics*. I will come back to this point in Chapter 8.

5.5. CONCLUSION

Let us wrap this chapter up. I began this chapter by presenting what appear to be the three most salient features distinguishing expressives from descriptives.

(a) *Hot vs. Cold.* Expressives inherit their meaning from mental states which are phenomenologically 'hot' – the feelings of affects include positive or negative hedonic tones, various felt reverberations from changes in the body, as well as felt action tendencies. By contrast, descriptives inherit the coldness of the doxastic attitudes they communicate. Think of the difference between someone stating 'Someone has covered my car with graffiti.' and the same person yelling 'Shit!!!'.

(b) *Appropriate vs. True.* Expressives can be assessed as more or less appropriate to the situation (merited by it, deserved by it), but we do not usually qualify them as literally true or false: a 'Yuk!!!' would be deemed inappropriate if it is directed at a delicious dish, but it wouldn't count as literally false.

(c) *Direct vs. Indirect.* Expressives can directly show the affects they express because they constitute part of their manifestation, belonging to the motor expression and/or action tendency components of affects. By contrast, even when descriptives are about affects, they indirectly report them.

I have thus explained how these three features of expressives – hot, (in)appropriate, direct – derive naturally from a picture of affects depicting them as felt bodily reactions to stimuli evaluated as relevant to the concerns of the person undergoing the affect. I have spelled out how understanding speech acts as attitudinal intentionalism does enable us to explain this matter of fact by seeing expressives as utterances that inherit their meaning from what is distinctive of the attitudes expressed, i.e. from what is distinctive of affects.

This is why a proper analysis of how language expresses emotion, of what expressives are, requires an in-depth analysis of emotions themselves and of the other kinds of affects conveyed by expressives.

6. Understanding Expressives by Understanding Emotions

« If you do not feel a thing, you will never guess its meaning. »
– Emma Goldman, *Letter to Alexander Breckman (May 24, 1929)*

« La joie innocente est la seule dont les signes flattent mon cœur. Ceux de la cruelle et moqueuse joie le navrent et l'affligent quoiqu'elle n'ait nul rapport à moi. Ces signes, sans doute, ne sauraient être exactement les mêmes, partant de principes si différents … »
– Jean-Jacques Rousseau, *Rêveries du promeneur solitaire*

Abstract. This chapter discusses three possible accounts of what understanding expressives amounts to. The first account, doxasticism, claims that the audience must merely attribute a doxastic attitude (propositions believed, supposed, doubted, etc.) to the utterer. The second view, moderate affectivism,[118] claims that the audience must believe that the utterer undergoes (or is disposed to undergo) emotions, highlighting the specificities of affective as opposed to doxastic attitudes. The third view, radical affectivism, claims that instead of believing that the utterer expresses an emotion, the audience must resonate affectively with the expresser to properly understand the expressive utterance. I discuss some advantages and disadvantages of these three views, arguing that moderate and, especially, radical affectivism are in a better position to explain the distinctive features of expressives. If this is correct, then a significant portion of the literature on expressives seems to be mistaken insofar as it accepts or presupposes doxasticism. Overall, this is an example of how the philosophy of emotion may inform and constrain the philosophy of language.

6.1. Introduction

The goal of this chapter is to present some of the constraints that bear on a satisfactory account of the meaning of expressives by focusing on what it

[118] As we shall see, despite what the names seem to indicate, affectivism differs in several respects from emotivism, the famous meta-ethical views defended by A. J. Ayer and C. L. Stevenson. One main difference is that affectivism is a cognitive view in the sense that, contrary to Ayer and Stevenson, I defend that affects, including emotions, involve a cognitive element – the appraisal process, see Chapter 9 – which makes them evaluable for their semantic correctness, and that utterances expressing emotions and other affects inherit this cognitive element and can thus be evaluated as semantically correct or incorrect. I know too little about the literary theory that is also called 'affectivism' (Fish, 1972; Pater, 1873; Richards, 1926) to judge its similarity with the view I will discuss, but it may have commonalities with what I call radical affectivism.

takes for an audience to understand expressives. I will focus on the expressives that express emotions. As I mentioned in the previous chapter, when it comes to affects, the central concept is that of emotion, and, arguably, the other affective states may be understood derivatively (Deonna & Teroni, 2012, Chapter 8; Prinz, 2004, Chapter 9). The same applies to expressives: first, let us focus on those expressing emotions, and then we will see how to adapt what we found to the expression of other affects. The distinctive contribution of this chapter, together with the preceding one, comes from focusing on the nature of the emotions expressed, and doing so more carefully than is usually the case in the relevant literature.

In the last chapter, I highlighted the main features that distinguish expressives from descriptives and how these features derive from the nature of emotions and other affects. We also saw that expressives are only one type of speech act whose illocutionary point is to communicate affects.

Here are the examples of expressives we used:

(1) Outrageous!
(2) Ouch!!!
(3) The frogs won it again!

Recall examples of descriptives (4)–(9) which were also clearly in the business of conveying affects:

(4) What the government did was wrong.
(5) I feel outraged by what the government did.
(6) This boiling oil has burned my hand and this is bad for me.
(7) I feel great pain.
(8) The French won the world cup again and I believe that the French are contemptible.
(9) The French won the world cup again and I feel contempt towards the French.

Examples (4)–(9) are not considered expressives because, although they convey affects, they don't express them,[119] contrary to examples (1)–(3). This is so even though (4)–(9) are construed as being about the same states of affairs as (1)–(3).

I will use the three main features that distinguish expressives from descriptives which we discussed in the last chapter as benchmarks to be

[119] Remember that they were to be conceived as uttered in a context where they would not directly show any affective states (i.e. said in a neutral tone, with a neutral facial expression, etc.).

met by the accounts discussed in this chapter. As a reminder, here they are:

(d) Hot vs cold. Expressives always express affects (emotions, desires, moods, sentiments, pleasures, pains, whims, etc.). In contrast, descriptives express beliefs or other doxastic attitudes (doubt, supposition, conjecture, etc.). To use a usual metaphor, descriptives seem to communicate mental states which are 'cold' as opposed to the alleged 'hotness' of affective states.

(e) Appropriate vs. true. Expressives, like the affects that they convey, seem to answer to normative standards of (in)appropriateness, (un)meritedness, or (un)deservedness, while descriptives, like the doxastic attitudes they communicate, appear to answer to the normative standards of truth and falsity. Compare for instance (2) and (7) while imagining that the person doesn't feel pain. We would say of (7) that it is literally false, but not of (2) (Kaplan, 1999).

(f) Direct vs indirect. While descriptives can and do sometimes convey affects, expressives' distinctiveness is to do so directly. If one describes one's affects, as in (4)–(9), one in fact communicates a thought one has about an affect, a thought that cannot be perceived directly by the audience. When expressing an affect, however, as in (1)–(3), it seems as though one directly shows the affect, or at least some of its components (e.g. facial, vocal, and gestural expressions, certain action tendencies, certain verbal behaviors).

In this chapter, the goal is to further explore the meaning of expressives by comparing the merits of three theories about what it takes for an audience to understand what speakers convey with expressives. The rationale for such an angle is grounded in the idea that to develop a theory of meaning, one must develop a theory of understanding because understanding an utterance is understanding its meaning – an idea that finds strong support in the philosophy of language since Frege, and especially Frege's interpretation by Dummett (1973: 92).[120]

Let me note already that understanding comes in degree, so that one may partially understand what an expressive means although one does not

[120] This focus on understanding is also supported by influential considerations from Wittgenstein's *Philosophical Investigations* (Wittgenstein, 1953: §142ff). See also (Searle, 1969: 42ff).

optimally understand it. By 'optimal understanding', I mean that the relevant communicative functions have been achieved optimally. This claim thus piggybacks on the idea that communication has several functions and that we can evaluate how well these functions are fulfilled. The task of ascribing functions to communication is not straightforward, but solid theories exist (see Green, 2007; Millikan, 1984; Scarantino, 2013; Skyrms, 2010 in philosophy, as well as Dawkins & Krebs, 1978; Hauser, 1996; Maynard Smith & Harper, 2003; Scott-Phillips, 2008 in biology). The main two communicative functions that are discussed in the literature are (1) transmitting information and (2) influencing others' behavior. I follow Scarantino (2013) in hypothesizing that these two functions usually go hand-in-hand: communication is fully functional when it successfully transmits information (thus influencing others' cognitive states) and that, in the relevant cases, this allows influencing others' behavior. I will concentrate mainly on information transmission because, firstly, this function is primary in the sense that influencing others through communication must make use of information transmission (Scarantino, 2013) and, secondly, because the philosophical literature on expressives is concerned more with information transmission than with influencing others' behavior. However, in §6.3, I will also discuss how expressives, beyond the transmission of information, have the function of influencing others, namely: to generate affects in them.

The plan now is to present and discuss three views about what it takes to understand expressives – doxasticism, moderate affectivism, and radical affectivism – and to evaluate how well they capture the communicative functions of expressives, that is, how well they capture what it takes to optimally understand an expressive.

6.2. Doxasticism about Expressives

6.2.1. Introducing Doxasticism

This first view – doxasticism about expressives – claims that, to optimally understand the meaning of an expressive, it is necessary and sufficient to attribute to the communicator the right doxastic attitude (belief, doubt, supposition, conjecture, etc.) towards the right propositions. In other words, understanding expressives consists in recovering the relevant propositions believed (doubted, supposed, etc.) by the communicator.

A view along these lines is put forward by Schlenker (2007) who argues that the meaning of expressives is to be found in presuppositions concerning the evaluative beliefs of the signaler. A similar view is defended

by Sauerland (2007). I believe that many other accounts of the meaning of expressive utterances may be classified as doxasticist as well, or as having a doxastic tendency, even if this view is more in the background than explicitly argued for (see for example the account of thick terms by Cepollaro & Stojanovic, 2016).[121] I will rely on Schlenker's account to present doxasticism because it concerns any expressive (e.g. not only interjections, or only slurs) and it is explicit that a belief attribution is both necessary and sufficient to understand the meaning of expressives (in particular, the literal use of expressive expressions).

Note that the putative upholders of doxasticism I have mentioned in the last paragraph, and in the preceding footnote, are concerned with providing a *semantic* theory of expressives, but doxasticism is defined with respect to speaker-meaning, not word meaning. In this respect, we may hesitate to classify the aforementioned authors as supporters of doxasticism. However, as far as I know, nothing in their work indicates that moving from word meaning to speaker-meaning would change their views concerning what kind of mental state needs to be attributed to speakers using expressives. For the sake of argument, I will thus suppose that their view on word meaning indicates what their view on speaker-meaning is.

Presuppositions are propositions that are taken to be part of the common ground between participants to a conversation, they constitute the body of information that all participants in the discussion are assumed to take for granted (Stalnaker, 2002). For instance, when you say 'It was John who broke the computer', you presuppose that someone broke a computer and that your audience knows it, i.e. the proposition <Someone broke a computer> is assumed to be in the interlocutors' common background. Now, according to Schlenker, and putting aside certain technicalities (see

[121] For instance, several philosophers (Diaz-Leon, 2020; Hom, 2008; Hom & May, 2013, 2018; Lycan, 2015) concur with Schlenker insofar as they argue that what is derogatory in the meaning of slurs can be accounted for by regular truth-conditional semantics. These views may be understood as doxasticist insofar as the derogatory meaning of slurs is typically understood as what makes a sentence using it an expressive, i.e. a sentence illocutionary intent is to express affects. In other words, what speakers mean by slurs in such cases is determined by what is said with slurs, by the semantic meaning of these words. So, these philosophers may well presuppose the view that understanding an expressive speech act that derogates a certain group by using a slur only requires understanding the evaluative judgment that is expressed by the sentence containing the slur. However, the philosophers in question may also have a different background theory and, for instance, argue that the semantic of slurs is insufficient to determine any kind of expressive meaning. By contrast, Schlenker seems to accept doxasticism more obviously insofar as he is clearly in the business of giving an account of the meaning of expressives and opposes the view of Potts (2007) which itself may be classified as a moderate affectivist.

the next footnote), someone using the word 'frog' instead of 'French' makes the presupposition that she believes that French people are despicable or contemptible. In other words, using this word is considering that the proposition <I believe that French people are contemptible> is part of the common ground between the interlocutors. Within this account, we can suppose that understanding the sentence 'I met a frog', what you need to retrieve as an audience are the following two propositions: (i) The speaker met a French person, and (ii) the speaker believes that French people are contemptible.[122]

Doxasticism doesn't need to rely on a presuppositional account of expressives. It could make use of other traditional linguistic devices to analyze it (e.g. conventional implicatures as in Williamson (2009), or regular lexical entry as in Hom (2008), Hom and May (2013, 2018)). We won't focus on these different linguistic mechanisms by which expressive meaning may be conveyed. What is important for us is the kind of psychological *attitudes* that are at stake in understanding expressive meaning.

In all its forms, doxasticism would analyze the meaning of

(1) Outrageous!

along the lines of

(4) What the government did was wrong.

given the discourse context we have assumed for (1). Similarly, assuming as we did, that the utterer of

(2) Ouch!

expresses the pain of a burn caused by boiling oil, doxasticism makes the meaning of (2) very close to that of

(6) This boiling oil has burned my hand and this is bad for me.

If we accept the idea – presented in the last chapter – that emotions are value-tracking attitudes and take into account the different ways in which philosophers have captured this insight, it appears quite plainly that doxasticism tends to equate the meaning of evaluative judgments to that

[122] The account is here simplified to avoid technicalities. The original account claims that the lexical entry for 'Frog', based on a two–dimensional framework where 'w' stands for world, 'c' for context, and '#' for presupposition failure, is the following: [[Frog]](c)(w) ≠ # iff the agent of c believes in the world of c that French people are despicable. If ≠ #, [[Frog]](c)(w) = [French](c)(w)

of expressives, *modulo* the way in which the meaning is brought into the common ground, e.g. through presuppositions, conventional or conversational implicatures, explicitly, etc.

One might of course hold doxasticism and still argue that the meaning of, e.g. (2) and (6) cannot be equated because, say, the connotation is different: one is implicit and the other explicit, one is a sentence and the other isn't, or that there is only a partial overlap between propositions expressed. For instance, Kaplan (1999) argues that 'Oops' has the same informational content as 'I just observed a minor mishap', but that these two utterances differ in their meaning because they are not syntactically, and thus not logically, equivalent. For Kaplan, because 'Oops' is not a sentence, it cannot be true, false, or logically derivable, and this implies that its semantics are different from that of 'I just observed a minor mishap', even if the two convey the same informational content.

Nevertheless, putting aside grammatical differences, doxasticism would not see any discrepancies in terms of the attitude communicated by expressives and descriptives and this is what is critical here. In all cases, the meaning can be understood by retrieving a certain range of *propositions believed by the utterer*. The usual candidate for these propositions is an evaluative judgment, c.f. the examples above from Kaplan (1999) and Schlenker (2007).[123]

[123] Doxasticism could also say that the relevant propositions, i.e. those believed by the speaker and that the audience would need to infer, are not about evaluative properties, but about the affects felt by the speaker. We could call this version of doxasticism 'affective doxasticism' as opposed to the 'evaluative doxasticism' of Schlenker et al that I discuss in the main text. Kaplan (1999) may be interpreted as proposing an affective doxasticism when he claims that 'Ouch' has the same content as 'I am in pain' because he may think that 'Ouch' expresses the belief that the person is in the affective state of being in pain. Alternatively, Kaplan may not want to say that the person who says 'Ouch' expresses the *belief* that she is in pain, but rather transmit information about *the fact* (or the event) that she is in pain, a fact which the audience would need to understand. If this alternative interpretation is the correct one, then Kaplan is an evaluative doxastic about 'Oops' (since 'I have just witnessed a minor mishap' clearly expresses a doxastic attitude) and a moderate expressivist about 'Ouch' (see §6.2 for the presentation of moderate expressivism). To be charitable to Kaplan, I will suppose this alternative interpretation. I take affective doxasticism to be even more problematic than evaluative doxasticism. The reason is that an utterance of 'I am in pain', as well as utterances (5), (7), and (9) are cases where speakers express an explicit self-ascription of affects, but expressives such as 'Ouch' or (1)–(3) need not express explicit self-ascriptions. Explicit self-ascription of affects requires judgment that one is undergoing an emotion, but uttering an expressive does not. This is clear in the case of a newborn baby screaming: we may say that the baby expresses pain even if the baby does not master the concept PAIN and so cannot have a belief about the fact that she is in pain. Because it requires this supplementary belief, affective doxasticism runs into the problems of evaluative doxasticism which I discuss below. Furthermore, it runs into more problems because it implies that the attitude expressed by expressives is a meta-representation: a representation (a belief) about an affective

Observe that this means that, if we preserve a taxonomy of speech acts such as that proposed by Bach and Harnish (1979), a taxonomy (entirely or largely) based on what kind of attitude figures in the illocutionary intent (what kind of attitude is expressed, in the sense used by Bach and Harnish), then doxasticism may be taken to imply that expressive and descriptive utterances belong to the same speech act category. This would be the case because, if doxasticism is true, these two types of linguistic expressions do not require attributing different illocutionary intents to the speaker: in both cases, their meaning is understood by attributing to the speaker the expression of a doxastic attitude. So expressive speech acts would actually be a sub-class of descriptive speech acts. I will come back to this point in the conclusion.

Before I turn to arguments against doxasticism, let me do a little preemptive detour, which will also highlight some positive qualities of doxasticism. Doxasticism says that, for the audience to understand an expressive, it is necessary and sufficient that it attributes the doxastic attitude expressed by the utterer. One could reject doxasticism while accepting the following claim (N): in some cases, it is necessary for the audience to attribute a doxastic attitude to the utterer to understand an expressive.

A reason for rejecting doxasticism while accepting (N) can be that (N) applies only to *some* expressives (and, one may add, that this subset is not representative of expressives, is not paradigmatic). Another reason could be that (N) is about a necessary condition but one may want to reject the sufficiency condition required by doxasticism. I highlight this because some could be tempted to adopt doxasticism because of the famous Frege–Geach problem (for an overview of this problem and its consequences, see Schroeder, 2016a). However, I take this problem to only show that (N) is plausible.

Take this example from Hom (2010), adapted from Geach (1965):

(10) If Joe fucked up his presentation again, he will be fired.
(11) Joe fucked up his presentation again.
(12) Therefore, he will be fired.

This is a perfectly valid inference. Now (11) certainly qualifies as an expressive which expresses a negative affect toward Joe's presentation.

attitude, which is itself a representation of the situation. This surely needlessly overintelectualizes expressives. By contrast, evaluative doxasticism says that the attitude expressed is a simple representation (an evaluative judgment). Finally, I don't see any advantage that affective doxasticism would have over affectivism.

(10) on the other hand, doesn't express a negative attitude toward Joe's presentation: it states a conditional. Someone sincere and honest can state (10) with or without having a negative attitude toward Joe's presentation, but someone honest and sincere who states (11) must have a negative attitude (evaluation, emotion, desire, etc.) about Joe's presentation; so let us assume that (11), unlike (10), is an expressive. Now, for this argument to be valid, the antecedent in (10) should carry the same semantic content as (11). Since logicians and epistemologists generally agree that premises of inferences need at least to be supposed (if not believed) for someone to operate logical operations over them, and that supposition is a doxastic attitude, someone understanding the inference (10)–(12) should entertain doxastic attitudes toward these propositions. So, this example shows that, at least in some cases, the audience must entertain a doxastic attitude to understand an expressive.

From this, we may argue that (N) is the case as follows: a person who sincerely utters (10)–(12) must herself possess doxastic attitudes about (10)–(12) and intend to convey these doxastic attitudes. To understand what is meant, the audience must understand that the speaker has this intention. Thus, since (11) is an expressive, in some cases it is necessary for the audience to attribute a doxastic attitude to the utterer to understand an expressive (N).

But (N) cannot be used, as such, to defend doxastism and that is true even if we could extend (N) to all expressives, and not only to some cases. To evaluate whether doxasticism is true, we need to ask: is this doxastic attribution sufficient? For instance, can we understand all that is meant by (11) *qua* expressive if we only attribute doxastic attitudes to the utterer?

6.2.2. Evaluating Doxasticism

It is now time to evaluate doxasticism in light of the discussion up to here, including that of the preceding chapter.

Let us begin by stressing what doxasticism seems to be doing right. First, it correctly identifies what the audience needs to understand: a mental state having an intentional structure. Moreover, it correctly identifies the intentional state in question as relating the speaker to the evaluative nature of the object in relation to her concerns. In the case of 'The frogs won it again', the mental state is (among other things) about the French people's supposedly contemptible nature. By the same token, it correctly

treats the mental state as having correctness conditions. [124] It thus allows us to spell out why using the expression 'The frogs' is normatively defective: because French people aren't in fact contemptible or despicable. This is good news for doxasticism since it adequately explains an important feature of certain pejoratives (slurs, derogatory expressions, etc.): that we tend to blame people using them (Hom, 2008; Williamson, 2009). Similarly, if one says 'Outrageous!' about something completely innocent and innocuous, we may blame this person because the correctness conditions of her exclamations do not obtain.

But this seems to be the extent to which the account satisfyingly renders what was put forward above. In fact, doxasticism fails to satisfyingly meet the three benchmarks that distinguish expressives from descriptives – (a) hot vs cold, (b) appropriate vs true, and (c) direct vs indirect. As such, even if doxasticism captures some of what is meant to be conveyed by expressives, it fails to account for what it is to understand expressives optimally, and so can, at best, only tell us part of what the meaning of expressives is.

(A) HOT VS COLD

The account seems to ignore the intuition according to which there is a distinctively affective phenomenal character to the state that is expressed by the speaker in uttering expressives. Doxastic attitudes do not have the adequate phenomenology to capture the 'hotness' of emotions (Goldie, 2000) and thus we cannot account for the intuition that expressives carry information from 'hot' mental states. Undergoing an emotion about X is importantly different from having a belief about X, even when the emotion and the belief track the same evaluative properties. We can elaborate on this from different perspectives.

First, let us take a purely epistemic, information-retrieval perspective. If you merely believe that walking on ice is dangerous, but you are not afraid of walking on ice, you don't approach walking on ice in the same way as if you really were afraid of walking on ice (see the description of Mary the ice-scientist in Goldie, 2002). If you merely believed that it was dangerous without being afraid, you probably wouldn't be as cautious in the exploratory movements of your feet, you would pay less attention to the appearance of the ground, you wouldn't spot certain light reflections that your fear would allow you to distinguish by focusing your eyes on potential

[124] These two conditions (intentionality and evaluative correctness) would not be met by non-intentionalism about expressives. The second condition would not be met by non-cognitivism about expressives (a view defended by e.g. Searle 1979: Chapter 1).

threats, etc. If you merely believed that walking on ice is dangerous, some information would escape your attention. Being afraid modifies information retrieval tendencies, it focuses attention, it reinforces associated memories, it feeds the whole organism with new energy, it renders one more careful, etc. In sum, affects help gather and treat information relevant to your goals, needs, and values, information that would otherwise be unavailable or unused. Expressing affects is thus quite different from expressing doxastic attitudes, even if we restrict the difference to an epistemic perspective. This is one way of cashing out the 'hotness' metaphor in cognitive terms.

One should think further about the felt bodily engagement which I described in the last chapter (e.g. feeling one's body as ready to run, the muscles tensed, the eyes wide open, the heart racing, …). What would it mean to translate these felt bodily engagements into doxastic states? Is it possible to account for them with a set of affirmative sentences expressing the propositions believed by the expresser? Or is it impossible to translate this phenomenological feel – the 'hotness' of the state – into a set of believed propositions? Think of how effective expressives can be at conveying an affective experience: you may readily understand how one is feeling thanks to the appropriate use of an expressive (e.g. a stringent, loud 'Shiiiiiiit!!!!!') – the 'hotness' of one's affect is effectively conveyed.

Finally, the hotness of emotions also appears in their intimate link with action-tendencies, a link which beliefs by themselves lack. This is, once again, apparent in expressives, but it is hard to see how doxasticism would account for it. If someone screams 'Outrageous!' and another says, in a neutral tone, 'What the government did was wrong.', we may expect the first person to be more readily engaged in actions against the government, to possess a stronger inclination or desire to change things, and we may expect her to react more strongly to someone expressing a divergent point of view (e.g. someone saying 'What the government did is not so bad').

Expressives seem to have the ability to convey the hotness of the state one is undergoing in ways that are much more potent than through affirmative sentences expressing doxastic attitudes. Doxasticism is in quite a bad position and is perhaps incapable of rendering the features of expressives which derive from the 'hotness' of affects. These features are indeed very hard to capture, and perhaps cannot be captured, merely by propositions believed by the utterer.

(B) Appropriate vs True

Because doxasticism requires us to identify the attitude expressed by the speaker with a doxastic attitude, it forces an understanding of the state in question as one that represents its object as true. Furthermore, if one is a representationalist about doxastic attitudes, one normally understands them as propositional attitudes which require possession of the concepts that make up the proposition entertained (Mandelbaum, 2016; Quilty-Dunn & Mandelbaum, 2018; Schwitzgebel, 2019, sec. 1.1). Under this conception, in the 'frogs' example, doxasticism entails that we must attribute to the utterer a representation of French people as contemptible, a representation that it is correct to have if, and only if, the concepts of FRENCH and CONTEMPTIBLE are entertained in a propositional form which makes the proposition entertained true. When I discussed arguments against judgmentalism about emotions in the last chapter, I gave reasons to doubt that expressives really recruit the normative standards of truth and falsity. I defended that the normative standards of correctness for expressives instead inherit those of emotions, which are the standards of (in)appropriateness, (un)deservedness, or (un)merit rather than truth and falsity. I also presented strong arguments for the claim that the content of emotion need not be conceptual.

This is linked to the fact that we can have emotions, and express them, without having the corresponding beliefs. For instance, you can be persuaded that it is not dangerous to fly by plane but still be afraid of flying, and you can express this fear by saying 'Mama miaaaaa!' when the plane starts the engine. This doesn't mean however that you have two contradictory beliefs: believing at the same time both that flying is dangerous and that flying is not dangerous. Even if affects and doxastic states are both evaluations, and even if they seem to attribute the same evaluative property to their object (danger in this case), they seem to differ in how they evaluate their object (e.g. through a conceptual representation or not).[125]

(C) Direct vs Indirect Display

Third, doxasticism forces on us the idea that the utterance is a proxy for something internal to the mind of the utterer and that cannot be directly seen or heard, since doxastic attitudes do not involve a bodily component

[125] Note however that if one holds the view that implicit biases are unconscious beliefs and that beliefs may be fragmented so that one may unproblematically believe the propositions that p and ¬p at once (Mandelbaum, 2016), then the last objection seems not to apply.

that shows on one's face or in one's voice.[126]You cannot guess whether or not I believe that the French are contemptible just by looking at my face, studying my posture, or by hearing the sound of my voice. Of course, you may infer from my behavior that I do not believe that the French are compatible, but this is not a direct perception of my belief, it is an indirect inference. This is in stark contrast to how you can access evidence for my undergoing contempt: you can see or hear it directly because you can see or hear a proper part of the emotional episode: its expression. Indeed, emotional expressions are usually considered as proper parts of emotional episodes, as has been noted by philosophers such as Scheler, Wittgenstein, Austin (and more recently by e.g. Green (2010)), as well as by many psychologists (see reviews by Sander et al., 2018; Scherer & Moors, 2019). For instance, a wrinkled nose may partially constitute an episode of disgust because it is part of the motor component of this emotion, which itself is a part of the emotion episode. And so you seeing the wrinkled nose constitutes a case where you directly perceive a proper part of the emotion, just like you may directly perceive someone running by seeing only her feet.

The fact that doxastic attitudes cannot be directly seen or heard goes against the intuition that expressive utterances themselves seem to be constitutive of the emotions they express, being a proper part of the emotion. As a consequence, doxasticism cannot explain the intuition that expressives can directly acquaint us with the affect they express.

Additional objections

We see that doxasticism is in tension with the three benchmarks with which we started. To these issues, we can add two final worries: that doxasticism implies judgmentalism about emotions and that it requires the state expressed to have a propositional content. I will address these worries in turn.

According to doxasticism, one understands the meaning of an expressive if, and only if, one attributes to the communicator the right doxastic attitude (and in particular the right evaluative belief) ($p \leftrightarrow q$).[127] Now, we started this chapter with two claims defended in the last chapter. First, the nature of expressives is to express affects and in particular emotions. Second, one understands the meaning of an expressive only if one

[126] Attitudes such as surprise, strong incredulity, 'hot' doubts, are construed as affective attitudes here. They indeed involve the typical affective components (appraisal, action tendencies, physiological changes, motor expression, subjective feeling) and this explains why they can show on one's face, voice, posture, etc.
[127] More precisely, this is evaluative doxasticism as opposed to affective doxasticism, see the footnote about Kaplan (1999) above.

understands what attitude is expressed. Putting these two starting points together, we get that one understands the meaning of an expressive only if one understands the affect expressed (and in particular the emotion expressed) (p→r). Supposing that these starting points are correct, doxasticism entails that if one attributes to the expresser the right doxastic attitude (and in particular the right evaluative belief), then one understands the affect expressed (and in particular the emotion expressed) (q→r).

How should doxasticism explain this entailment (i.e., q→r)? Well, an obvious way would through judgmentalism about emotions, i.e. the idea that emotions are nothing but evaluative judgments (Nussbaum, 2001; Solomon, 1993) coupled with the idea – presented in the last chapter – that understanding affects amounts to understanding emotion-related attitudes. If these were true, this would indeed explain why attributing a doxastic attitude to the expresser is sufficient to understand the affect she expresses (q→r). If there are no other explanations available, and I don't see that there would be any, we are led to the conclusion that doxasticism implies judgmentalism about emotions. [128]

This should not come as good news for doxasticism, because, as I noted in the last chapter, judgmentalism about emotions is defective in many ways (Deigh, 1994; Deonna & Teroni, 2012, 2014; Prinz, 2004; Scarantino, 2010, 2014; Tappolet, 2000, 2016). If doxasticism implies judgmentalism about emotions and if judgmentalism about emotions is not a good theory, then doxasticism is not a good theory.

Let me note however that many of the critics against judgmentalism about emotions are directed at views of judgments and beliefs that are not held by everyone. For instance, if one holds that beliefs can be inaccessible to consciousness, fragmented, and gradable (Mandelbaum, 2016; Quilty-Dunn & Mandelbaum, 2018), then several criticisms made against judgmentalism won't hold. Nevertheless, even with such a view of beliefs, judgmentalism won't be safe since several potent arguments against it remain (e.g. that emotions need not be propositional attitudes, that they need not require the possession of the evaluative concepts corresponding

[128] Another way to make my point is that the following seems very plausible: If attributing to the expresser the right doxastic attitude implies that one understands the affect expressed, and if understanding affects amounts to understanding emotion-related attitudes, then emotion-related attitudes are doxastic attitudes (((p→r)&s)→t). We have seen that doxasticism and our starting points imply that p→r and, in the last chapter, we have defended that s is true (or at least very plausible).

to their correctness conditions, that they involve a conative aspect absent from beliefs, and that they involve a phenomenology absent from beliefs).

Let us now turn to our final worry: that doxasticism implies that the content of the mental state ascribed to the expresser is always propositional. This is the case if we follow the usual view of beliefs (supposition, conjectures, …) as propositional attitudes, a view held by proponents of very different views on beliefs (Quilty-Dunn & Mandelbaum, 2018; Schwitzgebel, 2019). This is a problem insofar as there may be expressives where well-formed propositions are to be found neither in the signal nor in the mind of the utterer and thus where the illocutionary intent is not to express a propositional attitude at all. For instance, we may well express our contempt, disgust, or admiration with expressions such as 'Pfff', 'Yuk!' or 'Wow!' without expressing affective attitudes whose contents are propositional.[129] I find this view plausible if one uses 'proposition' as a mental representation structured through syntactic operations such as predication, Merge, or functional application (for the relevant notion of proposition, see Camp, 2018b), as opposed to sets of possible worlds (Stalnaker, 1976).[130]

This is in essence the point made by Piero Sraffa to the young Wittgenstein. The latter was trying to explain that all meaningful entities have the logical form of propositions and Sraffa asked him 'What is the logical form of that?' while mimicking the famous Neapolitan gesture of spite (Malcolm, 2001: 58-59). My interpretation of Sraffa's point is that the meaning expressed by this Neapolitan gesture is not propositional.

The problem in question is compounded by the fact that doxasticism implies that, if the audience understands the doxastic attitude expressed but does not recover the propositional content, then barely anything at all is understood: understanding the fact that the speaker expresses a doxastic attitude in absence of any propositional content is not satisfying. This is in sharp contrast to understanding that the speaker expresses an affective attitude. Consider for instance asking how Joe is and being answered 'He believes', 'He judges', or 'He supposes' versus being answered 'He is sad', 'He is angry', or 'He is joyful'. Attributing doxastic attitudes without

[129] For the view that some mental states, including emotions, may have a sub-propositional content, see Montague (2007).

[130] One may respond to this objection by claiming that there are unconscious mental states which make up these affects that are propositional (e.g. unconscious appraisals or belief-desire pairs). However, it seems problematic to require the audience to recover these unconscious propositional attitudes instead of what the speaker consciously experiences and intends to express. One may also respond by saying that, despite appearances, there are no affective states whose content is non-propositional.

contents is very strange, but this is not the case with affective attitudes, probably because understanding that someone is experiencing an affect is already to know a lot, even in the absence of any inkling as to what the intentional content of the affect is, which is not the case for doxastic attitudes.

We have seen several reasons to doubt that doxasticism can explain what is special about expressives, what it takes to understand expressives optimally, and so what is special about the meaning of expressives. We are thus led to abandon the view.

Rescuing Doxasticism?

Before we move on, let us note that one might nevertheless defend doxasticism against all these criticisms by denying either one of two of our assumptions. First, one can try to hold onto the view that affects are doxastic attitudes and find counter-arguments for each point that was made against judgmentalism. As we have seen in the preceding chapter, this idea is not popular.

Second, one can maintain doxasticism if one holds that expressives do not have the function of expressing affects. Indeed, the defender of doxasticism could argue that the use of expressive is a convenient and efficient way of conveying one's evaluative judgments. For instance, by using the word 'yuk' one could intend to communicate that the food is disgusting irrespectively of whether one feels or is disposed to feel disgust toward the food. The doxasticist can further argue that the use of conventionalized words points to the fact that there is a certain intellectual distance between the affect and the verbal expressives. The affect would have been 'translated' into a doxastic attitude and the latter, not the former, is what we should understand in an expressive.

Such a defense is coherent, but has an important consequence: it treats expressives as descriptives. In the terminology of speech acts, the two would thus be taken to have the same illocutionary act. Such a doxasticism would not be a theory of how to understand the expression of emotions and other affects in language, but rather a theory of how to understand descriptives whose content involves evaluations, which might or might not come from affects. This may be considered a fatal consequence if one is convinced that expressives convey affects and that this is how expressives differ from descriptives, which was our starting point, a starting point around which benchmarks (a)–(c) revolve. But if one rejects this important premise, doxasticism should be a viable option, with one important

condition: that the doxasticist give an alternative account of how and why expressives (such as (1)–(3)) appear to differ from evaluative judgments or reports of one's emotions (such as (4)–(9)), an account that doesn't rely on the differences between affective and doxastic attitudes. Alternatively, the doxastist may defend that expressives and descriptives don't differ after all, which begs the question of why researchers have drawn this distinction in the first place.

Another problem for such a doxasticism would be the following dilemma: either to explain why there doesn't exist any kind of speech act whose illocutionary intent is to express one's affects, or to explain why they exist but aren't to be considered expressives.

6.3. Moderate Affectivism

I have offered some reasons to doubt the viability of doxasticism, the view that understanding the meaning of expressives consists in understanding the relevant propositions believed (doubted, supposed, etc.) by the speaker. In this section, I articulate the view according to which understanding expressives implies recovering the relevant affects of the speaker. A view along these lines is defended or suggested, in different ways, by several authors (Bach & Harnish, 1979; Dorit Bar-On, 2017; Copp, 2009; Croom, 2011; García-Carpintero, 2017; Green, 2007; Jeshion, 2013; Marques & García-Carpintero, 2020; McCready, 2010; Potts, 2007; D. J. Whiting, 2008). Note that, among these authors, we find several researchers who have paid more attention to the nature of affects than any of the putative defenders of doxasticism I mentioned above.

6.3.1. Introducing Moderate Affectivism

The distinctiveness of affectivism – moderate or radical – lies in the importance it puts on the fact that when one understands an expressive, it is *affects*, with all their specificity, that are ascribed to the speaker, and that these affects are expressed directly.[131] *Moderate* affectivism about expressives claims that, to understand sincere expressives optimally, it is necessary and sufficient that the audience believes that the speaker is expressing the relevant affect. By contrast, *radical* affectivism claims that this is indeed necessary, but is not the full story, because it is also

[131] As opposed to mediated by the expression of a doxastic attitude which one has about one's affective state. This latter view is what I have called 'affective doxasticism' in a footnote above. Roughly, it identifies the meaning of an expressive such as 'Outrageous' with something like 'I believe that I feel outraged'. I do not discuss this view in the main text because, for reasons I outlined in footnote 124, I do not find the view plausible and, furthermore, I do not know of anyone that defends such a view.

necessary that the audience *undergoes* an affect about the speaker's affect to qualify as optimally understanding an expressive, as we shall see in §6.4.

Let us now flesh out moderate affectivism about expressives by going back to our first examples and comparing pairs of expressives and descriptives that are about the same states of affairs.

> (3) The frogs won it again!
>
> (8) The French won the world cup again and I believe that the French are contemptible.

Contrary to doxasticism, moderate affectivism has it that (8) does not come close to being a good rendering of what (3) expresses. Ascribing to the speaker the corresponding beliefs would not count as understanding utterance (3). To count as understanding (3), says the affectivist, the audience needs to ascribe to the speaker the relevant affect, which we assume is contempt for the French, i.e. an experience of the French as contemptible or, even more to the point, this felt bodily attitude towards the French that is appropriate if, and only if, the French are contemptible (see Chapter 5 for more on this). This is what the audience needs to believe about the speaker in order to qualify as understanding what he or she means by (3). Moreover, the affectivist will insist that the audience's belief must be acquired on the basis of being directly aware of the contempt in the voice or the behavior of the speaker. By uttering (3), the speaker has displayed, as opposed to reported to the audience, the kind of value the French have for him. Thus, the affectivist will argue that (9) is not a good rendering of (3) either:

> (9) The French won the world cup again and I feel contempt toward the French.

Although (9) fulfills the affectivist's requirement that an affect is ascribed to the speaker, (9) reports the affect instead of displaying it. (9) in fact implies that what is conveyed is that the speaker believes she feels contempt toward the French, and so (9) expresses a belief about an emotion instead of the emotion directly.[132]

Much the same can be said about the following trio:

> (2) Ouch!!!
>
> (6) This boiling oil has burned my hand and this is bad for me.

[132] See the preceding footnote.

(7) I feel great pain.

In this case, the affect is a pain that is intentionally directed towards a burn. So understanding (2) requires understanding that the speaker is having a pain and not that the speaker is having a belief about a pain of his, as in (7). Even if pains, let us assume, constitute forms of evaluation, being experiences of various forms of tissue damage in various parts of the body and that tissue damage is bad for the subject (Bain, 2017), affectivists would argue that they cannot be equated with non-affective forms of evaluations, such as the doxastic attitude expressed in (6), even if the evaluation in question is directed at the same event or state of affairs. This is because the attitudes expressed in (2) and (6) differ: the attitude in (2) is a hedonically salient bodily reaction, while that in (6) is a cold, distanced, intellectual consideration.

In the terminology of speech acts, (2) should be understood as having a content very close to that of (6), but a different illocutionary force. Analyzing an expressive would require not only making explicit the propositional or objectual content (what is evaluated as more or less correct), but also the attitude which gives the particular illocutionary force of expressives, their affective stance toward their content (for a similar conclusion, see García-Carpintero, 2015, 2017; Marques & García-Carpintero, 2020).

According to affectivism, a further and important difference between (2) on the one hand and (7) or (6) on the other is that a sincere 'Ouch!' is not just a signaling of the presence of a pain or evaluation towards bodily damage. Screaming in reaction to bodily damage is arguably constitutive of the affect expressed. As I have noted already, along with many psychologists (see Scherer and Moors 2019 for a review), we may say that the action tendencies and motor reaction of affects are proper parts of the affects (along with, at least, appraisals, physiological changes, and subjective feelings), so that if one comes to believe that the speaker has a pain on the basis of such an utterance, one forms the belief on the basis of a direct awareness of the pain.

6.3.2. Evaluating moderate affectivism

Let us now evaluate moderate affectivism by assessing how it can deal with the benchmarks (a)–(c). I will present benchmark (a) last because it will allow a better transition toward radical affectivism.

(B) APPROPRIATE VS. TRUE

Unsurprisingly perhaps, moderate affectivism is doing quite well with respect to the 'appropriate vs true' benchmark. Given a view of affects which always involve a cognitive component (see previous chapter and the next), it correctly predicts that the audience needs to understand a mental state that has an intentional structure (e.g. an emotion directed at the French or a pain directed at the burn), it correctly identifies the intentional state in question as relating the speaker to the evaluative nature of the object in relation to her concerns (French people's contemptible nature, one's bodily damage), and it correctly treats the mental state as having correctness conditions. As such it has the same virtues as doxasticism. But then it seems to do much better with respect to the way in which it captures these correctness conditions. Indeed, it conceives of the mental state ascribed as an affective engagement with, or attitude towards, the world which is amenable to an assessment in terms of appropriateness rather than truth. This is also what is captured by saying that expressives differ from descriptives in the illocutionary force they carry. That seems right and corresponds to benchmark (b) above.

(C) DIRECT VS. INDIRECT DISPLAY

Another strong advantage of affectivism over doxasticism is that it very naturally captures the fact that expressives seem to directly show what they express. We have seen that the expressive component of emotions may be directly observable and that this importantly and relevantly distinguishes them from doxastic attitudes. If what one needs to retrieve from the utterer of an expressive is an occurring affect, as opposed to a doxastic attitude, then it makes sense to say that expressive utterances seem to partially show what they mean rather than report on it.[133] Affectivism easily meets benchmark (c).

[133] But what about expressives which express non-occurring affects, such as affective dispositions? In such cases, I believe that the 'direct vs. indirect' benchmark actually does not hold anymore. In a sentence like 'There are 177 Swiss, 123 Italians, 27 Germans, and 54 frogs living in this building' where one of the communicative intents is to express francophobia (an affective disposition), but where there is no expression of occurring emotions (by contrast to the use of 'frogs' in (3) which we construed as expression of occurring contempt), it does not seem to be the case that the mental states expressed are directly displayed or shown as opposed to indirectly conveyed. Such expressives (if they really are expressives) thus seem to be similar to descriptives with respect to benchmark (c).

(A) HOT VS COLD

How does moderate expressivism deal with benchmark (a), the 'hotness' of expressives? From one perspective, it seems to deal quite well with this feature of expressives since the attitude ascribed to the speaker is phenomenologically hot (or a disposition thereof). Here as well, it fairs better than doxasticism. But, from another perspective, moderate affectivism, just like doxasticism, fails to capture the special 'hotness' of expressives. Indeed, what is ultimately striking about moderate affectivism is the complex belief that the audience must form in relation to the utterer's emotion in order to count as understanding her. The central insight of radical affectivism, to which we will now turn, is that the audience does not need to have complex beliefs about the utterer, at least not in all cases. But, in all cases, it needs to undergo an affective response to the affects expressed in order to qualify as optimally understanding the expressive. A main advantage is that, by allowing an understanding that bypasses doxastic attitudes altogether, radical affectivism eschews the threat of divorcing expressives and pre-expressives, as we will see.

Before I move on, let me mention that a virtue of affectivism (moderate or radical) is that the attitude ascribed to the speaker need not be a propositional attitude. Affects expressed by 'Ouch!', 'Yuk!', 'Pfff', or 'Wow!' may well have an objectual content (e.g. a location in one's body, a taste, an action, a painting) rather than a propositional one, i.e. are directed at objects rather than states of affairs (Montague, 2007).

6.4. RADICAL AFFECTIVISM

6.4.1. INTRODUCING RADICAL AFFECTIVISM

Radical affectivism, just like moderate affectivism, claims that, for the audience to qualify as optimally understanding an expressive, it is necessary that the audience retrieves the affects intended to be communicated. Contrary to moderate expressivism, it claims in addition that it is necessary that the audience understands the affect expressed *through its own affects*. In short, radical affectivism claims that *a proper understanding of expressives requires affective resonance*. According to this view, the audience would not optimally understand an expressive if it reacted coldly. Its understanding would be incomplete unless it reacted by being in the appropriate affective state. Observe that if one reacts by having the appropriate belief about the affective state of the expresser and undergoes the appropriate affective state (as a downstream consequence, an upstream cause, or for causally independent reasons), then the

conditions for both moderate and radical affectivism obtain. If, on the other hand, the audience reacts with the appropriate affective state, but without a belief, then only those for radical affectivism may obtain.[134]

The audience could resonate affectively with the expresser either by sharing the same affect, in a mirror-like response, or by undergoing a different affect that is an appropriate, complementary reaction. Different mental mechanisms could be at work in the first case – the mirror-like case. The audience could empathize with the speaker, undergo emotional contagion, simulate the speaker's affect through imagination, revive an episodic memory of the affect in question, undergo an affect for the other through another social cognitive mechanism (Theory of Mind, mirror neurons, participatory sense-making, vicarious perception, etc.). In the second type of case – the complementary-affect case – the audience should undergo an affect congruent with what the expresser intends to do with her expressive speech act. For instance, if one is angrily reprimanded, a complementary affect of the target audience could be feeling guilt. If one apologizes, sincerely expressing regret, a complementary response of the target audience could be forgiveness, relief, or gladness. If one expresses excitement about an unexpected piece of news, beside the mirroring response of also feeling excitement, a complementary response could be feeling surprised. In any case, radical affectivism claims that, to optimally understand the expresser, the audience need to be related to the affect expressed *through its own affects*.

Let us illustrate. If one utters 'Ouch!' in reaction to a burn, radical affectivism would argue that, for communication to be optimally successful, it would not be adequate to merely believe that the speaker suffered some pain. Instead, or in addition, the audience would need to resonate affectively with the pain, to be attuned to the pain through, for instance, a mirror-like response like empathic reaction.

[134] Note though that, for radical affectivism, there may be many cases where it is required to ascribe to the speaker both an affective and a doxastic attitude, as in the Frege–Geach problem. Such sentences would have both an expressive and a descriptive illocutionary intent, as was mentioned at the end of the last chapter. Note also that, because the meaning of expressives (as opposed to that of pre-expressives) belong to speaker-meaning, and because our definition of speaker-meaning and the speech-act framework presented in the last chapter requires that the speaker's intention to express an affect be *mutually recognizable*, which requires being disposed to have an attitude with collective intentionality (WE recognize), then radical affectivism cannot defend that it is sufficient to merely react with an affect which lacks collective intentionality without in addition being disposed to have another attitude which possesses collective intentionality. The audience must be disposed to have another attitude with collective intentionality in addition to the affective attitude required by radical affectivism (e.g. I feel empathy toward your outrage and it is recognizable for US that you are expressing outrage).

In this example, that the audience understands the relevant affect in an intimate way is easily understandable, since the audience undergoes the same type of affect as the speaker. It thus makes sense that the audience possesses a more complete understanding of what the speaker expresses than if the audience did not undergo this affect. But what about cases in which the audience reacts to the expressive not through a mirror-like response, but through a complementary response? How would that help the audience to understand the affect expressed by the speaker? How can one be acquainted with a certain affect (say, anger) by feeling a different affect (say, guilt)? Indeed, the pairs of complementary affects I have given above (anger–guilt, regret–forgiveness, regret–relief, regret–gladness, excitement–surprise) involve two emotions which feel (very) differently, have (very) different action tendencies, etc.

The idea that complementary response nevertheless allows you to better understand what the other person is expressing than responding merely with a doxastic attitude is that, even if the two emotions are very different, these emotions nevertheless acquaint the relevant people with the exact same state of affairs – and, more precisely, with the same evaluative state of affairs – only from two different perspectives. For instance, if you are angry about my behavior, because you apprehend it as offensive, and that, in response, I feel guilty, then, if all goes well, I feel guilty about the exact same thing that makes you angry: my offensive behavior. Both your anger and my guilt thus acquaint us with the same state of affairs, although we have different roles in this state of affairs and thus different perspectives about it: you are the victim, I am the perpetrator. A similar story could be told for all complementary affective responses. Even if we have different perspectives, our complementary emotions give us special, privileged, access to the evaluative state of affairs. This access is through a 'hot' psychological state which, arguably, cannot be attained by 'cold' psychological states such as doxastic attitudes (Deonna & Teroni, 2012, Chapter 10; Goldie, 2002; Prinz, 2007; Tappolet, 2000, Chapter 7).

As said in the introduction, understanding comes in degree and, in the sense in which we are concerned, it depends on how well the communicative functions are fulfilled. Let us distinguish between 'optimal communication' and 'good enough communication'. By 'good enough' communication, I mean to describe those cases where we would say that communication did take place and that, from the point of view of the communicators, it was sufficiently successful to be deemed minimally valuable. By contrast, as said above, optimal communication only happens when all the communicative functions are fulfilled. I believe that there are cases where good enough communication is nothing less than optimal

communication, but others where the standards for good enough communication are much less stringent.

For communication to take place, it need not be perfect. The communication channel may be 'noisy', but still count as a communication token and, in circumstances where the stakes of having optimal communication are not very high, this would be *good enough*, even if not optimal. Think for instance of someone listening to a football match on the radio. Even if the radio signal is defective and the listener loses 10% of the broadcast, this may well be good enough for the listener's purpose because (as is usual in football) nothing interesting happened during the 9 minutes where the signal was lost. By contrast, in other circumstances, losing 10% of the information would not be good enough (e.g. a text message saying 'Let's meet at the grocery store called ****' where 9.3% of the information sent is missing).

The defender of radical affectivism can, on the one hand, agree with moderate affectivism that if the audience merely *believes* that the speaker has undergone pain, this is good enough in certain situations, and, on the other hand, add that if the audience only reacts by entertaining a mere cold belief, devoid of any affective tone, the audience is unable to grasp all the information transmitted, information which only affective attitudes can recover.

Furthermore, besides information transmission, radical affectivism can highlight that if the audience does not react to the expressive with an affect, but reacts with a cold belief instead, the *influence* that the expresser will have on the audience's behavior may be good enough, but it will not be optimal. (You may remember from the introduction that the two main functions of communication generally are taken to be transmitting information and influencing others (or 'manipulation').) That is so because if the audience had reacted with an affect instead, the audience would be more disposed to deploy the action tendencies intended by the expresser, in virtue of the essential link that affects have with action tendencies. For instance, if you angrily reprimand me about my offensive behavior, but I merely form the belief that I have committed something offensive without feeling guilty about it, I may tend to reproduce this kind of behavior more than if I had felt guilt, such that, from a communicative point of view, it is not optimal that I merely form a belief without feeling guilt. In certain situations, this may be good enough (because the stakes of me not reproducing my behavior are not very high) while in other cases this non-optimal communication won't do.

Whatever is the communicative function which we evaluate, a communication token can be good enough without being optimal, because, in the relevant context, it meets some minimum standard for communication while failing to optimally fulfill the function in question.

To give more of an intuitive appeal to radical affectivism, let us note how we sometimes require our interlocutors to be affectively attuned with us. During an enraged argument, when your partner says 'No, you don't understand what I feel!', what is important for him or her is not that you *believe* he or she undergoes an emotion, but that you are in a much more intimate relation with his or her feelings. Of course, you know *that* he or she is angry, you also know *why* he or she is angry, but what your partner would want you to understand is *how* he or she really feels and *what* he or she really feels, and that you react accordingly, by changing your behavior and being disposed to act differently.

I find it intuitive that what is required for communication to be optimal in such situations is that the audience resonates affectively with the interlocutor – in fact, in this situation the communication may not be good enough unless you resonate affectively with your partner. This is so because of the two communicative functions discussed. First, concerning information transmission, if one merely believes that the interlocutor undergoes this or that emotion, one transforms the action-ready, bodily, phenomenologically hot information expressed into an intellectual, passive, phenomenologically cold representation. Information transmission is not optimally successful; what is special about emotions has been partially lost in translation from the 'language' of affects to that of beliefs. Second, concerning behavioral influence, if you undergo the relevant affect as opposed to merely forming a belief, because of the intimate link between affects and actions, you will be much more disposed to act in the ways intended than if you merely *believed* that he or she is expressing this or that affect. This explains why affective resonance is better placed to fulfill one of the functions of expressive communication, i.e. the aim of influencing the behavior of the addressees.[135]

But if we concentrate on information transmission, we see once again how affects cannot be epistemically reduced to doxastic attitudes, a claim which we have already encountered in the preceding (especially when discussing

[135] Here is another example given to me by Julien Deonna. If you say to your child "Understood?!' after having expressed how important it is that her behavior must not occur again, you not only ask her if she grasped what you want from her, but also that her behavior will conform to what you ask from her. See also the remark made by Bar-On (2017, p. 304) on the Janus-faced function of expressive communication quoted in the last chapter (§3) and on how its purpose may be to influence ensuing behaviors.

doxasticism above as well as the arguments against judgmentalism about emotions in the last chapter). Thus, a first set of reasons for holding radical affectivism are identical to the ones one would have to reject doxasticism. From an epistemic perspective, undergoing an emotion about X is importantly different from believing that X, even if the belief is about the same evaluative state of affairs (remember the example above from Goldie 2002 on believing that the ice is dangerous vs being afraid of the ice).

The radical affectivist can argue along these lines that a cold belief about the emotion of a speaker is epistemologically less accurate than an affective understanding of the speaker, a reaction where the affect communicated resonates in the audience, where the audience is affectively attuned with the speaker. The emotion of the audience is not some dispensable downstream effect of their beliefs, it is the psychological attitude which allows them to really grasp the information that is meant to be communicated. To empathize with someone screaming 'Ouch!' or 'Outrageous!' may be the only way to really understand what this person feels and thus what the person wants to communicate. Once again: remember how affects modify information retrieval tendencies, focus attention, reinforce associated memories, feed the whole organism with new energy or drain it out, bias reasoning (virtuously or not), and thus play a fundamental role in our acquiring and processing the information that is meant to be communicated through expressives.

The purely doxastic, cold, intellectual understanding of the speaker's affect demanded by moderate affectivism might be good enough in many everyday interactions, especially with strangers, but such an affectless understanding of expressives is not sufficient for optimal understanding and is not good enough in other scenarios.

Let us further illustrate the distinction between 'good enough' and 'optimal' understanding through a couple of examples. First, imagine that you are driving at the legal speed limit and that an impatient driver behind you wants to overtake you, but cannot do so for a few minutes. When the driver finally succeeds, she opens the window and angrily shouts 'Freaking snail!'. According to radical affectivism, it may be good enough to just form the belief that this person is angry, but it would not be an optimal understanding of the expressive. To optimally understand it, the audience could either undergo a complementary response by e.g. feeling sorry for making this driver so angry, or by feeling a mirror-like affect for the driver, undergoing empathy for instance or by reviving affective memories (who hasn't been mad at someone driving too slowly?).

Now, since this driver is a total stranger that you will probably never see again, the pressure to optimally understand her is extremely low, and you would have all the excuses in the world to not care about optimally communicating with her. In fact, if you index the value of 'good enough communication' on how well the communication allows the sender and the receiver to achieve their goals, then a good enough communication in this case may be close to zero communication: given the situation, very little information transfer and very little behavioral influence may be considered to be good enough communication in the sense that what is communicated by the angry driver may not be very useful in allowing you or her to fulfill your goals. Especially from your perspective, that of the receiver, there may be nothing problematic if you ignore her (remember that you were driving at the legal speed limit). And even from the perspective of the sender, it is not clear what important goals would fail to be reached if only very little information is transmitted and if no behavioral influence ensues. In other words, in this case, there is little pressure, normatively speaking, to optimally understand her.

By contrast, in the argument with your partner, the communicative stakes are much higher, because it is someone you care about and because you have such an important role to play with respect to each other. What is 'good enough' communication is in this case much more demanding. The difference in stakes may (partially) explain why there is an intuitive sense in which one ought to optimally understand one's partner, but not the insulting driver.

One could also imagine a possible world where you always ought to optimally understand anyone who addresses you. In such a situation, radical affectivism would imply that you ought to undergo mirror–like or complementary affects in response to all expressives addressed to you.

Now, let us also observe that there are many cases where expressives are not directed at you. Take for instance the passionate speech of a politician. This speech may be addressed only to her followers and her potential followers. Consequently, the communicative functions only concern a limited number of people, and, fortunately, radical affectivism would not require everyone in the world to undergo the relevant affects in response to the politician's speech. Now, if the politician aims to direct her passionate speech at everyone in the world, it is understandable that this crazily ambitious communicative intention would be far excessive, it would be an illocutionary act that is practically impossible to realize, and so it would not be the case that everyone in the world really is the appropriate

audience, despite the existence of this (megalomaniac) communicative intention.

These examples thus point toward the fact that, although radical affectivism may seem too demanding (how often do we undergo affects in response to expressives directed at us?) there are at least two reasons why one is not normatively required to have an affective response to all expressives: (i) sometimes it is not the case that we ought to optimally understand expressers and (ii) we often are not the appropriate target audience of an expressive.

6.4.2. Evaluating Radical Affectivism

Let us now turn to evaluating radical affectivism with respect to the three benchmarks with which we begun: it succeeds in meeting the same benchmarks as moderate affectivism – (b) appropriate vs true and (c) direct vs indirect – since it also requires that we detect an affect expressed, but it is better at explaining the fact that the meaning of expressives is 'hot' – benchmark (a). It captures this by claiming that the way an audience optimally understands an expressive is by directly being acquainted with what is 'hot' about it, i.e. by undergoing an affect similar or complementary to the one expressed.

As mentioned above, another reason to defend radical affectivism is to avoid an artificial divorce between expressives and pre-expressives[136], which is a risk for both doxasticism and moderate affectivism since these two theories require doxastic states that are often conceived as cognitively quite demanding. They indeed require a belief about the others' mental states which may well require passing the false-belief task. By contrast, the mental mechanisms which radical affectivism requires the audience to possess – i.e. affects and a minimal form of social cognition – are not cognitively demanding in the sense that they seem to be shared by many nonhuman animals.[137] Think of animal alarm calls, such as those of birds' or monkeys'. It is probable that what happens in many cases is that one

[136] Pre-expressives are communicative acts lacking speaker-meaning (e.g. because the creature producing it does not have the sufficient mindreading abilities) and so lacking a full-blown illocutionary intent, but being nevertheless produced with the pre-illocutionary intent to express an affect about a content, see last chapter and the next

[137] But see the two previous footnotes: if we preserve the definition of speaker-meaning and the speech act framework presented in the previous chapters, then more than a minimal form of social cognition would be required to understand an expressive, as opposed to a pre-expressive. There would be a need to understand the expressive illocutionary intent as being *mutually recognizable*, which arguably is not possible for many nonhuman animals nor infants below a certain age (about 9 months according to Tomasello (2008)).

animal is afraid of a predator, expresses its fear through a call (similarly to a human screaming from fright), its group members resonate affectively with the caller, and thus display the fear action-tendencies which enhance the chances of escaping the predator (e.g. fleeing or freezing) (Andrews, 2015, sec. 5.2.2, see also Bar-On, 2017). There is no need for a complex mental ascription here, no need for a propositional doxastic attitude about a mental state of the expresser, unlike what doxasticism and moderate affectivism require. Radical affectivism thus naturally builds a bridge between expressives and pre-expressives: in both cases, the core ingredient for expressive communication is affective resonance.

Indeed, an interesting consequence of radical affectivism is that it leads to the view that there can be entirely successful expressive communications where there is no need for doxastic states at all: not in the speaker's mind – as long as the affect expressed is not based on a doxastic state but, e.g., on a perception – nor in the audience's mind – as long as the affect is understood through an affective resonance which is itself non-doxastic, e.g. through emotional contagion. This is a strong advantage if one believes that all kinds of expressives (whether or not their meaning belong to speaker-meaning as this notion was defined in the preceding chapters) should be explained through at least some of the same mechanisms and that there are creatures who can communicate through expressives although they lack the conceptual ability to entertain the propositional thoughts required for the relevant doxastic states.[138]

6.5. Conclusion

I have presented three theories of what it takes for an audience to understand expressives, following the Fregean insight that a theory of meaning should be informed by a theory of understanding. The first account is doxasticism and claims that understanding an expressive requires understanding propositions believed (supposed, doubted, …) by its utterer. The second theory is moderate affectivism and claims that the audience must attribute to the utterer the specificities of affects. In light of the discussion in the previous chapter, I have highlighted how moderate affectivism diverges from doxasticism, insisting that attributing an affect to someone cannot be reduced to attributing a doxastic attitude. The third theory is radical affectivism and claims that the audience must not only attribute to the utterer an affect, but that the audience must resonate

[138] For the claim that non-human animals may well lack the relevant semantic and syntactic abilities to entertain the propositional representations that doxastic states are often thought require, see Davidson (1982) and, outside philosophy, see Hauser, Chomsky, and Fitch (2002), Jackendoff and Pinker (2005), and especially Reboul (2017, Chapter 3).

affectively with the utterer, either through mirror-like affective responses (empathy, affective contagion, simulating an affect through imagination, reviving an affective memory) or through complementary affective responses (examples given included: anger–guilt, regret–forgiveness, excitement–surprise).

I have offered reasons to doubt the viability of doxasticism because it failed to account for the key features that distinguish expressives from descriptives – (a) hot vs cold, (b) appropriate vs true, and (c) direct vs indirect display. These reasons compound with those discussed in the previous chapter – regarding the nature of affects and non-natural meaning – against doxasticism. I have concluded that one can maintain doxasticism if one rejects the idea that expressives express affects. We have seen that moderate affectivism successfully meets the three benchmarks, but I have argued that radical affectivism does so even more successfully.

In light of these conclusions, it is unfortunate that most researchers studying expressives today are closer to doxasticism than affectivism. The main reason for this state of affairs, I surmise, is that they try to cash out the meaning of expressives with the traditional tools of semantics and pragmatics. This is problematic because these traditional tools force an analysis of meaning as propositions believed by the utterer (for instance Stalnaker, 2002 define 'common ground' through beliefs). Nevertheless, there also are advances in the development of semantic and pragmatic tools which would allow an analysis of expressives to occur within an affectivist framework, see Potts (2007), Garcia-Carpintero (2017), or Marques and Garcia-Carpintero (2020) for semantic analysis of expressives consistent with affectivism (at least with its moderate version).

I hope that these considerations, strongly inspired by developments in the philosophy of emotion from the past decades, will motivate the readers to take affectivism seriously – whether it is its moderate or radical version, with a preference for the latter – and, most importantly, to see how the philosophy of emotion can inform and constrain the philosophy of language.

7. AFFECTIVE MEANING, NATURAL MEANING, AND PROBABILISTIC MEANING

« … I imagined my god confiding his message to the living skin of the jaguars
… May the mystery lettered on the tigers die with me. »
– Jorge Luis Borges, *The God's Script*

Abstract. In this chapter, I discuss two main ways in which philosophers have argued that information can be naturally encoded in stimuli: *natural meaning* and *probabilistic meaning*. I evaluate how useful they are when it comes to analyzing affective meaning. I argue that natural meaning, because it is a factive relation, is too strict for this purpose. The notion of probabilistic meaning seems adequate for analyzing non-communicative signs, but it faces several difficulties when it comes to analyzing *signals,* i.e. communicative signs. In the next chapter, we will see how notions from teleosemantics may overcome these difficulties.

7.1. INTRODUCTION

When studying affective meaning, we are very often confronted with cases that are produced not only without communicative intentions but without any control over the effects they have on the audience. This is true if we focus on non-communicative signs as well as on cases of affective communication – cases where the affective meaning has the function to be transmitted.

Think about betraying one's emotion, despite one's will, through a facial expression or a scream. Think about nonhuman animals communicating their affects through alarm calls, food calls, or hormonal releases, where the senders cannot control the effects on their peers. Even though intentions to communicate are absent from these cases, they belong to the realm of affective communication insofar as both the sending (e.g. a scream) and the receiving of the information (hear a scream and understanding that the screamer is afraid) are designed to transmit the information that is so transmitted. In these cases, there is no intentional design, but supposedly the design was done through natural selection.

Outside the realm of communication, we can think about the physiological manifestations used to detect one's emotion which one cannot help but produce, the kinds of manifestations that could be used in lie detectors: sweating, pupil dilatation, heart rate, or brain activity. Such cases aren't

communicative because information transfer is not the purpose of the manifestation: the manifestation (e.g. heart rate) is not designed for transferring the information that is so transferred.

All of these are examples of what I call *affective non-Gricean meaning*. They are examples of *meaning*, because we may say, for instance, that Sam's blushing *means* that he is embarrassed: it carries this piece of information. The meaning is *affective* because they are examples where the information carried is about someone's affective state. And I classify them as a *non-Gricean* type of meaning because what is distinctive about Gricean models of information transmission does not apply to them; in particular, such cases don't involve implicatures of any kind.

They don't because the signs are produced in such a way that pragmatic principles such as Grice's Cooperative Principle don't apply to them. Such principles only apply to cases where the information is transmitted either in an intentional way or, at least, with a certain degree of control (see Chapters 1 and 2). When the effects produced on the audience, including what information is transmitted, are not controllable, we cannot infer that certain implicatures are carried by the signs, because there are no goal-oriented reasons why the creature has produced these signs (or: omitted to refrain from producing them).

As we have also seen in Chapter 1, when there are no implicatures of any kind, it should be possible to account for information transmission through some code model. So, the examples of non-Gricean affective meaning given above should be analyzable through a code, an established pairing between stimuli and the information they transmit.

A main theoretical construct put forward to analyze non-Gricean meaning is that of *natural information* (Drestke, 1981, 1986; for a review, see Stegmann, 2015), itself based on Grice's notion of natural meaning (1957). At the heart of this notion is the idea that lawful relations hold between states or events in the world – e.g. between fire and smoke – and that this is what explains that information is encoded in these states or events. The code, the established pairing, is that between a stimulus and that with which it is lawfully related. If an observer can somehow use this established pairing, it can allow her to decode what information is encoded in the stimulus. We can learn from lawfully related entities, and they thus have a meaning.

In this chapter, I will present several ways to exploit the idea that lawful relations explain some cases of information transmission and see how they apply to cases of non-Gricean affective meaning. First, I will present Grice's

(1957) notion of *natural meaning* and explain why it fails to apply to most cases of affective meaning.[139] I will then present how Dretske (1981, 1986) elaborated upon Grice's notion, but explain why, once again, it is hard to see how it applies to affective meaning. Finally, I will present more recent developments of the notion of probabilistic meaning (Millikan, 2004; Scarantino, 2015; Scarantino & Piccinini, 2010; Shea, 2007; Skyrms, 2010; Stegmann, 2015).[140] We will see that probabilistic theories seem to apply well to the affective meaning of non-communicative signs, but face three problems with respect to affective signals, signs whose function is to carry the affective information they carry.

7.2. Affective natural meaning

7.2.1. Introducing natural meaning

I understand natural meaning as *factive indication,* which is how Grice (1957) introduced this notion. This means that if x means naturally that p (in context C), then whenever x is the case, it is also the case that p (in C). Only if this entailment holds may we say that x naturally means p or that x is a natural sign of p. Here is a classic example: if the number of growth rings of a tree naturally means how many winters the tree has lived through (in a context in C_W), then whenever a tree has 28 growth rings, this tree must have lived through 28 winters (in C_W). For this strong constraint to be realistic, we may restrict C_W to a context where winters are the coldest season of the year and where trees grow as they do on Earth. Now, if there is only one tree in a million that has 28 rings in C_W but has lived only 27 winters, then growth rings are *not* a natural sign of a tree's age in C_W. They don't possess this natural meaning, at least in this context. Because only one exception to the rule is enough to rule out natural meaning, it has a very strict necessary condition.

7.2.2. Natural meaning and affects

Because of this strictness, it is hard to find rock-solid candidates for affective natural meaning. This is so although philosophers often claim that facial expressions of emotions may have natural meanings, starting with Grice himself, who talked about frowning as a natural sign of displeasure (1957). However, *pace* Grice, I don't think that there are good

[139] Natural' here is not to be contrasted with 'super-natural'. I take it to have the same meaning as in 'natural sciences', which we oppose to e.g. social sciences or humanities.
[140] Most of these views, just like Dretske's, are based on (Shannon, 1948); for a relevant case outside philosophy see (Hauser, 1996). See also Denkel (1992) for an early move in this direction.

candidates for affective natural meaning to be found in the domain of facial expression (see also the discussion in Chapter 3).

For one thing, facial expressions may be faked, and so one may display a typical 'angry face', but not be angry. Secondly, the muscles of our face may, in certain situations, inadvertently form a pattern that is normally associated with a certain emotion, even though we do not undergo this emotion. Take the following picture as an example:

Fig. 7.1. Giannis Antetokounmpo displaying the facial expression associated with fear according to Ekman and Friesen's code model (Ekman & Friesen, 1971),

Here Giannis Antetokounmpo is displaying the facial expression associated with fear (Ekman & Friesen, 1971), but, of course, he is not afraid of slamming the basket.

Here is another example, borrowed from Barrett, Mesquita, and Gendron (2011), where Serena Williams displays an 'Ekmanian angry face' even though she certainly is not angry, as she has just scored an important point.

Fig. 7.2. Serena Williams displaying a facial expression associated with anger according to Ekman and Friesen's code model (from Barrett et al., 2011)

Similarly, the Duchenne smile or 'real smile' has been thought to naturally mean a positive affective state, because people usually cannot control the relevant activation of the zygomatic muscles. But it actually doesn't. First, the Duchenne smile – i.e. activation of lip corner puller (AU 12) together with a contraction of the zygomatic muscles or 'cheek raiser' (AU 6) – may be faked by a substantial portion of the population (Gunnery et al., 2013). Secondly, people also display Duchenne smiles when frustrated (Hoque & Picard, 2011). Thirdly, we may inadvertently activate the muscle configuration of the Duchenne smile, for instance when we are screaming certain words (personal communication from psycho-physiologist Sylvain Deplanque).

So if facial expressions are not natural signs of affects, where should we find examples? Well, an obvious place to look is in the affective neurology literature. Here is an example: it appears that whenever humans' amygdala is activated above a certain level, they have (unconsciously) detected a stimulus that is emotionally relevant (Murray et al., 2014). At least this appears to be so given a context C_A where the people in question don't suffer from emotional or neurophysiological malfunctions, where their amygdala is not artificially manipulated through chemical or electric stimuli, where their brains work as that of typical humans, and other 'normal' contextual conditions. So, it seems, this amygdala activation means naturally that something is appraised as affectively relevant by the subject in C_A, and so that the person is in an affective state (for the link

between appraisals and affects, see Chapter 9). Many other correlations may be proposed between neurological activations and affective states, even if most of them are debated in the field (see Nummenmaa & Saarimäki, 2019 for a review). However, neurologists are generally interested in *statistically significant correlations*, not in the strict natural meaning relation defined by Grice, since the latter requires that 100 percent of cases are correlated, something quite foreign to neurological reality. Even the amygdala results, which appear to be pretty robust, do not give enough evidence for such a strong demand: we don't know whether there is a one-one correlation between amygdala activation and affective states, i.e. that there is a probability of 1/1 that the one indicates the other, even if we restrict the context to C_A.

Is there an example where such results would be established? Where a brain signature naturally means an affective state? Perhaps, but since we are especially concerned with communication, and since we don't communicate by using brain signatures, the affective natural meanings that we may potentially find in neurology are not central to our inquiry. Maybe someday we will be walking around with portable fMRI and EEG scanners connected to some screen which will allow us (or force us) to communicate our emotions by displaying the activity of our amygdala. Since we are not there yet, I will now put aside neurological affective meaning.

Other candidates for affective natural meaning are the cues discussed in the psychophysiology literature. Such cues may be used in affective communication since at least some physiological signs are observable with the naked eye. What are the psycho-physiological candidates? When it comes to affects, signs used by the psycho-physiologists concern in particular affective *arousal* and *valence*. In a context C_B where we know that the person has not been subjected to physical efforts (e.g. has not run in the last hour), has not been artificially manipulated through electric stimulation of their brain or chemical injection, and where we put aside extraordinary scenarios (e.g. the person is a very realistic android), then the following set of physiological changes are credible candidates for being physiological states which mean naturally that the subject is undergoing emotional arousal: blushing, sweat increase, pupil dilatation, heart rate modification, increase in respiratory rhythm, decrease in the deepness of breath (panting), and exacerbated muscular tonus. We may even consider each of these stimuli taken individually to naturally mean arousal if we restrict the context C_B further (e.g. if we take pupil dilatation as a sign, we need to restrict C_B to a context where there is no change in light intensity). And indeed, each of these signs is used by affective scientists to measure

arousal (Bradley et al., 2008; Cacioppo et al., 2017, Chapter 20). Concerning valence, some evidence shows that a decrease in heart rate is correlated with negatively valenced affective states (Cacioppo et al., 2017, Chapter 20).

When it comes to candidates of physiological cues observable with the naked eye which can be taken to possess an affective natural meaning, that's it. Psycho-physiologists do not (or at least should not) infer more than negative valence and increase in arousal from physiological cues.[141]

And, in fact, these candidates may not be cases of natural meaning. Some psycho-physiologists dispute the claim that these changes (pupil dilatation, increased sweating, deceleration of heart rate, etc.) naturally mean an increased arousal and a negative valence (Bradley & Lang, 2007; Turpin, 1986). This is because these changes seem to also be correlated with *information-seeking behavior*, where the person is alert but is not reacting strongly enough to genuinely be affectively aroused by the situation or be undergoing a negatively valenced affect (Bradley & Lang, 2007; Turpin, 1986). Imagine for instance that you hear an unfamiliar noise in your apartment and try to listen and figure out what it is, but you are not afraid or undergoing any other arousing or negatively valenced affect. The signs mentioned above may well be present in this situation.

So, after due consideration, it seems that there are no rock-solid candidates for natural affective meaning. There are two main solutions to this problem. The first one is to sufficiently restrict the set of contexts so as to reach factive indication, which is the path followed by Dretske as we will shortly see, and the second one is to relax the notion of natural meaning to allow for non-factive but probabilistic indication, a path which we will explore in §7.3.

7.2.3. Contextualizing natural meaning

Dretske (1981, 2008) proposes to focus on contexts where we do have a one-to-one correlation between the sign and what it means. This allows us to secure a perfect epistemic relation and have cases where we can *learn* from signs, where we can acquire knowledge from observing them. According to Dretske, a sign carries the *natural information* that p just in case the probability of p given the occurrence of the sign is one, and less than one

[141] Personal communication with Sylvain Deplanque (April 2020). Note that psycho-physiologists may infer more from signs that are not observable with the naked eye. Note also that some evidence shows that when we look at pleasant pictures – but not when we listen to pleasant sounds – an increase in heart rate is correlated with a positively valenced affective state (Cacioppo et al., 2017, Chapter 20).

otherwise. So, we can say that blushing naturally means that someone is aroused if we concentrate only on those contexts, given background knowledge, where blushing is indeed a factive indicator of higher arousal and not an indicator of physical efforts or high body temperature.

This theory of natural information may well be sound, but it is not an appealing solution for someone concerned with communication: how are communicators supposed to know which contexts are such that there is a 100 percent chance that a sign means an affective state? There does not appear to be an answer to this question. For this reason, many commentators consider Dretske's natural information, just like Grice's natural meaning, to be too restrictive to be useful in analyzing communication or even knowledge acquisition (Godfrey-Smith, 1992; Millikan, 2000; Scarantino & Piccinini, 2010; Stegmann, 2015; Suppes, 1983).

Another solution that is more implementable and certainly more popular nowadays is to move away from factive indication and define a new theoretical construct that is less strict and allows x to mean that p even if p is not the case. We will now turn to such a notion – that of probabilistic meaning. For the record, let me observe that Dretske did move to another notion as well, a teleosemantic one called 'functional meaning'. We will turn to it in the next chapter.

7.3. AFFECTIVE PROBABILISTIC MEANING

7.3.1. INTRODUCING PROBABILISTIC MEANING

To allow for cases where a sign x can mean that p even when p is not the case, but without appealing to intentions and mindreading, several authors have parted from Grice's factive notion to define what we may call 'probabilistic meaning' (Millikan, 2004; Scarantino, 2015b; Scarantino & Piccinini, 2010; Shea, 2007; Skyrms, 2010; Stegmann, 2015). What I call probabilistic meaning includes several theoretical constructs that are defined in different manners, but which all make use of *less-than-1 correlations* between the sign and what the sign is a sign of, i.e. statistical correlations that are not factive.[142] There are two main ways to define probabilistic meaning: using objective or subjective probabilities. I will

[142] To make a clear distinction between natural meaning and probabilistic meaning, I will only talk of natural meaning and not of probabilistic meaning when the relation between the sign and what it indicates is factive, although, according to the definitions I will give, natural meaning is a species of probabilistic meaning, since factivity is a correlation where the probability is 1.

present a preeminent example of each type, that of Shea (2007) and that of Scarantino (2015b), which should be representative of the general notion.

Here is Shea's definition of 'correlational information', a type of probabilistic meaning defined in terms of 'chance', or objective probability, i.e. an ontologically mind-independent kind of probability:

> « R carries the correlational information that condition C obtains iff for a common natural reason within some spatio-temporal domain D: chance (C | R is tokened) > chance (C | R is not tokened). » (Shea, 2007)

The formula 'chance (C | R is tokened) > chance (C | R is not tokened)' reads as follows: the chance of C being the case given that R is tokened is greater than the chance of C being the case given that R is not tokened. The restriction to 'common natural reason within some spatio-temporal domain' is meant to eliminate accidental correlations, i.e. eliminate 'spurious correlations' which are neither directly nor indirectly causally related.[143]

If we put aside spurious correlations, the idea is that if some sign x (e.g. dark clouds approaching) raises the objective probability that a state of affairs p obtains (it will rain), then x gives a probabilistic indication that p. Contrary to natural meaning, we don't have a relation of factive indication; if information is understood as what reduces uncertainty (Adriaans, 2019), then some information is indeed carried by the sign. Consequently, organisms may make beneficial use of such probabilistic indication.

Another kind of probabilistic meaning, called *Incremental Natural Information (INI)*, is defined by Scarantino, who uses Bayesian, ontologically mind-dependent probabilities.[144] In the following quote, 'r' stands for a sign, i.e. an entity which can carry information, 's' for the source of information in the world, i.e. that which we can learn about, and 'G' and 'F' for properties or states:

> « Incremental Natural Information (INI): r's being G carries incremental natural information about s's being F, relative to background data d, if and only if p(s is F | r is G & d) ≠ p(s is F | d). » (Scarantino, 2015: 423; see also Scarantino & Piccinini, 2010)

[143] I will discuss spurious correlations further below. For nice examples, visit https://www.tylervigen.com/spurious-correlations

[144] I put aside the complementary notion of 'Degree of Overall Support', which is defined as so: "the degree of overall support provided by a signal r's being G about s's being F, relative to background data d, is equal to p(s is F | r is G & d)." (Scarantino, 2015: 423).

In English, 'p(s is F | r is G & d) ≠ p(s is F | d)' reads as 'the probability that s is F, given that r is G and given background data d, is not equal to the probability that s is F, given (only) background data d'.

Scarantino's definition of INI is a way to formalize the Bayesian/Shannonian idea that the information carried by a cue corresponds to the number of hypotheses one can eliminate once one has become appropriately acquainted with the cue, given one's initial hypotheses or background data. In other words, perceiving the cue, or otherwise becoming appropriately acquainted with it (e.g. through memory), changes the subjective probability that s if F, and so it transforms one's prior *credences* to one's *posterior credences*.

Simplifying somewhat, we may say that if s probabilistically means that p, then knowing s modifies one's credence about p, given some background data. For instance, if black clouds on the other side of the lake mean that it is raining in Evian, then knowing that there indeed are black clouds on the other side of the lake modifies one's credence about whether it is raining in Evian (given relevant background data, such as that when there are black clouds over someplace, it is usually raining there).

An advantage of probabilistic meaning over natural meaning is that it seems to account for the following cases (examples adapted from Scarantino 2015: 431):

(1) The eagle alarm call of the vervet monkey means that an eagle is present (to the other vervet monkeys).
(2) John's frown means that he is angry (to John's wife).
(3) The burning sensation in Mary's fingers means that her hand is in contact with something very hot (to Mary).
(4) The sound of the bell means that an electric shock is forthcoming (to the fear-conditioned rat).
(5) The shape of the tracks in the snow means that a deer has walked by (to a zoologist).
(6) The structure of the bee dance means that the nectar is 100 feet to the right (to other bees in the group).

The word 'means' does not denote a factive relation in any of these examples because, for instance, a monkey may produce an eagle alarm call when it is another species of bird or even a drone that is flying around. Similarly, Mary may have a burning sensation on contact with something extremely cold – those who have had a wart removed through cryotherapy will know that being frozen with liquid nitrogen feels like one's skin is burning.

By contrast, in (1)–(6), Shea's or Scarantino's definitions can be applied and we may thus hypothesize that 'means' refer to 'probabilistic meaning' in each of these cases. Whether we understand 'probability' objectively (as Shea) or subjectively (as Scarantino), we may say that, relative to the relevant domain for Shea or to the relevant background data for Scarantino, a vervet monkey's eagle alarm calls raise the probability of an eagle being nearby, one's frowning raise the probability of one's anger, one's burning sensations raise the probability of one being in contact with something hot, and so on (Scarantino, 2015: 431).

It is easy to see how probabilistic meaning may include affective meaning: John's frown was already a case in point. Even if John also frowns when he plays chess, when he sneezes, or when he faces a very bright light, if tokening his frown makes it more probable that he is angry, we may say that the frown probabilistically means that he is angry.

A nice application of this concept is the following: the famous Ekmanian facial expressions (Fig. 7.3) may probabilistically mean that the person undergoes the corresponding emotion, although actors can fake them and although we may inadvertently put on these faces while not undergoing the corresponding emotions (see the Antetokounmpo and Williams' examples above). In fact, it seems that such probabilistic relations are what Ekman is going after in his conception of facial expression of emotions, which allows him to respond to critics who point out that there are no one-to-one correlations between Ekmanian facial expressions and the emotions they express (Ekman & Cordaro, 2011). However, we will soon see reasons to doubt that this application of the concept works as well as it may seem. Furthermore, we will see in the next chapter that a teleosemantic notion deals better with these cases. Nevertheless, probabilistic meaning certainly does better than natural meaning.

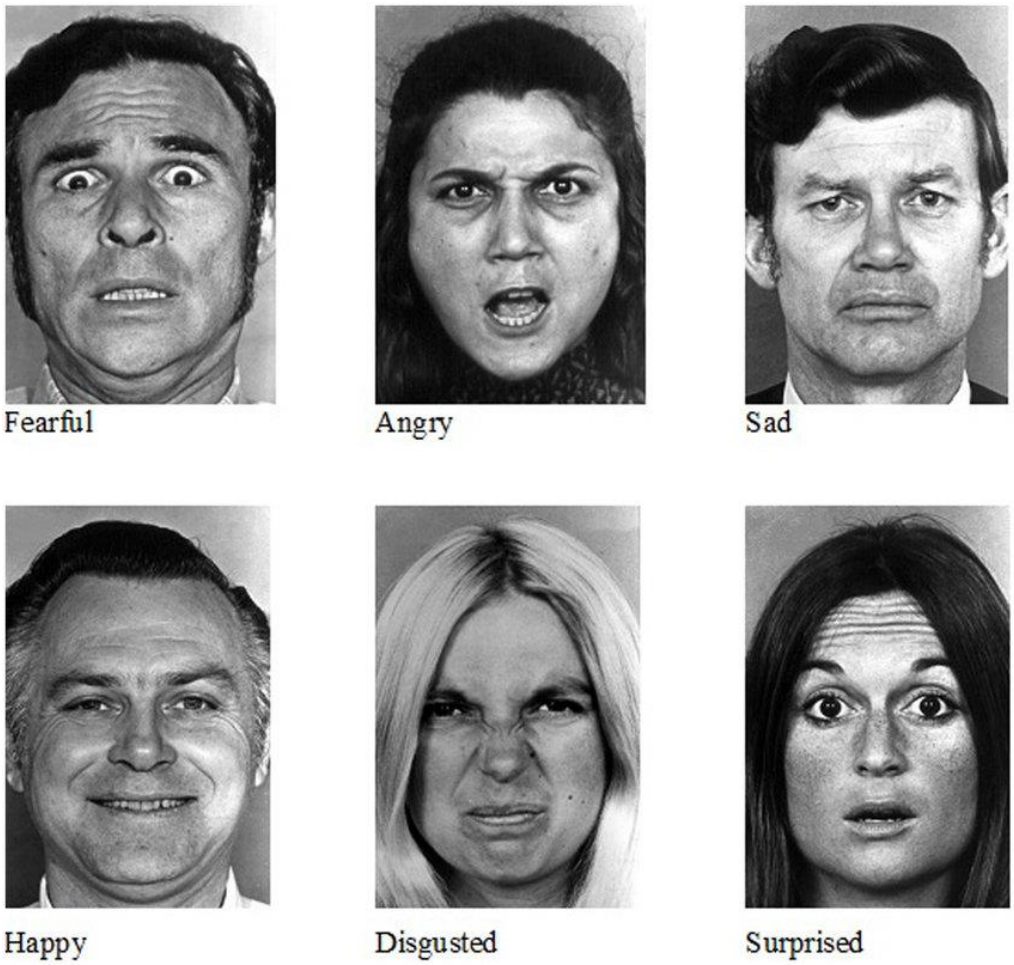

Fig. 7.3. Ekmanian facial expressions, from (AW et al., 2002)

As is quite obvious from the examples given by Scarantino, probabilistic meaning, just like natural meaning, may be instantiated by signs that are not signals, i.e. by information vehicles that do not have any communicative functions. For instance, the shape of the tracks in the snow left by the deer is a sign but not a signal.[145]

[145] By the way, this is something that teleosemantic accounts – which we will discuss in the next chapter – cannot or cannot easily deal with, because, on the one hand, it is not the function of tracks in the snow to serve as information vehicles and, on the other hand, it is dubious that the mechanisms leading to the zoologist's belief that it is a deer track have the biological function of making her infer this piece of information. It is at least plausible that the mechanisms in question were rather selected for other, more general, purposes, but were used by the zoologist in a way that cannot be predicted by her evolutionary history – although Millikan (1984) or Papineau (1984) would argue otherwise since they defend the thesis that our token beliefs have the biological function of having true content. This commits them to a special, controversial view of evolution; they belong to what Sterelny (1991, sec. 6.7) and Godfrey-Smith (1996, p. 174) after him call 'immodest' teleological accounts. In any case, the explanation given by probabilistic meaning for the meaning of non-communicative signs seems much more economical and powerful than those which could be proposed by teleosemantic accounts.

Probabilistic meaning seems perfectly fitted to cash out the affective meaning of the non-communicative signs. Among those are the physiological signs used by affective scientists to measure affects and which we discussed in the preceding section (§7.2.1). Take as an example the following signs: increased sweating, pupil dilatation, heart rate, breathing rate, breathing deepness, and the context C_B discussed above (e.g. the person has not done important physical efforts in the last hour). This combination of signs probabilistically means that the person is affectively aroused. And we can indeed quantify such probabilistic meaning using Bayesian tools such as Scarantino's formula. In fact, some of the best physiological studies on affects use similar Bayesian statistics (Cacioppo et al., 2017, Chapter 27).

However, as we will shortly see, probabilistic meaning seems inadequate to deal with the meaning of affective *signals*, which is the kind of meaning that interests us most in this dissertation, since we are primarily concerned with how emotions are communicated.

Before we move on, and now that we know what probabilistic meaning is, let me point to the fact that speaker-meaning cannot be accounted for by probabilistic meaning because the latter, even though it is less strict than Grice's natural meaning or Dretske's natural information, can only account for *encoded* meaning, i.e. for the information that is associated with stimuli through a pre-established pairing. To see the difference, take the following anecdote: a car was sitting half on the road, half on the pavement to let a passenger out. A bus was about to pass it, but the bus driver noticed that the car driver had finished dropping the passenger and so she stopped the bus to let the car go first. The car driver hesitated, noticed the bus driver's hesitation, went ahead, and turned both its blinkers on a couple of times. Although I had never seen such use of the blinkers, I immediately understood that the car driver said 'Thanks' to the bus driver. From a purely probabilistic point of view, it is inexplicable that I understood what the car driver meant. The probabilistic association between the blinkers and a 'thank you' was, from my point of view, exactly equal to 0. Nevertheless, because I assumed that the car driver was respecting some pragmatic principles and was being relevant with its use of the blinkers, I could immediately infer what the driver speaker-meant. This example, I believe, illustrates why probabilistic meaning is insufficient for understanding speaker-meaning and, more generally, Gricean meaning (see Chapters 1 and 2).

The question for us now is: can it account for non-Gricean meaning? We saw that it does well with non-Gricean meaning of non-communicative signs, but we will now see that it cannot do so well with that of all signals.

7.3.2. First difficulty: All that a call means

To evaluate how probabilistic meaning can deal with signals, we will focus on an example given by Scarantino (2015) and assume, as he does, that the vervet monkey alarm call for eagles ('eagle call' for short) means the following:

(i) An eagle is present.

As mentioned, given that an eagle call is tokened, there is more probability, objective or subjective, that (i); so the eagle call does probabilistically mean that (i). Even if in some cases a monkey utters an eagle-call in the absence of the predator, meaning (i) is not lost because the probabilistic correlation remains. This is a big advantage that probabilistic meaning has over natural meaning.

However, a first difficulty that one can see with probabilistic meaning is that there are many other states of affairs whose probability is greater given that the eagle call is tokened, so it seems to probabilistically mean lots of things which we may not want it to mean.

To establish what else it probabilistically means, let us first take Shea's definition. Here is a short list of some of the states of affairs which, I take it, have a greater (non-accidental, objective) chance of being the case given the fact that the eagle call has been tokened compared to a scenario where the eagle call has not been tokened:

(ii) The user of this call does not believe that Pythagoras' theorem is true.
(iii) The sky is green.
(iv) The theory of evolution is true.
(v) The speed of sound in air is about 343 meters per second.

(ii) is made more probable because it is much more probable that (ii) is the case if the signal sender does not speak a human language, and the fact that the vervet monkey has emitted an eagle alarm call raises the probability that the monkey does not speak a human language. So if the eagle alarm call is tokened, it raises the probability that (ii) is the case.

(iii) is made more probable because a tokening of the call raises the probability that a monkey has seen an eagle, and so raises the probability

that the sky is any color other than brown since there is a greater probability that an eagle is seen by a monkey if the sky is any color other than brown. The fact that (iii) is false is irrelevant since probabilistic meaning is non-factive.

To see why (iv) is made more probable if the eagle call is tokened, imagine a world where vervet monkeys act in ways that are in contradiction with evolutionary principles. In this world, there would be less chance that the theory of evolution is true than in our world. So, it seems, the fact that this vervet monkey acts following evolutionary principles makes it more probable that the theory of evolution is true. Of course, we may say the same for every biological fact that corroborates evolutionary principles (for a review see Coyne, 2010). They all probabilistically mean that evolution is true.

Similarly, (v) is made more probable given that the eagle call is tokened because, I take it, the probability that the speed of sound in air is 343 m/s is made greater every time a sound is tokened that travels in air at this speed.

Of course, we need not stop at (v): we may extend indefinitely the list of what the eagle call means consistently with Shea's definition.

The difficulty with this consequence is that, intuitively, we don't want to say that the meaning of the eagle call is that an eagle is present *and that the sky is green*, but this is what is implied by Shea's definition.

Roughly, the same remarks can be made concerning Scarantino's definition except that we should always relativize to whom the eagle call means (i)–(v). For him, the call means (i)–(v) for those who have the background knowledge which allows them to modify the strength of their credences about (i)–(v). For instance, the call means (iv) to evolutionary biologists (not to monkeys), (v) to sound physicists, and so on.

Another difference between Shea (or Millikan 2004) and Scarantino is that the latter does not require that there is any causal link between the call and what it means. Since he accepts that any kind of accidental correlation may be taken into account, there certainly is much absurd information that the call is carrying given certain background data. Indeed, absurd correlations are legion, although I have no specific example to give when it comes to vervet monkey eagle calls. I do have an example with Nicolas Cage though, illustrated by the following graph:

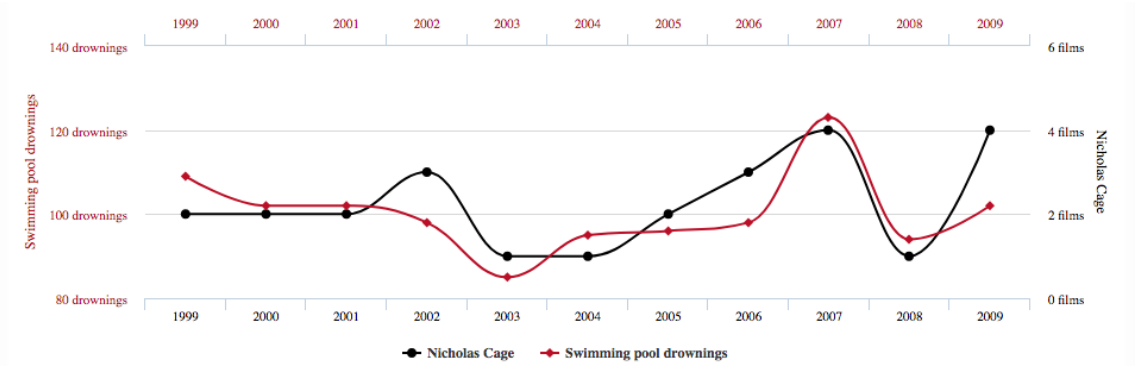

Fig. 7.4. A spurious, accidental correlation: Number of people who drowned by falling into a pool (red) correlates with films Nicolas Cage appeared in (black). Correlation: r=0.666004.[146]

Now that you have seen this graph and learned about the 0.666 correlation between the number of films Cage appears in and the number of people who drown by falling into a pool, you possess sufficient background knowledge for the following to hold according to Scarantino's definition of probabilistic meaning: to you, if 100 people have drowned in a pool in 2020, it means that Nicolas Cage featured in two movies in 2020.

Such results may seem risible, but Shea and Scarantino can bite the bullet. After all, it is not unreasonable to make inferences based on statistical correlations. Of course, sometimes such correlations are trivial (e.g. (v)) and sometimes they are spurious (Cage's example), but, overall, correlations *are* informative. So, even though we find examples where the common use of the word 'meaning' does not apply, defenders of the probabilistic account may say that their purpose was not to find a definition that reflects the common use of this word (contrary to what Grice was doing) but to define a useful theoretical construct, a technical notion, and so it is fine that it is odd to use the word 'meaning' in certain cases.

I actually agree with this response and don't think that (ii)-(v) or spurious correlations should lead us to leave probabilistic meaning behind. Nevertheless, I find it reasonable to consider the present difficulty to be a genuine objection to using a definition of probabilistic meaning to analyze the meaning of eagle alarm calls, especially if there is a better technical concept around. The fact is that teleosemantic notions which I will introduce in the next chapter do seem to do the job better. In a nutshell, the main disadvantage of probabilistic meaning when it comes to analyzing signals, as opposed to non-communicative signs, is that it does not take

[146] Retrieved from https://www.tylervigen.com/spurious-correlations on April 13, 2020.

into account the *function* that signals have, contrary to teleosemantic notions.

Let us move on to further difficulties for a probabilistic analysis of the meaning of affective signals (let us not forget that these problems don't arise for non-communicative signs).

7.3.2. Second difficulty: The rare eagle scenario

Let us now turn to a second difficulty. Certain emotion episodes, such as fear and disgust episodes, very often happen to be 'false alarms', in the sense that we are often afraid of something that is not dangerous or disgusted by something that is not contaminated. From an evolutionary perspective, at least for certain species and certain stimuli, it makes sense that animals react with fear and disgust more often for non-dangerous and non-contaminated stimuli compared to dangerous and contaminated ones, because the cost of 'false alarm', i.e. of undergoing these emotions when there is nothing to be afraid of or to be disgusted by, is significantly less than the cost of 'misses', i.e. of not undergoing these emotions even though there is something to be afraid of or to be disgusted by (Breznitz, 1984; Nesse, 2019, Chapter 5). So a large number of false alarms relative to the number of genuine alarms make sense for these emotions, especially in vulnerable animals, since it is better to be safe than sorry. The same can be said about other negative emotions like shame, guilt, perhaps anger, and, plausibly, about positive emotions such as hope, admiration, or hilarity. [147]

Now, take the following imaginary but realistic scenario. Let us assume once again that the meaning of the vervet monkey eagle call is adequately captured by (i). Now, imagine a group of vervet monkeys that happens to live in an area where most eagles have disappeared. Maybe eagles have recently been dispersed because of safari hunts. Assume furthermore that vervet monkeys are quite faint-hearted creatures, whose fear episodes often are false alarms (which is not at all unrealistic from what I gather). Although there rarely are any eagles present, they often are afraid that

[147] Thus far, this argument is similar to the 'better safe than sorry' argument as it can be found in Millikan (1989) and especially in Godfrey-Smith (1991), i.e. the idea that the meaning of signals cannot be accounted for by Grice's natural meaning or Dretske's natural information because certain signals, especially those signaling danger, can successfully convey what they mean despite often being false alarms. However, my argument diverges from those insofar as Scarantino's notion can deal with the false alarm cases presented by Millikan and Godfrey-Smith. As the example goes thus far, Scarantino can argue that the background data of the monkeys (their priors) allows them to raise their credence that an eagle is present although, objectively, there are more false alarms than correct alarm calls.

there is one, for instance when another type of big bird flies around. This happens a lot because, now that eagles have nearly disappeared, the species of birds which used to compete with eagles are thriving in this area. As a result, the monkeys use their eagle call more often when there are no eagles around than they do when there are. Nevertheless, the monkeys still seem to understand the call as an eagle call as they react with exactly the same behavior as before: they gaze up and anxiously get down from the canopy as quickly as possible, to find cover, and wait there terrified. This is so even for monkeys that were born during this 'rare eagles' era and so have never experienced a high correlation between calls and eagle attacks.[148] For these youngsters, what the call communicates seems not to have changed at all: everyone reacts as though an eagle was present every time.

The probabilistic account of meaning, however, cannot explain how it is that the meaning of this call still is eagle-related. If we take Shea's definition, the probabilistic meaning of the call is the *opposite* of what the call intuitively means: since there is a greater probability that there are no eagles when the monkeys make this call, this implies that the eagle call probabilistically means that there are no eagles around. And if we take Scarantino's definition, the call also means that there are no eagles around for the young monkeys whose background data (priors) is only constituted by a negative correlation between eagles and eagle calls

It is hard to swallow that what the call means is that there are no eagles around. In a sense, sure, we can infer from the call that there probably are no eagles around. But the monkeys do not make this inference, since they react to the call exactly as they did when eagles were frequent predators. They still perceive the call as an eagle alarm call. There is a clear sense in which the meaning of the monkey calls has remained the same despite the quantitative drop in the probability of eagle attacks.

A defender of probabilistic meaning can tell us: okay, fair enough, we cannot account for the fact that the call means (i) 'An eagle is present' anymore. But what the eagle call really means is rather the following:

(vi) I am so afraid (of an eagle)!!!

And/or:

(vii) Get down from the canopy!

[148] This is what is different from the classical 'better safe than sorry' argument, see preceding footnote.

The probability that the monkey uttering the call is afraid (of an eagle) and the probability that the caller desires that the other monkeys get down from the canopy because of the call have not changed. And the correlations between these two facts and, respectively, (vi) and (vii) allows us to account for these meanings of the eagle call. So, in the rare eagle scenario, even though the meaning of the call cannot be (i), it can still be (vi) and/or (vii), and that's fine. (I will discuss messages (vi) and (vii) in more detail in §2.4.)

I am not convinced by such a response. Even if it is true that, in the rare eagle scenario, the call is still probabilistically correlated with (vi) and (vii), it seems artificial to claim that the meaning of the call is not (i) anymore. I believe that this answer is guilty of the 'No true Scotsman' fallacy:

- A: 'No Scotsman puts sugar on his porridge.'
- B: 'But my uncle Angus is a Scotsman and he puts sugar on his porridge.'
- A: 'But no *true* Scotsman puts sugar on his porridge.'

Indeed, we could interpret the aforementioned defense of probabilistic meaning as follows:

- A: 'The meaning of the eagle call can be understood through probabilistic correlations.'
- B: 'But in the rare eagle scenario, message (i) cannot be understood through probabilistic correlations.'
- A: 'But the *true* meaning of the monkey call is (vi)–(vii), not (i)'.

This response seems *ad hoc*. I believe that my counterexample really is one and that probabilistic meaning cannot account for the meaning of the monkey call. And we may find counterexamples like the one I have presented for plenty of other cases. For instance, we may imagine that a new plant tricks honey bees into thinking that there is plenty of nectar when there is none, and we have a 'rare nectar' scenario for the bee waggle dance.

Another line of response for the defender of probabilistic meaning is the 'add a belief' strategy. The idea is the following: (i) is just a short cut for '(i*) The user of the call believes that an eagle is present' and even in the rare eagle scenario, there is a correlation between the eagle call and the monkey making the call having this belief. And the explanatory role of the hypothesis that the meaning of the call is (i) can also be played by the hypothesis that the meaning is (i*). For instance, the fact that the other monkeys look up to the sky when they hear the eagle call, and get down

from the canopy is due to their inferring (i) (An eagle is present) from (i*) (The user of the call believes an eagle is present).

I see three problems with this response. First, it is quite clear that organisms such as bees or trees cannot attribute beliefs to one another. However, bees and trees do seem to send signals whose meaning is similar to (i) – messages such as 'there is nectar at a certain location' for bees (Riley et al., 2005) and 'there is a threat' for trees (Gorzelak et al., 2015). Thus, a scenario similar to the 'rare eagles' but adapted for bees or trees – e.g. a 'rare nectar' or a 'rare threat' scenario – could not be dealt with by an appeal to *believes* as is done by the substitution of (i) with (i*). This is problematic for the probabilistic meaning account insofar as a general solution would be preferable to the 'add a belief' solution since the latter can only work for a restricted number of cases – and there are reasons to think it does not even work for the monkey case, as we will now see.

Second problem: it is not clear that vervet monkeys can attribute beliefs to others and infer the relevant conclusions as is required in the inference from (i*) to (i) because they may well lack the cognitive capacities to attribute beliefs and make inferences on this basis. Furthermore, even if they can, such inferences may still be a cognitively challenging task, which may take some time, attention, and energy to achieve, and be performed with a low success rate. Such cognitive difficulties are far from being ideal in situations of emergency such as when a predator attacks. Thus, it would be much more efficient to have a call that means (i) rather than (i*). But then, why would (i*) be naturally selected instead of (i), a message which appears to be more fit to the function of the signal?

A third, related, problem of the add-a-belief response is that the primary function of indicative acts of communication is to inform about the external *world* as opposed to *beliefs* or other mental (or internal) states of the communicator. This has been remarked by critics of Gricean definitions of speaker-meaning. Here is a nice illustration:

> « [T]he primary point of making assertions is not to instill into others beliefs about one's own beliefs, but to inform others – to let them know – about the subject matter of one's assertions (which need not be, though of course it may be, the asserter's beliefs). » (McDowell, 1980: 38, quoted by Green, 2009)

This constraint cannot be respected by the move from (i) to (i*).

In sum, I don't see how the rare eagle scenario can be adequately dealt with probabilistic meaning.

7.3.4. THIRD DIFFICULTY: IMPERATIVE PRE-ILLOCUTIONARY FORCE

Let us now turn our attention to another type of difficulty for the probabilistic account of the monkey call: it is hard to understand how probabilistic meaning can have an imperative force, but human and nonhuman signals seem to be replete with imperative messages – even if we focus only on non-Gricean meaning. To develop this objection I will use the notion of pre-illocutionary force, which I have introduced in Chapter 4.

Remember that, following Frege and the speech act tradition, I distinguish between two components of speaker-meaning: the force and the content (Austin, 1962; Bach & Harnish, 1979; Fogal et al., 2018; Frege, 1956; Lewis, 1979b; Searle, 1969; Strawson, 1964a). I reserve the expression 'illocutionary force' for speaker-meaning (as defined in Chapter 1 and the Appendix). For the meaning of signals which does not belong to speaker-meaning, but where it nevertheless makes sense to distinguish a force and a content component, I use the expression *pre-illocutionary force*.

For instance, we may analyze the meaning of tree stress signals as having different pre-illocutionary functions: that of *informing* other trees vs. that of *making them react* to better protect themselves.[149] Indeed, we know that trees and other plants may receive stress signals and adapt their behavior accordingly, through rapid changes in physiology, gene regulation, and defense response (Gorzelak et al., 2015). We may interpret this phenomenon as showing that trees exchange information which includes either a representation of the world, or a representation of the behavior to adopt, or both, with 'x represents p for the individual S' understood as 'x has the function of providing information about p to S' (Dretske, 1995: 2). Even if biologists do talk of 'tree behavior', we may refrain from saying that a tree can perform a pre-illocutionary *act*, because trees don't perform actions as these are usually defined. Nevertheless, their signals may be analyzed through what originally is a speech act formalization: F(p), where F represents the force and p represents the content. So we may interpret the meaning of tree stress signals as !(other trees protect themselves) or as ⊨(there is a threat) where '!' represents the tree equivalent of an imperative force and '⊨' that of an assertive force.

As said above, a third difficulty that I see with probabilistic meaning is to deal with imperative pre-illocutionary force. I will stick to the vervet

[149] Another plausible example could have been that pain signals possess an imperative force, e.g. the pain of an ingrown nail may be something like 'Protect your nail!' (Martínez, 2011).

monkey eagle call. Remember that we assumed quite reasonably that this call sends the following message:

(i) An eagle is present.

Now, when monkeys see an eagle, they are really scared. Thus, as we briefly discussed in the preceding section, it is not unreasonable to assume that besides the descriptive content of (i), a message sent by the eagle alarm call is something like the following:

(vi) I am so afraid (of an eagle)!!!

Furthermore, given that monkeys evidently address their calls to other members of the group, and given how the other monkeys react to the call, it may well also mean something like the following:

(vii) Get down from the canopy!

In the following, I will assume for the sake of the argument that the meaning of the eagle call involves messages (i), (vi), and (vii). If this turns out to be empirically incorrect, we could change the example and talk about the bebet monkeys, a fictional but realistic species of primate whose eagle call *does* mean (i), (vi), and (vii).[150]

These three kinds of messages have different pre-illocutionary forces (F), and they have different contents (p). (i) has an informative force and its content is that an eagle is present, (vi) has an expressive force and its content is that the monkey is afraid (perhaps: afraid of an eagle), and (vii) has an imperative force and its content is that the other group members get down from the canopy.

We have seen how the notion of probabilistic meaning is supposed to explain that the eagle call means (i). How does it deal with (vi) and (vii)? Well, one strategy, briefly addressed above (§7.2.3), would be to pursue the so-called 'flattening scheme' (García-Carpintero, 2015): to flatten all forces to only one, the indicative force. In our case, this would amount to claiming that the meanings of (vi) and (vii), even though they appear to have a different force from that of (i), are re-describable through, and reducible to, a message with an indicative force, one that can be accounted for in probabilistic terms. So, (vi) may be understood as 'The call user is afraid (of an eagle)' and (vii) as 'Other monkeys will get down from the canopy (as result of hearing this call)'[151] or perhaps as 'The call user desires that other

[150] A similar assumption is made by Millikan (1995, p. 190).
[151] It may be more intuitive to use 'should' or 'ought' instead of 'will' here, but I don't see how we can quantify the probability of 'should' or 'ought' as opposed to that of 'will' and

monkeys get down from the canopy (as a result of this call)'. Vervet monkeys can communicate about the presence of an eagle (indicative force) as well as how they feel about it, and how others will, or are desired to, react to the presence of an eagle, thanks to a probabilistic co-variation between these three states of affairs and the uttering of the call. This allows us to 'flatten' the expressive and imperative forces to the indication of feelings and (future/desired) behaviors.

I am not convinced by this 'flattening scheme' line of response.[152] To see why let us first concentrate on (vii) 'Get down from the canopy!' and its flattening interpretation 'Other monkeys will get down from the canopy (as result of hearing this call)'. Sure: there is a correlation between a monkey making an eagle call and the members who heard the call getting down from the canopy. We can thus establish that the probabilistic meaning of the call involves this information about the future behavior of the other monkeys. However, a message carrying this information has an indicative function: it fulfills its communicative point just in case it is accurate, just in case it represents the world as possessing certain features, as being such that other monkeys will get down from the canopy after hearing the call. But this is not the (pre-)illocutionary function of imperatives. This is the (pre-)illocutionary function of a prediction. Imperatives do not have the function of representing the world as it is or as it will be, their function is to *get others* to do things and to do them as is specified by the content of the imperative.[153]

The defender of probabilistic meaning may want to answer as follows: actually, it is the second flattening option which is the good one, i.e. 'The call user desires that other monkeys get down from the canopy (as a result of this call)'. Imperatives express desires, and desires themselves have an

thus cannot see how probabilistic meaning can account for a 'should' or an 'ought' as opposed of a will, other than, perhaps, through desires. Hence my two 'flattening' interpretations.

[152] In general, I am very skeptical about the flattening scheme (defended by (Davidson, 1979; Lewis, 1979) for reasons related to objections given by García-Carpintero (2015); see also my arguments in the Appendix, §A.5.3, as well as the arguments against a reduction of the expressive force to an indicative force in ch. 5–6.

[153] I prefer avoiding the concept of 'direction of fit' because it may be inherently incoherent (Frost, 2014). However, if it can be used coherently, we may paraphrase my use of 'imperative' and 'indicative' forces by saying that the function of the signals in question is not that the content fits the world, but to change the world so that it fits the content. (A main reason why this concept would be incoherent is that the 'ought' in 'The content of a belief ought to fit the world' is different from the 'ought' in 'The world ought to fit the content of desires'. Frost doubts that we should use the word 'ought' in the latter case, and even if we may, its meaning is different from that of the 'ought' of the first case. Thus, the fitting relation is not the same in the two cases, but this is something that is presupposed by the notion of direction of fit.)

imperative force, this is how imperatives inherit their (pre-)illocutionary force. So, in the monkey case, the imperative force of 'Get down from the canopy!' actually comes from the indication of the monkey's desire that others get down from the canopy. The desire has the function of representing the world in such a way that it must change to fit its content. In this case, the content of the desire is that the other monkeys get down from the canopy. And probabilistic meaning can easily account for how this desire is conveyed: through a correlation between a monkey using the eagle call and this monkey desiring that the others get down from the canopy. The call inherits the force and the content of the desire, which is how it becomes an imperative, as opposed to a descriptive (such as the prediction that others monkey will get down from the canopy).

This is basically the move that is made by Ekman (1997) and adopted by Scarantino in his 'Theory of Affective Pragmatics' (Scarantino, 2017). Ekman remarks that a facial expression of emotion may not only carry information about what is felt by the person, but also plenty of other, related, pieces of information. These pieces of information may play different communicative roles and have different pre-illocutionary forces. He takes the example of a photo he took:

Fig. 7.5. What information is conveyed by the facial expression of the woman looking at the camera? (from Ekman, 1997)

Focusing on the woman that looks right at the camera, this is what Ekman writes:

> « Consider the diverse information that someone who observes this expression, totally out of context, just as it appears on the page, might obtain.
>
> - Someone insulted/offended/provoked her.
> - She is planning to attack that person.
> - She is remembering the last time someone insulted her.
> - She is feeling very tense.
> - She is boiling.
> - She is about to hit someone.
> - She wants the person who provoked her to stop what he/she is doing.
> - She is angry. » (Ekman, 1997: 316)

Scarantino (2017), drawing on Ekman, argues that these diverse pieces of information correspond to different (pre-)illocutionary forces. With her emotional expression, the woman is producing a signal with a descriptive force ('Someone insulted/offended/provoked her'), an expressive force ('She is boiling'), an imperative force ('She wants the person who provoked her to stop what he/she is doing'), and a commissive force, i.e. that of promises ('She is planning to attack that person') (Scarantino, 2017: 3).

This answer makes very interesting points, but it involves a move that is critical and about which I am skeptical: moving from the fact that a signal informs about a mental state to the claim that the signal thus inherits the force of the mental state. So, if we go back to our monkey example, I am skeptical about the move from the fact that a signal conveys information about a desire to the claim that this signal thus is an imperative. The reason is that there can be signals which inform about desires that are not thereby imperatives. Here is an example: I inform you that I desire to finish this paragraph and eat lunch. Through this sentence, I thereby convey information about a world-to-mind mental state. Does this suffice to give to the sentence itself an imperative force? Clearly not. Even if I inform you about a desire which concerns your actions, for instance in the sentence 'I hereby inform you that I desire that you wash your hands frequently', this is not sufficient to give to this sentence an imperative force.

In such cases, I have made *assertions* about my desire. I have produced a signal with an indicative force, even if its *content* is a desire, i.e. a mental state with an imperative force. The force of the *content* does not determine the (pre-)illocutionary force of the signal. That is why even though I conveyed my desire as part of the content of my speech act, I have not produced an imperative. But then, why would it be the case for the monkey using the eagle call? Why wouldn't he or she convey a desire that is not an imperative? How can this be captured by the probabilistic theory of meaning? I don't know how the defender can respond to these questions.

Another direction in which the defender of the probabilistic account may go is to refuse the starting point of this section and to deny that there is such a thing as pre-illocutionary imperatives. The defender of this strategy may argue that the way the eagle call works is that the probabilistic, indicative information about the eagle is plugged into some kind of consumer system of the receiver, which turns something non-imperative into action. 'Get down from the canopy!' is not part of the meaning of the signal, but monkeys infer it from the fact that there is an eagle around, just like they would if they had seen an eagle instead of having heard the call.

This solution seems to work for the vervet monkey case, but it has the disadvantage of giving up on explaining how can pre-imperatives exist. This is problematic insofar as some primatologists argue that apes do perform imperatives through 'request gestures', something that can be observed independently of how the addressee reacts (Gómez, 2007; Hopkins et al., 2013). In other words, there is not only a downstream, pre-perlocutionary effect of the request gestures, but we can observe the pre-illocutionary intent to request something. Similar observations are made concerning human infants (Van der Goot et al., 2014). Outside the realm of animal communication, imperative pre-illocutionary force is a construct that, for instance, serves to explain what information pain signals carry (Klein, 2007; Martínez, 2011; Martínez & Klein, 2016). In these examples and others, denying the existence of pre-imperatives or giving up on an explanation of their nature would be unfortunate. Fortunately, we are not forced to go in this direction because, instead of trying to account for all non-Gricean meaning through probabilistic meaning, we may account for the meaning of signals such as pre-imperatives through teleosemantic notions, as we will see in the next chapter.

7.4. Conclusion

In this chapter, we have reviewed various ways to capture how non-Gricean affective meaning may be encoded in both communicative and non-communicative signs. First, we reviewed two constructs based on factive indication: Grice's notion of natural meaning and Dretske's natural information. We then discussed two kinds of probabilistic meaning, one making use of objective probabilities (Shea's) and the other one making use of subjective probabilities (Scarantino's).

We have seen that, when we compare it to natural meaning and natural information, the notion of probabilistic meaning seems to be in a much better position to account for affective non-Gricean meaning. This is because most, if not all, affective signs are not factive or, at least, we don't know in what contexts they would be. We may nevertheless understand a lot by detecting them. In other words, we can learn from affective signs that do not possess a natural meaning.

Probabilistic meaning, on the other hand, seems to make accurate predictions about what information we may learn from non-communicative signs, for instance how we may learn about the arousal and the valence states of a person by studying physiological changes such as pupil dilatation or heart rate. However, we have seen that a probabilistic account of *signals* is faced with three kinds of difficulties.

The first difficulty is that, if we understand the meaning of signals to be defined in probabilistic terms, there is too much information that is predicted to be part of their meaning, especially because some of it seems absurd. For instance, probabilistic meaning arguably predicts that the vervet monkey's eagle call means 'The sky is green'.

The second problem comes from the fact that certain representations, such as the appraisals involved in emotional reactions, may be functional even if they are not statistically correlated with what they have the function to indicate. For instance, in many species (and perhaps all species), it is a good thing that fear episodes are more often false alarms than misses, i.e. that many animals, including humans, are more often afraid of something that is not dangerous than indifferent to something dangerous. This leads to a problem for an objectivist probabilistic account (such as Shea's) because, if false alarms are more frequent among alarm calls than veridical alarm calls, then an objectivist probabilistic account predicts that alarm calls mean the opposite of what they are supposed to mean, i.e. that they mean 'No danger' instead of 'Danger'. We have also seen how to extend this argument to show that a subjectivist probabilistic account (such as Scarantino's) also cannot cope with the 'rare eagle' scenario.

The third problem was that it is hard to see how probabilistic meaning can account for pre-imperative force. It seems restricted to the indicative force, i.e. that of informing others about something, as opposed to, e.g., getting them to do things.

In the next chapter, we will see how teleosemantic notions may overcome these difficulties. But before we move on, let me highlight again that the three difficulties presented here concern *communication* because it concerns the function of signals – e.g. to not transmit the claim that the sky is green, to involve more false alarms than misses, to get others to do things as opposed to inform them. When it comes to the meaning of non-communicative signs, probabilistic meaning seems to do a good job – a much better one than that of Grice's natural meaning and Dretske's natural information. Furthermore, since the teleosemantic notion which we will defend in the next chapter applies only to the meaning of communicative signs, we can conclude that, among all the kinds of meaning analyzed in this dissertation, probabilistic meaning, under its objective or subjective guise, certainly is the best account of the affective meaning of non-communicative signs.

8. TELEOCODED MEANING AND AFFECTIVE MEANING

« La Nature est un temple où de vivants piliers
Laissent parfois sortir de confuses paroles;
L'homme y passe à travers des forêts de symboles
Qui l'observent avec des regards familiers. »
– Charles Baudelaire, *Correspondances*

Abstract. In the last chapter, we saw that, contrary to the notion of natural meaning, the notion of probabilistic meaning seemed to be in a good position to analyze the affective meaning of non-communicative signs, but that it faces three kinds of difficulties when it comes to affective *signals,* i.e. signs whose function is to communicate affects. In this chapter, I present and develop the notion of *teleocoded meaning*, a teleosemantics akin to existing proposals such as Green's *organic meaning* (2019) or Shea, Godfrey-Smith, and Cao's *functional content* (2018). Like them, it is a modest teleosemantic account (Sterelny, 1990, Chapter 6), in contrast to more ambitious teleosemantic notions such as Millikan's (1984), Dretske's (1986), or Papineau's (1984). It is indeed restricted to information *encoded* in signals through a pre-established pairing, and this pairing must be best explained through evolutionary processes. I argue that it can overcome the difficulties that we saw probabilistic meaning was facing in the last chapter while preserving its advantages over natural meaning.

Another conclusion reached in this chapter is that code-based meaning can be inherited from properties of affects that need not be consciously accessible, contrary to Gricean meaning. This last conclusion will lead us to examine what is unconscious in affects in the next chapter.

8.1. INTRODUCING TELEOCODED MEANING

In this chapter, I will introduce what I mean by 'teleocoded meaning' and then show how it helps us solve the problems faced by probabilistic meaning in the last chapter. We will see how it helps us account for cases of affective communication which we could not explain with the notions introduced in the preceding chapters.

8.1.1. TELEOCODED MEANING AND SOME OF ITS ANTECEDENTS

Just like two of its closest antecedents Dretske (1986)'s 'functional meaning' and Green (2019)'s 'organic meaning', teleocoded meaning may be introduced as a notion that sits in between Grice's (1957) natural and non-natural meaning.[154] Like natural meaning and unlike non-natural meaning, it does not require communicative intentions, pragmatic principles, or mindreading abilities. Like non-natural meaning and unlike natural meaning, it may misrepresent and it is always communicative.

Just like Dretske's and Green's notions, and as its name indicates, teleocoded meaning belongs to the teleosemantic tradition, where representation ('–semantic') is understood through the notion of biological function ('teleo–'), which is itself often defined in evolutionary terms (or through selection by learning processes) (Dretske, 1986, 1988; Godfrey-Smith, 1991, 1996; Green, 2019b; Millikan, 1984, 1989; Nanay, 2014; Neander, 1991, 2018; Papineau, 1984, 1993; Shea et al., 2018; Sterelny, 1990).

Unlike Dretske's and Green's notions and the vast majority of its teleosemantic antecedents –with the notable exception of Shea, Godfrey-Smith, and Cao's 'functional content' – teleocoded meaning is restricted to information *encoded* in signals, i.e. information that can be transferred from a sender to a receiver thanks to pre-established rules which allow them to pair pieces of information and types of signals. As such, teleocoded meaning, just like natural meaning, probabilistic meaning, and conventional meaning, can be accounted for by a code model of information transfer. No need to postulate the mindreading abilities and the respect of pragmatics principles required by Gricean models.

Fig. 8.1 illustrates how the kinds of meaning discussed in this dissertation hang together.

[154] More precisely, it is meant to fill an explanatory gap left by probabilistic meaning (see Chapter 7) and what is not coded in Gricean meaning (see Chapter 1 and Chapter 2).

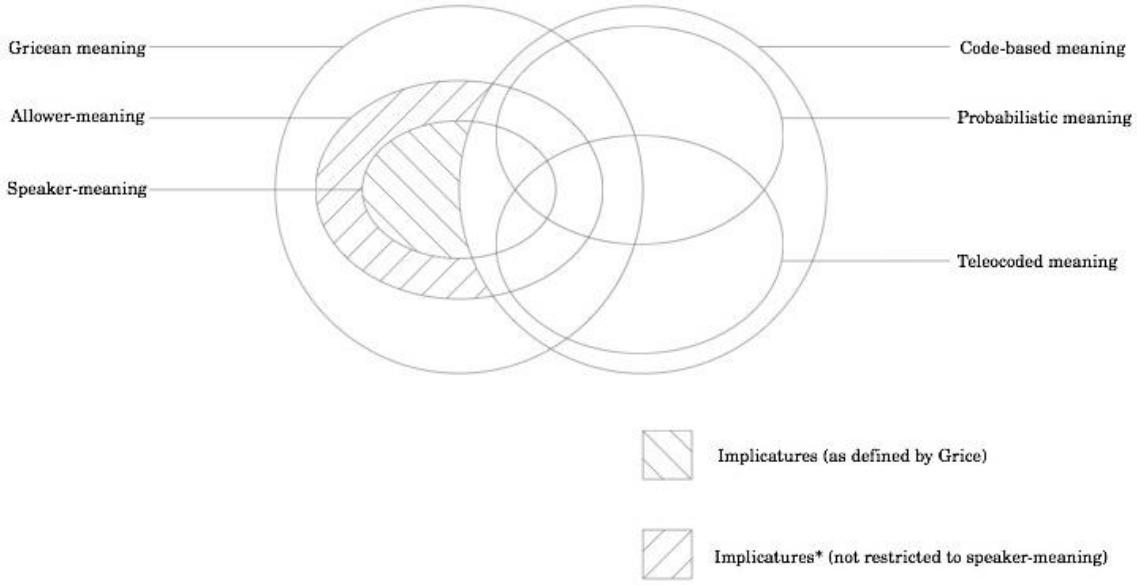

Fig. 8.1. Our typology of meaning.

Teleocoded meaning differs from the two kinds of code-based meaning we have discussed in Chapter 7 because, unlike natural meaning, it is not factive and, unlike probabilistic meaning, it is not defined through correlations. Rather, it is defined through the communicative *function* of signals.

For this reason, it is a subspecies of what Dretske calls 'functional meaning' ('meaning$_f$'), a paradigmatic teleosemantic notion. Here is how he defines it:

> « (Mf) d's being G means$_f$ that w is F = d's function is to indicate the condition of w, and the way it performs this function is, in part, by indicating that w is F by its (d's) being G. » (Dretske, 1986: 22)

So, for instance, the bark of a vervet monkey that has the acoustic features of the eagle call means$_f$ that an eagle is present because the function of this bark is to indicate that an eagle is present by its having the acoustic features of the eagle call.

However, Dretske's functional meaning is not equivalent to teleocoded meaning because functional meaning includes indicative speaker-meaning (as opposed to, for instance, imperative speaker-meaning) while teleocoded meaning does not. Signals carrying indicative speaker-meaning indeed have the function of carrying information and do so in part by indicating certain things. They acquire this function notably through the successful intentions of speakers. These intentions allow speakers to mean more than

what is encoded in the signal; speaker-meaning includes implicatures. So functional meaning goes beyond encoded meaning.

Similar remarks can be made about other classic teleosemantic notions, such as Millikan's 'representational proper functions' (1984, 1989) and Papineau's 'normal purposes' of 'states with representational powers' (1984, 1993, Chapter 3). Just like Dretske's functional meaning, these are not restricted to code-based meaning. On the contrary, they were designed with the ambition to explain all kinds of representations in biological, non-intentional terms.

As such, Dretske's, Millikan's, and Papineau's teleosemantic notions are what Sterelny (1990, Chapter 6) calls 'immodest' accounts: they attempt to give a fully general analysis of representation (see also Godfrey-Smith, 1996, Chapter 6). Modest teleological accounts, by contrast, aim at explaining only some kinds of representations. Because teleocoded meaning is restricted to the meaning of signals with an encoded meaning, it is a modest notion.

Another notion that is very similar to teleocoded meaning is Green's 'organic meaning'. The two are similar insofar as both are defined through evolutionary notions and both are modest teleosemantic notions (neither aim to account for all kinds of meaning). Green defines organic meaning as the meaning of signals that are sent by organisms (animals, plants, bacteria, cells, …) without necessitating the complex communicative intentions defining speaker-meaning. Instead, signals possessing organic meaning acquire their communicative function thanks to natural selection (he also mentions cultural selection as a possible extension of his definition of 'organic meaning'). More precisely, the signals in question must be 'a behavioral, physiological, or morphological characteristic fashioned or maintained by natural selection because [in part] it serves as a cue to other organisms.' (Green, 2019b, p. 215) A 'cue' is an information vehicle. So, a signal is an information vehicle that some organisms have because it makes them more fit to their environment (otherwise the organism would lose this trait, *ceteris paribus*) and because it makes other organisms more fit.[155]

Although this definition comes very close to how I will define teleocoded meaning, an important difference is that organic meaning is not limited to the information *encoded* in signals. For this reason, even though Green

[155] A misleading cue thus doesn't fit the definition of having organic meaning because although it benefits the organism producing the cue it doesn't benefit the receiver; it doesn't *serve* as a cue to other organisms.

explicitly excludes speaker-meaning, his definition does not exclude other kinds of non-code-based meaning and, in particular, it does not exclude the implicatures that are allower-meant (see Chapter 2 for these notions). For instance, Green considers that the meaning of humans' spontaneous expressive behavior is a kind of organic meaning because it is not produced with communicative intentions. But, as I have argued in the first part of this dissertation, it is not because a signal is produced without communicative intentions that its meaning is restricted to what it encodes (Chapter 1). I have argued instead that accounting for the information transferred by some signals produced without communicative intentions nevertheless requires a Gricean kind of explanation, as opposed to a code-based explanation (Chapters 2–4). Organic meaning differs from teleocoded meaning because the latter is restricted to code-based meaning and thus excludes any kind of implicature.[156]

Below, I will compare teleocoded meaning to another modest teleosemantic notion: Shea, Godfrey-Smith, and Cao's 'functional content' (2018). But first, let me give a definition of teleocoded meaning and explain it.

8.1.2. Defining teleocoded meaning

Here is how I define teleocoded meaning:

> Teleocoded meaning – definition:
>
> The teleocoded meaning of a signal consists in
>
> (a) the information encoded in the signal
> (b) that the signal has the function to convey,
> (c) where the performance of this function is best explained as being the result of genetic or cultural evolutionary processes.

First of all, note that the definition is restricted to the meaning of signals. You will remember that I define signals as stimuli that have the function of transferring information from a sender to a receiver. Non-communicative stimuli cannot have a teleocoded meaning. Note also that 'information' is understood broadly to include misinformation: a signal doesn't stop being a signal because what it 'says' is false.

Clause (a) implies that there is a pre-existing set of rules – a code – used by the sender and the receiver and which pairs the information and the

[156] A further difference is that a signal possessing teleocoded meaning not only needs to be the product of evolution, but its having this meaning because of evolutionary processes must be the *best explanation available*. Finally, it is not clear whether Green wants to extend the source of organic meaning to cultural selection.

stimuli carrying it in such a way that the receiver can get the information encoded in the stimuli by the sender. In other words, the sender encodes the information in the signal by following pre-established sender's rules, and the receiver decodes the information encoded in the signal thanks to the receiver's rules. The encoding and decoding rules must thus correspond to each other. A simplistic example of a sender's rule may be something like 'if you see an eagle around, emit bark A' and the corresponding receiver rule 'if you hear bark A, assume that there is an eagle around'.[157] These rules constitute a code that pairs information (an eagle is around) and signals (bark A) and explain how senders and receivers communicate. This clause thus excludes any kinds of implicatures.[158] Another way to put it is that teleocoded meaning is restricted to (super-)semantic as opposed to pragmatic meaning (see Chapter 1).

Clause (b) is to be unpacked as such: the communicative function of the signal is performed through the following two steps: (i) a sender produces a signal, thereby encoding (using the sender's coding rules) the information that is to be conveyed into a shareable vehicle which makes up the signal's cues, and (ii) the signal is detected by a receiver who can decode it (using the receiver's rules) and thus access the information in question. The information of steps (i) and (ii) is *meant* to be transferred; it is the function of this signal to do so. Even in cases where the information is not transmitted (because of problems on the sender's side, on the receiver's side, or in between).

Clause (c) adds that the signal having this function and successfully performing it must be best explained through an *evolutionary process*. By 'is best explained' I mean 'is best explained *given the explanations available*', and not, for instance, 'would be best explained if we had more information'. We may contrast an explanation based on evolutionary processes with one based on intentional design. For instance, to explain how a word of Sindarin – an Elfish language devised by J.R. Tolkien – has acquired its communicative function, we will not give an evolutionary explanation, but talk about how Tolkien devised the lexicon of this fantasy language. Clause (c) thus excludes the conventional meanings which are best explained by intentional design, but it does not exclude the

[157] We can ignore the further rule which would say something like 'If you assume that an eagle is around, go down from the canopy'.
[158] But, as we will see below, if someone can design a code model of information transfer that can account for speaker-meaning and/or allower meaning, and can give an evolutionary explanation for this code model which is better than an explanation based on intentional design, then teleocoded meaning will 'swallow' speaker-meaning and/or allower meaning, including their implicatures.

conventional meaning that is best explained through a cultural evolutionary process. Indeed, clause (c) states that the evolutionary process may be genetic or *cultural*. Let me say more about this.[159]

I will follow Sterelny (2006) and the dual inheritance theory by which he is inspired (notably Richerson & Boyd, 2005) in considering that there are three general properties of inheritance that are central to how selection mechanisms generate adaptation, and so how evolutionary processes work. These properties are shared by both genetic and cultural evolution.

First, there must exist inheritance systems that allow a *stable transmission* of traits over generations. Second, despite this stability, the inheritance systems must also allow the generation of *variations*. Third, what is inherited makes the individual or the group fitter to the environment, and thus more likely to transmit what is inherited further.

Teaching what plants are edible may be such an inheritance system. It can be stable: e.g. thanks to repeated feedback from teachers during a long period of acquisition, until pupils can become trustworthy teachers for the next generation. It can allow for variations: new plants may be found to be edible and become part of the curriculum, while other plants disappear from the region or are forgotten. And this can obviously have an impact on the fitness of the individual and especially the group.

Other examples of cultural variants – i.e. what is culturally reproduced within cultural evolution, a more recent alternative to Dawkins' 'memes' – include tools, melodies, or, indeed, non-genetically determined signals.[160]

That cultural evolution is included in the definition of teleocoded meaning implies that signals whose meaning aren't intentionally designed (unlike Tolkien's Elfish words), but that are not genetically determined may nevertheless possess a teleocoded meaning. Many human emotional expressions belong to this category: being non-intentionally designed, but

[159] We may wonder where to classify functions resulting from a genetic process of selection that is intentional, i.e. results from 'artificial selection', as in breeding. For instance, the 'pointing' behavior of certain hunting dogs may be considered as a teleocoded meaning, because the dog presumably has not rationally or intentionally designed the function of her pointing. On the other hand, this function may be considered as having been intentionally designed by human breeders. Maybe it is best to see it as a signal with different kinds of meaning, where different notions give complementary explanations. I will come back to mixed signals in the general conclusion of the dissertation.

[160] Instead of cultural variants, a theoretical construct that may have been used is Millikan's 'learned proper function' (Millikan, 1984), but this notion is much less flexible than that of cultural variants as it presupposes a whole lot of other definitions, 'proper function' being defined through a series of concepts which belongs to Millikan's specific framework.

not genetically determined. Such signals would possess a teleocoded meaning if their communicative function is best explained through cultural evolutionary processes (along with genetic ones).

Here is a sketch of how such an explanation may go: we often tend to imitate (unconsciously) the facial, vocal, or gestural expressions of our peers. Mastering how to interpret and how to produce these expressions increases one's fitness (as well as the group's fitness). These expressions evolve over time, through mutation and selective reproduction, so that two cultures will have different ways of expressing certain emotions, even if these expressions come from a common ancestor. This whole process can be entirely unintentional, in the sense that producing and imitating expressions can be wholly unintentional. Just like I never intended to have the accent that I have, I may express myself through signals that I learned from my peers but which I never meant to imitate, e.g. the way I laugh comes, at least in part, from a mixture of unconscious influences.

I contrast evolutionary explanations with explanations based on intentional design. Now, of course, everything that is intentionally designed is *ipso facto* the product of evolution, at least on our planet, because intentionality itself has been designed through evolutionary processes. But if we focus on the particular signals and their particular function, i.e. their function to carry this or that particular piece of information, then the best explanation available, the most sophisticated and satisfying one, is not always an evolutionary explanation.

For instance, if we take the sentence that I am presently writing, we can explain the fact that it has the communicative function that it does by talking about my intentions to express certain ideas to a certain audience, and thus to have chosen those words in that order. Such an explanation, i.e. one based on an intentional design, will be much more satisfying than an evolutionary explanation, either genetic or cultural. Of course, there are evolutionary processes that we can use to explain the communicative function of this sentence: we can form illuminating hypotheses about human language evolution, and isolate the genes that allow us to write down sentences. We can also study the cultural evolution of English words and syntax, e.g. the shift from early Middle English /d/ to modern /ð/ and how this cultural variant spread. But such evolutionary explanations won't explain the communicative function of *this particular sentence* that I was writing, as opposed to many other English sentences. There are no satisfying evolutionary explanations for how each English sentence has acquired its particular function. But there are intentional explanations: those using the concept of speaker-meaning and the Gricean model.

For these reasons, I find it quite uncontroversial that explaining the communicative function of speaker-meaning is best done through intentional design than through evolutionary processes. At least, this is so given the current explanations available, given our current knowledge, and limited cognitive capacities. A super-intelligent being who can compute an evolutionary process that yields a more precise, satisfying, and sophisticated explanation than any intentional explanations for each token function of speaker-meaning may consider that speaker-meaning is a kind of teleocoded meaning, and that's fine with me! Given that evolution appears to be the ultimate source of all signals we know of, it is reasonable to except that an omniscient being considers them as possessing teleocoded meaning. But for us mortals of the 21st century, the distinction between teleocoded and speaker-meaning remains.

This, by the way, points to a big advantage that teleocoded meaning has over its more ambitious teleosemantic antecedents. A common criticism of teleosemantics is that it cannot account for representations that clearly have no evolutionary functions whatsoever, such as random sentences we may have about useless stuff. Such representations don't contribute to our fitness and never have, so it is hard to see how they could possess a communicative function, where 'function' is understood in evolutionary terms. Millikan (1984), Papineau (1984), or Dretske (1988) have proposed answers to this challenge, but none that has been widely accepted. This challenge is indeed one of the main motivations, if not the main one, to 'go modest' (Sterelny, 1990, p. 129).[161]

In sum, if the meaning of a signal is best explained through the extended Gricean model, then it is not a teleocoded meaning – it is either speaker-meaning or allower meaning (or another kind of Gricean meaning of which I am not aware). If it is best explained through a code and where its encoding-decoding function is best explained through an evolutionary process, then it is teleocoded meaning. If the meaning of the signal in question is best explained through something else than either the extended

[161] Furthermore, we can imagine that in other possible worlds, or in Swampman scenarios, there are creatures who speak just like we do but who don't have an evolutionary history. Teleocoded meaning won't apply to such cases, but Gricean meaning can. This may be considered as an advantage that teleocoded meaning has over more ambitious teleocoded meanings. Nevertheless, it also points to a potential problem of teleocoded meaning: a similar scenario – such as a Swampmonkey scenario – may be considered as a strong objection. I won't discuss this classic objection which has been responded to in many ways by different authors (for a review, see Neander (2018, sec. 4.2)). I will only note that we may define evolutionary functions and thus teleocoded meaning without etiology, without the evolutionary history of a trait, but through its potential increase in fitness (see Nanay, 2010, 2014), which allows avoiding Swampman problems, among others.

Gricean model or the evolutionary processes behind teleocoded meaning, then this signal may possess a probabilistic meaning, a natural meaning, or it may possess a kind of meaning of which I am not aware.[162]

I summarize the main features differentiating these five kinds of meaning in Table 8.1.

	Requires Gricean communicative intentions	**Requires mindreading based on Gricean principles**	**Always is communicative**	**Is factive**
Speaker-meaning	Yes	Yes	Yes	No
Allower meaning	No	Yes	?[163]	No
Teleocoded meaning	No	No	Yes	No
Probabilistic meaning	No	No	No	No
Natural meaning	No	No	No	Yes

Table 8.1. Five kinds of meaning

8.1.3. Applying teleocoded meaning

We can find examples of teleocoded meaning in a wide diversity of cases, from trees to monkeys, bees, neurons, and icons, but I will concentrate on some of the examples we have encountered in the last chapter.

First, let us observe that teleocoded meaning does not apply to many cases where probabilistic meaning does: non-communicative signs don't possess the communicative function which defines teleocoded meaning. So many of

[162] I cannot think of an example where the affective meaning of a signal is best explained by either natural or probabilistic meaning compared to either teleocoded or Gricean meaning, so that affective signals seem to always possess either a teleocoded or Gricean meaning (or both). Remember nevertheless that, as we saw in the last chapter, when it comes to non-communicative signs, their affective meaning seems best explained through the notion of probabilistic meaning. This fits well within our general picture since, by definition, teleocoded meaning does not apply to such non-communicative cases.

[163] See the discussion in chapters 1 and 2 on whether allower meaning always is communicative or not.

the examples discussed in the last chapter, such as the signs used by psycho-physiologists to infer affective states, won't have a teleocoded meaning, even if they do have an affective probabilistic meaning. For instance, sweating, pupil dilatation, or blushing do not appear to be signals, as they seem not to have the function to convey the information that they may convey. Indeed, these signs appear to be affective reactions that are, in the eye of evolution, *accidentally* informative about affective states.

By contrast, it is widely recognized that certain responses have evolved to convey information about the affective state of the expresser, that they can fulfill this function through a code shared by expresser and their receivers, and so that they do have an affective teleocoded meaning. Alarm calls are a prime example to which I will come back. Another example is unintentional facial expressions.

Ekman and his colleagues have argued that a certain number of human facial expressions have evolved through natural selection to acquire the meaning they have for us (Ekman, 1992, 1997; Ekman & Cordaro, 2011; Ekman & Friesen, 1971). This claim is disputed by a number of affective scientists, such as Russell and Barrett, who argue that most facial expressions discussed by Ekman are in fact socially constructed (Barrett et al., 2019; Barrett & Russell, 2015; Russell, 1994). Because the evolutionary process behind teleocoded meaning can be cultural, whether or not the expressions discussed by Ekman are social constructions, at least part of their meaning may be considered as teleocoded meaning (another part may be a non-code-based meaning, as we have seen in Chapter 3). In the preceding chapter, we have seen why the hypothesis that facial expressions can be explained in terms of natural meaning or of probabilistic meaning is not satisfying. Above, I gave a brief sketch of how an explanation based on unintentional imitation and cultural selection may be more adequate. But instead of focusing on this example, I will come back to the vervet monkey eagle alarm call which we discussed at length.

Let us suppose, as we did in the last chapter, that the function of the eagle call is to carry (i)–(iii):

(i) An eagle is present.
(ii) I am so afraid (of an eagle)!!!
(iii) Get down from the canopy!

Now, I would argue that the best explanation of how this alarm call has obtained its communicative function is an evolutionary process, as opposed to an intentional or intelligent design. The question is not whether vervet

monkeys emit the call intentionally – they may well do so (Zuberbühler, 2018). The question rather is how the eagle call obtains the meaning it has. And the best answer available, I believe, involves genetic evolutionary processes. A main reason for thinking so is that vervet alarm calls are largely innate. For instance, if we compare the calls given by two subspecies of vervet monkey, from Eastern (*Chlorocebus pygerythrus pygerythrus*) and Southern Africa (*Chlorocebus pygerythrus hilgerti*), we find only marginal acoustic differences for the three predator calls. Furthermore, only marginally more pronounced differences are found between the alarm calls of vervet monkeys and that of another species of the same genus, the West African green monkey (Price et al., 2014).

How should we explain the signal's meaning through a genetic evolutionary process? It may well go along the lines of an evolutionary signaling game. Such explanations are 'naturalizations' of classic 'game theory' explanations, such as Lewis signaling game (1969) and have been developed by, e.g., Maynard Smith and Harper (2003), Skyrms (2010), or Shea, Godfrey-Smith and Cao (2018). Although Lewis restricted signaling games to interactions between rational agents with mindreading abilities (a very Gricean framework), these authors have shown how signaling games can be defined in purely unintelligent, evolutionary terms.

A signaling game is a system where the sender has access to information about states of the world, but cannot act on it except by sending a signal. The receiver can only see the signal, but can subsequently act on it in ways that are more or less beneficial to both sender and receiver. What signal is sent given the state of the world – the sender's rule – may change and what action is performed by the receiver given what the signal is received – the receiver's rule – may change as well. Depending on the different benefits yielded by the different rules they follow, the sender and receiver's behaviors will be modified until the system reaches stabilization or equilibrium, i.e. a state where changing the behaviors yields fewer benefits.

Now, here is a rough and simplistic presentation of how signaling games may help explain how the vervet's alarm call acquired its meaning. We think of the ancestors of vervet monkeys as participants to the game. A first participant observes the world and produces a certain behavior in response to it, a response which has, as of yet, absolutely no function, and which has emerged randomly. E.g. she sees an apple and moves her tail. Other participants observe the response, and respond to it by their own behavior. E.g. one sees the tail moving and jumps, another approaches the first monkey. The responses either benefit the participants, are neutral, or

are damaging. Now, if the responses are determined by *genes* (e.g. genes determine the apple–tail, the tail–jump, and the tail–approach responses), even though the responses emerged randomly, through gene mutations, a proportion of these responses will get passed on the next generation. The beneficial responses will tend to be preserved over generations – especially if they benefit the whole group – because they have helped the participants of the game to be more fit to their environment and thus have helped them reproduce their genes. Damaging responses will tend to disappear because they are detrimental to the participants so that their genes will spread less than those of the participants whose responses are beneficial. Despite its oversimplified nature, this sketch helps understand how we may explain that vervet monkeys possess genes that make them respond to certain stimuli in certain ways (e.g. see-eagle–make-eagle-call) and respond to such responses in certain ways (hear-eagle-call–assume-an-eagle-is-present/go-down-from-canopy); how we may explain the emergence of code-based meaning of the vervet monkey alarm calls.

Another reason to believe that the way in which the eagle call has acquired its communicative function is best explained through evolutionary processes is that the alternative explanations are worse. As far as I know, alternative explanations may be based on the notion of probabilistic meaning, natural meaning, and Gricean meaning. We have seen that the first two are unsatisfying in the last chapter. And an explanation based on Gricean meaning would be cognitively too demanding. It would probably need to go along the lines given by Grice (1967/1989), Lewis (1969), Schiffer (1982), or Sperber (2000) (e.g. Reboul, 2017; Scott-Phillips, 2015) and thus require mental processes that are out of reach to vervet monkeys. Such explanations are based on a process of conventionalization (also called 'fossilization') of signals created with the communicative intentions of the prevailing Gricean models. This process necessitates that participants possess a sophisticated amount of cognitive skills, especially mindreading abilities, which the vervet monkeys appear to lack (Zuberbühler, 2018).

For these reasons, among others, it appears that the evolutionary explanation is in a better position to account for the meaning of vervet monkey's alarm calls. It thus makes sense to hypothesize that these calls possess a teleocoded meaning. In §8.2, we will see how this hypothesis avoids the problems yielded by the probabilistic account.

8.1.4. TELEOCODED MEANING, SIGNALING GAMES, AND FUNCTIONAL CONTENT

Let me end the introduction of teleocoded meaning by comparing it with a similar teleosemantic notion, called 'functional content', which was recently proposed by Shea, Godfrey-Smith, and Cao (2018). The two are similar insofar as both are defined through evolutionary notions, both are modest teleosemantic notions, both are restricted to encoded meaning, and both are defined in ways that make them complementary, rather than in competition with, probabilistic meaning. They differ insofar as Shea et al. restrict their notion to signals whose communicative function must be explained through a signaling game such as has been presented in the last subsection. By contrast, teleocoded meaning may make use of, but is not restricted to, such a framework.

Here is how Shea et al define 'functional content' (remember that an equilibrium is a state where neither sender nor receiver can change their rule unilaterally and be better off, given what the other is doing):

> « The messages in a sender–receiver system have functional content only if the system is at an equilibrium maintained by some selection process. If it is, then for each signal M, we ask whether there is a behaviour (or distribution over behaviours) of the receiver specific to M, in the sense that the receiver responds differently to M than it does to some other available signal. … If so, we look at whether there is a specific state of the world that obtains on some occasions when the message is sent, where the relation between that state of the world and the behaviours produced by the message contributes to the stabilization of those sender and receiver behaviours. If so, that state is the content of M. » (Shea et al., 2018, pp. 1015–1016)

This is not the place to explain in detail this complex definition. I will only note that a signal may have a functional content only if it benefits senders and receivers because equilibria are reached only in cases where both sender and receiver overall benefit from their respective behaviors (both agents receive above-baseline payoffs).[164]

[164] More precisely, here is how the relation between functional content and baseline payoff is defined: 'The baseline for each agent is the agent's average payoff in a situation where the receiver adopts the best strategy available to it without conditioning its behaviour on any signals … the functional content of a message correspond to states in which the message is sent and both agents receive above-baseline payoffs, given the receiver's rule for that message.' (Shea et al., 2018, p. 1016).

Even though I see a lot of value in the notion of functional content – especially because it can be captured through a mathematical formulation which may allow quantitative empirical tests – I don't want to define communicative functions as such, because doing so prevents me from capturing some cases where teleocoded meaning apply.

Such cases include those where a signal may preserve its communicative function despite not being beneficial to senders and receivers and so not contributing to the stabilization process. One example is the rare eagle scenario presented in the last chapter (see also below). In this situation, it is overall harmful to both senders and receivers to produce the signal, since it makes them waste energy and has no benefits. Consequently, it cannot have a functional content according to Shea et al.'s definition, because it does not lead to a stabilization of the sender-receiver system. However, it still has the teleocoded meaning 'An eagle is present'.

Shea et al. may respond that this is not a counterexample to their definition because this scenario is not at an equilibrium maintained by some selection process (their first necessary condition) – it is unstable: the call will eventually disappear in such a situation due to evolutionary constraints. The scenario thus falls outside the scope of their definition. We may happily agree and remark that this shows in any case that their definition of functional content is not co-extensional with the notion of teleocoded meaning, and that it doesn't apply to cases we want to capture.

Their definition also implies that if a signal in a signaling game tends to hurt the audience, then it has no functional content, because the payoff of the receiver is not raised and so this does not contribute to the stabilization of the system as they define it (for the same reason that deception has no functional content according to their definition, which is a welcome result). But there may be signals with a teleocoded meaning which tend to hurt their audience. For instance, certain insults may tend to hurt their receiver and so lack a functional content although they may possess a teleocoded meaning. This example is hypothetical: I don't have any evolutionary explanation for insults – but I don't want to close that door either.

Finally, if a signal tends to be hurtful to the sender (maybe using a taboo word which one is not supposed to use), then it cannot have functional content either, although it may have a teleocoded meaning.

For these reasons, but especially the first, I don't want to restrict teleocoded meaning to functional content.

8.2. Advantages of teleocoded meaning over probabilistic meaning

Let us now review the advantages that the notion of teleocoded meaning has over that of probabilistic meaning, which already seemed superior to natural meaning for studying affective code-based meaning (last chapter). Let me note by the way that the advantages that this notion has are the advantages that teleosemantic notions have in general. But it can account for the cases we need it to account for without facing the problems faced by immodest notions such as Dretske's, Millikan's, or Papineau's.

8.2.1. First advantage: A functionnally restricted meaning

The first difficulty with the notion of probabilistic meaning was that, if we understand the meaning of signals to be defined in probabilistic terms, there is a lot of information that is predicted to be part of their meaning, but which we would rather not attribute to the meaning of a signal. Remember that we found that the vervet eagle call probabilistically means the following:

- The user of this call does not believe that Pythagoras' theorem is true.
- The sky is green.
- Evolution is true.
- The speed of sound in air is about 343 meters per second.

Quite obviously, it is not the function of the eagle call to transmit any of those. Thus, they are not part of the teleocoded meaning of the signal. This is a first advantage that teleocoded meaning has over probabilistic meaning.

By contrast, we can very plausibly understand that it is part of the evolutionary function of the eagle call to communicate (i), (ii), and (iii):

(i) An eagle is present.
(ii) I am so afraid (of an eagle)!!!
(iii) Get down from the canopy!

Maybe it is less obvious that (ii) has an evolutionary function, but, anticipating what we will discuss in the next chapter, if we consider that emotions involve appraisals, then along with (ii), or as part of (ii), the signal may well transmit the following:

(iv) The user of the call is appraising a stimulus as goal-unconducive and very hard to control.[165]

It may well be the case that the function of the eagle alarm call is to convey (iv), even if monkeys don't consciously represent (iv). At least, such information is the kind of information whose transmission would be beneficial for both the monkey producing it and the audience.

Such remarks point to the fact that, even in creatures who can be conscious of the messages they send, teleocoded meaning may well be inaccessible to consciousness. As such, it is very different from Gricean meaning. Even if we can imagine that monkeys can be conscious of (something like) (i), (ii), and (iii), they certainly are not conscious, and cannot be conscious, of (iv). This is also how it is with us: when we utter the equivalent of an alarm call, e.g. scream out of fear because we saw a snake or a burglar, we may well be conscious of something like the equivalent to (i), (ii), and (iii), but we are not conscious of appraisals such as (iv) which underlie our fear, as we will see in details in the next chapter. Nevertheless, from an evolutionary point of view, it makes a lot of sense that an alarm call or a scream is designed to convey (iv), and so that it is part of their teleocoded meaning.

8.2.2. Second advantage: Safe from stats

Perhaps the main advantage that teleocoded meaning has over probabilistic meaning is that it allows signals to have, and preserve, their communicative function even in cases where what the signal means is probabilistically less frequent than what it does not mean.

Remember the rare eagle scenario which I presented in the last chapter where a group of vervet monkey happens to live in a place where eagles have become so rare – we can imagine they are extinct – that there is no probabilistic link between the presence of an eagle and their uttering the eagle call, which they nevertheless still utter when they see something that vaguely resembles an eagle. Probabilistic meaning cannot account for the fact that, in this scenario, the eagle call preserves the meaning (i) 'An eagle is present' and does not acquire the meaning 'There is no eagle'.

[165] We may be more precise and hypothesize that the louder and the noisier (inharmonic and non-linear) a call is, the more the event is appraised as unconducive to the monkey's goals. If a vervet sees an eagle in the distance as opposed to very close to her, she would emit a softer, less noisy call (for related claims, see Bar-on, 2013; Blumstein & Récapet, 2009; Grandjean et al., 2005).

Now, if we understand 'meaning' as 'teleocoded meaning', we may justifyingly say that, no matter what is the proportion of false alarms, even in the rare eagle scenario, the eagle call preserves the meaning (i) and does not acquire the meaning 'There is no eagle'. Indeed, the function of the eagle call is preserved despite the lowering of statistical correlations. The code used by the monkeys (i.e. the sender's and receivers' rules) remains the same and the information transmitted (in this case: the systematic misinformation) remains the same. The code was established when the eagles weren't extinct and it has not changed. The signal encodes information that has so to say passed its use-by date.[166]

Let me observe by the way that even though teleocoded meaning does not essentially depend on statistical correlations, correlations can certainly help the evolutionary process in many cases. In other words, the fact that certain non-communicative signs correlate with certain information certainly can be a factor that helps explain how the non-communicative sign evolves to become a signal whose function is to indicate the information in question.

A rather clear example of this process is what is usually called 'ritualization' in evolutionary studies. This process is nicely illustrated by what is often proposed as the evolutionary history behind the baring of teeth in wolves or primates (among others). In the first step of this history, ancestors bare their teeth when threatened in preparation for attack. At this point, there is a correlation between 'I will attack' and bared teeth. Over time, other creatures evolve to respond to bared teeth by treating it as a cue to a forthcoming attack. They thus tend to keep the distance from creatures baring teeth, and especially from those which emphatically show their teeth in an 'exaggerated' way that goes beyond mere preparation for attack. Avoiding fights is beneficial for both the 'exaggerated teeth barer' and the 'distance keeper' so that their genes spread comparatively more than those who do not follow these sender-receiver rules. At the end of the process, baring teeth is a 'ritual' which means 'I will attack' even though there is no longer any strong correlation between baring one's teeth and

[166] Maybe this particular function of the signal (i.e. to encode information about eagles) is determined by the monkeys' DNA, as evidenced by the innateness of the call–responses mechanisms in vervet monkeys and related species. The change in context thus would not affect what information is carried by the signals – at least not until the monkeys' DNA adapt and evolve accordingly. In our example, the function remains despite the absence of eagles. This may be comparable to our hypnic jerks: when falling asleep, 70% of us sometimes twitch. This may have been a reflex useful for our primate ancestors who were sleeping in tree branches to secure that their sleeping position was stable enough before they fall deeply asleep (Coolidge & Wynn, 2006). In our context, this reflex has no more evolutionary value, but, arguably, its function remains the same.

attacking – even if teeth are bared more often when the creature actually will *not* attack. The initial probabilistic correlation between bared teeth and attacks helps explain how the ritualized display acquired its communicative function and its teleocoded meaning.

8.2.3. Third advantage: Non-indicative teleocoded meaning

Let us now turn to the third problem that probabilistic meaning had: its incapacity to account for different kinds of pre-illocutionary forces, being restricted to an indicative pre-illocutionary force.

You will remember that, in the last chapter, I mentioned how signals sent by non-human animals, and even by plants or intra-organism cells, may be analyzed through what originally is a speech act formalization: F(p), where F represents the force and p represents the content. For instance, we may want to interpret the meaning of tree stress signals as !(other trees protect themselves) or as ⊨(there is a threat) where '!' represents the tree version of an imperative (or directive) force and '⊨' that of an indicative (or descriptive) force, to account for the 'behavior' which trees adopt in response to stress signals (Gorzelak et al., 2015).

We have seen in the last chapter that probabilistic meaning seems to always have an indicative force, because it only tells you how the world probably is like, and not, for instance, what others *ought* to do. So if non-Gricean signals with an imperative force exist, as is argued by e.g. Millikan (1984, Chapter 3, 1989, 1995) and more recently by philosophers working on the meaning of pain signals (Klein, 2007; Martínez, 2011; Martínez & Klein, 2016), then probabilistic meaning is in a difficult position as it seems incapable to account for this fact.

Through a teleocoded meaning framework, however, there is no problem with accounting for the claim that signals may have different pre-illocutionary forces.[167] Indeed, it is entirely reasonable that the evolutionary function of certain signals is to convey contents with different forces.

So, for instance, if we go back to the tree stress signals, it is an open question whether the signals in question have the function of encoding !(other trees protect themselves) or ⊨(there is a threat), or neither, or both – like Millikan's 'pushmi-pullyu representations' (1995). What information

[167] See also Millikan (1984, Chapter 3, 1989, 1995) on indicative and imperative representations for concepts slightly different than that of teleocoded meaning, but which also allow a naturalized perspective on what I call 'pre-illocutionary forces'.

these signals have the function to carry will determine their pre-illocutionary force. We may find out about their force by, for instance, discovering that the relevant molecules composing tree stress signals, which travel through the soil from trees to trees, are the same molecules that are used internally by a tree to carry information about a threat from one part of its 'body' (e.g. the roots) to another (e.g. the leaves), and so that the function of this type of molecules probably is to carry the information ⊨(there is a threat) rather than !(other trees protect themselves), since the tree seems to use the molecules internally, in the absence of other trees, for what is apparently the same 'communicative' function (supposing that communication takes place between the roots, which are the 'sender' of the signal, and the leaves, which are the 'receiver'). Another hypothesis compatible with such an empirical finding would be that the message is !(leaves go in protection mode), something which we could also try to falsify through various empirical maneuvers.

Of course, these are complex questions and maybe we don't need to posit other pre-illocutionary forces than that of indicatives, which a probabilistic theory of meaning can account for. For instance, Rescorla (2012) argues against Millikan (1995) that we don't need to posit that honeybees send signals with imperative forces. Following Carruthers (2004) and honeybee researchers Menzel and Giurfa (2001), Rescorla argues that honeybees have the cognitive ability to transform a signal with an indicative force ('The nectar is over there') into a representation with an imperative force ('Let us go over there!') thanks to their ability to perform something like practical syllogisms ('There is nectar over there, I want nectar, Therefore let us go over there!').

Nevertheless, even if these researchers are correct about honeybees, as I noted in the preceding chapters, we can give many other candidates for communicative acts with a pre-illocutionary imperative force which don't seem to be so easily eliminable. I will mention three.

First: non-human primates. There is much empirical evidence that chimpanzees make gestures with a commanding/directive force (Leavens et al., 2005; Tomasello, 2008, Chapter 2.3.1). In fact, many researchers discussing attempts to teach sign language to apes (gorillas such as Koko or chimps such as Nim Chimpsky) note that they seem able to learn mostly, and perhaps only, imperative signals. In a review of a wide corporus of videos of chimpanzees who were taught sign language, 3'448 chimp signs were analyzed and the authors concluded that 'Requests for objects and actions were the predominant communicative intentions of the sign utterances, though naming and answering also occurred.' (Rivas, 2005, p.

404). We also saw that it is at least plausible to understand vervet monkey alarm calls to mean something like 'Get down from the canopy!', and not merely 'An eagle is present.' and that this is also the conclusion of several researchers working on animal communication (Andrews, 2015, Chapter 5; Dorit Bar-On, 2017).

Second: babies. It is entirely plausible that certain communicative behaviors performed by babies or young children can have an imperative meaning, and that is so before they can speak or have the mindreading capacities for speaker-meaning (Gómez, 2007). For instance, a hand gesture toward the toy, which looks like an attempt to grab it, paired with an imploring glare directed at the closest adult and a frustrated, impatient vocalization. We can understand this behavior as sending a message close to the following '!(adult gives the toy to baby').

Third: adult emotional expression. Our emotional expressions often seem to involve an imperative component as part of their meaning. We have seen this in the last chapter and discussed how Ekman and Scarantino try to account for it. We considered a photo (taken by Ekman) of a woman with an angry face. We saw that Ekman considered that her facial expression meant that 'she wants the person who provoked her to stop what he/she is doing.' (Ekman, 1997: 316). Scarantino (2017) explicitly interprets such message as having an imperative force. That is: he considers that an angry facial expression may (and perhaps always does) have the function of carrying a message such as '!(the offender/provocateur stops what he/she is doing)', where '!' represents the imperative force. This seems entirely plausible to me. But we have seen that Scarantino's own definition of probabilistic meaning seems incapable to account for the production of such messages.

In these three cases, the imperative force is ascribed based on how the sender behaves and her goals in producing these signals. In other words, the focus is on what the sender is trying to do as opposed to the downstream consequences of the signal on a receiver. For this reason, the counterargument given by Rescorla (according to which the signal has an indicative force but the receiver transforms it to an imperative force) seems not to apply. The three cases are unlike the tree or the bee examples (or even the vervet monkey example) where we cannot ascribe to the sender a definite indicative or imperative goal in sending the signal based solely on the sender's behavior.

In the three cases – primate gestures and vocalizations, baby communicative behaviors, adult emotional expressions – it is plausible that

we find communication without full-blown speaker-meaning, because it is plausible that senders don't produce such signals with the intentions that define speaker-meaning. So if researchers are correct in interpreting these messages as having an imperative force, we are led to posit a pre-illocutionary, non-Gricean, force which the notion of probabilistic meaning cannot account for, but which is readily understandable through that of teleocoded meaning or, more generally, through teleosemantic notions.

8.3. Conclusion

In sum, it seems that when it comes to the code-based meaning of affective signals, the best account at our disposal is given by the notion of teleocoded meaning. Like probabilistic meaning, it is not constrained by the factivity defining natural meaning, but, unlike it, teleocoded meaning avoids the three problems presented in the last chapter: (1) teleocoded meaning is restricted to a reasonable amount of information and avoids predicting absurdly irrelevant meanings; (2) teleocoded meaning is safe from statistical issues, such as what happens in the rare eagle scenario; (3) teleocoded meaning easily makes room for non-indicative pre-illocutionary forces.

Furthermore, teleocoded meaning presents advantages over other teleosemantic notions. First, by being a modest notion – being restricted to encoded meaning – it avoids some of the main problems faced by immodest teleosemantic notions. In particular, it leaves it to Gricean models to account for signals whose meaning seem to not contribute, and to have never contributed, to our fitness. Second, unlike other modest teleosemantic notions, such as Shea, Godfrey-Smith, and Cao's functional content, it can apply to the cases where we need it, such as the rare eagle scenario.[168]

Before I close this chapter, let me nevertheless mention an important drawback of the notion of teleocoded meaning compared to that of probabilistic meaning, one important disadvantage that it shares with other teleosemantic notions. In many cases, the evolutionary processes that would need to be put forward to explain how a signal has acquired its function are not clear. Worst still, evolutionary explanation for the communicative functions of signals may be just-so stories, providing only untestable narratives, pseudo-explanations. That is so even if we can build

[168] Another teleosemantic notion that seems too restricted to apply to all the cases we discussed is Nanay's (2014) since he restricts his proposal to pragmatic representations (representations of action-properties). As such, his teleosemantic account could not apply to a content such as 'An eagle is present'.

a precise mathematical model of a signaling game representing the evolution of a trait. Such 'just-so story' criticism is justifyingly widespread when it comes to evolutionary explanations of mental, behavioral, or cultural features, since one often needs to postulate unknown and unprovable hypothesis when giving such evolutionary explanations, by contrast with those based on DNA or fossil records.[169]

The notion of probabilistic meaning on the other hand, especially its Bayesian version, does not postulate unclear or un-evidenced mechanisms: positing that creatures understand probabilistic meaning only requires that the creature is capable of updating its credence about states of affairs when confronted with new relevant evidence. No need for complicated and hard-to-test evolutionary explanations. In this respect, probabilistic meaning is advantageous: its theoretical posits are much less demanding.[170]

Remember however that the two kinds of explanations – probabilistic and teleocoded – are not exclusive. I illustrated with the process of 'ritualization' how a probabilistic meaning can help establish a teleocoded meaning.

Furthermore, not all evolutionary theories about communication are just-so stories. On the contrary, more and more solid, fruitful, and wide-ranging literature on the evolution of communication is produced every year, both outside and within philosophy. Such literature shows that despite the difficulty of giving evolutionary theories about the function of signals, it is nevertheless possible to make progress in this direction.

And in any case, we have seen that probabilistic meaning seems incapable of accounting for several aspects of affective communication which seem accountable by the notion of teleocoded meaning. We have seen that by focusing on vervet monkey cases, but I believe that it is easy to guess how our discussion extends to many other types of affective signals. I have mentioned for instance non-Gricean facial expressions and could have easily taken examples from non-Gricean vocal, postural, gestural, or behavioral expressions.

Another remark is worth making in this conclusion, a remark which will also serve as a transition to the next chapter. Even though teleocoded meaning is restricted to communicative signs (i.e. signals), it is broad

[169] For an engaging presentation of solid evolutionary evidence see (Coyne, 2010).
[170] Furthermore, the recent progress in associative artificial intelligence or cognitive science based on Bayesian paradigms brings grains to the Bayesian-cognition mill. For a pre-2013 overview, see Clark (2013).

enough to encompass the communication information communicated that is either accessible or not to consciousness. This is an important contrast with Gricean meaning, since what is speaker-meant and allower-meant must be consciously accessible to the person sending the signal (Chapters 1–4). In some cases, the fact that the information communicated is not conscious is quite obvious: with the tree example for instance since I assume that they can send and receive messages without any awareness. Less obviously, we have also seen an example with the vervet monkey eagle call, which may well transmit information about what emotions represent unconsciously (appraisals). I have also mentioned that this is very similar for us humans: it may well be the function of our screams to transmit information about how we appraise an event, even if we don't have conscious access to this information. And the same applies to any non-Gricean emotional expression.

Because teleocoded meaning seems to be the best construct to analyze the code-based meaning of signals, we are led to conclude that the properties of affects inherited by code-based meaning may be conscious and unconscious, as long as it is the function of the signals in question to transmit the information that they do. This is an important difference between code-based and Gricean meaning.[171]

We will consider what unconscious information may be represented by affects in the next chapter, and so what kinds of information may be communicated through teleocoded meaning without awareness.[172] In particular, I will argue that emotions represent evaluative properties unconsciously. Together, the claims of the present chapter and the next thus make it plausible that signals with a code-based meaning transmit representations of evaluative properties even if participants to the communication are unaware they do.

[171] By the way, natural and probabilistic meaning may also be transmitted without conscious access.

[172] Here is a potentially important link between this chapter and the next: If it is a function of emotions to represent evaluative properties, and if it is the function of affective signals to convey affective states, then it may well be the function of affective signals to convey the representation of evaluative properties. For a similar point concerning empathy, see Deonna (2007).

9. Emotions Represent Evaluative Properties (unconsciously)

« Does the body rule the mind or does the mind rule the body? »
S. M. Morrissey, *Still III, Hatful of Hollow*

Abstract. In this chapter, I argue that, if we accept widespread views of emotions (§9.2.1), of representation (§9.2.2), of evaluative properties (§9.2.3), and of consciousness (§9.3.1), then emotions involve a component – the appraisal process – that represents evaluative properties unconsciously, from which we may justifiably say that emotions represent evaluative properties *tout court* (§§9.3.2–9.3.7). This seems to clash with several philosophical theories of emotion which claim that emotions don't represent evaluative properties and to bring support to the theories which claim they do. However, I also argue (§9.3) that since most philosophical theories do not distinguish explicitly between the conscious and unconscious representation of evaluative properties, and since the relation between unconscious and conscious representations in emotion is not straightforward, the matter is more complex than it seems (§9.4.). In particular, the consensus in the affective sciences is in fact compatible with a charitable interpretation of several theories that officially claim that emotions do not represent evaluative properties, notably the attitudinal theory and the non-intentional feeling theory. These theories focus only on *conscious* representation and so we may interpret them as saying that emotions don't represent evaluative properties consciously, although they may do so unconsciously. The distinction between conscious and unconscious representation thus reveals hidden compatibilities between (charitable interpretations of) different philosophical theories of emotions on this question.

Beside my conclusion that emotions do represent evaluative properties (unconsciously), I thus also conclude that the debate surrounding this question must make the distinction between conscious and unconscious representations and render explicit what the relation between conscious and unconscious representations is so that participants in the debate don't risk talking past each other.

In the final section, I offer a sketch of how I think we should think about the relation between the putative unconscious representations of appraisals and the flow of our consciousness during an emotional episode. As we will see, I see it as a relation of partial causal determination. The

causal determination is only partial because the flow of our consciousness is also determined by many other variables, including how our body feels (physiological changes), how we feel poised to act toward the stimulus (action tendency), and how the stimulus appears to us besides how it is appraised.

9.1. INTRODUCTION

Within the philosophy of emotion, the question tackled in this chapter – 'Do emotions represent evaluative properties?' – is a sharply polarizing, contentious issue, as this short list of quotes illustrates:

> « … emotions … represent their object as having specific evaluative properties. » (Tappolet, 2016: 15)

> « [Our] strategy consists in preserving the idea that emotions relate to values while rejecting … that values are represented by the emotions. » (Deonna & Teroni, 2014: 25)

> « … some of the qualities represented by emotional experiences, on the account I am proposing, are evaluative qualities of things in the world. » (Tye, 2008: 47)

> « … we don't think it is accurate to say emotions represent at all. » (Shargel & Prinz, 2018: 110)

Of course, this list could be greatly extended: all the main contemporary emotion theories seem to answer this question either explicitly or implicitly. By 'implicitly', I mean that some theorists do not say in so many words that emotions do or do not represent evaluative properties, but either use different terms or say enough that it is clear what their explicit answer would be. Table 9.1 summarizes the position of seven of the main philosophical theories of emotion (theories which I will present in §9.2.4) and illustrates the lack of consensus.

Yes, emotions do represent evaluative properties (expected or official answers).	No, emotions don't represent evaluative properties (expected or official answers).
(Quasi-)judgmentalism about emotion (Nussbaum, 2001; R. C. Roberts, 2003; Solomon, 1977, 1993)	Attitudinalism about emotion (Deonna & Teroni, 2012, 2014, 2015).
(Quasi-)perceptualism about emotion (De Sousa, 1987; Deonna, 2006; Deonna & Teroni, 2008; Döring, 2007; Goldie, 2002; Helm, 2009; Prinz, 2004; Ratcliffe, 2005; Roberts, 2003; Tappolet, 2000, 2016)	Non-intentionalism about emotion (Shargel, 2015; D. Whiting, 2011).
Motivationalism about emotion (Scarantino, 2014, 2015a)	Enactivism about emotion (Hutto, 2012; Shargel & Prinz, 2018).[173]
Representationalism about emotional experiences (Mendelovici, 2014; Tye, 2008)	

Table 9.1. Do emotions represent evaluative properties? Official or expected answers.[174]

However, as we will see, the 'yes/no' answers summarized in Table 9.1 will differ once we start asking a more precise question, namely: 'Do emotions represent evaluative properties consciously, unconsciously, both, or neither?'. Besides emphasizing the importance of the conscious/unconscious distinction, the main goal of this chapter is to contribute to the debate by focusing on unconscious representation.

The structure of the chapter is as follows. In §9.2, I will detail the idea that emotions represent evaluative properties by characterizing more precisely what is meant by 'emotions' (§9.2.1.), 'representation' (§9.2.2.), and 'evaluative properties' (§9.2.3.), which will lead us to a brief presentation of how the seven emotion theories listed in Table 9.1 answer our question (§9.2.4.).

In §9.3, I will introduce the distinction between A-consciousness and A-unconsciousness (Block, 1995, 2002) (§9.3.1.) and argue that, if we accept the characterizations of emotions, representations, and evaluative properties presented in section 9.2, as well as a certain, broadly accepted,

[173] Here I only discuss one kind of enactive theory of emotion: other kinds seem compatible with the view that emotions represent evaluative properties (unconsciously) (e.g. Colombetti & Thompson, 2007) and which appear to be quite close to attitudinalism by emphasizing how the action tendency component of emotion determines the subjective feeling component and how this makes emotions different from traditional representational states such as beliefs.

[174] Observe that several authors appear on both sides of this table: this is because they have changed their views.

characterization of 'appraisals' in contemporary affective sciences, we must conclude that emotions involve a component which represents evaluative properties unconsciously (§9.3.2–9.3.7). Let me note by the way that whether there can be unconscious emotions will have no bearing on my argument and that I leave it as an open question. This is so because, I will argue, there is an unconscious representational component in emotions whether or not emotions themselves can be unconscious.

In §9.4, I will discuss how the conclusion that emotions do represent evaluative properties (unconsciously) relates to the seven emotion theories of Table 9.2. This will reveal that some of the apparent inconsistent answers of Table 9.1 are in fact not inconsistent if they are charitably interpreted. In particular, the motivational theory's position is about unconscious representation, while the attitudinal and the non-intentional feeling theories' talk about conscious representation. Among the seven theories considered, only the enactivism of Hutto (2012) and Shargel & Prinz (2018) really is incompatible with my conclusion that emotions represent evaluative properties unconsciously.

Another take-home message of this chapter then is that apparently antagonistic positions may be talking past each other. To avoid misunderstandings, emotion theorists must distinguish between conscious and unconscious representations.

9.2. Emotion, representation, evaluative properties, and the debate

In this section, I will detail what is meant by 'emotion' (§9.2.1), 'representation' (§9.2.2), and 'evaluative properties' (§9.2.3) and then briefly present the philosophical debate over the central issue: whether emotions represent evaluative properties (§9.2.4).

9.2.1. Emotion

Even though research on emotions has boomed in the past 40 years, there is no precise definition on which everyone agrees, either within or outside the philosophy of emotions (Gendron & Feldman Barrett, 2009; Sander, 2013; Scarantino & De Sousa, 2018). However, there is a rather broad consensus within the affective sciences (the interdisciplinary endeavor on emotions and other affects) concerning some of the *components* that emotions are paradigmatically made of.

In the section 'Common Ground Among Emotion Theories' of a recent review from the more than well-established *Annual Review of Psychology*, Klaus Scherer and Agnes Moors write:

> « There is also substantial agreement [among emotion theorists] that emotional episodes comprise different components such as [1] appraisal of the situation, [2] action preparation, [3] physiological responses, [4] expressive behavior, and [5] subjective feelings. » (Scherer & Moors, 2019: 721)

I will follow this consensus in calling 'emotions' the physio-psychological episodes which paradigmatically comprise these five components. It is important to highlight that I am not claiming that these five components are *essential* to all emotional episodes. The consensus rather is that *paradigmatic* emotions are made of these five components. So, this consensus is not incompatible with the claim that someone can undergo an emotion while only four of the five emotion components are present. This can be the case for instance when scientists artificially manipulate someone's brain chemically or electrically so that the appraisal process is not present but the other four components are (Izard, 1993).

Let me say a few words about each (I have changed slightly the terminology).

1. *Appraisal of the situation*. This is generally considered to be a fast, automatic, unconscious categorization of whether and how a stimulus relates to our goals, understood broadly to include needs, concerns, values, desires, and more. For instance, if we see a mouse in the kitchen and react with fear or with anger, in both cases we would have somehow categorized the mouse being in the kitchen as an event that is relevant to some of our goals. We somehow consider this event to be significant, to be relevant to our concerns, otherwise we would not have reacted emotionally.[175] I will say more about what appraisals are in a few paragraphs as well as in §§9.3.2–9.3.5.
2. *Action tendency* (also called 'action preparation'). Emotions typically involve a modification in one's action tendencies: in how one will deal with the situation, in how one is prepared to react (including

[175] Some theorists talk of different cognitive levels of appraisals. The most basic would be rapid, automatic, unconscious, nonconceptual (or nonsymbolic) and more prone to error, while less basic ones can be slower, more flexible, partially conscious, and make use of conceptual/symbolic resources (Leventhal & Scherer, 1987; Scherer, 2001). This distinction however is not universally made and most theorists nowadays focus on what can be considered as the 'basic' appraisals, which are fast, automatic, unconscious, prone to errors, and widely spread among the animal kingdom.

reacting by *inaction*, which is linked with sadness). This can mean being ready to flee, to fight, to give in, and more. Such action preparation involves modulations of the level of activity (animation, relaxation, …), the direction of movement (approach, freeze, withdraw, …), kind of adaptation (destroy the stimulus, protect it, protect something from it, …), and more.

3. *Physiological responses*. Action preparation is inherently linked with physiological changes such as modifications in heart rate, blood pressure, skin conductance, digestion, pupil dilatation, and more. These can be measured more easily than action tendencies and constitute an important source of data for empirical studies on emotions.
4. *Expressive reactions* (also called 'expressive behavior' or 'motor reaction'). Another component of emotion that is rather easily measurable and on which much empirical research has been based (especially since Ekman & Friesen, 1971) is the changes in *expression*, whether they are facial, vocal, postural, or from another modality (e.g. shivering or crying). Expression, in this context, is used quite broadly as a change, which can be more or less controlled (unlike physiological changes), that displays one's emotion.[176]
5. *Subjective feelings* (also called 'experienced feeling' or, in philosophy, 'phenomenal character'). This is the experience typical of undergoing an emotion. What it is like to experience this and that fear, anger, or sadness. Subjective feelings might be split up into different components including variations in valence (how pleasant or unpleasant, bad or good an emotion feels) and arousal (how activated or deactivated one feels). The feeling component might also include *awareness* of components 1-4. For instance, in anger, one might feel one's body as poised to react (action preparation) to a goal-obstructive stimulus (appraisal), feel one's hand clench into a fist (expressive reaction), or one's heart rate accelerates (physiological responses).

Some theorists believe human emotions are also paradigmatically made of a sixth component: emotion *labeling* or *categorization* (Barrett, 2006; Schachter & Singer, 1962). This would be an essentially culture-dependent component since emotion labels and categorizations can differ quite drastically from one culture to another (Jackson et al., 2019). This feature however cannot be held to be paradigmatic of emotions in general, because the emotions undergone by cultureless animals or by newborn babies can

[176] This use differs from a more precise, philosophical use such as that defined in Green (2007).

be paradigmatic even though these creatures lack any capacity to label their emotions – think of a terrified monkey desperately screaming or of a four-month-old baby laughing. Although we can accept that this sixth component is paradigmatic of *human adult's* emotions, I will stick to the five components listed above in order not to exclude nonverbal creatures from the discussion.

Fig. 9.1 schematizes the five components discussed.

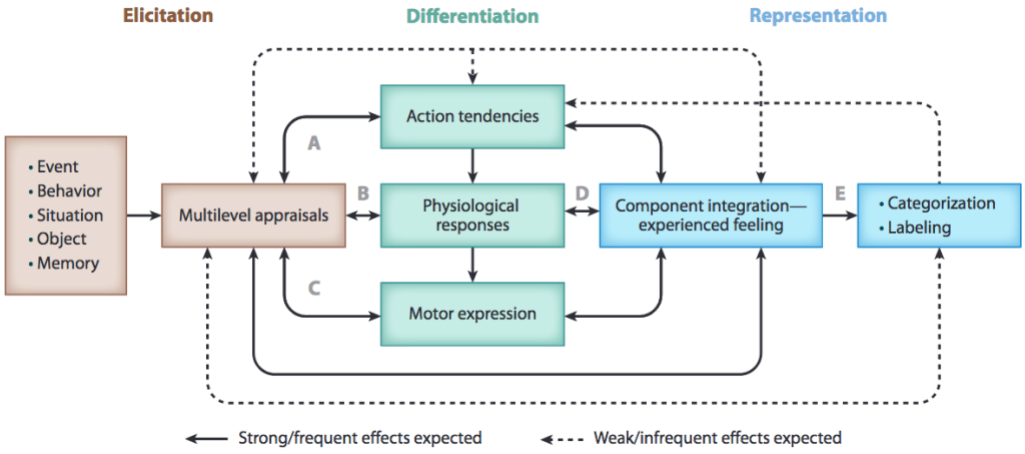

Fig. 9.1. Emotions' paradigmatic components, reproduced from (Scherer & Moors, 2019).

The fact that emotions are comprised of these five components is generally accepted in the affective sciences. Any controversy regards the *relation* between them. For instance, appraisal theorists think that the appraisal of the situation *causally determines* changes in the other components (Moors et al., 2013; Moors & Scherer, 2013; Sander et al., 2018). For others, such as constructivists, appraisals typically are the *result* of some of the other components (Barrett, 2006, 2017). We don't need to decide who is right on this question, nor do we need to decide what precise relations obtain between the other components. As Scherer and Moors put it:

> « Most emotion theorists do not fundamentally disagree about the emotion process as conceptualized in Fig. [9.]1, but they differ in the components on which they focus. » Moors & Scherer (2019: 722)

'Emotion theorists' here include philosophers (who often focus on the subjective feeling component).

Once again: the consensus is not that an emotion is a modification in these five components. This would imply that unconscious emotions don't exist since there couldn't be emotions without the subjective feeling component. The consensus is that emotions *paradigmatically* involve these five

components. This is quite a weak claim but it will be sufficient for our purpose and will allow us to avoid commitment to any particular theory. I will come back to how philosophical theories of emotions differ in the way they define emotions once we have seen more precisely what representation and evaluative properties are.

Before I move on, let me come back to the notion of appraisal and to why there is a consensus among emotion theorists that we need this construct. Appraisals have been posited to explain a simple phenomenon: the absence of one-to-one relations between kinds of stimuli and kinds of emotional responses. This is illustrated by the fact that the same kind of stimuli – e.g. a mouse running in the kitchen – can elicit different kinds of emotional episodes in different individuals or the same individual at different times – e.g. fear vs anger. It is also illustrated by the fact that different kinds of stimuli – e.g. seeing a snake vs seeing a red rose – may elicit the same kind of emotional episode in different individuals, or in the same individual at different times – e.g. aesthetic admiration. So, different kinds of stimuli can elicit the same kind of emotions and the same kind of stimuli can elicit different kinds of emotional responses. The many-to-many relation between kinds of stimuli and kinds of emotional responses cannot be explained by the simple, automatic, 'mechanical' stimulus-response analysis that we can give for other biological phenomena, such as reflexes, white blood cells' behavior, or plant growth. In the latter cases, we can find this simple, mechanistic reaction, where we have the one-to-one relation between kinds of stimuli and kinds of responses, e.g. given a certain position of a functioning leg, if you hit at this place with this force (stimulus) you will have this reaction (response); given this type of white blood cell circulating in one's blood, if it detects this kind of DNA signal (stimulus), it will react in this way (response); given a certain plant in certain soil and temperature context, if you give it this much light and water (stimulus) it will grow, approximately, this much every day (response).

The discrepancy between types of stimuli and types of responses is what led researchers to posit an intermediary categorization between the stimuli and the emotional responses: this is what appraisals are.[177] Two individuals, or the same individuals at different times, may exhibit two different kinds of response (fear vs anger) to the same kind of stimulus (a mouse in the kitchen) because they *categorize* the stimulus differently. This is what explains the different emotions. Similarly, that two individuals, or

[177] For similar reasoning concerning other domains, in particular vision, see Sterelny (1990, p. 21ff).

the same individuals at different times, can undergo the same kind of emotion (aesthetic admiration) with respect to two very different kinds of stimuli (a snake; a rose) is explained by the fact that there is something in common to the way they categorize the stimulus – they categorize it as, say, fascinatingly beautiful. This is represented in Fig. 9.2.

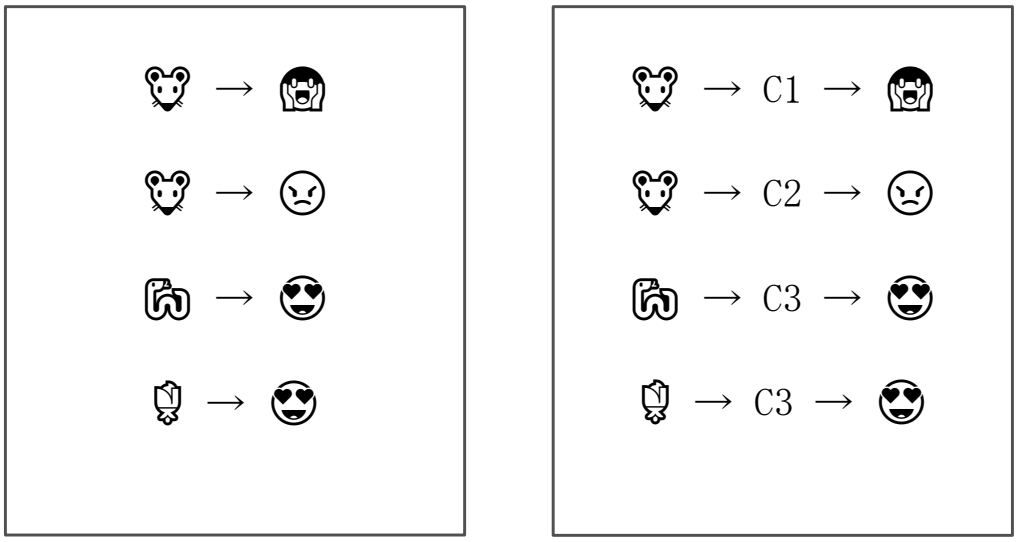

Fig. 9.2. The basic rationale for postulating appraisals: to explain why the same kinds of stimuli may elicit different kinds of emotions and why different stimuli may elicit the same kind of emotions (box on the left), researchers were led to posit intermediary categorizations that are linked in one-to-one relations to emotions (box on the right).

So, appraisals are the kinds of categorizations that have a one-to-one relation to the different kinds of emotional episodes and which thus serve to explain different emotion elicitations. These categorizations can diverge for different individuals even though they are about the same stimuli, because, for instance, these individuals don't have the same beliefs – e.g. one individual believes that mice very often spread dangerous diseases but the other one does not. They may also diverge because different individuals don't have the same goals – e.g. one individual is an employee of a pest removal company but the other is not. Similarly, the same kind of categorization – e.g. these visual details are worthy of the deepest attention – can apply to very different stimuli, which explains why these different stimuli can cause the same emotions – e.g. aesthetic admiration of a rose vs a snake.

9.2.2. Representation

I now turn to what is usually meant by 'representation' and so what it means that emotions represent evaluative properties.

Here is a quote that helps give a first clarification:

> « Many of our mental states are representations: my belief that it is raining outside represents a putative state of affairs: that it is raining outside. If I am afraid of a tiger, this fear is also directed at, or is about, something: a tiger. In other words, many mental states refer to something, they are about something: they have content. » (Nanay, 2015: 153)

Even though not everybody agrees with the examples given by Nanay – for instance, enactivists would surely disagree that the fear in question must represent the tiger – everyone should agree that what is described here is the target phenomenon: representation.

Can we be more precise about what representation is without presupposing an answer to the question discussed in this chapter? I believe so. Take this definition by Dretske (this is based on the notion of 'functional meaning' defined in Dretske (1986)):

> «… a system S represents a property F if and only if S has the function of indicating (providing information about) the F of a certain domain of objects. … A speedometer (S) represents the speed (F) of a car. Its job, its function, is to indicate, provide information (to the driver) about, how fast the car is moving (F). » (Dretske, 1995: 2)

Roughly, 'having a function of indicating' means having the purpose to indicate or being designed to indicate, whether the design is the result of an intentional agent or that of cultural or natural selection. One is free to define 'function' in different ways (e.g. Boorse, 1977; Dretske, 1995, Chapter 1; Millikan, 1984; Nanay, 2010). Indication too may be defined in different ways, e.g. as natural meaning (Grice, 1957), natural information (Dretske, 1981), or probabilistic information (Millikan, 2004; Scarantino, 2015; Scarantino & Piccinini, 2010; Shea, 2007; Skyrms, 2010; Stegmann, 2015). We don't need to be committed to Dretske's way of defining these terms to accept that representing is having the function of indicating (for a discussion of these notions, see Chapter 7). As such, this notion of representation may well avoid the typical problems faced by teleosemantics (e.g. the swampman and the functional indeterminacy) by defining 'function' differently than Dretske does. We may well interpret the definition given by Dretske as corresponding to claims made by theories that are incompatible with teleosemantics, but that are nevertheless broadly functionalist, such as the Asymmetric Dependency Theory (Fodor, 1987) or the Conceptual Role Semantics framework (Block, 1998) (e.g. by

cashing out 'function' in terms of the semantic and syntactic roles that the indicative signs play in the cognitive life of the agent).

Let me give a couple more examples so that we have a better idea of what this definition entails. If one thinks that chromatic perception represents colors, then, according to this definition, one should agree that the function of chromatic perception is to provide information about the colors of certain objects (e.g. the ones I look at from a certain distance). If one denies that chromatic perception represents colors, then one should disagree that chromatic perception has this function. A common hypothesis is that pitch perception represents a certain range of frequencies (for humans, about 20 to 20'000 hertz) with which the surrounding air vibrates (or with which water vibrates if one is underwater, etc.). If one agrees with this hypothesis and Dretske's definition, then one should agree that the function of pitch perception is to indicate, provide information about, these vibrations. If one thinks that the function of pitch perception is to indicate *pitch height* and that this is not identical with the frequencies with which air vibrates, then one might well accept Dretske's definition and thus disagree that pitch perception represents vibrations.[178]

So, philosophers can accept Dretske's definition – that a system S represents a property F if and only if S has the function of indicating the F of a certain domain of objects – while disagreeing on whether or not emotions represent evaluative properties because they can accept the definition while denying that emotions have *the function of indicating evaluative properties*. This could be the case even if they agree that emotions might be said *to correlate* with evaluative properties. So, it seems to me, all participants in the debate on whether emotions represent evaluative properties can agree with Dretske's definition of representation, as long as we leave enough room for different definitions of 'function', of 'indication', 'properties', etc.[179]

[178] Obviously then, accepting Dretske's definition does not force one to accept his Representationalism, i.e. the thesis that all phenomenal properties are representational properties. For instance, we can accept the definition and reject that, say, the phenomenal character of hearing a sound with a certain pitch represents the frequencies of the vibrations of the medium (air, water, etc.) surrounding the person. We may say that the phenomenal character in question (e.g. a high pitch) may correlate with certain physical states of affairs (e.g. the air vibrates at 15'000 Hz) while denying that it is *the function* of the phenomenal character in question to indicate something about the world.

[179] Some philosophers working on emotion follow Searle (1983) in distinguishing 'representation' from 'presentation'. Thus, some might say that emotions *present* the world as possessing evaluative properties or present it in an evaluative light, instead of saying that emotions represent evaluative properties. However, as Searle makes it clear, what he calls 'presentation' is a subset of what he calls 'representation' (1983: 45-6). For instance, beliefs require concepts while perception, according to him, doesn't. Emotions

Let me observe by the way that a main advantage of Dretske's definition is that it very naturally allows for misrepresentation. If something has a function, it can also malfunction, so a definition of representation in terms of function generates a definition of misrepresentation.

Let me also observe that Dretske's definition does not presuppose that appraisals are representations: appraisals may be categorizations that have a one-to-one link with kinds of emotions without having the function of indicating anything (although, as we will see, this is very implausible).

9.2.3. Evaluative properties

Now, what are evaluative properties? I will not give a bi-conditional definition, but only say that if object O possesses (or we apprehend O as possessing) evaluative properties (relative to subject S), then O is positive or negative morally, hedonically, aesthetically, epistemically, cognitively, politically, instrumentally, socially, prudentially, or in another axiological way (relative to S). So, evaluative properties are properties that are good or bad in some general way. By 'general way' I mean that we allow 'good/bad' to designate either 'good/bad simpliciter' or 'good/bad for O' where O is either some individual or a kind (for these distinctions, see Schroeder, 2016).

Here is a small list of putative evaluative properties:

> Absurd, admirable, annoying, beautiful, caring, charming, civilized, clever, comic, cool, courageous, cowardly, cute, dangerous, delicious, despairing, disgusting, disturbing, dreadful, egoist, elegant, fair, foolish, gentle, graceful, groovy, guilty, harmonious, honest, hopeless, horrifying, hypocritical, ignoble, inappropriate, incapable, incongruous, ingenious, inspiring, intelligent, ironic, lazy, majestic, melancholic, nice, noble, offensive, picturesque, pleasant, polite, rational, reckless, relevant, repugnant, rigorous, satiric, sexy, shameful, smart, solemn, splendid, stupid, sublime, sumptuous, superb, tragic, ugly, unjust, unjustified, useful, vile, ...

I will remain entirely neutral concerning the ontological nature of evaluative properties. Whether evaluative properties should be accounted for by subjectivism, cultural relativism, error-theory, realism, fitting-attitude analysis, or still another theory remains open.

may not require concept and many philosophers thus say that emotion 'present' the world as evaluatively loaded. However, if we follow Searle in this distinction, this is just a way of saying that emotions represent evaluative properties in a special way.

By the way, let me announce here already that I consider that the best candidates for evaluative properties represented by emotions are not the usual garden-variety evaluative properties usually discussed by philosophers – such as loss for sadness, slight for anger, danger for fear – but are more fine-grained, less common properties – such as being unconducive to one's safety (which *is* bad for an organism and so is an evaluative property). We will come back toward the end, in §9.3.7.

Let us take stock. We have seen that emotions are considered as episodes paradigmatically constituted of (at least) five components: (1) appraisal of the situation, (2) action tendency, (3) physiological response, (4) expressive reaction, and (5) subjective feeling. We have seen that 'representing' is to be understood as having the function of indicating. Finally, evaluative properties, roughly, are positive or negative properties.

9.2.4. THE DEBATE: HOW THE MAIN PHILOSOPHICAL THEORIES OF EMOTION ANSWER OUR QUESTION

We have now sufficiently clarified our initial question – 'Do emotions represent evaluative properties?' – to introduce the philosophical debate around it. To do so, I will briefly present seven philosophical theories of emotion and how they answer the question or are taken to answer the question.

Our first philosophical theory of emotions is (quasi-)judgmentalism. This theory claims that emotions are (quasi-)judgments that certain evaluative properties are instantiated (Nussbaum, 2004; Solomon, 1977, 1993). For instance, sadness is a (quasi-)judgment that something constitutes a loss. Anger is a (quasi-)judgment that someone has offended (has committed a slight toward) the emoter. Even though, as far as I know, Nussbaum and Solomon have not explicitly answered our question, it is clear that, given the definitions above, they would say that emotions do represent evaluative properties, because they would surely agree that judgments represent their objects, e.g. a judgment that it rains represent the weather as rainy.

The second philosophical theory is (quasi-)perceptual theory: emotions are (quasi-)perceptions of evaluative properties (De Sousa, 1987; Deonna, 2006; Deonna & Teroni, 2008; Döring, 2007; Goldie, 2002; Helm, 2009; Prinz, 2004; Ratcliffe, 2005; R. C. Roberts, 2003; Tappolet, 2000, 2016).[180]

[180] I include what Scarantino and de Sousa call 'evaluative feeling' and the 'patterns of salience' theories as quasi-perceptualist theories because of their affinities with the perceptualist theory, affinities discussed by Scarantino and de Sousa (2019) and mentioned in chapter 5.

Sadness is a (quasi-)perception of a loss. Anger is a (quasi-)perception of an offense (or of a slight). (Quasi-)perceptualists usually explicitly defend the claim that emotions represent evaluative properties, even though they might not agree on how exactly this is so. For instance, Prinz (2004) has a teleosemantic theory of how emotions represent evaluative properties, while Tappolet (2000, 2016) pursues an account of how emotions represent evaluative properties based on phenomenology.

Our third theory is motivationalism: emotions are prioritizing action control programs guiding our actions in response to detected evaluative properties (Scarantino 2014). Sadness is an action control program that prioritizes certain (in)actions, such as giving in rather than fighting – a response motivated by the appraisal of the situation as constituting a loss. Anger is another type of motivated response, an action control program that prioritizes certain actions, such as fighting, to deal with a detected slight or offense. Scarantino states that emotions represent evaluative properties through their appraisal process, and espouses a type of explanation that mixes psychological theory (especially Frijda, 1986) and a teleosemantic framework close to Dretske (1986) and Prinz (2004).

Fourth, representationalism: this is a theory about emotional experiences rather than emotions as a whole. Its *explanandum* is the subjective feeling component of emotions and not the other components, but, for our question, its perspective is interesting to consider. Representationalism says that emotional experiences consist of, or supervene on, representations of situations involving evaluative properties to which one is (bodily) reacting (Tye 2008, Mendelovici 2014). So the experience of sadness consists of, or supervenes on, a representation of the situation as involving a loss, a loss to which one is bodily reacting. Described as such, it may be compatible with, and complementary to, at least, perceptualism and motivationalism.

Fifth, attitudinalism: emotions are felt evaluative bodily attitudes (Deonna & Teroni 2012, 2014, 2015). According to this theory, the representational content of emotions, i.e. the propositional or sub-propositional content of the emotional attitudes,[181] does *not* involve evaluative properties. Deonna and Teroni claim that, consequently, emotions do not need to represent evaluative properties as part of their content, just like beliefs wouldn't need to represent truth as part of their content and desires wouldn't need

[181] They use 'content' as what is 'inside' the psychological mode/attitude, as opposed to the broader notion of content (c.f. e.g. Tye (2008)) which does not differentiate the mode/attitude and its inside content but rather corresponds to what Recanati (Recanati, 2007, p. 21) calls 'the complete truth-conditional content', which inherits properties not only from the (sub-)propositional object but also from the type of mode/attitude.

to represent desirability as part of their content. A creature may believe something without possessing the concept of 'truth'. Similarly, they argue, a creature may be afraid of something even if it cannot represent what it is afraid of as dangerous. However, emotions are essentially evaluative attitudes insofar as they are the kinds of attitudes that are appropriate only if they are about something that instantiates the relevant evaluative property, just like beliefs are appropriate only if their content is true. So sadness, for instance, is a felt bodily attitude whose representational content consists of, say, the death of someone dear, an attitude that it is appropriate to have only if this death constitutes a loss.

Sixth, non-intentionalism: emotions are non-intentional feelings, not different from moods (Whiting 2011, Shargel 2015). Because they are non-intentional, emotions are not about anything and thus cannot represent anything. They are mere feelings. The impression that they are about something comes from their causes, which could be evaluative judgments for instance, but their causes are not the emotions. Observe that this theory appears to be in contradiction with the consensus in the affective sciences according to which emotions are paradigmatically made of four other components besides subjective feelings.

Seventh, enactive theory: emotions create action possibilities through their bodily changes which in turn create non-static, state-dependent, motivating affordances (Hutto 2012, Shargel and Prinz 2018, for a discussion, see Hufendiek, 2018). According to enactivism, emotions do not represent evaluative properties because they do not represent at all. For Shargel and Prinz (2018), this is mostly because emotions *create* positive or negative affordances rather than target or indicate pre-existing evaluative states-of-affairs. These evaluative affordances are not classical affordances as Gibson (1977) describes them because they are (a) inherently motivating and, more importantly, they are (b) state-dependent, which means that they come into existence when the token emotional episode starts and disappear when the emotional episode stops. As such, emotions do not represent a pre-given feature of the world, nor do they represent a response-dependent property (since response-dependent property can exist at a time t even if nobody has the relevant occurring response at t, so that if red is a response-dependent property, something stays red even if everybody closes their eyes).

This short presentation of the seven theories is based on what their authors claim or on what I hope are reasonable interpretations. Taking what they say at face value leads us to classify them into two polarized camps, as represented in Table 9.1. The first four theories officially are 'yays': (quasi-

)judgmentalism, (quasi-)perceptualism, motivationalism, and representationalism explicitly or implicitly claim that emotions do represent evaluative properties. The last three officially are 'nays': attitudinalism, non-intentionalism, and enactivism explicitly or implicitly claim that emotions do not represent evaluative properties.

As we will see, the picture yielded by these 'yays' and 'nays' will get much blurrier, much less polarized, much less clear-cut, once we start asking more precise questions.

9.3. How emotions unconsciously represent evaluative properties

In this section, after having clarified what I mean by consciousness, I will argue that emotions represent evaluative properties unconsciously based on the definitions given above and evidence from affective sciences. I will further argue that this should lead us to accept the claim that emotions represent evaluative properties *tout court*.

To make this argument, I will focus mainly on the psychological literature. This is so for three reasons: first, it will allow me to make the argument without presupposing any philosophical theory, which will then allow a discussion of philosophical theories from a more neutral standpoint. Second, relatedly, philosophers usually want their theory to be compatible with empirical results and with the best theories from the empirical sciences, so discussing the best psychological theories around should allow us to propose constraints for philosophical theories. Third, as I will argue, in the last decades, psychological theories of emotions have given detailed accounts of the appraisal process and, I will argue, the appraisal process should be considered a representation of evaluative properties. Of course, philosophical theories have also given very detailed accounts of the appraisal process, if we understand this notion as broadly as I have introduced it above (to take just two examples, I will mention Robert Godron's *The Structure of Emotions* where a sophisticated wish-belief account of the appraisal process is developed and Robert Roberts' *Emotions* where the concern-based notion of construals is applied to numerous kinds of emotions).

In section 9.4, we will see the implications of these conclusions for the seven philosophical theories of emotion we have encountered. To anticipate a little, since emotions might (a) represent evaluative properties unconsciously, but not consciously, or (b) represent them consciously, but not unconsciously, there can be pairs of theories which (appear to)

contradict one another in answering the unqualified question with which we started – i.e. 'Do emotions represent evaluative properties?' – but which don't contradict each other anymore once we ask the disjunctive question: 'Do emotions represent evaluative properties consciously, unconsciously, both, or not at all?'.

9.3.1. A-CONSCIOUSNESS AND P-CONSCIOUSNESS

According to Block (1995, 2002), when we talk of consciousness, we can mean two different things: phenomenal consciousness (P-consciousness) or access conscious (A-consciousness). To make my argument, it will be useful to concentrate on what Black calls A-consciousness. However, my argument can be extended to P-consciousness, and I will come back to this toward the end.

Here is how Block describes P-consciousness:

> « P-consciousness is experience. P-conscious properties are experiential properties. P-conscious states are experiential states; that is, a state is P-conscious just in case it has experiential properties. The totality of the experiential properties of a state are "what it is like" to have it. Moving from synonyms to examples, we have P-conscious states when we see, hear, smell, taste and have pains. P-conscious properties include the experiential properties of sensations, feelings and perceptions, but I would also include thoughts, wants and emotions. » (Block 2002: 206)

And here is how he describes A-consciousness:

> « A-consciousness is access-consciousness. A representation is A-conscious if it is broadcast for free use in reasoning and for direct 'rational' control of action (including reporting). … I see A-consciousness as a cluster concept in which reportability is the element of the cluster that has the smallest weight even though it is often the best practical guide to A-consciousness. » (Block 2002: 208).

Commenting on this passage, Block adds:

> « The 'rational' is meant to rule out the kind of automatic control that obtains in blindsight [e.g.] the thirsty blindsight patient would not reach for a glass of water in the blind field. » (2002: 208).[182]

[182] So 'rational' must refer to instrumental rationality, i.e. the capacity to adopt suitable means to one's end.

Let me elaborate on this example. Blindsight patients possess a blind spot in their visual field, a 'hole', and thus cannot consciously see objects situated in this hole. So, for instance, if there is a glass on the table in front of them but that the glass is in this hole, if you ask them what is on the table, they will report not seeing anything. This is why the thirsty blindsight patient won't reach for the glass of water. However, surprisingly, if they are forced to guess what is in front of them, they will guess correctly above chance level. They will think that they are just randomly guessing, but in fact, the forced-choice tasks show that they somehow do have unconscious access to objects in their blind spot.

In Block's terminology, blindsight patients are not A-conscious of the objects situated in the holes of their visual fields, but they nevertheless have A-unconscious representations of at least some properties of these objects. This explains why they won't reach for the glass but can nevertheless guess above chance level what is in front of them.

In the following, I will concentrate on A-consciousness as opposed to P-consciousness. Let me observe that I don't thereby commit to any particular theory of consciousness. Maybe A-consciousness is distinct from P-consciousness as Block (1995, 2002) argues. But maybe he is mistaken and the two concepts refer to exactly the same phenomena, as e.g. Tye (2000) argues. I am not claiming that A-consciousness must be different from P-consciousness. Furthermore, maybe A-consciousness is a higher-order representation (Armstrong, 1981; Carruthers, 2004; Lycan, 1996; Rosenthal, 1986), but maybe it is not (Dretske, 1995; Prinz, 2011). We need not commit to one view of consciousness here.

What is important is that we can speak of A-unconscious representations. These are the representations that are not freely available for use in reasoning or for the rational control of action, representations that are not accessible in a direct, non-inferential way.

Observe that such A-unconscious representations are in principle accessible through the methods of cognitive sciences. Take for instance how we localize sound sources. When we hear a sound, we can normally perceive whether it comes from our left or our right. We notably do so thanks to so-called 'Interaural Time Delay', i.e. the difference in time that the sound waves take to reach our right or left ear. If the sound comes from the left, it will reach our left ear first. Our brain computes the time difference to calculate the position of the sound source. According to Woodworth's Formula for Interaural Time Delay (see Fig. 9.3), the computation is the following: $\tau = r\,(\theta + \sin\theta)/C$, where τ is the time difference of sound

detection between the two ears, r is the radius of the head, θ is the bearing of the source, and C is the speed of sound.

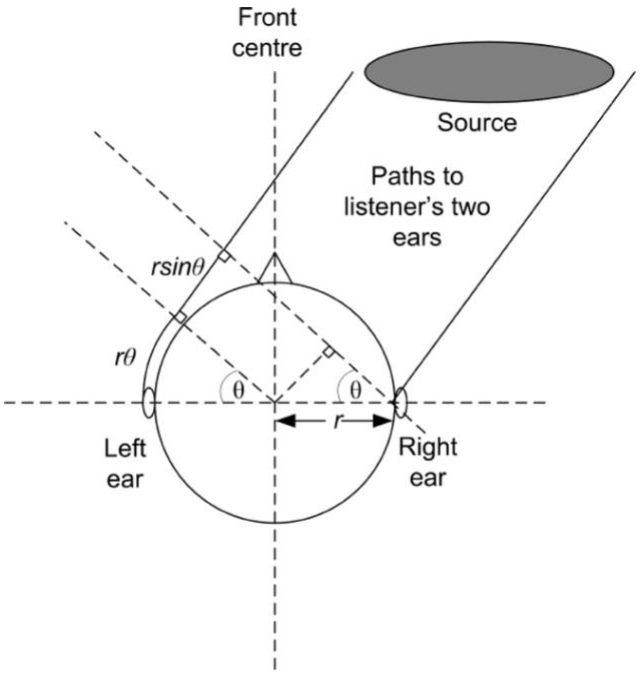

Fig. 9.3. Woodworth's Formula for Interaural Time Delay ($\tau = r(\theta + \sin\theta)/C$). A typical example of A-unconscious representations.

If Woodworth's Formula for Interaural Time Delay approximates what happens in our brain, then we have a solid reason to postulate representations that are A-unconscious, i.e. all the variables of the formula. We cannot access such representations directly, by introspection, in a non-inferential way. What we can access directly is whether a sound source appears to be more or less on the left or right-hand side. Some scientific exploration must be achieved to know *how* our mind determines the location of a sound source. This endeavor consists in data collection, model building, deductive, inductive, abductive inferences, and more. That is not the direct, non-inferential way in which we access the representations of which we are conscious, such as my representation that it is Friday or that what I see is a computer.

In the following, if I talk of (un)consciousness without qualification, I will refer to A-(un)consciousness.

9.3.2. Unconscious processes in emotions

Before we consider whether emotions represent evaluative properties consciously or unconsciously, let me observe that, if one accepts the consensus in affective sciences according to which emotions are

paradigmatically made of (at least) the five components I have presented above, then one must accept that emotions are paradigmatically composed of at least some unconscious processes.

Here, again, are the five components typical of emotions: (1) appraisal, (2) action tendencies, (3) physiological responses, (4) expressive reactions, and (5) subjective feelings. Even if we consider that (5) is a necessary component of emotions and that it is necessarily conscious, emotions would nevertheless partially consist of four components that can be unconscious.

Now, it is quite uncontroversial that (3) physiological responses can be unconscious. For instance, we don't have direct access to whether or how much our adrenal glands secret adrenaline. It should also be uncontroversial that (4) expressive reactions can be unconscious. For instance, when I smile, I do not have conscious access to whether I am displaying a Duchenne or a non-Duchenne smile. The main difference between these two facial expressions consists in an activation of zygomatic muscles, which I cannot control or monitor consciously. So, there are at least two components of emotional episodes that can be unconscious. But our question is whether there can be unconscious representations of evaluative properties. Physiological responses such as adrenaline secretion and expressive reactions such as Duchenne smiles are not great candidates for representations of evaluative properties. Although one would have strong arguments that these bodily changes correlate with the subject's apprehension of a stimulus as possessing evaluative properties, there are no strong arguments that these changes have the *function* of indicating that the stimulus possesses these evaluative properties. The function of expressive reactions rather is to communicate one's affects and the function of physiological changes such as hormones secretion rather seems to be to prepare the reaction to a stimulus that is already apprehended as possessing evaluative properties.

The emotional component that is our best candidate for the unconscious representation of evaluative properties is the appraisal process. We saw that the consensus in affective sciences was that they are proper components of emotions. So if (a) they are (or can be) unconscious, and (b) they represent evaluative properties, then we must accept that a component of emotions represents evaluative properties. We will first see that the best contemporary theories lead us to consider them as unconscious, that they are representations, and finally that they represent evaluative properties.

9.3.3. Appraisals as unconscious: Theoretical evidence

Remember the way I introduced appraisals above: they were postulated to explain how diverse stimuli may elicit the same emotion and how different emotions may be elicited by the same stimuli. For instance, the mouse in the kitchen may elicit fear in one person and anger in another. This is explained by the fact that they categorize this stimulus differently; they appraise it differently. One may appraise it as something dangerous, because, say, this person believes that mice often transmit dangerous diseases. The other may appraise it as a non-dangerous but undesirable intruder that must be chased at once. Or take an even more banal case: if the Lakers win and the Blazers lose, I am happy and Damian is sad, but if the Lakers lose and the Blazers win, Damian is happy and I am sad. I appraise the Lakers' victory as being positive and their defeat as an unfortunate event, while Damian appraises these two events the other way around.

In the early 20th century, theorists postulating such intermediary categorizations between stimuli and emotions saw them as rather simple representations that are accessible through consciousness. Typically, they saw these representations as evaluative judgments.[183]

Today, however, the view that the appraisal process is entirely accessible through consciousness is largely, if not wholly, abandoned.[184] One reason for this is that the appraisals posited by contemporary theories just do not correspond to the cognitive processes to which we have direct conscious access during emotional episodes. They do not correspond to what the flow of our consciousness is like, nor to what we can access through introspection if we turn our attention 'inward' when we undergo an emotion.

It is worth giving a partial presentation of some theories which have tried to give a detailed analysis of what the appraisal process consists in, so that we can compare what they postulate in addition to the subjective feelings undergone in emotional episodes, and thereby illustrate why the appraisal process, as is posited by some of the best theories there are today, is not

[183] This was the case in particular in the work of Brentano's students Alexus Meinong and Carl Stumpf, but also in that of the first psychologist defending appraisal theorist, Magda Arnold (1960), who was indirectly influenced by the Brentanian school (for a history of early appraisal theory, see Reisenzein, 2006).
[184] However, as I will discuss below, especially in the conclusion, the appraisal process may nonetheless be partially reflected into consciousness.

conscious. My argument will be focused on A-unconsciousness, but we can extend it to P-consciousness, given certain assumptions, as we will see.

Reviewing the literature on the postulated appraisals involved in emotions, Sander, Grandjean, and Scherer state that the following are widely agreed upon among appraisal theorists:

> « How relevant is this event for me?; Does it directly affect my social reference group or me? (goal relevance); What are the implications or consequences of this event and how do these affect my well-being and my immediate or long-term goals? (goal congruence); Did I expect this event and its consequences and how certain are they (novelty, expectation, certainty); Who caused this event, am I responsible or someone else? (agency, causation); How well can I cope with or adjust to these consequences? (coping potential, control, power). » (Sander et al., 2018, p. 226)

This list makes it quite obvious that these appraisals were not hypothesized on the basis of what is available through introspection or, more generally, what is A-conscious, nor were they hypothesized on the basis of what it feels like to undergo an emotion (P-consciousness). These categorizations do not, and are not supposed to, correspond to what goes on in the flow of our consciousness when we undergo an emotion (P-consciousness), nor to anything accessible through our consciousness (A-consciousness).

Sander et al. list the most widely recognized appraisals, but some psychologists argue that this list is not sufficient, because other variables must be taken into account to predict the whole range of possible emotional reactions, i.e. to predict what type of appraisals (what type of representations) correspond to or cause a given type of emotions. According to Sander et al. (2005), we need to know the values assigned for 16 appraisal dimensions to make accurate predictions as to the kind of emotional episode that will be elicited:

> Suddenness; Familiarity; Predictability; Intrinsic pleasantness; Goal/need relevance; Cause: agent; Cause: motive; Outcome probability; Discrepancy from expectation; Conduciveness; Urgency; Control; Power; Adjustment; Normative significance: internal standards; Normative significance: external standards.

Once again, it is quite obvious that this list is not determined by what is accessible through consciousness during an emotional episode, nor by how it feels to undergo an emotion. Rather, it is based on theoretical hypotheses

that are empirically tested through various experimental methods, including behavioral and (neuro-)physiological experiments (see for instance Scherer & Meuleman, 2013; Skerry & Saxe, 2015; Smith & Lane, 2015). This list of appraisals is supposed to correspond to representations that we would need to postulate in order to predict emotions based on appraisal values, not to the A-conscious or P-conscious mental states.

To illustrate further, according to them, an episode of rage (hot anger) will be caused by an appraisal process that yields the following values.

> High suddenness; low familiarity; low predictability; high goal relevance; other agent causation; intentional causation; very high outcome probability; dissonant discrepancy from expectation; obstructive; high urgency; high stimulus controllability; high self-ascribed power; high self-ascribed adjustment; low external standards.

An episode of sadness on the other hand will be elicited by the following appraisal values:

> Low suddenness; low familiarity; high goal relevance; causation by chance; very high outcome probability; obstructive; low urgency; very low stimulus controllability; very low self-ascribed power; medium self-ascribed adjustment.

According to Sander et al (2005), these appraisal values, which are computed unconsciously, initiate other unconscious components: action tendencies, physiological changes, and expressive reactions. Ultimately, these four kinds of unconscious modifications in the organism may result in a subjective feeling, which possesses a certain what-it-is-likeness (P-consciousness) which may very well be accessed consciously (A-consciousness).[185]

For instance, if a monkey appraises a stimulus as very obstructive to her goals, as very sudden and very urgent, and appraises her own power over the stimulus as very low, this can result in an attempt to get away from the stimulus (action tendency). To achieve this action tendency, more blood will need to circulate (physiological changes). Furthermore, the appraisal can cause the monkey to scream, an expressive reaction whose function is to alert group members that the monkey is in a difficult, urgent situation, and/or that they should help control the stimulus which the monkey cannot

[185] The feeling is at least accessible even if it is sometimes not accessed, i.e. is not part of the flow of our consciousness. It may be accessible without being accessed because, for instance, our attention is captured by something else.

control by herself. All these changes and more may reach consciousness and constitute the feeling of fear. Importantly, it may well be the case that the monkey is not conscious of her own fear before she reacts with fear, before she screams, before her heart rates accelerate, etc.

My point then is that the appraisals postulated by emotion theorists are like the variables in Woodworth's Formula for Interaural Time Delay: they are completely and always A-unconscious and, plausibly, are also completely P-unconscious. I say 'plausibly' because we have no ways of knowing whether mental states that are always A-unconscious possess P-consciousness or not, since we have no conscious access to them and that conscious access is the only way we have of knowing whether something possesses a phenomenal character, i.e. a P-conscious property. But unless there is a reason to postulate that some mental states which are always A-unconscious nevertheless are P-conscious and that appraisals belong to this category,[186] we should follow the economical hypothesis and assume that the appraisals postulated by recent emotion theories do not possess P-conscious properties.

I have used concepts coming from so-called appraisal theorists (e.g. Scherer, Sander, Grandjean), to illustrate what appraisals are because these theorists have given the most detailed accounts of appraisals. But not all emotion theorists agree with the latter. I claimed above that the existence of appraisals, as I introduced them, was a consensus among very different emotion theories, both in psychology (e.g. in psychological constructivism or basic emotion theory) as well as in philosophy (although many philosophers do not focus on this aspect of emotion). Let me thus observe that, to make the point that I am making here, I could have used other theoretical frameworks.

In Russell's psychological constructivist's theory, we find that the appraisal process is causally antecedent to the core affects; the latter being what is consciously perceived, the appraisal is thus considered as an antecedent to what can be consciously perceived. Similarly, so-called basic emotion theorists consider the appraisal process as 'immediate, unbidden, opaque, *unconscious*, and automatic' (Matsumoto & Ekman, 2009, p. 107, my italics). If we look at the more recent Integrated Theory of Emotional Behavior proposed by Moors (2017), we also find that the appraisals are

[186] One such reason would be that the program known as 'Phenomenal Intentionality' – a program which tries to ground intentionality in phenomenal characters – is so successful that it forces us to accept that, because appraisals possess an intentional content, they must possess a phenomenal character.

antecedent to consciousness and are A-unconscious (and most probably P-unconscious).

In sum, the best recent theories that try to determine precisely what appraisals are consider them a kind of categorization that is not consciously and non-inferentially accessible. Nor are they thought as mental states which possess a what-it-is-likeness, as possessing P-conscious properties. We find out about them not by experiencing them, but through experiments, data collection, model building, inferences, etc. Appraisals, contrary to the subjective feeling component of emotions, are best conceptualized as A-unconscious and most probably P-unconscious.

Let me note however that emotion theorists working on the appraisal process usually consider that, despite being unconscious by default, appraisals may be partially reflected into, or surface in, consciousness (Frijda, 2007, Chapter 8; Grandjean et al., 2008; Lambie & Marcel, 2002; Moors, 2017; Sander et al., 2018; Scherer & Meuleman, 2013). What 'reflected into' or 'surface in' means exactly depends on the different theory and it is a contentious topic.[187] However, one thing that appears to be quite widely agreed upon by these researchers is that the conscious component of emotion (the subjective feeling) is at least partially determined by the four other emotional components: the action tendency, the physiological changes, the motor reaction and, indeed, the appraisal process. In the conclusion, I will come back to the relation between consciousness and the appraisal process and present the account that I find most plausible.

9.3.4. Appraisals as unconscious: Experimental evidence

Besides the fact that the most sophisticated accounts of appraisals conceptualize them as unconscious mental states, several experiments support this idea.

In one influential study (Öhman & Soares, 1994), negative moods were elicited through subliminal pictures of snakes and spiders. Pictures of snakes and spiders as well as of flowers and mushrooms, were projected for a time too brief to be consciously accessible (20 and 30 milliseconds).

[187] The way we may want to think about how the unconscious components of emotions can be 'reflected into' consciousness would be to take it as similar to the fact that Woodworth's Formula for Interaural Time Delay is postulated as an unconscious mental calculation that is nevertheless partially 'reflected into' consciousness through our ability to consciously hear where a sound source is situated. Or, to take another example, how the syntactic trees postulated by linguistics are unconscious representations that may be partially reflected into consciousness through our ability to assess whether a sentence is syntactically well-formed or not, e.g. by consciously hearing or feeling grammatical mistakes.

Subjects couldn't guess above chance which pictures they saw when asked in a forced-choice task. Nevertheless, subjects who unconsciously saw pictures of snakes or spiders reported feelings that were, overall, significantly more negative, more aroused, and less dominant – feelings which are paradigmatically associated with fear – than subjects presented with pictures of mushrooms or flowers. This strongly indicates that even though subjects in the subliminal trials didn't have conscious access to what they were seeing they nevertheless appraised the stimulus unconsciously, an appraisal causally linked to the unconscious affective reactions that resulted in the feelings in question.

Another kind of experiment shows that subliminal exposure to pictures can influence action tendencies without influencing subjective feelings. This indicates that appraisals may influence behavior without being reflected in consciousness. In one of these experiments, participants were subliminally exposed to happy, neutral, and angry facial expressions (Winkielman & Berridge, 2004). Some participants were asked to rate their feelings (through negative/positive and high arousal/low arousal scales as well as through multi-item scales asking about specific emotions, such as contentment or irritation). Some other participants were asked to taste a beverage and evaluate it. Participants' feelings were not influenced by the subliminal exposure. However, participants' consumption and rating of the beverage were. They drank more after being exposed to happy faces than to angry faces and they were willing to pay twice as much for the drink after happy than after angry expressions.

This is how I interpret these results: participants were A-unconsciously appraising the situation as more or less positive depending on whether they were subliminally exposed to positive stimuli (happy faces) or negative stimuli (angry faces). This is why they were acting more positively toward the beverage when shown happy faces and more negatively when shown angry faces.[188]

Other kinds of experiments that tend to show that there can be unconscious appraisals involve blindsight subjects. For instance, De Gelder et al. (2005)

[188] Note by the way that my interpretation of the results of such experiments doesn't commit me to the strong claim that all action tendencies are the result of, or always happen with, an unconscious affective episode. The experiment involves positive and negative facial stimuli and corresponding reactions that reveal a positive or negative evaluation: these stimuli and the resulting action tendencies are typically linked with affective episodes. Action tendencies that are not noticeably linked to affects, e.g. acting on the basis of a cold deliberation about a neutral topic, may without any problem be considered as resulting from non-affective episodes. However, see Nanay (2017) for the claim that no actions are completely non-emotional.

displayed pictures of faces with emotional expressions to a blindsight subject. The subject showed no sign of having consciously perceived the faces, but the exposure to these faces nevertheless influenced subsequent emotional tasks. Similarly, Hamm et al., (2003) conditioned a blindsight subject to have aversive responses to pictures of airplanes. Once presented with airplane pictures in his blind spot, he displayed such aversive responses even though he reported no conscious awareness of the airplane pictures.

One can easily find many other (more recent) experiments showing how stimuli that are not consciously accessible can influence affective states, and in particular how they modify action tendencies, physiological changes, and neuronal processes associated with emotions (e.g. R. L. Blakemore et al., 2017; Mumenthaler & Sander, 2015, 2019; Peláez et al., 2016; Shi et al., 2018; Smith & Lane, 2016; Vetter et al., 2019).

Now, a skeptic might claim that, in all these experiments, since there are no conscious intentional objects – i.e. since the subjects are not conscious of something which their mental states are about – there might be no intentional objects at all, because unconscious intentional objects may not exist. And since emotions are always intentional states, the skeptic would conclude that the affective feelings (Öhman & Soares, 1994) or action tendencies (Winkielman & Berridge, 2004) displayed by the subjects in these experiments are not emotions, but perhaps another kind of affective state like moods, which can lack intentionality. If that is so, how could these experiments tell us anything about the appraisal process that defines *emotions* as opposed to other kinds of affects?

Against the sceptic, observe that it is highly dubious that the participants in these experiments are in non-intentional states. Indeed, subjects respond to the unconscious stimuli in a way that is identical or very similar to how they respond when they consciously perceive the pictures of snakes, spiders, or facial expressions. And these responses share the properties of normal intentional responses. For instance, action tendencies are not just preparedness for responding to anything, but for responding to certain types of things. I don't see any strong reason to refuse the simple explanation of this phenomenon according to which the unconsciously perceived pictures constitute unconscious intentional objects.

Furthermore, even if one grants to the skeptics that these experiments do not feature emotional episodes, but another kind of affective episode (such as moods), these experiments nevertheless constitute evidence that the appraisal process involved in emotions can be unconscious, because it is

more than reasonable to hold that the relevant mental mechanisms involved in the affective episodes in question are the appraisals involved in emotions. Indeed, since evolution usually is parsimonious, it would be strange that we possess two cognitive mechanisms whose functions are the same: detecting and reacting to goal-relevant situations (e.g. threats such as snakes, spiders, and angry faces) by modifications of physiological changes, action tendencies, and/or affective feelings (e.g. valence, arousal, feeling of dominance). It is theoretically possible, but not really plausible, that there are two distinct mental mechanisms with identical functions that respond to the same stimuli with very similar outputs. The hypothesis that the relevant mental mechanisms involved in these experiments are the appraisal processes that we have been discussing is much more parsimonious.

In the previous subsection, we have seen above that the appraisal processes postulated by (neuro-)psychologists such as Sander et al (2005), Moors (2017), or by the researchers reviewed by Scherer and Moors (2019) involve categorizations that are not hypothesized to be consciously accessible. It would make a lot of sense that, in the experiments mentioned, the mental mechanisms that detect the stimuli and that are involved in the affective responses are the unconscious appraisals postulated by the aforementioned (neuro-)psychologists.

In conclusion, I believe that the theories I have briefly reviewed as well as the experiments I have mentioned constitute evidence that the appraisals postulated by emotion theories are unconscious. Let me observe once again that I don't take this to show that there can be unconscious *emotions*; we are discussing a component of emotions.

9.3.5. Appraisals as representations

Now that we know more about appraisals, it should become clear that these categorizations very plausibly are representations: i.e. states of a system that have the function of indicating that something is the case. Remember that these categorizations are postulated to explain why we may have different kinds of emotional reactions toward the same kind of stimuli and the same kind of emotional reaction toward different kinds of stimuli. The idea that I want to make plausible now is that these categorizations exist for a reason, they have a purpose: classifying stimuli along the different appraisal variables and thus indicating that these stimuli possess these goal-relevant properties is the function of appraisals because it helps us react to these stimuli in useful ways, *ceteris paribus*.

Let me illustrate the idea. When one undergoes fear about a stimulus, one appraises it, let us assume[189], as unconducive to one's safety and as hard to control. This is a fast, automatic, error-prone, unconscious categorization. But why do we perform this categorization? Well, the idea is that if the stimulus is unconducive to one's safety and hard to control, one would better do something about it: one should react to this stimulus accordingly. Furthermore, since there are many different kinds of stimuli that may be unconducive to one's safety, possessing a mechanism whose purpose is to yield a fast, automatic categorization of different kinds of stimuli as having these properties is an enormous advantage for the organism, even if the mechanism is error-prone because one better be safe than sorry. This categorization is especially advantageous because there is a general type of reaction that is generally useful toward the different kinds of stimuli that are unconducive to one's safety and hard to control: avoid them. Fear allows such a reaction in a very efficient way: our blood flow circulates more quickly so that we may deploy our muscles more efficiently; our eyes and nostrils open up to capture more information; adrenaline rushes make us more alert, focus our attention, and they stop our digestive process to save energy; etc. Countless other physiological changes prepare us to avoid the threat.

Simpler organisms (e.g. bacteria) also possess mechanisms that make them avoid stimuli that are unconducive to their safety and hard to control. However, these mechanisms function through one–one correlations: the reactions are limited to one type of stimuli. By contrast, the appraisal process applies to various kinds of stimuli. Not only can these stimuli possess very different perceptual properties (e.g. a snake vs a cliff vs the sound of an explosion), but they may not be perceptible at all (e.g. a financial crisis). Furthermore, the kinds of stimuli are expandable since completely new stimuli may be appraised (e.g. the coronavirus or a new individual that we learn to fear). Finally, the kinds of stimuli change depending on one's goals, and one's goals vary across individuals and over time.

The appraisal process, by classifying various kinds of stimuli under the same category serves an important function: it attributes to them a kind of (non-perceptible) property to which we better react in a certain way given our goals; indeed, we better react to it with the emotion associated with the appraisals in question. This is why it makes so much sense to postulate

[189] This assumption is shared by a very large number of emotion theorists, whether they are appraisal theorists or not.

that there is a one-to-one relation between types of emotions and types of appraisals, but not between emotions and kinds of stimuli.

The appraisal process as it has been described here thus is a mechanism that serves the function of indicating that stimuli of different kinds possess certain non-perceptual properties (e.g. being goal-unconducive). In other words, it has the function to *represent* that something is, say, unconducive to one's safety. According to this plausible description of an appraisal process, it thus becomes apparent that appraisals are representations.[190]

By the way, observe that defining representations through functions means that representation goes hand in hand with misrepresentation, because if something has a function, it means that it can malfunction. And this is very often the case with appraisals: they often misrepresent, because they are fast, automatic, error-prone processes. For instance, the appraisal process behind my fear of heights misrepresents jumping from a diving-board as something goal-obstructive for my safety (even though I know that

[190] I believe that this description of appraisals, which is entirely standard among appraisal theorists, allows us to meet the three challenges put forward by Schroeter et al. (2015) for appraisal theory. Their first challenge is that « it's not clear that appraisal theory posits any processes that require stable internal representations » (370). However, it seems clear to me that if each emotion kind (fear, anger, etc.) is associated with a kind of appraisal, then the latter is a stable representation. Their second challenge is that « it's not obvious that appraisal theory is committed to attributing underived accuracy conditions » (370), as opposed to « merely register[ing] incoming information ». However, from the way I have here presented appraisals, it is apparent that the appraisal process does construct representations of non-perspectival features of the environment that do not merely register incoming information. This is shown by the fact that the same emotion can be caused by very different kinds of stimuli (e.g. fear of height vs fear that the financial market will crash). The same representation applies to very different kinds of incoming information, to kinds of information that are independent of sensory registrations. Of course, given an emotion kind (e.g. fear), there is sensory information with which the emotion is associated, such as the perception of the physiological changes which define this emotion kind (e.g. perceiving one's muscles being tensed when fearful). But appraisals do not have the function of registering this sensory information. Rather, according to appraisal theory, these physiological changes are *caused* by the emotion elicitation process, which is itself explained by the putative appraisals. This point is linked to their third challenge which is the following: « assuming that appraisals have underived accuracy conditions, why think they pick out evaluative properties? Why not think that they represent proximal states, like physiological changes? » (371). Positing that appraisals represent the physiological changes brought by emotions is misunderstanding what role appraisals are supposed to play within appraisal theory. They are supposed to explain how the same stimuli can trigger different emotional reactions and how different stimuli can trigger the same emotional reactions. If appraisals represented physiological changes, appraisals would not explain these two problems, because the same stimuli can trigger different physiological changes and different stimuli can trigger the same physiological changes. The physiological changes involved in emotions are *caused*, and are to be explained, by emotion elicitation mechanisms. So, since appraisals are meant to explain emotion elicitation, it would make no sense to claim that appraisals have the function of representing the proximal states happening during emotional episodes, such as physiological changes.

it is not). There are also cases where the affective system is manipulated so that there can be an emotional episode where the appraisal process is entirely bypassed, which is another way for it to fail to fulfill its function. These cases include elicitation of emotion through chemical induction, direct brain stimulation, or facial feedback (Izard, 1993). They are no counterexamples to what has been argued here: on the contrary, they show that there can be cases of misrepresentations, which itself shows that there must be cases of successful representations since misrepresentations cannot exist without representations.

9.3.6. Appraisals as representations of evaluative properties

The argument so far has been the following: first, we saw that unless we reject the consensus that (1) appraisals, (2) physiological changes, (3) expressive reactions, and (4) action tendencies can be components of emotions, then there are unconscious emotion components. Then, we saw that the appraisals postulated by recent psychological theories should be conceptualized as unconscious categorizations because what goes on in consciousness during an emotional episode does not correspond to the postulated appraisals. Finally, we saw that these unconscious categorizations should be considered as representations, having the function to indicate goal-relevant properties of stimuli.

The issue I said I would tackle in this section (§9.3) is whether emotions represent evaluative properties unconsciously. To collect the final ingredients needed to decide, we need to ask the following question: are appraisals representations of *evaluative properties*?

It seems that the answer to this question must be 'yes' if one agrees that being goal-conducive is good and being goal-obstructive (a.k.a. goal-unconducive) is bad, which should be uncontroversial. Let me elaborate.

Remember that goals are understood widely to encompass concerns, needs, desires, ideals, etc. which may be innate or acquired. Goals are representations of states of affairs that we try to reach, toward which we aim, or more generally that we wish would be instantiated. These representations may not be consciously accessible. Newborn babies have goals which they don't need to entertain in their conscious thoughts, but which nevertheless guide their behavior: such goals may include being safe, resting, being well nourished, making sense of sensory stimuli, etc.

Another way to put things is to say that goals are what determines what is *significant* or *relevant* to the organism. This is how Sander et al (2018: 225) put the matter:

> « Whether a theory refers to stimulus significance primarily in terms of (a) pleasure and arousal (e.g., Bradley, 2009); (b) biological and evolutionary considerations (e.g., LeDoux, 1989; Öhman & Mineka, 2001); (c) primary appraisal (e.g., Lazarus, 1991); (d) dynamics of appraisal checks (e.g., Scherer, 2009b); or (e) concerns (e.g., Frijda, 2007), there seems to be a consensus that emotions need to have objects that the organism, at some level, considers relevant—even if this relevance is not always explicitly accessible to the subject. »

Borrowing this terminology, we could say that a stimulus X is significant for an organism if and only if X is conducive or obstructive to the organism's goals, goals being conceptualized through concepts (a)–(e) or still others (e.g. through desires as in Reisenzein, 2009). Defined as such, it is a consensus among psychological theories that emotions involve an appraisal of goal-(un)conduciveness.

As far as I know, all theories which participate in the consensus that emotions involve appraisals postulate an appraisal that we can call 'goal-conduciveness', where 'goal' is understood in a broad way as comprising concerns, needs, desires, ideals, etc. This is true of philosophy where emotions are typically conceived as reactions to events that are somehow apprehended as significant by the subject (Scarantino & De Sousa, 2018).

Let me note by the way that the main figures in the three main psychological theories of emotion – appraisal theory, constructivism, and basic emotion theory – postulate what we can call a goal-conduciveness appraisal. We have already seen that this is the case for appraisal theory (cf. the review by Agnes Moors, Phoebe Ellsworth, Klaus Scherer, and Nico Frijda, 2013). Concerning psychological constructivism, James Russell, one of its leading advocates, notably describes a typical emotional episode in terms of *goal relevance* (Russell, 2003: 150). In a recent update of basic emotion theory, Paul Ekman (its main champion) and Daniel Cordaro describe how "each basic emotion prompts us in a direction that, in the course of our evolution, has done better than other solutions in recurring circumstances that are relevant to our goals" (Ekman & Cordaro, 2011, p. 364).

The goal-(un)conduciveness appraisal is supposed to represent those goal-conducive and goal-obstructive properties of stimuli. Now, it seems rather obvious that properties represented as goal-conducive are represented by

the organism as good, and properties represented as goal-unconducive (a.k.a. goal-obstructive) are represented as bad, at least if we use 'good' and 'bad' in a general way. By 'general way' I mean that we allow 'good/bad' to designate either 'good/bad simpliciter' or 'good/bad for O' where O is either some individual or a kind (for these distinctions, see Schroeder, 2016). We should allow that the goal-conduciveness appraisal represents certain properties as good/bad for the organism (e.g. not listening to the news is good for me), good/bad for a kind (e.g. the corona crisis is good for many wild species), as well as good/bad *simpliciter* (e.g. that Sam does not change his habit in response to the corona crisis is bad *simpliciter*, even if it ends up not being bad for anybody). Furthermore, we should also leave open whether properties are represented as being good or bad morally, aesthetically, cognitively, socially, vitally, evolutionarily, etc.

Now, this corresponds entirely to how I have characterized evaluative properties above (§9.2.3.). We should thus accept that being goal-conducive is a positive evaluative property and being goal-obstructive is a negative evaluative property. Since the appraisal process is always supposed to represent whether stimuli as goal-(un)conducive, we should conclude that they represent stimuli as being good or bad in some way, as possessing positive or negative evaluative properties.

We have gathered all the ingredients to answer the initial question. Given the consensus in affective sciences about what emotions are (§9.2.1.), given the way I have defined representation (§9.2.2.), evaluative properties (§9.2.2.), and (un)consciousness (§9.3.1.), and given the further characterization of appraisals (§9.3.6.), we are led to conclude that emotions involve proper parts that do represent evaluative properties unconsciously.

9.3.7. Intermediary conclusion and some objections

From the conclusion that emotions involve proper parts that do represent evaluative properties unconsciously, we are led to the claim that emotions represent evaluative properties *tout court*, because when philosophers discuss whether emotions represent x, they are interested in whether there is at least one proper part of emotional episodes that represent x, and not in the claim that all proper parts of emotion represent x. It thus seems reasonable to conclude that, given the evidence reviewed here, emotions do represent evaluative properties (unconsciously).

Before I move on, let me address a few potential objections. First, someone may argue that the appraisal process, as it is described in the literature I

have relied on, is not part of the emotion, but an external cause of it. What if the relation between an emotion and the appraisal is a bit like the relation between a perceptual belief and a perception? We wouldn't want to say that the perception is part of the belief, so why would we classify the appraisal as part of the emotion?

First of all, I don't think that the comparison is good. Perceptual beliefs are, I take it, a kind of belief and you can have regular beliefs without perceptions. By contrast, if the literature reviewed above is correct, you can't have regular emotions without appraisals (by 'regular emotion', I am excluding emotions triggered artificially through e.g. brain manipulation or chemical induction). Allegedly, even in cases where we are 'hardwired' to react emotionally – candidates involve fear-like responses to spiders, snakes, heights, and spiders (LoBue & Adolph, 2019) – there is, as far as I can see, no reason to deny that we appraise the object as goal-(un)conducive. So, the relation between emotions and appraisals is very different to the relation between beliefs and perceptual beliefs. Appraisals and emotions go hand in hand, but that is not true for beliefs and perceptions.

Secondly, why would one want to insist, against the widespread consensus in the affective sciences, that the appraisal process is not part of the emotional episode? If there are no strong reasons not to agree with the consensus, shouldn't we go along and accept it? Let me present and reject two reasons.

One reason to think that appraisals are not proper parts of emotions is that it is possible to have 'irregular emotions', emotions which are caused by something else than appraisals.[191] However, this is only a reason to think that they are not an *essential* proper part of emotions. As said at the beginning of this chapter, I consider that emotions are paradigmatically made of the five components discussed and I have not claimed that any of them is essential. I am ready to accept that there can be emotional episodes that don't involve subjective feelings, or action tendencies, or physiological responses, or motor reactions, just like I am ready to accept that there can be emotional episodes without appraisals.

Another reason to think that appraisals are not proper parts of emotions would result from the combination of two claims: (a) only what is specific to emotions can be part of emotional episodes and (b) the appraisal process is not specific to emotions. Claim (b) may be made plausible if, e.g., one successfully argues that they are regular belief-desire pairs. However, even

[191] Thanks to Tristram Oliver-Skuse for remarking on this.

if we accept (b), I don't find (a) convincing. Claim (a) would exclude from emotional episodes all the action tendencies and physiological changes that are not specific to emotions. Since there are arguably no action tendencies or physiological changes that are specific to emotions (see Chapter 7 on why we can't infer from an observable motor, behavioral, or physiological changes that one is undergoing an emotion), one may be led to hold that emotions are mere positive and negative feelings, because this would be the only one of the five components discussed above that is unique to emotions (this seems to be the reasoning given by James (1984) and behind Whiting's (2011) claim that emotions are just non-corporal, non-cognitive feelings). Furthermore, claim (b) is, to say the least, controversial. A reason to think that appraisals are not regular belief-desire pairs is recalcitrant emotions. When one strongly believes that one is out of danger, one may still unconsciously appraise the situation as dangerous and be afraid. If appraisals just are belief-desire pairs, (i.e. I believe the situation is dangerous and I don't want to be in danger) this would mean that one holds two contradictory beliefs at the same time. By contrast, if one holds that appraisals are cognitive mechanisms whose function is dedicated to affective episodes (that they are closer to Fodorian modules than to his Language of Thought), then one could claim, as many do, that recalcitrant emotions are not cases where one has two contradictory beliefs, but instead that a belief is in tension with a modular, information-encapsulated cognitive state: just like when we look at a Müller-Lyer illusion while knowing that the two lines are of the same length.

So, I don't see any reason to hold that appraisals are external causes to emotions. Instead, I see plenty of reasons to consider them proper emotional components. And if they are to be considered as a causal force, as appraisal theorists argue, then we should consider them as a partial cause to the other emotional components.

Another potential objection would be that the evaluative properties in which philosophers have been interested are different from those represented by unconscious appraisals, and so that I have changed the subject.

I entirely agree that philosophers, beginning with Aristotle, have generally discussed rather complex properties such as slight (for anger), loss (for sadness), injustice (for indignation), danger (for fear) which are quite different from more specific properties such as being unconducive to one's safety. However, it is not true that I have changed the subject, since I have never claimed that the discussion would be restricted to evaluative

properties such as slight, loss, etc.; I have stuck to the definition of evaluative properties given in the first part (§9.2.3).

Third potential objection: there may be kinds of emotional reactions that are linked in a one-to-one relation to kinds of stimuli because evolution has hardwired our brain in this way. In these cases, there is no appraisal process going on, the mechanism does not involve any representations, but we react to the stimuli in a mechanical, reflex-like way, a bit like white blood cells reacting to foreign DNAs.

First of all, the evidence for hard-wired emotional reactions is very much debated. Candidates for stimuli that would trigger such emotional reactions include spiders, snakes, heights, and foreigners because newborn babies supposedly reacted with fear to such stimuli even when they were presented with them for the first time. But recent reviews show that the experiments leading to these hypotheses may not be trustworthy, notably because the babies' reactions cannot reasonably be assumed to be fear (LoBue & Adolph, 2019). There may in fact be no stimuli that are linked to emotions in a hard-wired, one-to-one way.

Furthermore, the plasticity of goals and beliefs enables humans, as well as many other animals, to learn not to have certain emotions toward all kinds of stimuli. For instance, one may learn not to be afraid of snakes or spiders, because one knows that they are not dangerous (think for instance of someone working in a vivarium). Also, if one really wants to be bitten by a snake and has no desire whatsoever to avoid being bitten, one might not be afraid of being bitten by a snake, but one would hope to be bitten by a snake and act accordingly. We can imagine for instance that the person believes that if she is bitten by a snake, she will win 10 billion dollars. Our emotional reactions depend on what goals and beliefs we have. The fact that all humans may react in the same way to certain stimuli may be explainable by the fact that some goals and beliefs appear to be universal – e.g. the goal of surviving and the belief that falling from a certain height is deadly. [192] The fact that goals and beliefs can modify emotional reactions is easily explainable if emotions are mediated by representations such as appraisals, since the latter is inherently linked to goals and beliefs. However, I don't see how hard-wired, reflex-like reactions may be modified by beliefs and goals.

[192] But think about the movie *Groundhog day*, where a person relives the same day over and over again. The character realizes that even if he dies, he starts the same day again. He then starts experimenting with different ways of committing suicide to end the day and start it over again in a better way. In this scenario, the character is not afraid of deadly stimuli, and the reason is, I believe, that he now finds them goal-conducive.

Now, even if there were emotional reactions that would be hardwired to certain stimuli, they would constitute an exception and my conclusions would still apply to the vast majority of emotional episodes.

9.4. Implications for philosophical theories of emotion

In the first part of this chapter, I have briefly presented seven philosophical theories of emotion and given their official or expected answer to the question 'Do emotions represent evaluative properties?'. Their answers were summarized in Table 9.1. which I reproduce here. The reasons for this classification were presented in §9.2.4.

Yes, emotions do represent evaluative properties (expected or official answers).	**No, emotions don't represent evaluative properties (expected or official answers).**
(Quasi-)judgment theory (Nussbaum, 2001; R. C. Roberts, 2003; Solomon, 1977, 1993)	Attitudinal theory (Deonna & Teroni, 2012, 2014, 2015)(Deonna & Teroni 2012, 2014, 2015).
(Quasi-)perceptual theory (De Sousa, 1987; J. Deonna, 2006; J. Deonna & Teroni, 2008; Döring, 2007; Goldie, 2002; Helm, 2009; Prinz, 2004; Ratcliffe, 2005; Roberts, 2003; Tappolet, 2000, 2016)	Non-intentional feeling theory (Shargel, 2015; D. Whiting, 2011).
Motivational theory (Scarantino, 2014, 2015a)	Enactive theory (Hutto, 2012; Shargel & Prinz, 2018).
Representationalism about emotional experiences (Mendelovici, 2014; Tye, 2008)	

Table 9.1. Do emotions represent evaluative properties? Expected/official answers.

At the beginning of the second part (§9.3.), I said that emotions may (a) represent evaluative properties unconsciously, but not consciously, or (b) represent them consciously, but not unconsciously, and so there are pairs of theories which (appear to) contradict each other in answering the unqualified question with which we started – i.e. 'Do emotions represent evaluative properties?' – but which actually don't contradict themselves once we take consciousness into account and ask the disjunctive question: 'Do emotions represent evaluative properties consciously, unconsciously, both, or not at all?'. I will now elaborate on this point and we will see that, in particular, once we ask the disjunctive question, motivationalism no longer contradicts with (charitable reconstructions of) attitudinalism and non-intentionalism.

First of all, let us observe that, to the best of my knowledge, most of the authors mentioned in Table 9.1 don't specify whether emotions represent evaluative properties unconsciously but focus only on conscious representation. Only four of them give enough evidence about their position on this question.

Among those who give sufficient evidence, Scarantino (2014) and Prinz (2004) endorse the view that emotions represent evaluative properties unconsciously while Hutto (2012) and Shargel and Prinz (2018) refuse it (see also Schroeter et al., 2015 for a skeptic/negative answer). The other theorists we have encountered above, to the best of my knowledge, don't specify whether they take emotions to unconsciously represent evaluative properties or not.[193] So, as far as I can tell, the answer to our question is unspecified by (quasi-)judgementalism, (quasi-)perceptualism besides Prinz's (2004), representationalism, non-intentionalism, and attitudinalism.

Most theorists focus on whether emotions *consciously* represent evaluative properties. Here, we find a positive answer in (quasi-)judgementalism, (quasi-)perceptualism, and representationalism. Among those who claim that emotions do not consciously represent evaluative properties, we find enactivism, non-intentionalism, and attitudinalism. Only one of the seven theories discussed here – motivationalism – does not seem to specify whether emotions consciously represent evaluative properties.

These positions are presented in Table 9.2.

[193] For instance, Deonna and Teroni make it clear that they consider that there are A-unconscious emotions although there are no P-unconscious emotions, and that emotions don't represent evaluative properties P-consciously. Nevertheless, they don't discuss whether emotions represent evaluative properties A-unconsciously, as far as I know. However, they agree that appraisals may involve sub-personal mechanisms that determine our emotions (Deonna & Teroni, 2012, p. 103 n.6). They may thus agree that emotions represent evaluative properties A-unconsciously even if they don't do so P-consciously. In fact, Julien Deonna agrees that this claim is plausible (personal communication).

Do emotions represent evaluative properties...	**Yes**	**Unspecified**	**No**
... without qualification? (cf. Table 9.1.)	(Quasi-) judgementalism (Quasi)perceptualism Motivationalism Representationalism		Non-intentional feeling Enactivism Attitudinal
... A-unconsciously?	Motivationalism Prinz's perceptualism	(Quasi-) judgementalism Other (quasi-) perceptualism Representationalism Attitudinalism Non-intentional feeling	Enactivism
... A-consciously?	(Quasi-) judgementalism (Quasi-) perceptualism Representationalism	Motivationalism	Non-intentionalism Enactivism Attitudinalism

Table 9.2. Do emotions represent evaluative properties consciously or unconsciously?

What I find especially notable about Table 9.2 is the following: When we ask the unqualified question: "Do emotions represent evaluative properties?", we find a strong opposition between the yays and the nays (first line of Table 9.2). However, once we take into account consciousness as a variable, the debate is much less polarized. There are still oppositions of course: enactivism is incompatible with both motivationalism and Prinz's perceptualism (second line of Table 9.2), and we also find that judgmentalism, perceptualism, and representationalism are incompatible with non-intentionalism, enactivism, and attitudinalism (third line of Table 9.2). However, some hidden compatibilities are revealed: we see that motivationalism is not in contradiction with (a charitable interpretation of) either attitudinalism or with non-intentionalism, contrary to what the answers to the unqualified question suggested.[194] Indeed, it may well be the case that emotions represent evaluative properties unconsciously but

[194] It is a charitable interpretation in the sense that we could instead take their claims that emotions don't represent evaluative properties literally and so as implying that emotions neither represent them consciously nor unconsciously (just like if I say that there are no bears in Geneva, I am committed to there being neither Belgian nor non-Belgian bears in Geneva). Thanks to Tristram Oliver-Skuse for this point.

not consciously, thus respecting the claims made by these three theories. Of course, these theories still differ on other issues.

Another interesting point about Table 9.2 is that, of the seven theories discussed, only enactivism conflicts with the evidence I have brought forward in this chapter, i.e. the evidence supporting the claim that emotions represent evaluative properties unconsciously. I find this to be good news for everybody: the six other theories don't want to conflict with contemporary cognitive sciences, only enactivism – or, rather, only the radical form of enactivism defended by the authors in question – seems happy to take the rebellious attitude of refusing to accept what the immense majority of affective scientists tell us about the mind.

Finally, I believe that the discussion up to here shows that it is of foremost importance to take into account the consciousness variable when discussing whether and what emotions represent. This is especially true for those who argue against the claim that emotions represent evaluative properties, because when they say, without qualification, that emotions don't represent evaluative properties, then their claim, taken as face value, is wrong according to what I have argued in this chapter (§9.3.6.).

9.5. Conclusion

In this chapter, I have explained why we have good reasons to conceptualize emotions as psycho-somatic episodes involving a proper part which represents evaluative properties unconsciously, and why we can thus say that emotions as a whole represent evaluative properties.

Although this seems to be in contradiction with philosophical theories of emotion which officially claim that emotions don't represent evaluative properties – such as attitudinalism (Deonna and Teroni, 2012; 2014; 2015) and non-intentionalism (Shargel, 2015; Whiting, 2011) – I have argued that it actually is not, because these theories in fact concentrate only on conscious representation. Furthermore, taking the consciousness variable into account shows that certain apparent antagonisms are dissolved: motivationalism (Scarantino, 2014; 2015) is not, despite the appearance, in contradiction with attitudinalism and non-intentionalism on the present issue. Therefore, to avoid talking past each other, I believe that philosophers of emotion need to distinguish between conscious and unconscious representation.

FROM APPRAISALS TO CONSCIOUSNESS

Let me end this chapter by coming back to a question that has appeared implicitly in several places above: What is the relation between the putative unconscious representations of appraisals and the flow of our consciousness during an emotional episode? I will briefly elaborate on some points I have already made above which may constitute the first steps toward an answer.

Here is the picture that I find most attractive: First, the emotional episode begins with an unconscious appraisal process which causes, as a second step, an action tendency. An action tendency is a representation (Moors, 2017, p. 10) or at least it involves a representational component: it represents what sort of actions need to be implemented to respond to the goal-(un)conducive event represented by the appraisal process, the general way in which one should react to this stimulus to maximize the probability of attainting or preserving one's goal (I know that this is a controversial claim that would need to be defended, but I am just presenting the picture that I find attractive here). Action tendencies thus are like action plans: not wholly determined, but supplying a general direction (Bratman, 1987). As some would put it, action tendencies are (or involve) representations that possess an imperative force as they 'demand a certain kind of action to be satisfied' (Klein, 2007, p. 519) just like imperative sentences demand action from the addressee. By contrast, appraisals possess an indicative force: as argued above, their purpose is to indicate how the world is like, rather than to direct behavior.[195]

These first two steps – the appraisal process and the formation of an action tendency – can be conceptualized in many ways and as such are compatible with many psychological theories, including those mentioned in §9.3.4. These two steps will be reiterated over and over again during the emotional episode in a way that will allow evaluating the situation and responding to it appropriately over time, taking into account all the changes that happen during the episode, whether they are external changes, cognitive changes, or changes in our bodily configuration.

The appraisal and action tendencies cause multiple kinds of modification. A main function of appraisals is to represent evaluative properties so as to

[195] This is similar to saying that action tendencies are (or involve) representations with a world-to-mind direction of fit while appraisals have a mind-to-world direction of fit, but the coherence of this notion has been doubted (Frost, 2014), which is why I prefer to talk about imperative and indicative forces. Observe also that it is not obvious that Dretske's definition of representation applies to representations with an imperative force, since he defines representations through indications.

determine the most appropriate action tendencies and maximize one's goals, but they also cause modifications in other kinds of representations, such as beliefs (e.g. how we will henceforth conceive the emotional stimulus) and desires (e.g. whether we will henceforth be more or less attracted by the stimulus). The action tendency, on the other hand, causes many physiological and motor changes that are necessary to deploy it. These involve modifications in processes as diverse as digestion, muscular contraction, hormonal release, pupil dilatation, etc.

Among all these modifications, some become consciously accessible – or, more probably, a more or less distorted representation of some of these changes becomes consciously accessible. Let me give some examples: In anger, we may feel our body as poised to aggress (action tendency). We may feel the tension in many of our muscles, or the increased blood flow in some parts of our body (physiological changes). We may also experience the object of our anger as obstructive and feel ourselves as having the power to reduce this obstructiveness (appraisals).

However, many of the modifications caused by the appraisal and action tendency are not accessible through consciousness. I gave several examples of changes that are not accessible to consciousness in §9.3.2. (e.g. hormonal release, the difference between Duchenne vs non-Duchenne smiles, at least some appraisals, etc.).

Finally, and importantly, during an emotional episode, our attention is usually directed at the object of our emotion, which means that we usually don't focus on the internal changes, whether they concern our physiological changes, action tendencies, etc. The flow of our consciousness may be wholly 'filled' with thoughts and feelings about the *external object*, the stimulus of our emotion, and not what goes on inside us. Even though we could turn our attention to how our body is reacting, to how our beliefs have changed, to the kinds of action we are ready to take, the situation often requires that we do not do this but focus our attention instead on the intentional object of our emotion so as to deal with it appropriately. Thus, many emotional modifications are accessible through consciousness but need not, and often are not, actually accessed.

In sum then, the way I see the relation between the unconscious representations of the appraisal process and what goes on in our consciousness is a relation of partial causal determination. The causal determination is only partial because the flow of our consciousness is also determined by many other variables, including how our body feels (physiological changes), how we feel poised to act toward the stimulus

(action tendency), and how the stimulus appears to us besides how it is appraised.

10. General conclusion

« What do you mean? Ooh
When you nod your head yes
But you wanna say no »
– Justin Bieber, *What do you mean?*

« Et des confins du monde, à travers des milliers de kilomètres, des voix inconnues et fraternelles s'essayaient maladroitement à dire leur solidarité et la disaient, en effet, mais démontraient en même temps la terrible impuissance où se trouve tout homme de partager vraiment une douleur qu'il ne peut pas voir. »
– Albert Camus, *La Peste*

In the first part of this dissertation, I have illustrated why, contrary to a widespread assumption, the two main kinds of models of information transmission are insufficient. There are certain cases where stimuli carry information that neither the code models nor the prevailing Gricean models can account for. I have argued that the Extended Gricean Model is a solution to this problem and have shown how it applies to a variety of cases.

In the second part of the thesis, I have presented and analyzed different theories of meaning and how they can help us understand what information is carried by different kinds of emotional signs. I have thus investigated what the meaning of expressive speech acts amounts to – notably by distinguishing a force and a content component of meaning.[196] We have seen how the notion of probabilistic meaning can account for the meaning of non-communicative emotional signs, and how teleosemantic notions may help us understand what information is carried by non-Gricean emotional signals. I ended by considering what emotions in and of themselves mean and argued that they represent evaluative properties (unconsciously).

In sum, to the best of my knowledge, I have discussed all the different kinds of meanings which could help us understand what affective meaning is – in fact, I have felt the need to create a new notion: allower-meaning. Fig. 10.1 illustrates the relation between the kinds of meaning I have discussed.

[196] In the Appendix, I discuss further what speaker-meaning is.

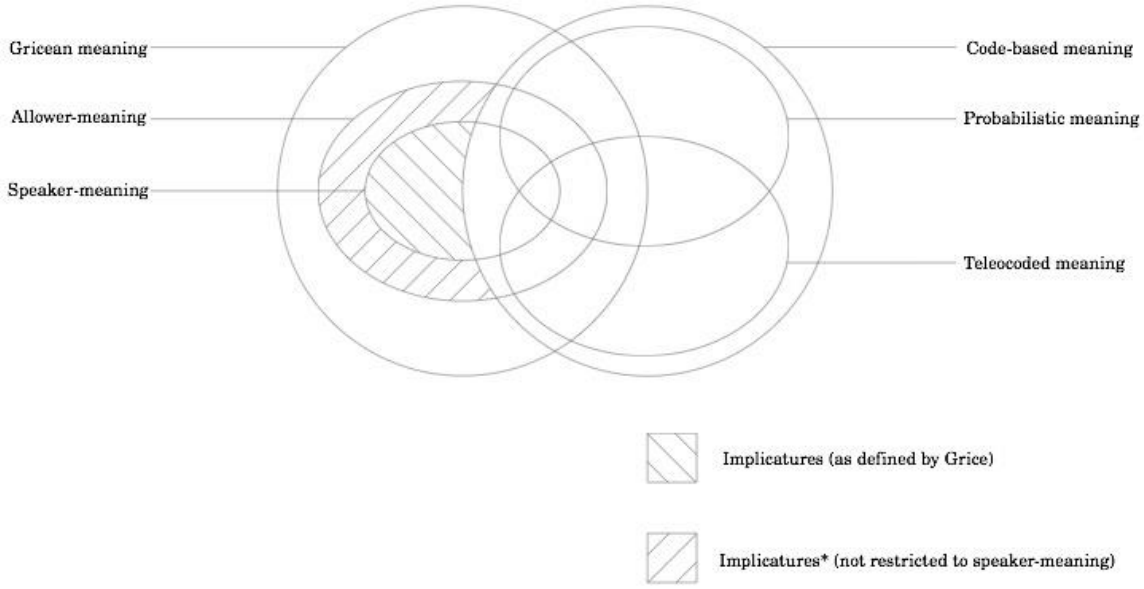

Fig. 10.1. The kinds of meaning discussed in this dissertation.

The whole process has been importantly structured around the distinction between Gricean and non-Gricean meaning and by further distinctions within these two broad categories. I have focused on each of these sub-categories, one after the other, and, as a consequence, I have not talked about the *overall* affective meaning of a sign, which can be a mixture of these different kinds of meaning. Although this subject merits much more attention than I can offer in this conclusion, it nevertheless seems to be the right place to say a few words about it.

When we express our emotions, we never produce just one kind of sign in a void, contrary to what analyses of meaning often seem to suggest, including those given in the present work. Instead, we produce a plethora of different kinds of cues, which mix to produce an extremely complex overall meaning. When, for instance, we perform an expressive speech act, we not only make a verbal utterance but produce many other kinds of emotional signs. The latter include signs which have a communicative function, i.e. that are supposed to carry the information that they carry, including our prosody, our facial expressions, our bodily position, the accompanying gestures we make, and so on. Some of these are intended to mean something, we just allow others to have a meaning – including implicatures* – and some have effects which are not mutually recognizable

as controllable, and so can only carry encoded information (or at least cannot carry any kind of implicature). Besides the meaning of these signals, a dizzying amount of information is to be found in non-communicative signs: the blushing, the stutter, the sweating, the pupil dilatation, the exacerbated muscular tonus, etc. This information is not meant to be communicated, but may nevertheless be transmitted.

The overall meaning of such an expressive act thus is a mixed bag of speaker-meaning, allower-meaning, teleocoded meaning, and probabilistic meaning (as well as perhaps other kinds of meaning which escape me). But it is more complicated than that because these different kinds of meaning interact and may vary depending on each other. For instance, if non-communicative signs, such as blushing or sweating, are in the common background, what we can intend or allow the verbal signs to mean is constrained by these nonverbal signs. In other words, the appropriate interpretation of what is meant with words sometimes depends on the context created by such non-communicative signs. The information sent verbally (e.g. through our choice of words) interacts with the information sent acoustically (e.g. the tone of our voice) and visually (e.g. whether our gestures look calm or aggressive) in such a way that the overall affective meaning is not the mere sum of the information sent by the different kinds of signs taken separately, in isolation. What could be interpreted as expressing a certain affective state when one concentrates on the verbal utterance is understood as expressing another when one takes into account the prosody. We are all too familiar with this phenomenon when we exchange text messages. This is when, for an optimal understanding, we need to call each other instead of texting, or at least need to use emoticons (i.e. facial expression icons) to disambiguate the verbal meaning. Another typical example is 'mixed signals', for instance when what is speaker-meant contradicts, or is in tension with, the information sent nonverbally, as when someone's voice betrays her insincerity.[197]

In this dissertation, following the typical methodology from the philosophy of language and linguistics, I have focused on analyzing the different signs

[197] Another example is in the immortal verses used in the epigraph: 'What do you mean? Ooh/ When you nod your head yes/ But you wanna say no'. Supposing that the narrator of the lyrics knows that the addressee wants to say 'no' from signs that are not produced with overt intentions, we here have an example of speaker-meaning, the information sent through nodding 'yes', which contradicts teleocoded or probabilistic meaning, the information sent through the cues used by the narrator to infer that the addressee wants to say 'no'. The quote from Camus illustrates a similar idea: when we know of someone's suffering through description but that we can't *see* someone suffering nor otherwise be acquainted directly with it, we lack crucial information and so are (often) incapable of optimally empathizing with her.

in near isolation. However, a complete analysis of the overall meaning of expressive acts must take the different kinds of meaning into account and do so holistically. I have attempted here and there to avoid oversimplification and have discussed the interaction between different kinds of meaning, in particular when I discussed allower-meaning, since the latter may include as components all the other kinds of meaning (see Fig. 10.1). But I must admit that I did not devote ample attention to how the different kinds of meaning interact holistically.

Here is a striking fact. It is apparent from the discussion up to here that overall affective meanings manifest an extreme complexity. Nevertheless, most of the time, we intuitively, effortlessly, and automatically understand a huge amount of it when we see and hear a person, or a nonhuman animal, displaying an affective state. The complexity of what we so effortlessly understand is part of what makes it so hard to put into words what we get from an affective display. It may be why so little research has been achieved on affective meaning compared to descriptive meaning – because of the sheer complexity of the subject matter while simultaneously appearing so natural or intuitive from a first-hand perspective. It is extremely hard to conjugate these two facts.

Relatedly, even though I have insisted on what distinguishes the different kinds of affective meaning, from a first-person perspective, affective meaning is *not* understood as made up of these different kinds of meaning. Think of two friends in a passionate discussion, a father looking at his playful child, or a student trying to decipher whether jury members like what she says. From such perspectives, the way it is like to understand affective meaning is not chopped-up and classified into the different boxes in which I have put them. The distinctions I have made are useful to make models of communication and to hypothesize about the mental mechanisms through which we can share information about our affective states. But when it comes to taking a first-person perspective on the phenomenon at hand, the different kinds of affective meanings all merge into a general, often ineffable impression.

Nevertheless, the distinctions made here and the analyses of the different kinds of meaning permit a new way to investigate what are the different sources of this impression: what is intended by the person performing the expressive speech act (speaker-meaning), what is allowed by her to be meant (allower-meaning), what her behavior has the function of encoding (teleocoded meaning), and what it betrays (probabilistic meaning). I hope that, even from an ordinary everyday first-person perspective, once we

think about these different sources, the general, ineffable impressions may be better comprehended and clarified.

APPENDIX – SYNTHESIZING A NEW DEFINITION FOR SPEAKER-MEANING

« Ma définition, avec des textes à prendre à 1 degré 5. »
– Booba, *Ma définition*

Abstract. In Chapter 2, I gave five different definitions of allower-meaning. Four were based on existing, competing definitions of speaker-meaning (Grice's, Neale's, Sperber and Wilson's, and Green's) and one that was original, the latter being my favored definition. In this chapter, I will explain why I favored the latter over the definitions of allower-meaning that were based on the existing definitions of speaker-meaning by discussing the advantages and disadvantages of the latter.[198]

A.1. INTRODUCTION

I have chosen the four definitions of speaker-meaning that I will discuss here for different reasons. Grice's because it is the original definition from which all others have stemmed, it remains widely used, and is probably the best known. Neale's because it is supposed to capture some later discussions of Grice, some possible responses that Grice could have made to his critiques as well as some ideas which he didn't formulate explicitly in another definition. Sperber and Wilson's definition of ostensive-inferential communication illustrates an alternative account of the phenomenon in which we are interested, a so-called post-Gricean one, which possesses many advantages and has been very influential. Green's, finally, is probably one of the most sophisticated definitions on the market today, which is the reason why I spend many pages discussing it. Another great advantage of Green's definition is that it aims to capture the critiques made to both the Gricean and post-Gricean kinds of definitions, including Sperber and Wilson, but also Davis (1992, 2003), whom I don't discuss in detail here because Green's definition allows making the relevant points.

[198] Many thanks to Mitch Green for having read, commented on, and discussed at length this Appendix.

A.2. GRICE'S DEFINITION

Let us start with the last explicit definition Grice gave, on which I have based the first variant of my definitions of allower-meaning. Note already that, Grice used the terms 'speaker', 'uttering', and 'audience' in a very broad way that extends beyond speech (e.g. to writing) or language (e.g. to pointing), and we will follow him in these uses.

> « 'By uttering x, U meant something' is true iff for some audience A, U uttered x intending:
>
> (1) A to produce some particular response r,
>
> (2) A to recognize that U intends (1), and
>
> (3) A's recognition that U intends (1) to function, in part, as a reason for (1). » (Grice, 1989, p. 99)

Among the disadvantages which have been discussed in the literature, I will mention two here.

A.2.1. A FIRST DIFFICULTY: SHOWING AND SAYING

The first problem concerns Grice's third clause. He added it to the other two because he wanted to exclude certain cases of showing from the definition of meaning (Grice, 1957, 1969). For instance, Grice famously discusses the case of Herod who presents to Salomé the head of John the Baptist. Herod does this with the following intentions:

(1) To produce in Salomé the belief that John is dead, and
(2) That Salomé recognizes that he intends that (1).

Herod thus respects the first two clauses for speaker-meaning. However, Grice has the intuition that Herod, by producing the severed head of John, doesn't thereby literally *mean* that John is dead. His intuition seems to come from the fact that Salomé can be informed that John is dead just by seeing the head and that Herod might as well have no intentions to communicate whatsoever. As Neale (1992, p. 548) notes, when Grice (1957) first introduced the distinction between speaker-meaning (at the time called 'nonnatural meaning') and natural meaning, he might have been worried that something approximating natural meaning interferes with the idea of Herod speaker-meaning that John is dead. Grice thus added his third clause to restrict cases of speaker-meaning to cases where the intentions of the speaker play a role in the inferential process of the audience. Furthermore, Grice had the linguistic intuition according to

which we don't want to say Herod literally *meant* that John is dead by acting as he did.

However, many commentators (notably Neale, 1992; Recanati, 1986; Schiffer, 1972; Sperber & Wilson, 1986) did not share Grice's linguistic intuition about the use of 'meaning' in such cases. They thought Herod could very well be said to literally mean that Herod is dead by presenting the severed head. More generally, they argued that we often mean things by showing them. For instance, in response to an invitation to play squash, Bill can very well mean that he can't accept the invitation by showing his bandaged leg (an example given by Schiffer). Grice's decision to filter out such cases from speaker-meaning seems overly restrictive.

In addition, the same commentators remarked that several problems awaited Grice's third clause down the line. For instance, I can very well mean that I can speak with a squeaky voice by uttering 'I can speak in a squeaky voice' said in a squeaky voice. The fact that my intention to communicate is not necessary for you to infer what I intend to mean is not a good reason to argue that I thereby failed to mean that I can speak in a squeaky voice (Neale, 1992, p. 548). The same conclusion applies to a case where I say 'I'm right here' to someone who is looking for me. Many other problematic cases have been presented which show that Grice's third clause is unsatisfying.

The obvious solution is to just discard it: without the third clause, we can say that Herod speaker-meant that John is dead and that Bill speaker-meant he can't play squash. This doesn't appear to be a harmful consequence, contrary to what Grice might have thought.[199] This solution has been widely accepted, which is why there is no equivalent of (3) in the other definitions discussed here.

A.2.2. A SECOND DIFFICULTY: THE RIVER RAT AND THE MOON OVER MIAMI

Let us now turn to a second problem faced by Grice's definition. This can be presented with an adaptation (based on Schiffer, 1972, pp. 17–18) of a counterexample first presented by Strawson (Strawson, 1964b, pp. 446–447), which we will call 'the River Rat case'. Instead of paraphrasing it, let me just quote Green's version:

[199] As Neale (1992, p. 548) notes, this might even turn out to be positive for one of Grice's projects in « Meaning revisited » (1982), which is to show the continuity between natural meaning and speaker-meaning.

« The River Rat: Homebuyer is inspecting a house for possible purchase, and his friend—call him Friend—is concerned to convince him that the home is rat infested. Friend arrives at the house at a time when he knows that Homebuyer is inside, and although he knows he is being watched, skulks around to make Homebuyer think that he, Friend, believes he is acting unobserved. Friend has a river rat that he places in a salient position for Homebuyer to see. He intends for Homebuyer to see the rat and reason as follows: 'Although the rat display was rigged, Friend would not have put it there unless he believed that the house really is rat infested; hence Friend, who is reliable and honest, must intend me to believe that the house is rat infested. » (Green, 2007, p. 63)

Friend displays the river rat in the house with the following intentions:

(1) To create the belief in Homebuyer that the house is rat infested
(2) That Homebuyer recognizes that he (Friend) intends (1)
(3) That Homebuyer's recognition that he (Friend) intends (1) to function, in part, as a reason for (1).

However, we don't want to say that Friend here literally *means* that the house is rat infested. Proponents of Grice's program have attempted to remedy to such counterexamples by adding a fourth clause to Grice's definition which would be similar to the second, requiring that U makes more of his intentions known to the audience (see Vlach 1981 for a review). We would thus have something like this:

U speaker-mean something iff, for some audience A, U uttered x intending

(1) to produce a response r in A,
(2) that A recognizes that U intends (1),
(3) that A's recognition of (1) functions as a reason for (1), and
(4) that A recognizes that U intends (2).

This fourth clause prevents the definition of speaker-meaning to apply to the River Rat case and it seems to otherwise not exclude normal cases of speaker-meaning. So we appear to have a solution. However, counterexamples can be found even for such a definition. One such counterexample is the 'Moon Over Miami' case presented by Schiffer (1972, pp. 18–19) (I have slightly adapted the example to the present discussion):

Sam, who has a hideous singing voice, (1) intends to make Maria leave the room by singing *Moon Over Miami*. Sam (2) intends that Maria recognizes that he intends her to leave the room. Furthermore, because he wants to make it clear that he intends to show his disdain for Maria's being in the

room, Sam also intends (4) that Maria recognizes that he intends her to recognize that he wants her to leave the room. Now, Sam intends that Maria will believe that he plans to get rid of her by means of his repulsive singing, but he also expects and intends (3) that Maria's reason for leaving the room will in fact be her recognition of Sam's intention to make her leave the room. However, Sam doesn't intend the intention (3) to be manifest to Maria. In other words, while Sam intends that Maria *thinks* that he intends to get rid of her by means of his repulsive singing, Sam really intends Maria to have as her reason for leaving the fact that Sam wants her to leave, an intention that he doesn't make manifest.

Although Sam respects clauses (1)–(4), he doesn't appear to really speaker-mean anything with his song. This is in stark contrast with a case where Sam tells Maria 'Get out of here!' (so long as he has the appropriate intentions). One could think that the problem is that Sam has somehow hidden his intention that (3). A way to remedy to this counterexample would be to add a clause to the definition so that it is required that U have the further intention:

(5) that A recognizes that U intends (3).

With this fifth clause, the Moon Over Miami case would not count anymore as speaker-meaning, but, with enough ingenuity, one could devise a further counterexample which would require to add a sixth clause, and then a seventh, and so on. We might thus want to look for a remedy elsewhere, instead of just adding more and more complicated clauses to the definition. The solution that Schiffer himself proposed is to include what he called 'mutual knowledge', but we will see below that this is problematic.

The reason why the River Rat and the Moon Over Miami cases don't involve speaker-meaning is that Friend and Sam act in a non-overt way.[200] Some of their intentions are hidden from the audience. As Strawson (1964b) rightly remarks, the notion of speaker-meaning seems to imply that the communication is done in a 'wholly overt' way.

Grice (1982) acknowledged this problem and proposed a solution. Instead of adding a fourth (fifth, sixth, …) clause similar to the second, he discussed adding a clause which would get rid of what he called 'sneaky intentions',

[200] Let me remark that if we understand 'covert stimuli' as stimuli whose production is intended to be hidden and 'overt stimuli' as stimuli whose production is intended to be public, then there are stimuli that are neither overt nor covert, because they are neither intended to be hidden nor public. Cases where one allows something to be public without intending it to be public for instance, cases where one inadvertently produces a stimulus, or where one produces a stimulus despite oneself.

the kinds of intentions that Friend and Sam display. The idea, roughly, is that the speaker must not intend that the audience is deceived by intentions (1)–(3).

A.3. NEALE'S DEFINITION

Grice (1982) doesn't explicitly add this clause to his definition of speaker-meaning, but Neale (1992) has proposed such a reconstruction of Grice's ideas which is accepted today by, e.g., Moore (2017, 2018). Neale also modified aspects of Grice's original version to take into account the problem of showing (e.g. in Herod's case or Bill's case), as well as a third problem with Grice's definition that I won't discuss here. Taking into account these remarks, Neale proposes the following definition:

> « By uttering x, U meant that p iff for some audience A,
>
> (1) U uttered x intending A actively to entertain the thought that p (or the thought that U believes that p)
>
> (2) U uttered x intending A to recognize that U intends A actively to entertain the thought that p
>
> (3) U does not intend A to be deceived about U's intentions (1) and (2). » (1992: 550)

This definition resolves both the problem of showing (by getting rid of Grice's original third clause) and problems coming from the 'sneaky intentions' found in the River Rat and the Moon Over Miami cases. You will observe that the first and the second clauses of this definition are rather different from Grice's: they are restricted to cases where the speaker utters something affirmatively, as opposed to when a speaker intends to mean something which is (or is the nonverbal equivalent of) an order, a question, an interjection, an excuse, a promise, etc. Richard Moore proposes a definition that is not restricted to affirmative speaker-meaning by going back to Grice's clauses (1) and (2):

> « A speaker S non-naturally means something by an utterance x if and only if, for some hearer (or audience) H, S utters x intending:
>
> (1) H to produce a particular response r, and
>
> (2) H to recognize that S intends (1).
>
> In addition to acting with intentions (1) and (2), it's also necessary that the speaker should not act with any further intention:

(3) that H should be deceived about intentions (1) and (2). » (2017: 305).

This definition corresponds to the second variant of definitions of allower-meaning proposed in Chapter 2.

Neale and Moore's definitions avoid the problems we have encountered earlier. Furthermore, an advantage that Moore (2017) stresses is that this type of definition can be interpreted in a way that makes it cognitively much less demanding than Grice's. Grice's original definition requires communicators to entertain high-order metarepresentations, i.e. thoughts about thoughts about thoughts about thoughts, etc. Moore (2017: 318ff) gives an interpretation of Neale's definition which only requires the audience to have beliefs about the speaker's intention, and the speaker to have intentions about the audience's behavior, thus remaining at the level of first-order metarepresentations (mental states about mental states and that's it). We will come back to this point when discussing Green's definition.

A.3.1. A DIFFICULTY: THE VIGILANTE

However, despite such advantages, this type of definition is not without counterexamples.[201] Take the following, which I will call the 'Vigilante case' (based on Grice, 1957). Maria is a sort of Batman character, a vigilante who rights wrongs without legal authority. Sam, a villain, has committed a murder. Maria decides to leave Sam's glove on the crime scene to give the detective a reason to believe that it is Sam who committed the murder. Furthermore, she decides to do something else: she purposely leaves some of her DNA on the glove, because she would like the detective to know that it was she who put Sam's glove on the crime scene. However, she is not sure that the detective will notice that there are traces of her DNA on the glove and she is even less sure that he will infer that she intended to tip him off with the glove. The fact that she has left some DNA on the glove is a sort of test for the detective: Will he be able to notice it? Will he rightly infer that she wanted to help? What she is pretty sure is that the detective won't think that she has purposely left her DNA on the glove to test him: he doesn't know her well enough to infer that.

It seems to me that, here, just as with the River Rat and Moon Over Miami cases, we won't say that Maria has speaker-meant that Sam is the

[201] Green (2007: 76-77) gives other reasons to reject Neale's proposal.

murderer. However, she has produced the stimulus (put the glove on the crime scene) intending:

(1) The detective to produce a particular response, i.e. to think that Sam is the murderer.
(2) The detective to recognize that she has produced the stimulus intending that (1).

And she didn't act with any further intention:

(3) That the detective should be deceived about intentions (1) and (2).

She doesn't intend to deceive him about (1) and (2) because she actually wants the detective to know that (1) and (2). This is unlike the River Rat or the Moon Over Miami cases. However, here, again, what seems to prevent Maria from really speaker-meaning anything is the *covertness* of her action. I believe that Neale's third clause fails to exclude cases such as the Vigilante because it focuses on deception as opposed to covertness.

This remark points to another type of solution which has been proposed for such problems as the River Rat and Moon Over Miami, a solution which has been championed in different ways by Lewis (1969), Schiffer (1972), and Sperber and Wilson (1986). Their idea is that the relevant information involved in the interaction must be commonly recognized (Lewis) by communicators, mutually known (Schiffer), or mutually manifest (Sperber and Wilson). Instead of cashing out Strawson's insight that the communication must be 'wholly overt' by adding clauses, the idea is to modify clause (2) so as to make clause (1) *mutually* recognizable. That is, communicators must not only share the relevant information of clause (1), but they must be disposed to mutually recognize that they share this information.

Take for instance the River Rat case. By displaying a rigged rat, Friend's intention was to produce the belief that the house is rat infested. If this intention was intended to be mutually recognized by him and Homebuyer, his display of the rat would surely count as a case of speaker-meaning (albeit quite a weird one). The same is true of the Moon Over Miami case, as well as of the Vigilante case: in all the problematic cases discussed, what is lacking is some sort of mutual awareness of what is going on. This condition has been defined by Lewis (1969) and Schiffer (1972) as 'common knowledge' and 'mutual knowledge' respectively. Michael Bacharach sums up Lewis and Schiffer's position as follows, discussing the case of people sitting at a table with a carafe in full view:

> « A normal human will not only see the carafe, but will also see the *normality* of the other co-present normals; lastly, normality has the reflexive property that it is part of being normal to know the perceptual and epistemic capacities of normal people. […] These characteristics of normality imply that in the carafe situation an inferential process is set in motion which leads asymptotically, if the agents are logically omniscient, to common knowledge. » (Bacharach, 1998, p. 309, quoted by Campbell, 2005, p. 291)

But, as Campbell rightly notes 'we have to acknowledge that agents are not usually logically omniscient' (2005: 291). Lewis and Schiffer's notions, which require perfectly logical communicators, have been interpreted as ideal states, states at which communicators should aim, for people should strive, as opposed to actually be in. This is not a problem for those concerned with the definition of an abstract model of communication. However, for those concerned with a psychologically realistic characterization, their definitions might remain unsatisfying. Let us note however that Bacharach's interpretation according to which Lewis's common knowledge requires logically omniscient agents is not a consensus: there are psychologically realistic interpretations of Lewis' common knowledge.[202]

A.4. SPERBER AND WILSON'S DEFINITION

This is what led Sperber and Wilson (1986) to their definition of *mutual manifestness*, which is meant to do the same job as Lewis and Schiffer's mutual/common knowledge while being at the same time psychologically more realistic.

This notion is defined through new notions brought by Sperber and Wilson: that of *cognitive environment*, which is itself defined in terms of *manifestness*. Here are the official definitions of these terms:

> « A fact [or assumption] is *manifest* to an individual at a given time if and only he is capable at that time of representing it mentally and accepting its representation as true or probably true. […]
>
> A *cognitive environment* of an individual is a set of facts [or assumptions] that are manifest to him. » (1986, p. 39)

[202] In fact, Lewis' original definition might itself be psychologically realistic, despite what Bacharach or Campbell state. Below, I discuss Lewis' notion of common knowledge (see also Paternotte, 2011).

They specify the first definition in the next sentence: 'To be manifest, then, is to be perceptible or inferable.' (39). [203] Now, here is how they define mutual manifestness:

> « Any shared cognitive environment in which it is manifest which people share it is what we will call a *mutual cognitive environment.* In a mutual cognitive environment, for every manifest assumption, the fact that it is manifest to the people who share this environment is itself manifest. In other words, in a mutual cognitive environment, every manifest assumption is what we will call mutually manifest. » (42)

So the carafe in full view is mutually manifest to the people at the table because (a) that there is a carafe is manifest to them, since they are all capable at that time to represent that there is a carafe as true, and (b) the fact that it is manifest that there is a carafe is itself manifest to everybody since they are all capable at that time to represent the following assumption as true: that the carafe being there is manifest to everybody.

The notion of manifestness is much weaker than that of knowledge, and this is a reason why mutual manifestness is psychologically more plausible than mutual knowledge (Sperber and Wilson 1986, p. 41). The notion of mutual manifestness doesn't require ideally logical agents as, supposedly, Lewis' definition does[204], nor the capacity to know infinitely higher-order thoughts (such as knowing that you know that I know that …) which Schiffer's definition of mutual knowledge requires.

Another advantage of mutual manifestness is that manifestness comes in degree (Sperber & Wilson, 1986, pp. 40–41), but knowledge doesn't. So the carafe in full view will be strongly mutually manifest, while the soft noise

[203] Perhaps this precision isn't without problems since there may be assumptions that are neither perceived nor inferred but that are still manifest. For instance, while reading a fantasy novel, it is manifest to me that a dragon has been killed, although I wouldn't say that I perceive this proposition, nor that I infer it since I perfectly know that it is pure fiction. Similarly, when I suppose that there is a possible world where donkeys speak Spanish, I am neither perceiving this proposition nor inferring it, but it is manifest to me in my reasoning. Perhaps, we should add other attitudes to perceiving and inferring, such as imagining, supposing, etc. This might require a change in the definition to include truth in other possible worlds. Green (2007, p. 79, n.4) however seems to think that it is good to be restricted to perceiving and inferring and that we shouldn't count any old propositions that we can represent as true or probably true as manifest. He gives the example of the representation that someone is walking toward him which he can represent as true without any evidence that there is such a person walking toward him. However, I don't see why this representation shouldn't count as being manifest to him when he is thinking about it.

[204] However, see below my discussion of Lewis' notion.

coming from the street, a piece of information which is much less accessible to our mind, will be mutually manifest, but less so.

The notion of mutual manifestness is not used to define speaker-meaning, but to define ostensive-inferential communication (Sperber and Wilson 1986, p. 63). However, a definition of speaker-meaning based on the latter may be developed (see also Green 2007, p. 79), which is the basis for the third variant of the definition of allower-meaning in Chapter 2:

> A sender S speaker-mean something by a stimulus x if, and only if, S produces x while
> (1) S intends x to make manifest or more manifest to a receiver R a set of assumptions[205] {I}, and
> (2) S intends x to make it mutually manifest to R and S that (1).

I will sometimes refer to this definition as 'Sperber and Wilson's definition' although one should keep in mind that this is just a shorthand for 'the definition of speaker-meaning adapted from Sperber and Wilson's definition of ostensive-inferential communication'. The adaptation, however, is minimal since (1) and (2) are completely unchanged, only the first line is adapted.

Despite the many advantages of this definition, some of which have already been discussed, a few difficulties may be raised.

A.4.1. A FIRST DIFFICULTY: THE FLATTENING SCHEME

The first difficulty concerns clause (1) and the fact that Sperber and Wilson replace the intention to generate *effects* (or responses) in the audience with the intention to make a set of assumptions *manifest*. As they rightly remark, this makes their intention (1) an *informative* intention. (1) is about making cognitively available a set of assumptions or an array of propositions (Sperber & Wilson, 2015). But, not all communicative acts aim to *inform* the audience. For instance, if I shout 'Keep going mate!', I might intend to encourage someone rather than to inform him of something. The real point of my speech act is not to inform him that I encourage him (although I do need, as a means, to inform him of that). Similarly, when I say 'Get out of my way!', I want this sentence to make the person move rather than to inform her about the fact that I desire that she moves. Conversely, I may intend to inform you about my desire without ordering

[205] In Sperber and Wilson (2015), they replace « a set of assumptions » by « an array of propositions ».

you to do anything. We may intend to do different things with words, as has been famously advocated by Austin (1962).

One of Austin's central claim was to debunk the 'age-old assumption … that to say something … is always and simply to state something' (1962, p. 12). This claim, in other words, is that what we intend to do with speech should not always be analyzed through intentions to inform or to describe the world in ways that are true or false. In Austin's terminology, we can have illocutionary intents (what we do *in* saying something) that are not informative intents. And what we speaker-mean, accordingly, can involve other intentions than informative ones.

So, the difficulty concerning Sperber and Wilson's definition is that these non-informative intents, which they don't take into account in their definition, are constitutive of what one means. Another way to put this is by saying that the representation conveyed by orders or encouragements, for instance, do not have the mind-to-world direction of fit, but rather the world-to-mind direction of fit (Searle, 1983, p. 7ff), so that their meaning cannot be captured by assumptions, which have a mind-to-world direction of fit. Still another way to put is is to say, as Millikan (1995) puts it, that orders express *directive representations* while assumptions express *descriptive representations*.

One could reply to this objection by remarking that even when we utter orders or encouragements, we do so by *informing* the audience of our different illocutionary intents. And it would be in this informative intention that the meaning lies, rather than in the intention to produce non-informative effects in the audience (making them move, encourage them, …). This is, roughly, the position taken by Lewis (1970), Davidson (1979), and Sperber and Wilson (1986), as well as, arguably, Green (2007).

This move is called the 'flattening scheme' by García-Carpintero (2004, 2015). Here is a quote which nicely presents the view and gives an argument against it:

> « Davidson (1979) and Lewis (1970) propose dealing with non-declaratives by taking them to be synonymous with explicit performatives. They propose that we should take the latter to have, from a semantic standpoint, their compositional truth-conditions. 'Take bus 44!' would just mean the proposition that the speaker thereby requests the audience to take bus 44. … This is what we call the flattening scheme, or simply flattening. We have argued that these views are unmotivated (García-Carpintero 2004, 2015). An assertion that a command is given or a question posed can occur

without the command being given or the question posed. Conversely, the non-cognitive attitude/act (the command or the question) can occur without the cognitive one (the belief/assertion that the command or the derogation was made), for instance because the thinker/speaker lacks the conceptual resources to describe the non-cognitive state/act. » (Marques & García-Carpintero, 2020)

We will discuss further the flattening scheme in 5.3. below while discussing Green's take on this move and explain why it is not satisfying.

A.4.2. A SECOND DIFFICULTY: NON-PROPOSITIONAL CONTENT

A second, related, difficulty is that there can be communicative acts that don't involve propositions at all. For instance, 'Wow!', 'Yuk!', 'Oi!', or 'Ouch!' seem not to express any propositions and the attitudes they express may well be non-propositional, as recognized by some relevance theorists (Saussure & Wharton, 2020).[206] So, when I utter 'Yuk!' I might speaker-mean something without intending to inform you about a set of propositions.[207] Propositions, in the context of the relevance theory literature, are syntactically structured representations made of conceptual constituents which are inputs and outputs of logical inferences (Reboul, 2017, Chapter 3; Sperber & Wilson, 1986, pp. 72–73). They are not sets of possible worlds or functions from possible worlds to truth-values (Lewis, 1970; Stalnaker, 1978). To discuss (my adaptation of) their definition, I will stick to how Sperber and Wilson and other relevance theorists use the word 'proposition'. However, the argument made here may be made with a less restrictive notion of proposition, for instance that developed in Camp (2018b).

The same reasoning applies beyond interjections to, e.g., cases where one wants to communicate about something that is not a complex state of affairs, but a simple object. For instance, when I point at a rainbow, I might want you to just look at the rainbow, an object, without intending you to entertain any proposition (such as 'there is a rainbow', 'a rainbow has been formed over there', 'this is a rainbow', etc.), I might refer to an object without predicating anything about it, without referring to a full-blown state of affairs (I take a reference to a state of affairs to minimally consist

[206] For other arguments as to why some emotions are non-conceptual and so, as the term is used here, non-propositional attitudes, see also e.g. Prinz (2004, Chapter 2) or Tappolet (2016, Chapter 1).

[207] Observe that I may also utter 'Yuk!' without the intentions necessary for speaker-meaning, and thus either non-intentionally allow this to mean something (Chapter 2) or produce a signal with non-Gricean meaning (Chapter 8). Here, however, we are interested in a case where I mean something by uttering 'Yuk!'.

in a reference to an object that is predicated). The content of my speaker-meaning in these cases arguably is objectual rather than propositional (Grzankowski & Montague, 2018; Montague, 2007).

Blakemore (1987, 2011) and Wharton (2009, 2016) have argued that non-propositional meaning, such as that of interjections like 'Wow!', 'Yuk!', etc., can still be captured by Sperber and Wilson's definition of ostensive-inferential communication. To do so, and to capture the non-propositional meaning of other expressions, Blakemore (1987) introduces the notion of *procedural meaning*. Her idea is that an expression either describes the world truth-conditionally and thus has a propositional meaning (or, for sub-propositional expressions, a conceptual meaning) or it allows the addressee to go through intended *procedures* that help her interpret the propositional meaning. For instance, according to her, a sentence like 'She is rich but unhappy' not only conveys the propositional meaning 'That she is rich & she is unhappy', it also conveys a procedural meaning through the expression 'but'. This would consist in information about *the inferential route* the audience should take to arrive at the intended propositional representation, which might be something like 'Note the conceptual tension between the fact that she is rich and the fact that she is unhappy', an inferential route which may lead to the retrieval of further, open-ended propositions. According to Wharton:

> « On this approach, the function of an interjection such as wow might be to facilitate the retrieval of a range of speech-act or propositional-attitude descriptions associated with expressions of surprise or delight. ». (2016, p. 16)

This means that, although interjections do not themselves carry propositional (or conceptual) meaning, they point to a way to acquire a propositional (or conceptual) meaning. This approach thus seems to allow keeping a definition of ostensive-inferential communication and of speaker-meaning that is, in the end, propositional, as the one proposed by Sperber and Wilson.

I agree with the claim that interjections can sometimes play this role, especially when they are coupled with other a sentence, as in 'Wow! He arrived on time.' or 'Yuk! I am not touching that.'. In such cases, the interjection can certainly play something like the procedural role described by Blakemore and Wharton. However, I think they also play other roles in determining speaker-meaning, ones which are not dependent on any propositions. First, as just mentioned, the content of an emotion may be objectual and so the interjections which express such emotions may, in

turn, have an objectual, rather than a propositional, object. Secondly, the content that is intended to be communicated in an expressive interjection may well be something that cannot be captured solely by propositions, because this content involves affective feelings that cannot be captured propositionally (as a set of syntactically structured conceptual representations). Third, if the flattening scheme does not work (see especially §A.5.3 below for why I think it doesn't), illocutionary forces may well unaccountable through propositions. The same remarks apply to other kinds of speech acts whose content seems to be non-propositional, for instance those which express non-propositional perceptive attitudes, which Green calls meaning-α (see §A.5 below).

Being restricted to an analysis of the meaning of non-propositional speech acts or non-propositional expressions that is strictly procedural is problematic as it forces on us a reading that is, in the end, propositional. This is problematic because the content of, say, emotional speech acts does not need to point to a proposition to be meaningful. Thus, the point that I made above remains: when I utter 'Yuk!' I might speaker-mean something without intending to inform you about a set of propositions. This difficulty cannot be solved through the notion of procedural meaning, however useful this notion otherwise is. It seems, then, that the definition adapted from Sperber and Wilson cannot account for this feature of speaker-meaning.

A.4.3. A THIRD DIFFICULTY: THE RIVER RAT AGAIN

A third difficulty, presented by Green (2007, pp. 79-80), is that this definition of speaker-meaning actually is not safe from the counterexamples such as the River Rat. Here is, again, the scenario:

> « The River Rat: Homebuyer is inspecting a house for possible purchase, and his friend—call him Friend—is concerned to convince him that the home is rat infested. Friend arrives at the house at a time when he knows that Homebuyer is inside, and although he knows he is being watched, skulks around to make Homebuyer think that he, Friend, believes he is acting unobserved. Friend has a river rat that he places in a salient position for Homebuyer to see. He intends for Homebuyer to see the rat and reason as follows: 'Although the rat display was rigged, Friend would not have put it there unless he believed that the house really is rat infested; hence Friend, who is reliable and honest, must intend me to believe that the house is rat infested. » (Green, 2007, p. 63)

This should not count as a case of speaker-meaning (nor of ostensive communication), but the definition given by Sperber and Wilson might not be able to exclude this case. To see why, remark first that Friend (F) produces a stimulus which makes a proposition manifest to Homebuyer (H), namely that the house is rat infested. So F has an informative intention in the sense defined in (1). F also intends to make it manifest to H that he has this intention: this is part of the reasoning he intends to produce in H. Furthermore, an argument given by Green (2007, pp. 80-81) purports to show that F also intends his informative intention to be *mutually* manifest, and so to respect clause (2).

Here is the argument. Let us call 'Inf' F's informative intention, i.e. the intention to inform H that the house is rat infested. Inf is obviously manifest to F because he knows what he is doing. Inf is presumably also manifest to H because, as the description of the case makes clear, F intends to make Inf manifest to H and surely can succeed. So Inf is manifest to both F and H. Now, since H presumably guesses that F is trying to get him to believe that the house is rat infested, it must be manifest to H that Inf is part of F's cognitive environment as well. Finally, because F believes that his behavior can be successful, he must also think that Inf can be part of H's cognitive environment. So it is manifest to both of them that Inf is manifest to both of them. This means that Inf is made mutually manifest by F's behavior according to Sperber and Wilson's definition of mutual manifestness because (a) Inf is made manifest to both of them, since they are both capable of representing Inf as true, and (b) the fact that Inf is manifest to them is itself made manifest to both H and F, since they both are capable to represent the following assumption as true: Inf is manifest to both of them. The last step of the reasoning is to argue that because F's intentional behavior makes it mutually manifest that Inf, then F must intend that Inf is made mutually manifest (in fact, Green doesn't argue for this last step). This would show that F's displaying of the rat respects clauses (1) and (2) of Sperber and Wilson's definition of speaker-meaning (or, more precisely, their definition of ostensive communication, and thus our definition of speaker-meaning adapted from it).

However, there are reasons to think that Green's criticism doesn't hold water because this last step is not warranted. Sperber and Wilson could argue that we cannot conclude from the fact that F's intentional behavior makes Inf mutually manifest to the fact that he intends to make Inf mutually manifest. To take a comparison (adapting an example from Searle, 1983, p. 99), it is not because Gravilo Princip's intentional behavior brought about World War I that he intended to bring about WWI. He certainly intended to kill Franz Ferdinand, to revenge Serbia, to shot his

gun, and more, including some doings that he might not have thought about while acting, such as stretching the muscles of his index finger. But there are facts brought about by his intentional behavior that he didn't intend, such as bringing about WWI.

One could reply that this is because he couldn't have known that he would bring about WWI with his intentional behavior, but F can know that his behavior will make his informative intention mutually manifest. Note however that there certainly are doings that Princip could have thought about which he doesn't intend either. For instance, Princip arguably didn't intend to displace H20 molecules in the air, to spill blood on Franz Ferdinand's carriage, to scare the pigeon passing by, or to make Franz Ferdinand's grandson the Emperor of Austria and King of Hungary, even though he could have thought that all of these facts would be brought about by his firing the gun. To take another example, Anscombe (1957, p. 10) discusses the case of a man sawing a plank. Even though he is aware that he makes squeaky noises by doing that, Anscombe argues that we cannot say that he intends to make squeaky noises because this is not part of the reason why he saws the plank.

Similarly, one can argue that F in the River Rat case didn't intend to make Inf mutually manifest to him and H, although this is what he ended up doing, and although it is something he could have thought about. Since clause (2) of the Sperber-and-Wilson-style definition requires that F *intends* to make Inf mutually manifest, according to this reasoning, we cannot conclude, unlike what Green argues, that Sperber and Wilson's definition forces us to say that F speaker-meant that the house is rat infested.[208]

This reasoning would thus save their definition from Green's argument. Let me note however that the exact range of actions which are properly called 'intended' given a certain intentional behavior is far from being a

[208] See also Bratman (1990)'s discussion of the difference in the intentions of the Terror Bomber – who intends to terrorize the war enemy by bombing a school – and the Strategic Bomber – who intends to bomb a munition plant at all cost and who is aware that doing so will also kill the children of the school next to the plant. Strategic Bomber worries a lot about this bad effect but concludes that the benefits will outweigh this tragedy. Bratman argues that the Terror Bomber intends to kill the children while the Strategic Bomber does not. A main reason why is that if the Strategic Bomber learns that there is a way for him to evacuate the school before his attack without compromising it, he will definitely change the course of his action, unlike the Terror Bomber. This shows, according to Bratman, that only the latter intends to bomb the school. A similar reasoning applies to River Rat scenario: if, for some reason, F learned that his action may not make his intention to inform H mutually manifest, he would nevertheless pursue. It is not part of his goals to make his informative intention mutually manifest.

consensus and that philosophers will be divided as to whether F did or did not intend to make Inf mutually manifest, especially perhaps if you take into account a broad view of what is intentional, such as O'Shaughnessy's (2008).

Another point should be made about Green's argument against Sperber and Wilson's definition. Putting aside the precise formulation with which Sperber and Wilson (1986, p. 39-41) define mutual manifestness, the way they put this concept to use (42ff) makes it clear that it is supposed to imply the following infinite hierarchy of mutual manifestness:

> If X is mutually manifest to A and B then:
>
> It is manifest to A that it is manifest to B that X
>
> It is manifest to B that it is manifest to A that X
>
> It is manifest to B that it is manifest to A that it is manifest to B that X
>
> It is manifest to A that it is manifest to B that it is manifest to A that X
>
> And so on...

Clark (B. Clark, 2013, p. 116), an advocate of Sperber and Wilson's relevance theory, in fact defines mutual manifestness with this infinite hierarchy. Now, if one agrees to define mutual manifestness with this infinite hierarchy, then Sperber and Wilson's definition of speaker-meaning is not subject to the River Rat counterexample, however one conceives of intentional actions, because F surely didn't intend his informative intention to be embedded in this infinite hierarchy.

Note that although this infinite hierarchy looks a lot like that of Schiffer's definition of mutual knowledge, Sperber and Wilson argue that mutual manifestness, unlike Schiffer's mutual knowledge, is psychologically plausible. A main reason is that manifestness is about cognitive environment – what is *available* for one's cognitive consumption – rather than about actual cognitive processes and states.

Note also that Lewis shows how his definition of common knowledge is not about actual cognitive processes, but rather about dispositions; dispositions to infer certain logical implications and dispositions to believe. This makes Lewis' notion very similar to Sperber and Wilson. Thus, the allegations of psychological unrealism are in fact not entirely warranted (see also Paternotte, 2011). Contrary to the definition of mutual knowledge

given by Schiffer (1972: 30-32), and contrary to some inaccurate interpretations, Lewis' common knowledge doesn't require that agents are able to know an infinite hierarchy of propositions (i.e. I know that you know that I know that … I know that p). Lewis (1969: 53) explicitly states that people don't need to actually entertain this infinite hierarchy, but that this hierarchy – which is indeed required by his definition of common knowledge – should be considered as a chain of implications which follows from a finite set of assumptions that people need to actually possess. And these assumptions themselves need not be actively entertained by the people who have common knowledge, but they should only be disposed to make these assumptions. Here is his definition:

> « Let us say that it is common knowledge in a population P that _____ if and only if some state of affairs A holds such that:
>
> (1) Everyone in P has reason to believe that A holds.
> (2) A indicates to everyone in P that everyone in P has reason to believe that A holds.
> (3) A indicates to everyone in P that ____. » (Lewis, 1969, p. 56)

A further background assumption is (4) everyone in P has a sufficient degree of rationality and has reasons to ascribe to everyone in P a sufficient degree of rationality. Together, these four premises can generate the potentially infinite hierarchy of common knowledge (see Lewis 1969: 54), or, rather perhaps, of common *belief*. What might be interpreted as psychologically unattainable, then is not the infinite regress itself, but the logical and rational capacities that agents must possess. However, one could be more charitable with Lewis in two ways: (1) one could interpret his supposedly logically perfect subjects as an idealization of what happens, just as when physicists assume that a rocket just is a point in space even though they are perfectly aware that it has a tridimensional shape, or (2) one could understand his 'sufficient degree of rationality' to not require ideally logical subjects, but only imperfect humans. With this in mind, Lewis' characterization of common knowledge is very similar to Sperber and Wilson's notion of mutual manifestness.

Nevertheless, there are reasons to prefer the notion of mutual manifestness. One is that 'knowledge' is a strong requirement. First, it is a propositional verb, so that no subpropositional object can be literally known (or believed). We saw that Sperber and Wilson only consider assumptions or facts as manifest but we could talk of a simple object as being manifest. Second, knowing is a so-called 'factive' or 'success' verb: if one knows p, then p is true. Sometimes, we might be in a situation where

an assumption is mutually manifest, but is not in fact true.[209] Third, if one knows something, then this is in one's 'head', so to say. This is not true for manifestness. Fourth, knowing-that seems not to admit of degrees. You know that p or you don't, but can you know that p more or less? Many would say 'no'.

Now, as I have remarked, although Lewis uses the expression 'mutual *knowledge'*, his definition only talks of *reasons to believe* and of *indication*, not of knowledge, even though it is the latter that he wanted to qualify. If we assume that he talks of mutual belief instead, half of the points I have just made don't apply, since beliefs come in degrees (we can more or less believe that p) and are not factive. Still, beliefs are propositional attitudes and are in one's 'head', contrary to manifestness. Furthermore, having *reasons to believe* might be too exigent. When two babies interact, some things can be mutually manifest to them, but they might not be cognitively sophisticated enough to have *reasons* to believe something – at least, that is what people like Sellars, Davidson, or McDowell would say.

Now, the main problem I find with common knowledge is that one cannot use it in a definition of speaker-meaning to exclude counterexamples such as the River Rat. To see why, let 'Inf' be that 'Friend (F) intends to make Homebuyer (H) believes that the house is rat infested' and let us replace the variables in Lewis' definition as follows: '_____' by 'Inf', 'A' by 'F's behavior', and 'population P' by 'H and F'. Now, the following seem to hold (despite some awkward formulations):

(1) H and F have a reason to believe that F's behavior hold.
(2) F's behavior indicates to H and F that H and F have reason to believe that F's behavior hold.
(3) F's behavior indicates to H and F that Inf.

So, according to Lewis' definition, it is common knowledge that Inf, i.e. that the intention of Friend to make Homebuyer believes that the house is rat infested. Since F doesn't speaker-mean anything here, we cannot use Lewis' mutual knowledge for the role that mutual manifestness is supposed to play in the definition of speaker-meaning.

So, in sum, we have seen two reasons why the argument given by Green to reject mutual manifestness would not hit its target, but why it hits instead Lewis' common knowledge. Since Schiffer's definition of mutual knowledge

[209] We will see below that I use the verb 'to recognize' to describe such phenomena and, unfortunately, and this might as well be a success verb. If it is one, then let me use it in an untraditional sense where we can recognize something that doesn't exist or isn't the case.

is definitely not psychologically realistic, Sperber and Wilson's mutual manifestness seems, at least until now, the best candidate available to avoid the River Rat example.

A.4.4. A FOURTH DIFFICULTY: THE TWO GENERALS

Before we move on, let me discuss another argument, given by Jankovic (2014), to the effect that mutual manifestness cannot achieve its task in all cases. I don't think that this argument hits its target. Instead, it allows us to precise what mutual manifestness is and highlight some of its advantages. This argument can also be used to successfully discard Lewis's common knowledge or Schiffer's mutual knowledge. The argument is based on the so-called 'Coordination Attack Problem' or 'Two Generals' Problem'. My presentation is adapted from Jankovic (2014).

Two armies are preparing an attack on a common enemy. They can easily defeat it if they attack together. If each attacks separately, on the other hand, it would lead them to a catastrophic defeat. The generals of the two armies thus have to communicate the time of their attack. The problem is that the only communication channel passes near the place where the enemy army seats. The enemy is thus able to intercept a certain proportion of the messages sent. Each message has the same probability π of being intercepted. When a message reaches one of the armies, a confirmation of the receipt is sent automatically to the other army, which can also be intercepted by the enemy. Furthermore, let us assume that if a message is intercepted, it is automatically destroyed before the enemies can read it. All of this is known by the generals of both allied armies.

If a general (let us call her Jette) sends the message saying '5 a.m. tomorrow' and that the second general (let us call him Henri) receives it, it would be natural to say that Jette *has* communicated the time to Henri (even if she is not sure she did) and *has* speaker-meant something. However, according to Jankovic, Sperber and Wilson's definition of mutual manifestness and ostensive-communication force us to say that Jette has *not* communicated anything to Henri, even though he has received her message. Furthermore, if she is right, our definition of speaker-meaning based on Sperber and Wilson would wrongly predict that Jette has *not* speaker-meant anything.

To see why this is so, let us consider a scenario where Henri has received Jette's message saying '5 a.m. tomorrow' (she has thus succeeded in both communicating and speaker-meaning something). Henri automatically sends a confirmation message back, but the confirmation has been

intercepted. For Jette, the fact that she didn't receive any confirmation message means either that (a) her message didn't reach Henri or that (b) Henri received the message but his confirmation message didn't get through. This is all manifest to Jette.

The following is also manifest to Jette: if (b) is the case, Henri has received only one message, and will thus think that either (c) his confirmation message didn't go through or that (d) Jette has received his confirmation, but her confirmation of the confirmation didn't go through. Furthermore, Henri is good with conditional probabilities and is thus capable to figure out that, no matter what is the probability π that a message is intercepted, (c) is more probable than (d).[210]

Now, Jankovic (2014) argues that the following claim is, or should be, held by Sperber and Wilson:

(e) a proposition is not manifest in an environment E if it is more likely to be false than true given the evidence in E.

This might seem rather intuitive (however, I will give reasons to reject this claim). For instance, if I see that it is dark outside, given my evidence, it is more likely that the proposition that it is daytime is false than true, and this would make the proposition that it is daytime *not* manifest to me. Furthermore, Jankovic argues that this works rather well with the definition given by Sperber and Wilson (1986, p. 39) according to which to be manifest is to be perceptible or inferable. If the evidence for a proposition p makes it more likely to be false than to be true, then one should not infer that p, which would presumably make p not inferable.

If this is correct, it leads us to conclude that, in this scenario, Jette will conclude that (d) – that she has received Henri's confirmation, but her confirmation of the confirmation didn't go through – is not manifest to

[210] See Jankovic (2014, n. 12 and n. 13) for the calculation of the probabilities. Here is a rather what I take to be an illustration of why that is the case: let us imagine there is a 10% chance that a message is intercepted. So, on average, if 100 are sent, 90 go through and, thus, 90 confirmations are sent back. 10% of these 90 confirmations are in turn intercepted, i.e. 9 confirmations. So, on average, for 100 messages sent, 10 of the first messages are intercepted and 9 of the confirmations are intercepted. So the probability that (a) Jette's first message didn't go through is 10%, while the probability that (b) Henri's confirmation didn't go through is 9%. If we continue the reasoning, we can observe that, for 100 messages sent, Jette will have received 81 confirmations from Henri, and will thus send 81 confirmations that she received his confirmations, 8.1 of which will be intercepted on average. So the probability that (c) Henri has received Jette's first message, that his confirmation has been received, but that her confirmation that she received his confirmation has been intercepted is 8.1%. In other terms, there is an 8.1% chance that the second message sent by Jette has been intercepted. Following the same reasoning, the probability that (d) Henri's second message has been intercepted is 7.29%.

Henri, since it is more likely to be false than true, as (c) is more probable. In this situation, according to Jankovic, Jette should thus conclude that her message is not mutually manifest, because she thinks either that (a) is true, and thus that Henri has not received her message, or that (b) is true, which would mean that the following is not manifest to Henri: that it is manifest to Jette that her first message is manifest to him. By definition of what mutual manifestness is (whether we define it through the infinite hierarchy as Clark (2013) or as Sperber and Wilson (1986, p. 39-42) do it), we should conclude that Jette's first message is not mutually manifest. And since Jette's message is the only way Jette could have made her informative intention manifest, her informative intention is not made mutually manifest.

If this were true, then, even though Henri *did* receive her message, Sperber and Wilson's definition would predict that Jette hasn't communicated anything. And, according to our definition of speaker-meaning based on their definition, Jette wouldn't have speaker-meant anything. These unfortunate conclusions would force us to abandon the concept of mutual manifestness.

However, I am not convinced by Jankovic's argument, because I don't agree with one of her premises, namely the claim (e) that a proposition is not manifest in an environment E if it is more likely to be false than true given the evidence in E. Instead, the way I interpret Sperber and Wilson's notion of manifestness entails that a proposition that is more likely to be false than true can still be manifest, even though it would be less manifest than a proposition that is more likely to be true (see Sperber and Wilson 1986, p. 41ff). If my interpretation is correct, we should say that, in the scenario I have discussed, where Jette has sent a message but received none in response, then, since she is good at doing probabilities, she should have the following reasoning:

> 'Either (a) Henri has not received my message or (b) he did but his confirmation didn't go through. If (b), he has received just one message, and should thus conclude that it is more probable that (c) his confirmation didn't go through than (d) his confirmation did go through, but my confirmation of having received his confirmation didn't. By consequence, if (b) is true, (c) is more manifest to him than (d).'

She thus wouldn't conclude that (d) is not manifest to Henri. By consequence, she could still think that, in case (b) is true, her informative intention is mutually manifest to them. So, according to my interpretation,

Sperber and Wilson's definition gives the correct prediction: if Henri didn't receive Jette's first message, then communication hasn't succeeded because Jette's informative information wouldn't have been made manifest to Henri. On the other hand, if Henri did receive her message, some communication would have been achieved, since this would be enough to make Jette's informative intention mutually manifest. Indeed, the following would be true:

(1) It is manifest to Jette that Jette has an informative intention.
(2) It is manifest to Henri (1)
(3) It is manifest to Jette that (2)
(4) It is manifest to Henri that (3)
(5) It is manifest to Jette that (4)
(6) And so on...

The proposition (3) is in tension with the fact that Jette doesn't know whether Henri has received her message. And the proposition (5) is in tension with the fact that Jette knows that Henri knows that it is more probable that his confirmation didn't go through, since his confirmation is supposed to give evidence for (3). However, neither (3) nor (5) are contradicted by these facts, contrary to what Jankovic seems to think.

As I said above, this discussion actually highlights some advantages of (mutual) manifestness: the fact that Jette's reasoning is in tension with (3) and (5) is just how things should be, because even though Jette's informative intention is manifest to both of them, there is a risk that they haven't communicated. What goes on in Jette's head *is* in tension with the fact that, in the scenario I have presented, they actually did communicate. The tension between the definition of mutual manifestness and what Jette knows thus seems to reflect Jette's psychological situation. Furthermore, another advantage of mutual manifestness illustrated by this discussion is that it is flexible enough to allow that two contradictory propositions are mutually manifest. In this case, for instance, the proposition that Henri's confirmation didn't get through. It is not possible to mutually know two contradictory propositions, but it happens very often that two contradictory propositions are credible enough to both be mutually manifest.

A.4.5. A FIFTH DIFFICULTY: THE ABSENCE OF AN AUDIENCE

Let us now turn to a last argument against Sperber and Wilson's definition of ostensive-inferential communication/speaker-meaning. This one actually applies to all definitions above: it is based on the claim that speaker-meaning can occur without any intention to produce an effect in

an audience and purports to show that we need to exclude a reference to an audience from a definition of speaker-meaning. This criticism has been raised by several researchers (Davis, 1992; Green, 2007, p. 60; Hyslop, 1977). It applies to our definition based on Sperber and Wilson's definition insofar as the first clause requires that the utterer intends to make something *manifest* to an audience, i.e. modify their cognitive environment, and thus to produce a cognitive effect on an audience.

To sustain this claim, Green (2007, pp. 60-61) gives several examples where the speaker need not intend to produce an effect in the audience that is present. For instance, a suspect might mean that she is innocent by saying 'I'm innocent!' while being fully aware that she won't convince anybody present with this claim and so, being realistic, not intending to affect anybody's thoughts. A parent might look at her newborn daughter and say 'All things valuable are difficult as they are rare.' meaning what she says, without the intention to produce a belief or any other effect in the newborn or anyone else. Finally, in Woody Allen's film *Sleeper,* a character comes across, while exploring alone, a genetically modified chicken the size of a house and utters 'That's a big chicken.' He probably means what he says but there is, once again, no audience to be affected.

Let us now see Green's definition and how he can deal with such cases.

A.5. GREEN'S DEFINITION

Green argues that what is essential to speaker-meaning is not about affecting an audience but, instead, is about doing something *overtly* (65), making everything of relevance publicly discernable, available out there in the open, but not necessarily discerned by anyone. What needs to be made overt includes the subject matter of the meaning (what we refer to, predicate, ask about, etc.) as well as the intention to make this overt. Another way to put it is that in speaker-meaning, the utterer intentionally *shows* something and that this intentional showing is itself intended to be shown. This doesn't require an intention to affect the audience, or, in fact, any kind of intention that is audience-directed.

One possibility to cash out the notion of overtness which Green is after is similar to Schiffer's (1972) by having the following structure:

(a) I intend to show my disgust,
(b) I intend that my intention (a) be publicly discernable,
(c) I intend that my intention (b) be publicly discernable,
(d) And so on infinitely.

However, this infinite hierarchy is not a psychologically realistic description, since one cannot have an infinite number of intentions at once. Instead, Green proposes the following solution. He defines an overt action as one which is done:

> « intending that (a) something be publicly discernible, and (b) this intention itself be publicly discernible as well. » (2007, p. 66)

Thus, doing something overtly requires only *one* intention, but this intention has a content which is referring, inter alia, to this very intention. The content of the intention mentioned in (b) will, of course, be re-referring back to itself. This creates a sort of reflexive loop. This loop, Green argues, has the same advantage as the infinite hierarchy of Schiffer (or Lewis or Sperber and Wilson) in the sense that it will allow Green's definition to avoid counterexamples such as the River Rat or Moon Over Miami (however, depending on how it is interpreted, we will see that there are reasons to doubt that).

Take the River Rat case: Friend intends to make publicly discernable that the house is rat infested, but he doesn't intend this very intention to be made publicly discernable. He doesn't want to show to Homebuyer the very intention that makes him put the rigged rat in the house. The same reasoning applies, mutatis mutandis, to the Moon Over Miami case. The intentions of the awful singer are not 'out there in the open', he hides an intention to his friend, or at least doesn't intend to make it publicly discernable.

Despite its playing the same role as Schiffer's infinite hierarchy of intentions, Green argues that the self-referentiality of the intention involved in his definition of an overt action is psychologically realistic (again, we will later see that there are reasons to doubt that). To show this, he proposes to consider as an analogy an example from Peacocke (2005): after a disastrous car accident, the person regains consciousness and thinks 'This thinking is miraculous!'. Although this thinking is reflexively referring to itself, just like the intention in Green's definition, this doesn't prevent this thought to be entirely graspable and psychologically realistic.

With these tools in hand and a few others that I won't present here, Green gives three different definitions of speaker-meaning, each corresponding to a kind of overt showing: overtly showing a fact (e.g. that it is raining), an object (e.g. a bird), or one's commitment (e.g. that I am committed to a proposition under the force of an assertion). Let me now reproduce his definitions:

« *Factual Speaker Meaning:* Where P is an actual state of affairs, S factually speaker-means that P iff

1. S performs an action A intending that
2. In performing A, it be manifest that P, and that it be manifest that S intends that (2). » (67)

« *Objectual Speaker Meaning*: S objectually speaker-means α iff

1. S performs an action A intending
2. α to be manifest, and for it to be manifest that s/he intends (2). » (68)

« *Illocutionary Speaker Meaning:* S illocutionarily speaker-means that P φ'ly, where φ is an illocutionary force, iff

1. S performs an action A intending that
2. In performing A, it be manifest that S is committed to P under force φ, and that it be manifest that S intends that (2). » (74)

Unfortunately, here is not the place to discuss the many virtues of these three definitions (see Green 2007, Chapters 3, 4 and 6). While acknowledging the important differences between these three definitions and the value of keeping them apart, we can nevertheless find a common structure to them,[211] which is the following:

S speaker-means something iff

(1) S performs an action A intending that
(2) In performing A, this something be manifest and it be manifest that S intends that (2).

The fourth variant of the definition of allower-meaning in Chapter 2 is based on a reformulation that is meant to respect the main distinctive features of Green-style definitions while being as close as possible to the formulation of the other definitions (Grice's, Neale's, and the adaptation from Sperber and Wilson). Here it is:

S speaker-means something by x if, and only if, S produces x intending
(1) x to make something manifest, and
(2) to make it manifest that S intends (1).

[211] Assuming that 'speaker-meaning that P φ'ly' is equivalent to 'speaker-meaning that S is committed to P under force φ'. This may be debatable, but let us put this issue aside for now.

Like Sperber and Wilson's definition, Green's possesses many advantages, and it avoids the problems that we have raised concerning the former. As I have already mentioned, I consider it as the most sophisticated definition of speaker-meaning that I know of, which is also the reason why I will discuss it at length. However, despite its virtues, I can see three difficulties.

A.5.1. A FIRST DIFFICULTY: THE MODIFIED RIVER RAT

The first difficulty concerns Green's reflexive loop. The issue would be that we can interpret it in two ways, each being problematic: Under the first interpretation – a *de dicto* reading – it would not be psychologically realistic, and under the second interpretation – a *de re* reading – it would not be able to achieve the job which it is supposed to achieve, i.e. to make everything of relevance be in the open, publicly discernable, so that the communication be wholly overt.

Since the counterexample is based on the distinction between *de re* and *de dicto* attitudes, let me introduce it with three examples. First example: When Sarah sees a spy without knowing that he is a spy, she sees a spy *de re*, but she doesn't see a spy *de dicto*. She wouldn't report that she has seen a spy, although she actually did. The *de re* reading of 'Sarah sees a spy' picks out the person in question without qualifying him as a spy, without attributing to the man the property 'is a spy'. On the *de dicto* reading 'Sarah sees a spy' implies that she attributes to the man the property of being a spy, she sees him under the description that is specified in the verbal object of the sentence.

Second example: If Joe wants to marry the tallest girl in town, on a *de dicto* reading, he wants to marry whoever fulfills the description 'the tallest girl in town', it is somehow important for him to marry someone who possesses this property. By contrast, on the *de re* reading, Joe wants to marry the person which this description happens to pick out even if Joe does not think about her under this description. On the *de dicto* reading, if a new girl arrives in town and is taller than the girl who Joe wanted to marry, Joe will want to marry this new girl, but, on the *de re* reading, the new girl arriving won't change who he wants to marry.

Third example: The intention to make something manifest can also be *de re* and *de dicto*. On the *de re* reading of 'You notice my intention to communicate about Bob's funny hat', you need to have picked out the actual intention which is about Bob's funny hat, but you don't need to have noticed it *under that description*. You don't need to know that the intention you have noticed is about Bob's funny hat. On the other hand, on the *de*

dicto reading, you need to have noticed my intention under that description, to have attributed to my intention the property of being about Bob's funny hat. Thus, if you see me moving my arm in a certain way and notice that I have an intention to communicate something without understanding *that I intend to communicate about Bob's funny hat*, you will have noticed *de re* my intention to communicate about Bob's funny hat – you will have noticed that very intention – but you will not have noticed *de dicto* that I had the intention to communicate about Bob's funny hat. So, if I intended to make my communicative intention manifest *de re*, I will have succeeded, but if I intended to make it manifest *de dicto*, I will have failed.

In all cases, what distinguishes the *de re* reading from the *de dicto* reading of these sentences, is that the *de re* attitudes only pick out the relevant object (*this* man over there, *this* girl, *this* intention) without any requirement to attribute to this object the properties that are described in the sentences. By contrast, the *de dicto* reading of these sentences not only require that the subject has picked out the relevant objects, but has also attributed to them the relevant properties, has seen, desired, or noticed the objects *under the description* specified in the sentence (the man *is a spy,* the girl *who is the tallest girl in town*, the communicative intention *is about Bob's hat*). I believe that once we make this distinction, Green's reflexive intention runs into trouble.

To see why, let us first notice that, in the example of the car accident, the reflexive thought is easily graspable only under a *de re* reading. Remember the example: a person has a car accident, wakes up, and has the following thought 'This thinking is miraculous!'. The thought (which happens to be miraculous) is reflexive because this thought is about itself, so it is a thought about a thought. And the latter thought about which the former is, is about itself, so about a thought. And the latter also is about a thought. And so on. As Green (2007, p. 67) writes: « the content of this thought token cannot be finitely rewritten in a way that eschews all reference to the thought token itself ». In other words, if we *had* to describe the content of the thought without using an indexical ('this'), if we could not be satisfied with a reference to a thought without a qualification of the object about which this thought is, we would need an infinite description.

Now, if we only require a *de re* reading of this reflexive thought, the thought need not be qualified under that description. The property of being a reflexive, infinite loop, one whose content 'cannot be finitely rewritten', doesn't need to be attributed to the thought. The person only needs to pick out *this thought*, which happens to be a reflexive thought, without thinking about that. This reading is credible: the person just had a car accident,

notices that she is thinking, and thinks about *this* thinking that it is miraculous.[212]

On the other hand, if we require a *de dicto* reading of the reflexive thought, the story is not credible anymore. In this case, the thinker must not only pick out the thought in her mind, she must also ascribe to it the relevant property, she must think about it under the description that this thought is about her thought. And, since we have a *de dicto* reading of the latter thought, we need to qualify it under the relevant description, which is that it is about her thought. So, if the person thinks *de dicto* about her reflexive thought, she must have a thought to which she ascribes the property of being about her thought, and ascribes to the latter the property of being a thought about her thought, and so on *ad infinitum*. Contrary to thinking *de re* about a reflexive thought, thinking in a wholly *de dicto* way about a reflexive thought is not psychologically realistic.

Now, the same reasoning applies to Green's reflexive intention. If it is to be read as a *de re* attitude, then it is psychologically realistic, but not if it is to be read as *de dicto.* Here is again the reflexive intention: When one speaker-means, one needs to intend that (a) something be publicly discernible, and (b) this intention itself be publicly discernible as well. If we are to read this intention as *de re,* clause (b) only requires that the intention be made manifest in a way that would allow someone to merely pick it out, to identify it, and token it as *this* intention that the speaker has. But if it is to be read as a *de dicto* intention, then clause (b) means that, to be fulfilled, the intention should be manifest in a way that allows the potential audience to ascribe to it the relevant property, which is that it refers back to an intention, and the latter intention must itself be qualified as being about an intention, which itself needs to be qualified as an intention about an intention, and so on *ad infinitum*. The *de dicto*

[212] Julien Deonna (personal communication) observed that, even under the *de re* reading, we may have doubts that such a reflexive thought is entirely possible. He suggested a 'higher-order thought' reading of 'This thinking is miraculous.' The person would wake up from the car accident, have some unqualified thought (or perception or feeling or …), and, at the same time, have a higher-order thought which is about the unqualified thought and which would qualify it as miraculous. The higher-order thought would happen at the exact same time and, from an introspective point of view, would not be (easily) distinguishable from the lower-order thought that is qualified as miraculous (just like the higher-order thoughts that define consciousness according to e.g. David Rosenthal). According to this reading, there wouldn't be any reflexive thoughts involved in the example because the higher-order thought is not identical to a lower order thought. Deonna seemed to think that only this reading of the exampled is realistic.

reading thus seems psychologically unrealistic. So it may seem that Green would be better off with a *de re* reading.[213]

However, a *de re* reading of the intention used in his definition of speaker-meaning runs into another kind of trouble. To see this, consider the *Modified River Rat* case. Let us assume that Friend has a first intention that (a) it be manifest that the house is rat infested and that (b) this very intention be itself manifest, and let us interpret this on a *de re* reading. Now, Friend has a second intention that Homebuyer *thinks* that the first intention is not manifest, but covert. This is not contradictory. Friend, in behaving as he does, intends, firstly, (a) to provide publicly discernable evidence that he acts in such a way as to make a fact manifest and (b) that this very intention (the intention token) is made publicly discernable as well, and, secondly, he intends (c) to fulfill this first intention in a twisted way: by acting *as if* he wants the first intention to be hidden, although he doesn't want it to be hidden, as per (b).

Although Friend doesn't have the same intentions as in the original River Rat case, Friend can, it seems to me, have and indeed fulfill these two intentions – i.e. the intention that (a) and (b), and the intention that (c) – by acting exactly as he did in the original scenario. He places a rigged rat in the house at a time when he knows that Homebuyer will notice him doing so, but pretending that he is acting unobserved. He intends Homebuyer to think as follows: 'Friend would not have displayed this rigged rat here unless he intended to make it manifest that the house is rat infested, so the house must be rat infested.' Thus, Friend fulfills his intention that (a) it be manifest that the house is rat infested – since Homebuyer has inferred that it is – and that (b) this intention itself be manifest – since Homebuyer has identified the intention of Friend to make him think that the house is rat infested. Homebuyer doesn't ascribe to this intention the property of being reflexive (being about itself), but that is ok on a *de re* reading of Friend's intention to make this intention itself manifest. Friend only wanted to make manifest *this very intention,* the intention token, without requiring that it be qualified as having this or that content. Thus, on a *de re* reading, Friend can have and fulfill an intention that (a) and (b).

[213] This is the interpretation that I thought Green would favor, in particular because of the way he introduces the 'This thinking is miraculous' example: « The content of this thought refers to the thought *token* (*a particular thinking* with a spatiotemporal location, or at least a temporal location, and a content), and says of *it* that it is miraculous. It will, then, be true just in case *that very thought token* is miraculous. » (Green, 2007, p. 67, my italics). However, he has told me that he thinks he favors a *de dicto* reading, although he has not given an alternative interpretation to the problematic one that I give here.

Friend thus satisfies the criteria for speaker-meaning according to Green's definition. However, what makes me want to exclude this case from speaker-meaning is the further intention that Friend has: to make Homebuyer *think* that he intends this first intention to be covert. And, by behaving as if he is acting unobserved, Friend succeeds in making Homebuyer think that he intends the first intention to be covert. Although there is an intention to make something manifest and that this intention itself be manifest (an intention absent from the original River Rat scenario), there is a further intention which somehow cancels it: the intention that one *thinks* that this intention is covert.

So, once again, we seem to be stuck between the same rock and the same hard place: either we have the problem of a psychologically unrealistic constraint (on the *de dicto* reading), or the requirements are not sufficient to exclude cases where not everything of relevance is in the open, where the communication is not 'wholly overt' (on the *de re* reading).

We have seen that there are ways to interpret Green's definition of overtness as well as Sperber and Wilson's definition of mutual manifestness, Schiffer's definition of mutual knowledge, and Lewis's definition of common knowledge in ways that make them problematic. But, despite the problematic interpretations that they allow, they all may be interpreted as referring to the same phenomenon, namely a sort of publicness that is essential to speaker-meaning. This phenomenon seems to be, pre-theoretically, rather well understood; it is the precise formulations of these notions which can lead to problematic interpretations. To be sure to avoid the potential threats faced by the ways these notions have been defined, but to still be able to refer to the phenomenon they intend to capture, I thus propose to hereafter use a neutral expression: '*mutually recognizable'*. I won't define it otherwise than by stating that it refers to the psychologically real phenomenon which is essential in speaker-meaning and which is meant to be captured by 'mutually manifest' and 'overt'.[214]

[214] This move might seem like I just lazily give up on giving a definition of the phenomenon. This is only partially true. I do give up on the definition project, but it is not just laziness. I am actually tempted by Campbell (2005)'s suggestion that this phenomenon might well be an undefinable primitive component of our experience. According to him, joint attention might be qualified at a sub-personal level (talking about brain mechanisms, for instance), but not a personal level. At a personal level, all we can say is that we *feel* that there is a joint attention with this or that other agent. If this is true, we might as well take joint attention as a primitive notion, as a way it feels like to be in certain situation with other people or animals, a primitive phenomenon of our experience which then allows us to define common knowledge or other such notions. Even Green's overtness which does not require the presence of an audience and thus does not

My choice of terms however is not innocent. I believe that Schiffer and Sperber and Wilson, by pointing toward *mutual* knowledge and *mutual* manifestness highlight what I find to be intuitively essential to overtness: the possibility of a completely open *triangulation* between (at least) two communicators and the subject of their communication. Each communicator is disposed to be aware of everything of relevance about the other communicator and the subject of communication. Intuitively, this strikes me as what we need for genuinely *overt* signaling. By talking about *mutual* 'recognizability', I want to highlight this intuition.

It might be the case that a certain reading of Green's reflexive intention, one which has escaped me, also succeeds in capturing the same phenomenon. However, I find that the explicit reference to something *common* or *mutual* among potential communicators captures the phenomenon of overtness more intuitively than Green's reflexive intention. What prevents Green from using this sort of notion and leads him to define overtness through a reflexive intention instead is that he doesn't want to appeal to an audience. This is this issue that we will now address.

A.5.2. A SECOND DIFFICULTY: THE PRESENCE OF AN AUDIENCE

The second difficulty with Green's definition(s) concerns the idea that the effects in the audience should not be part of the definition of speaker-meaning. The point – which may be considered as closer to an observation than an objection – is that Green's definition doesn't put all the cards on the table, so to say, because it would in fact need to implicitly refer to effects on an audience. This is because he offers an account of speaker-meaning as kinds of overtly showing, and his account of showing does refer to an audience, and effects on this audience. The audience required by showing might be conditional or virtual, i.e. not actually present, but a reference to them is still required in his account of showing. Green's definition of speaker-meaning thus seems not to escape an implicit reference to effects on the audience, and by excluding this reference from his definition, we lose track of important variables.

According to Green (2007, pp. 47ff), showing comes in (at least) three kinds: (a) *showing-that*, which is a conceptual (or propositional) type of showing, (b) *showing-a*, a perceptual showing, and (c) *showing-how*, an experiential showing. In each case, and this is what is important for us here, to analyze what it is to appropriately show something, we need to refer to an audience,

require a feeling of joint attention, might be cashed out through Campbell's joint attention: something would be overt to a conditional audience if, were the conditional audience present, we would have this feeling of joint attention.

be it present or not, that have the capacities to (a) conceptualize what is being conceptually shown, (b) perceive what is perceptually shown, or (c) experience what is experientially shown. Let me detail this point.

An example of showing-that is showing that there is a black hole in the center of our Milky Way through extensive calculation. Another is showing that I am brave through my behavior. In these cases, one provides compelling evidence for a conclusion. But this evidence must be compelling with respect to a certain audience. I cannot show that there is a black hole in the center of our galaxy to a horse, because the horse cannot understand the calculation used for this demonstration. This doesn't mean that an audience must be present or that the audience present must get that the evidence shows-that; it can be a conditional audience (called 'virtual audience' in Green 2007, p. 62, see also the discussion in Schiffer, 1972, pp. 73-80, and Hyslop, 1977). By conditional audience, I mean that, were the appropriate audience present, they would be shown-that through the display of the evidence. More specifically, the idea is that showing-that implies succeeding in displaying evidence which is such that, were an appropriate audience present (an audience with the right cognitive make-up, that is not too inattentive, not too tired, etc.), they would be compelled by the evidence. For instance, a math teacher might show to his class that $a^2+b^2=c^2$ in a right triangle by proving a theorem even though nobody in the class follows the proof (Green, 2007, p. 47). The fact that nobody is paying attention doesn't prevent the teacher to have succeeded in showing his class that $a^2+b^2=c^2$ because she succeeded in presenting compelling evidence with respect to this audience.[215] In fact, a student might remark: 'Oh, last week she showed us that $a^2+b^2=c^2$, I just didn't follow at the time.' (*idem*) Again, if the audience were composed only of horses, the teacher could not have shown them that $a^2+b^2=c^2$. This is because the proof could have produced the relevant cognitive effects in a normal class, but not in a class of horses. Similarly, a scientist writing by herself in her notebook can be said to have shown that there is a black hole in the middle of the Milky Way even though there is no audience, because she has succeeded in producing evidence such that, were the appropriate audience present, they would be compelled by the evidence. The appropriate audience here refers to the people to whom the scientist could be addressing such evidence, an audience of informed, curious astrophysicists, or perhaps a broader

[215] Some might disagree with Green's intuition here and thinks that the teacher has failed to show them that $a^2+b^2=c^2$. The student should thus think « Oh, last week the teacher has *tried* to show us that $a^2+b^2=c^2$, but we didn't follow. » The teacher would have shown them that $a^2+b^2=c^2$ *only if* they get it. If that is correct, it only brings more grind to my mill.

audience, depending on the knowledge required to be compelled by the evidence displayed in the notebook.

Let us now turn to showing-α, or perceptual showing. Examples include showing you my bruise (when you look at it) or showing you the texture of my shirt (when you touch it). Once again, one can show-α only so long as there is an audience that has the appropriate faculty, which is perceptual. As Green writes, « even if there are mice in the field, I don't show you them from an airplane passing two hundred yards above the field. On the other hand, if you had the visual acuity of a hawk, I might well do so. » (2007, p. 48). Or: « If you had electroreception like a hammerhead shark, I could show you the electrical activity in the body of a fish hiding under the sand. » (*idem*). Again, as for showing-that, Green's analysis of showing-α makes reference to the capacities of an audience and specifically to their capacities to be affected in a certain way.

Finally, the third kind is show-how, or experiential showing: by composing a song, I might be able to show you how I feel. The trepidation of my voice might also show how anxious I am. « Showing-how can provide qualitative knowledge for those with appropriate sensory capacities. It can also enable empathy for those with the capacity for empathy. » (*idem*). Again, showing-how is defined through a reference to effects achieved on an audience, these effects here being experiential.

So, in all three kinds of showing, the analysis requires, or, at least, makes use of, a reference to an audience and the capacities of this audience to be affected by what is shown, whether conceptually, perceptually, or experientially. In other words, showing is defined through effects on an audience, be it a conditional audience, in the sense that it is not present, but would be the appropriate audience.

Now, as he writes (2007, p. 82), he defines speaker-meaning as kinds of overt showing – overtly showing a fact, an object, or one's commitment. So, although a reference to the effects on an audience is explicitly excluded from his definition(s) of speaker-meaning, it seems that these return through the backdoor, since the notion of showing, which is used as an *explanans*, itself must be explained through effects on an audience.

Observe however that a conditional audience is not really a kind of audience, for the same reason that a prince is not a kind of king, even though he is, in some sense, a potential king. This is why my point here may be considered as an observation rather than an objection. We may well grant to Green that no actual, real audience needs to be present for one to speaker-mean anything and at the same time recognize my point that a

reference to an actual or a conditional audience is necessary. Remember the cases of the parent talking to her newborn baby, of the suspect which states 'I am innocent!' knowing perfectly well that she won't change the mind of her audience, or of the Woody Allen character who says 'That is a big chicken.' even though he is by himself. These cases constitute good reasons to exclude an audience from the definition of speaker-meaning if one means an *actual* audience, but not if one means a *conditional* audience.[216]

In fact, Green himself agrees that a notion of speaker-meaning which would refer to a conditional (which he calls 'virtual') audience might well be correct (2007, p. 62). The idea is that, even if there is nobody around which we intend to affect, *were* there an audience that could be affected appropriately, the utterance that we intend to produce would affect them. So the effect on the virtual audience is to be understood as a conditional. Thus, this is compatible with the claim that one doesn't actually intend to have effects on anybody present. Speaker meaning thus is not defined through audience *directed* intentions, but through intentions that are audience *restricted*. For instance, according to this hypothesis, the Woody Allen character speaker-means something partially because, even though he is by himself, by uttering 'That is a big chicken.', he intends to produce a stimulus which is such that, were there an appropriate audience, it would affect them. If he had said instead 'Blmadnahad…', even if he had the same intentions, he would not have speaker-meant that there is a big chicken, because the appropriate conditional audience would not have been affected in the relevant way.[217]

This move requires to qualify what we mean by a conditional audience that can be affected appropriately. One option is to account for it as the audience possessing the conceptual, perceptual, or experiential capacities required to be shown what is speaker-meant according to Green's own analysis.

In fact, Green sometimes seems to go in such a direction. For instance, when he comments on the intentions of the suspect who says 'I'm innocent', he seems to agree that even though she knows she won't produce the belief

[216] Thanks to Mitch Green for this point.

[217] We could imagine a conditional audience that would possess portable fMRIs and the knowledge sufficient to read his mind when he says 'Blmadnahad…'. Would he thus have speaker-meant that this is a big chicken? We will see that the following restriction should be made: the stimulus must make it *mutually* recognizable to the conditional audience and the speaker that the stimulus is intended in such a way that it would have the relevant effects in the conditional audience. So, unless the person in question intends to produce a stimulus while being disposed to have this fMRI audience in mind (or some other audience of this sort), we shouldn't say that he speaker-meant anything with 'Blmadnahad'.

in the actual audience that she is innocent, she might intend to produce an effect in a conditional audience. He writes: « … she may say what she does in order to make public, *for anyone who may be concerned with the matter*, her avowal of innocence. » (61, n. 4, my italics). The reference to 'anyone who may be concerned' can be understood as equivalent to the appropriate conditional audience, the audience to whom her statement would constitute the compelling evidence displayed by the stimulus she produces intentionally.

So in the case of meaning-that, the appropriate conditional audience must have the conceptual capacities to be compelled by the evidence presented by the speaker (that is true for both illocutionary speaker-meaning and factual speaker-meaning). In meaning-*a*, the appropriate conditional audience must possess the perceptual capacity to see, hear, or otherwise perceive *a*. (And, we might want to add, in meaning-how, the appropriate conditional audience must possess the experiential capacities. However, Green doesn't talk of 'meaning-how'.)

Our discussion of the role that the audience plays in speaker-meaning thus leads us to the following conclusion: even in cases where no audience is intended to be affected, one speaker-means something only insofar as one intendedly produces a stimulus that would affect an appropriate conditional audience. The reference to a conditional audience in cases where there is no actual audience present allows us to analyze what is being made manifest, what is being overtly shown, since something is always shown in reference to a potential receiver.

This conclusion has a consequence for our discussion of the preceding problem: that of the overtness of speaker-meaning. We said above that the notion of being mutually recognizable (which is equivalent to, depending on how they are defined, the notions of being mutually manifest or overt) is my favored option to refer to the overtness required by speaker-meaning. A potential problem of this option was that the 'mutualness' implies an audience with whom the recognition is mutual. We have now seen that this is in fact not a problem at all if we allow the reference to be about a conditional audience. By consequence, I will hereafter assume that mutual recognizability is the notion we need to define speaker-meaning.

Before we move on to another topic, I would like to ask: why would anyone speaker-mean something in the absence of an audience? What are the typical cases of solitary speaker-meaning? Answering these questions will allow us to better understand the respective roles of actual and conditional audiences. I will discuss three main cases: (a) utterances directed at a

future audience, (b) solitary affective expression, and (c) solitary talking as extended cognition.

(a) Let us begin with future audiences. At the moment, I am writing by myself, but I hope that this will be read by someone else. Among the cases discussed by Green, the suspect who claims 'I'm innocent' might fit this category. She might think that, even though she won't convince the people present that she is innocent, they might report her claim to a future audience, where someone might believe her. The Woody Allen character or the parent with her baby however do not belong to this category.

(b) Why would the latter two utter these meaningful utterances out loud then? As I interpret them, the Woody Allen character and the parent are in situations that are quite charged affectively, situations where one would typically express one's affects. Now, as we know, when we undergo an affect, especially strong ones such as poignant emotions or acute pains, we tend to express these affects in several ways. That is part of the nature of emotions. This could translate with bodily expressions such as a smile, a jaw dropping, raising eyebrows, certain bodily postures. And this could also translate through verbal expressions. In fact, when I first read the examples given by Green, I immediately thought of them as verbal expressions of affects. In a way similar to when we are by ourselves and we hum a tune corresponding to our mood, say 'Shit! The keys!', 'Ouch!', or 'Oops!'. The solitary expression of affects is common and is not too mysterious. Even though they normally are not directed at an audience – we don't have either a future audience or a conditional audience in mind – explaining why they exist, what is their proper function (to use Millikan's phrase) requires appealing to an audience: most evolutionary theorists would explain that the reason why we tend to express affects through detectable and easily interpretable cues is that, in social species such as ours, it is evolutionary fit to communicate emotions in a reflex-like, quasi-automatic fashion. I am not claiming that every emotional cue has evolved for social consumption, e.g. sweat increase indicates emotional arousal (given a context) but it certainly didn't evolve for communicative purposes. However, sweat increase isn't an emotional *expression*, merely a *cue* of emotional activity (given a certain context). Emotional cues have no communicative proper functions, but emotional *expressions* are designed for communication, whether the design is individual, cultural, artificial, or natural (see Green, 2007, Chapters 2 and 5). And the kind of solitary speaker-meaning which results from an emotion is certainly an emotional *expression*, as opposed to a non-communicative emotional cue.

We have an innate drive to express some of our emotions, even when we are by ourselves, although this drive is stronger when we are surrounded by people we trust. People watching the same video laugh 30 more times when they are with friends than when they are alone, but they nevertheless laugh when they are alone (Provine, 2004). Laughing typically isn't a case of speaker-meaning, but it illustrates our natural drive to express our emotions even when alone. This drive can lead us to express our emotions in much more sophisticated ways, as when we play music or utter a quote from Spinoza. Despite the sophistication of these cases, I find it plausible that they exist because of a rather primitive social function, which is to alert others of our emotional life, itself being a reliable sign of goal-relevant stimuli for our conspecifics, and especially our in-groups. So, although we might have no audience in our head when we utter solitary affective expressions, the evolutionary reason, the proper function, might well be to affect an audience.

Now, this evolutionary explanation does not give *primary reasons* for solitary affective expression, i.e. reasons that we have for acting as we do and which also are the cause of this behavior (Davidson 1963). But, I think that we often don't have any primary reasons to solitary express ourselves. We just do it in a reflex-like manner. (We might find ways to rationalize why we do that, and we might also find objective, impersonal reasons for acting as such, but neither count as primary reasons.)

(c) Besides the cases of solitary speaker-meaning directed at a future audience or of solitary affective expression, another typical case of solitary speaker-meaning is when we use language as an external aid to our cognition, whether it is to help our memory, our planning, our imagination, or to practice a performance. I may say out loud the phone number of my father to remember it. I might list the things I need to do to better plan them. Or I might practice some arguments out loud to get better at it. In all these cases, talking by oneself is an aid to a cognitive task. This is achieved, it seems, by discharging a cognitive load in an external medium: sounds. I think of this situation as similar to when we realize a calculus by writing it down instead of doing it in our head. They are cases where we extend our cognitive processes outside our head (see A. Clark & Chalmers, 1998). These cases seem much more foreign to regular examples of speaker-meaning than the affective expression ones. (In fact, some of these cases might not qualify as speaker-meaning at all: for instance, someone rehearses a speech is using sound production as an extended, extra-cranial, aid to achieve a cognitive task, but is usually not considered as a case of speaker-meaning.) On the face of it, these cases of solitary speech as extended cognition have not much to do with communication as they don't

seem to derive from a communicative habit, unlike the solitary expression of affects.[218] Rather, in the extended cognition cases, it seems like we have diverted the normal social function of a language for a non-social purpose.[219]

There certainly are other types of solitary speaker-meaning than the three I have mentioned, even though I cannot think of other typical ones. In any case, all of these instances of speaker-meaning seem to be derivative, piggybacking on the audience-directed cases of speaker-meaning. Cases of solitary speaking thus seem to derive, in one way or another, from the paradigmatic, audience-directed speaker-meaning.

A.5.3. A THIRD DIFFICULTY: THE FLATTENING SCHEME AGAIN

I now arrive to the third and final difficulty about Green's definition, a difficulty which was already raised with respect to Sperber and Wilson's. The concern is that the intention that we try to make mutually recognizable – i.e. the intention (1) – can be of another kind than an intention to make something manifest; it may not be an informative intention. (We saw above that the intention to make something manifest can be qualified as an informative intention.) Instead, it could be an intention to make the audience do something, to encourage them, to complain to them, etc. In other words, speaker-meaning might consist in making mutually recognizable intentions that go beyond informativeness. And the worry is that Green's definition does not allow for this. Let me elaborate. The meaning of 'Get out of my way!' *prima facie* is to be found in the intention to make the addressee *do something*, rather than in the intention to make something manifest to the addressee (e.g. one's desire that the addressee does something, or the commitment/norm under which speakers and hearers thereby place themselves). If this is correct, and taking into consideration what we have said about a reference to an audience, then the intention (1) should be about producing a stimulus that would affect an audience (conditional or actual) in a way different to making something manifest to it.

Now, one might argue that the difficulty mentioned is unfounded since Green, contrary to Sperber and Wilson (1986), or to Davidson (1979) and Lewis (1979), can account for different speech acts through his definition

[218] Unless one conceives of communication as an extension of our cognition (Lyre, 2018; Sterelny, 2012, Chapter 6).

[219] Chomskians claim that language is an essentially *solitary* tool, a non-social cognitive ability (Reboul, 2017). But such claims refer to a Language of Thought (*à la* Fodor) or something close to it, not to an externalized, conventional, phonological language such as English or French. It is to these external languages that I attribute a social function.

of illocutionary speaker-meaning, and so can account for intentions to get the audience to do things, to encourage them, complain to them, etc. (I will say more about the concept of illocution bellow and how it relates to these different speech acts.) However, the way he analyses the different speech acts is through the intention *to make something manifest,* namely to make manifest different kinds of illocutionary forces under which one is committed. So, his analysis of the meaning of 'Get out of my way!' would be that the speaker intends (a) to make manifest that she is committed under the illocutionary force of a command to the proposition that the addressee gets out of her way, and (b) that this intention itself is made manifest. Putting aside the question of what exactly is the illocutionary commitment of uttering an order, and whether – as Green (2007, p. 76) ponders – there really is a commitment undertaken by commands, the difficulty then is that the intention of the speaker which she tries to make (mutually) manifest seems to be the intention to make someone move rather than the intention to make manifest that she is committing herself to a proposition under an illocutionary force. Consequently, to understand what the person means, it would be essential to understand that she intends to make her addressee move, but not, or not only, that she is committing herself to a proposition under the illocutionary force of a command. Taking a stand on this question will require some detours.

It was Grice (1957)'s intuition that, to understand what a person means, one should understand what kinds of effects this person intends to generate in an audience (and we might add: whether this audience is actual or conditional). And these effects can be of another kind than making something manifest. This intuition was supported by Strawson (1964) who attempted to show that these different effects correspond to the different speech acts which Austin (1962) had recently theorized.

We have seen that Green, just like Sperber and Wilson, disagrees with Grice or Strawson here (as well as Neale and Moore, to cite the people we've mentioned so far), since Green as well as Sperber and Wilson believe that the intention (1) cannot be of another kind than an informative intention. Another of their ally is Searle. It is worth discussing his views since he was very influential in this debate. Searle characterizes speaker-meaning as follows:

> «Uttering a sentence and meaning it is a matter of (a) intending (i-I) to get the hearer to know (recognize, be aware of) that certain states of affairs specified by certain of the rules obtain, (b) intending to get the hearer to know (recognize, be aware of) these things by means of getting him to recognize i-I and (c) intending to get him to recognize

> i-I in virtue of his knowledge of the rules for the sentence uttered. » (Searle, 1969, p. 48).

We see here that the only effects on the hearer intended by the speaker should be that the hearer *knows* (recognize, be aware of) certain things. These all are kinds of *informative intentions*. He explicitly excludes from his definition the intention to affect the audience in other ways than by informing them of what the speaker is doing. When one says 'Get out of here!', the only intention that would be relevant to the meaning of this speech act would be the intention to get the hearer to understand that one utters an order. The intention to make someone move, which Searle, after Austin, calls the *perlocutionary* intention, should not be taken into account in an analysis of speaker-meaning.

More precisely, the only effect that determines speaker-meaning for Searle is what he calls the 'illocutionary effect'. Illocutionary effect consists in the audience understanding what illocutionary act is being performed, and an illocutionary act is an action that is done in saying that one is doing so. For instance, a promise (affirmation, order, etc.) is an illocutionary act as one can promise that p just by saying 'I hereby promise that p' (the same applies to affirmations, orders, etc.). Searle is quite explicit about this. In fact, he claims that Grice 'confuses illocutionary and perlocutionary' (1969, p. 44). The perlocutionary being the aspects of a speech act viewed at the level of its consequences. Following Austin, Searle calls 'perlocutionary' acts such as convincing, scaring, or inspiring, and distinguish them from illocutionary acts. The main contrast is that one cannot perform a perlocutionary act just by saying so: one cannot convince just by saying 'I hereby convince you'. Similarly, one cannot get somebody to move just by saying 'I hereby make you move.' nor by saying 'Get out!'.[220]

In sum, for Green, Sperber and Wilson, or Searle the intention (1) can only be an *informative* intention, i.e. an intention to affect the audience solely by making something manifest. For Grice, Strawson, Neale, or Moore, it should include other kinds of intentions as well. Let us consider some examples illustrating the difference between the two positions.

[220] The criticism that Grice confuses the perlocutionary and the illocutionary is not entirely adequate because not all perlocutionary effects intended by the speaker determine its meaning according to Grice: only the effects that are meant to be (mutually) recognized as such should be taken into account. So all the perlocutionary acts that are not intended to be (mutually) recognized by the audience (and the speaker) do not enter Grice's definition of meaning, something which Searle sometimes seems to forget. However, it is true that Grice's definition (or that of Neale and Moore above) implies that some perlocutionary intents determine speaker-meaning. And it is not obvious that Grice is totally wrong about this, despite what Searle claims, as we will see.

Searle discusses examples where a speaker utters something knowing that he will not change the mind of her audience. For instance, the person saying 'I'm innocent' to the police even if she knows she will not change their mind. However, we have already considered such cases above when we discussed the idea that no actual audience is needed to speaker-mean anything. The solution we have presented was to allow for a future or a conditional audience. This solution allows rejecting Searle's counterexamples as genuine counterexamples to the idea that perlocutionary effects need to be taken into account for a proper analysis of meaning. We will consider cases that seem more illuminating. They are not given by Searle, but they stem from his argument against taking perlocutionary effects as a component of speaker-meaning.[221]

Maria is finishing high-school. She wants to go to the university, but hesitates between philosophy and physics. Sally, her mother, would rather have Maria study physics because she is afraid that there are no jobs for philosophers. She thus tells Maria, with the intention to scare her: 'If you study philosophy, you'll end up being unemployed and will eventually have to accept whatever crappy, uninteresting job.' Furthermore, she intends Maria to notice her intention to scare her and that this intention be mutually recognized. Now, is the intention to scare Maria part of the meaning of Sally's speech act (I say 'part' because other effects are intended by Sally, such as informing her of the philosophy job market)? Searle's intuition here is clearly that it is not. Scaring someone is a perlocutionary act and thus cannot be counted as part of speaker-meaning. For Green or Sperber and Wilson, a similar conclusion should hold: intending to scare someone cannot be reduced not, it seems, to an intention to make something manifest. However, according to Grice and others, there is no reason to exclude this intention from what Sally meant. Grice could indeed argue that for Maria to fully understand what Sally wanted to communicate, she would need to infer that Sally intended to scare her (and, perhaps, thus make the dangers of the situation emotionally salient to Maria). By understanding this intention, Maria would get a fuller

[221] Other counterexamples given by Searle (1969, pp. 46ff) against Grice's willingness to take into account other effects than illocutionary effects do not hit their target because, contrarily to Searle's claim, not all perlocutionary effects count in Grice's definition (see the preceding footnote). Another counterexample given by Searle that doesn't hit the target is his remark that some speech acts don't aim at anything else than making understand the hearer what speech act has been performed. So, when one says 'Hello!', one might not expect this to have any other effect than to make the addressee understand he has been greeted. But this sort of effect, contrarily to what Searle suggests, is not a counterexample to Grice's model: the latter can perfectly take them into account, since it allows any kinds of effects to enter the definition, thus including illocutionary effects.

understanding of what Sally intended to convey by this sentence. The communication would thus be more successful.

Take another example: Sam and Maria are having dinner but Maria is in a bad mood. Sam intends to cheer her up and says 'This is delicious. You really are the best cook I know.' Furthermore, Sam knows that it won't really be the compliment that might cheer Maria up, but rather her recognition that he is trying to do that. Now, is it necessary to understand Sam's intention to cheer Maria up in order to understand all that Sam meant by these sentences? According to Searle: no, the meaning is restricted to the illocutionary act, which consists in an affirmation as well as perhaps an indirect expressive. Cheering someone up is a perlocutionary act which should be excluded from speaker-meaning. However, according to Grice and co, understanding that Sam intended to cheer Maria up is understanding part of what Sam meant with these sentences (although, of course, it is not the only meaningful component: he also means that the dish is delicious, etc.). Observe that if Sam had intended to hide his intention to cheer Maria up, if he had intended to cheer her up surreptitiously, then his attempt to cheer up Maria wouldn't be relevant to what he meant according to Grice and co. But since he intended this intended effect to be in the common ground, to be mutually recognizable, then, according to the Gricean intuition we are discussing, this should count as an implicature of the sentences uttered, and thus be included as part of what Sam speaker-meant.

Now, these considerations illustrating the rival accounts by, among others, Searle and Grice might only reflect a conflict of intuitions concerning the proper use of the phrase 'S meant this by doing so-and-so' which would be hard to settle. Thus, in order to decide whether we should allow other effects than informative ones to be part of the definition of speaker-meaning, let us study other considerations. As we will see, these make the balance weighs toward Grice's side.

The first concerns an apparently distant subject: that of the emergence of speaker-meaning. Moore, in a series of papers, (2017, 2018 and others) has defended that, in order to give a plausible story of the emergence of speaker-meaning, in both ontogeny and phylogeny, one needs to allow that there can be 'minimally Gricean communication'. By this, he means, roughly, that that speaker-meaning can already exist in exchanges which are cognitively less demanding than the scenarios usually considered. In particular, he argues creatures that cannot entertain higher-order metarepresentations may nevertheless speaker-mean. By contrast, Sperber (2000) has argued that Gricean communication can only emerge

among creatures that can entertain a fourth order metarepresentation (a belief about an intention about a belief about an intention about a belief about something). Moore argues that if Sperber were right, it would be hard to explain how babies can already understand Gricean communication, as experiments by, e.g. Michael Tomasello suggest, since they presumably don't possess the cognitive capacities to entertain fourth-order metarepresentations. In response to such worries, Sperber claims that this sophisticated capacity is actually hardwired in the human brain, so that babies actually are capable of entertaining representations about representations about representations about representations. Moore however has a more economic explanation.

The solution is his 'minimally Gricean' proposal (see Moore 2017, pp. 314-320). In a nutshell, the idea is that Gricean communication can happen in a two-time sequence which can serve the same function as some simple cases of speaker-meaning, but without requiring fourth-order metarepresentations. First, a speaker would produce a stimulus with the intention to signal signalhood, an intention that plays a role comparable to intentions (2) above: for instance, S intends that R attends and respond to her gesture. Subsequently, S would intend to fulfill something like the intentions (1) above: e.g. S intends that R looks at the ground by S's feet. These intentions, taken separately, only require first-level meta-representations – an intention about an intentional behavior of the addressee – but they can nevertheless achieve the effects minimally required for speaker-meaning. In the example given, the meaning would be something like 'Look at the ground by my feet.'.

Now, what is relevant to our discussion here is that Moore's 'minimally Gricean communication', as he presents it, can only work if the effects intended by the speaker extend beyond informative intentions and include intentions to affect the behavior of others. To see why, consider how the account given by Moore would be formulated if we could only refer to an informative intention. His intention (1) 'S intends that R looks by S's feet' would need to be instead 'S intends that R be aware (recognize, be informed, etc.) that R should look at S's feet' or perhaps 'S intends that R be aware (recognize, be informed, etc.) that S desires that R looks at S's feet'. Instead of having an intention (1) with a first-order metarepresentation, one would thus require a second- or third-order representation. One of Moore's argument is that his minimally Gricean account can explain how speakers can, in some circumstances, communicate in ways that are cognitively less demanding while achieving the same functions than in the cognitively more demanding accounts given

by his predecessors. This would help explain how Gricean communication can emerge both in ontogeny and in phylogeny.

So, to come back to our topic, my claim here is that, whether or not Moore is right about everything he claims, it would be wiser to have a definition of speaker-meaning that could include minimally Gricean communication, rather than to exclude it by definition. By consequence, we should allow for other types of intentions (1) than the intentions to make something manifest, e.g. we should not exclude *a priori* intentions to influence behavior.

I want to look at another kind of consideration that might lead us to the same conclusion: this one has to do with the understanding of expressive speech acts (see Chapter 6 for an elaboration on this case). The idea here is to pursue the Fregean insight that a theory of meaning is a theory of understanding, an insight on which Dummett (1973) has insisted but which is not at all foreign from Searle's theory. Now, consider the following example: During an enraged argument, your partner says 'You don't understand what I feel!' and continues with an intense series of expressive speech acts. What is important for her is not that it is manifest to you that she undergoes an emotion, but that you are in a much more intimate relation with her feelings. You know (in fact: you both mutually know that you know) *that* she is angry and *why* she is angry, but what your partner would want you to understand, and claims you don't understand yet, is *how* she feels, *what* she really feels.

Arguably, in order to really understand her, what is required in such situations is that you *resonate affectively* with her (see the arguments for 'radical affectivism' in Chapter 6). This could be achieved in different ways: through empathy or sympathy, through emotion contagion, imaginative simulation, or a vivid episodic memory. What your partner wants is that you understand her through an emotional reaction, an action-ready, bodily, phenomenologically hot attitude. This understanding couldn't be done through a cognitively cold, intellectual, passive representation. This seems reasonable: there is ample evidence that an intellectual and an emotional episode about the same object don't present you with the same kind of information. The epistemological import of these two kinds of attitudes cannot be reduced to each other (Deonna & Teroni, 2012, Chapter 10; Goldie, 2002; Prinz, 2004, Chapter 2; Scarantino, 2010). So, the effect intended by your partner's expressive speech act is that you undergo some kind of affective feeling, that you resonate with her affectively. For you to really understand her, you need to resonate with her affectively, otherwise the epistemological import of your psychological state won't be sufficient to

really understand her. The way the Fregean insight I mentioned applies here is by suggesting that if you haven't really understood your partner, you haven't really (completely, optimally) understood what she really meant. At least in some expressive speech acts, it seems that one cannot understand what is speaker-meant unless one undergoes some relevant affects.

Now, for Sperber and Wilson, 'something being manifest to someone' is a purely cognitive mental state, it cannot consist in undergoing an affect. They define it through cognitive effects such as strengthening the conclusion of an inference or verifying a belief. For Green, the concept of being manifest is more flexible but, as we have seen, the only three kinds of speaker-meaning he allows are factual meaning, illocutionary meaning, and objectual meaning. The first two make manifest states of affairs and illocutionary commitments respectively and belong to the conceptual kind of showing. The third makes manifest perceptual objects, requiring perceptual mental states. None of them require that the addressee undergoes an affect for her to understand what is speaker-meant.

In conclusion, it seems that your partner's intention that you resonate affectively with her, that you understand her through your own affective experience, is an aspect of speaker-meaning that cannot be accounted for by the definitions of speaker-meaning restricted to intentions to make something manifest as they are defined by Sperber and Wilson or by Green.

Let me note that this is so unless Sperber and Wilson or Green were to modify their intentions (1) so as to include making something *affectively* manifest, a kind of overt showing which would require that the addressee undergoes (or is disposed to undergo) the relevant affect to properly be shown what is being made manifest. This option is not foreign to Green's account since this might be captured by his notion of 'showing-how', which we have briefly mentioned above. So, if Green had allowed for a fourth kind of speaker-meaning, namely 'meaning-how', his account could probably have dealt satisfactorily with the example discussed here.

If that is correct, once again, we have a reason to not restrict the intention (1) in the definition of speaker-meaning to an intention to make something manifest as it is defined by Green (or by Sperber and Wilson). We should thus extend it to other types of effects, be they behavioral (as in the 'minimally Gricean' case) or affective (as in the preceding example).

Before we move to the conclusion, let me observe that even though we have reasons to allow non-informative intentions in clause (1), we also have to recognize that informative intentions, intentions to make something

manifest, should have a place of choice in this clause. This is comparable to the remark I made above about solitary speaker-meaning and how it should be allowed by a definition of speaker-meaning, although it should be recognized that it is a derivative form of speaker-meaning compared to audience-directed speaker-meaning. I am not here suggesting that non-informative speaker-meaning derives from informative speaker-meaning – on the contrary, I find that there are strong reasons to think that the opposite is true (see Bar-On and Green (2010) for the hypothesis that speaker-meaning comes from expressive meaning and Moore (2017) for the idea that it comes from intentions to affect behaviors). But it seems quite clear that the meaning of most utterances is to be found in intentions to make mutually recognizable informative intentions. There are several reasons to give informative intentions a place of choice in clause (1) of a definition of speaker-meaning.

For one, the most common speech acts are informative ones (also called assertive or representative speech acts). Ronan (2015) has found that they amounted to 64% of all speech acts in an extended linguistic corpus. It is not a coincidence that philosophers of language have mostly ignored non-informative speech acts until Austin (1962) (Reinach being one of the notable exceptions of a philosopher who has extensively discussed non-assertive speech acts before Austin).

Secondly, it is undeniable that truth-conditional semantics tells us a lot about speaker-meaning, and truth-conditional semantics works well with informative intentions, because an utterance which is supposed to make mutually recognizable an intention to inform about a state of affairs can usually be rather well understood as an utterance which is true just in case this state of affairs is the case. For instance, if I say 'Snow is white.', the speaker-meaning of this sentence should be analyzable through my intention to make mutually recognizable my intention to inform my audience that snow is white, and the truth-conditional meaning of this sentence consists in the conditions in which it is true, i.e. those where snow indeed is white.

Truth-conditional meaning doesn't work straightforwardly for non-assertive utterances since we cannot literally say that sentences like 'Keep rocking!', 'Get out!', and 'Oops!' have truth-conditions. However, some have argued that understanding these sentences is the same as understanding corresponding truth-conditional sentences, such as 'I hereby encourage you to keep rocking.', 'I hereby order you to get out.', and 'I have just witnessed a minor mishap.' (e.g. Lewis 1970). Green's discussion of illocutionary

speaker-meaning seems to go in this direction, at least for certain non-assertive speech acts.

If one is attracted by these moves (i.e. reducing non-truth conditional meaning to truth-conditional meaning), one should naturally take the speaker-meaning of non-assertive utterances as in fact being analyzable through intentions to make an informative intention mutually recognizable. Although I do not share this intuition, I have to acknowledge that many philosophers or language researchers think that it is the way to go (see also Lewis (1979b) and Davidson (1979)).[222]

Furthermore, besides the commonality of informative speech acts and the explanatory power of truth-conditional semantics, we have seen above that, e.g., Searle, Sperber and Wilson, or Green share the intuition that non-informative intentions in (1) can be ignored in a definition of speaker-meaning or can be satisfyingly translated into intentions to make something manifest. Even though I disagree, I must recognize that their intuition is common.

In sum, despite the reasons that I have given above to resist the reduction in clause (1) of non-informative intentions to informative intentions, they should certainly figure preeminently in a definition of speaker-meaning.

A.6. Conclusion: A new definition of speaker-meaning

We are now in a position to draw a few conclusions and to propose a new definition for speaker-meaning based on the advantages and disadvantages of the ones we have reviewed.[223] To do so, let us take stock and recapitulate the potential difficulties that I have underlined for each definition given above and how to dispel them.

[222] For an alternative account of different illocutionary forces, including expressive illocutionary force, that can also be implemented within a formal semantics framework, see García-Carpintero (2015). This account, based on commitments is itself compatible with the influential formal treatments of orders by Portner (2007) and of questions by Roberts (1996). This account may be made to work with Green's definition of speaker-meaning, also based on commitments. This may constitute a way for Green to argue that the flattening scheme objection does not apply to his account, but I won't pursue this idea here.

[223] Of course, I haven't reviewed all the existing definitions of speaker-meaning. However, several remarks that I made above apply to definitions by e.g. Vlach (1981) and Davis (1992, 2003). Further discussion of these authors' definitions as well as those of, e.g., Recanatti (1986) or Bach and Harnish (1979) would be desirable, but I don't think that it would change my account. For a complementary review, see Green (2007, Chapter 3).

A first problem raised by Grice's (1989) definition concerned its excluding from speaker-meaning cases of showing, such as Herod presenting to Salomé the head of John the Baptist, Bill with his bandaged leg, or 'I can speak in a squeaky voice' uttered in a squeaky voice. A second problem concerned counterexamples such as the River Rat, cases involving what Grice called 'sneaky intentions'.

Neale (1992) proposed a solution for each of these problems: the case of showing is solved by discarding Grice's third clause. The problem of the River Rat kind of counterexamples is solved by another third clause (inspired by Grice, 1982) requiring that the first two intentions are not deceptive, which guarantees that the speaker doesn't have the problematic 'sneaky intentions'. Moore (2017; 2018) adopted Neale's proposals.

However, we saw that further counterexamples can be given to Neale's definition. I illustrated this with what I called the Vigilante case. Despite the absence of deceptive intentions, the information transfer in this case is not wholly overt, a fact which prevents it from being a genuine case of speaker-meaning.

A solution for such problems, where not everything of relevance is wholly overt, is to use Schiffer's mutual knowledge or Lewis' common knowledge. However, Sperber and Wilson (1986) or Campbell (2005), argue that these notions are psychologically non-realistic because they require subjects that are ideally logical.[224]

Sperber and Wilson offer a similar solution to Lewis' or Schiffer's but which is psychologically more realistic, introducing the notions of being manifest and of mutual manifestness. They define ostensive-communication with these notions, from which a definition of speaker-meaning is derived.

I raised five potential problems for the definition inspired by Sperber and Wilson. First, the definition implies that we always intend to make mutually manifest an *informative* intention, but a flexible enough definition should make room for other types of speaker-meaning, including those with a world-to-mind direction of fit. Second, their definition implies that speaker-meaning is always propositional, but there are good reasons to think that expressions such as 'Wow!', 'Yuk!', 'Oi!', or 'Ouch!' only express non-propositional speaker-meaning. A solution to both these difficulties is to allow for other kinds of intentions to be made mutually manifest, intentions that need not be informative or propositional.

[224] I remarked however that Lewis' notion is considered as psychologically realistic by certain philosophers, see e.g. Paternotte (2011).

The third and fourth difficulties are that their definition can be interpreted in a way which makes it unable to avoid the River Rat counterexample (Green, 2007, p. 79-80) and the Two Generals one (Jankovic, 2014). I observed however that these problems might be avoided depending on how mutual manifestness is defined. Nevertheless, one should note that the precise way in which Sperber and Wilson have defined mutual manifestness might be problematic in allowing interpretations such as Green's or Jankovic's.

Finally, another difficulty presented by Green (2007, Chapter 3), one which is also raised by Grice and Neale's definitions, is that speaker-meaning need not be audience directed.

We thus analyzed Green's definition and, despite its advantages, have raised three difficulties. The first is the following. His definition of overtness allows a *de dicto* and a *de re* reading, but the *de dicto* reading seems no more psychologically realistic than Schiffer's mutual knowledge, while the *de re* reading cannot avoid the Modified River Rat counterexample.

A tentative solution was to stipulate that the new phrase 'mutually recognizable' refers to the phenomenon which is intended to be captured by Sperber and Wilson's mutual manifestness, and Green's overtness.[225]

The second difficulty with Green's definition is that a definition without a reference to effects in an audience seems not to put all the cards on the table, so to say. We saw that this is so because Green defines speaker-meaning as a kind of showing, and that his notion of showing requires a reference to an audience (actual or conditional) and the potential effects (cognitive, perceptual, or experiential) in the audience that the showing implies. We remarked however that Green (or Davis) gave convincing examples where one speaker-means something with no intentions to affect the audience present, or where there is no actual audience.

A reference to a *conditional* audience seemed to be a good solution: it didn't exclude the examples given by Green, but allowed the explicit reference to the audience required by the notion of making manifest. The intentions involved in speaker-meaning would not need to be *directed* toward an audience, but they would be *restricted* by a certain audience (e.g. by their cognitive, perceptual, or experiential capacities). We observed furthermore that all cases of speaker-meaning which don't involve intentions to affect

[225] We noted also that this stipulation is in line with Campbell's (2005) proposal that joint attention is a primitive phenomenon of our experience, one that we cannot define by using first-person concepts (only sub-personal causal mechanisms).

an actual audience seem to be derivative cases, whose existence needs to be explained by reference to the normal cases where an actual audience is intended to be affected.

The third difficulty was already raised with respect to Sperber and Wilson's definition: sometimes, when we speaker-mean, our aim is not to make mutually recognizable an *informative* intention, but rather another kind of intended effect, such as the intention that the audience does something, or the intention that the audience undergoes an affect. And speaker-meaning might as well be found in such intentions rather than uniquely in informative intentions.

However, we have also seen that there are reasons to give a preeminent place to intentions to make mutually recognizable informative intentions, in a spirit similar to giving a preeminent place to audience-directed speaker-meaning as opposed to solitary speaking meaning.

In sum, here are the features of a definition of speaker-meaning which appear as desirable after this discussion:

(1) To not exclude cases of showing.
(2) To exclude cases where the communication is not wholly overt (e.g. the River Rat).
(3) To do so in a psychologically realistic way (unlike e.g. Schiffer's mutual knowledge).
(4) To allow for kinds of speaker-meaning where the intended effects in the audience is not to inform them, but to recognize that informative intentions still are preeminent in speaker-meaning.
(5) To allow for non-propositional meaning.
(6) To give an explicit reference to an audience and how it would be affected by the speech act, but to do so in a way that doesn't exclude the derivative cases of speaker-meaning where there is no intention to affect an actual audience.

We saw that the solution to (1) is simply to ignore Grice's third clause.

A solution to (2) and (3) is to require that things of relevance be mutually recognizable.

A solution for (4) and (5) is to use a first clause where the effects intended in the audience are not restricted to informative ones or propositional ones, but where non-informative intentions have a second place.

A solution for (6) is to refer to an appropriate conditional audience, an audience that, would be affected by the stimulus intendedly produced if it

was present (a solution proposed by Hyslop, 1977). The '*appropriate* conditional audience' here refers to an audience with the cognitive, perceptual, or experiential capacities (and perhaps other capacities) that are necessary to be affected in the relevant way. The *relevant* way is the way in which the conditional audience would be affected by the kind of stimuli that is intendedly produced. So if one intendedly produces an assertive sentence in English, the relevant way to be affected, i.e. how an actual audience should react to this sentence, is to form a certain belief (whose epistemic strength depends on how trustworthy the utterer is). Since the cases of solitary speaker-meaning are derivative, the mention of the appropriate conditional audience should take less importance than the mention of an actual audience.

With all this in mind, here is my proposal for a definition of speaker-meaning:

> Speaker meaning - definition
>
> A sender S speaker-means something by the set of stimuli x if, and only if, S produces x while
>
> (i) S intends x to be a stimulus which makes (or would make) something manifest to the appropriate (conditional) receiver R, or generates (would generate) another kind of effect in R, and
> (ii) S intends x to be a stimulus which makes (would make) (i) mutually recognizable for R and S.

BIBLIOGRAPHY

Adriaans, P. (2019). Information. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2019/entries/information/

Anderson, L., & Lepore, E. (2013). Slurring words. *Noûs*, *47*(1), 25–48.

Andrews, K. (2015). *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*. Routledge.

Anscombe, G. E. M. (1957). *Intention*. Blackwell.

Armstrong, D. M. (1981). What is Consciousness? In J. Heil (Ed.), *The Nature of Mind*. Cornell University Press.

Arnold, M. B. (1960). *Emotion and Personality*. Columbia University Press.

Austin, J. L. (1962). *How to Do Things with Words*. Clarendon Press.

AW, Y., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). Facial expressions of emotion: Stimuli and tests (FEEST). *Thames Valley Test Company (TVTC)*.

Ayer, A. J. (1936). *Language, Truth, and Logic*. London: V. Gollancz.

Bach, K. (2004). Pragmatics and the Philosophy of Language. In *The Handbook of Pragmatics* (pp. 463–487). Blackwell. https://doi.org/10.1002/9780470756959.ch21

Bach, K. (2018). Loaded words: On the semantics and pragmatics of slurs. In D. Sosa (Ed.), *Bad Words: Philosophical Perspectives on Slurs* (pp. 60–76). Oxford University Press.

Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. MIT Press.

Bacharach, M. (1998). Common knowledge. In *New Palgrave Dictionary of Law and Economics* (pp. 308–313). Macmillan.

Bain, D. (2017). Evaluativist Accounts of Pain's Unpleasantness. In J. Corns (Ed.), *The Routledge Handbook of the Philosophy of Pain* (pp. 40–50). Routledge.

Bar-On, D., & Green, M. (2010). Lionspeak: Expression, meaning and communication. forthcoming in E. Rubenstein. In E. Rubenstein (Ed.), *Self, Language and World* (Ridgeview).

Bar-On, Dorit. (2013). Origins of meaning: Must we 'go Gricean'? *Mind & Language*, *28*(3), 342–375.

Bar-On, Dorit. (2017). Communicative Intentions, Expressive Communication, and Origins of Meaning. In *The Routledge Handbook of Philosophy of Animal Minds* (pp. 301–312). Routledge.

Barrett, L. F. (2006). Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements: *Psychological Science in the Public Interest*. https://doi.org/10.1177/1529100619832930

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290.

Barrett, L. F., & Russell, J. A. (2015). *The Psychological Construction of Emotion*. Guilford Publications.

Ben-Ze'ev, A. (2000). *The Subtlety of Emotions*. MIT press.

Blakemore, D. (1987). *Semantic constraints on relevance*. Blackwell.

Blakemore, D. (2011). On the descriptive ineffability of expressive

meaning. *Journal of Pragmatics*, *43*(14), 3537–3550.

Blakemore, R. L., Neveu, R., & Vuilleumier, P. (2017). How emotion context modulates unconscious goal activation during motor force exertion. *NeuroImage*, *146*, 904–917.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247.

Block, N. (1998). Semantics, conceptual role. In E. Craig (Ed.), *The Routledge Encylopedia of Philosophy*. Routledge.

Block, N. (2002). Some Concepts of Consciousness. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 206–219). Oxford University Press.

Blumstein, D. T., & Récapet, C. (2009). The sound of arousal: The addition of novel non-linearities increases responsiveness in marmot alarm calls. *Ethology*, *115*(11), 1074–1081.

Bonard, C. (2018). Lost in Musical Translation: A cross-cultural study of musical grammar and its relation to affective expression in two musical idioms between Chennai and Geneva. In F. Cova & S. Réhault (Eds.), *Advances in Experimental Philosophy of Aesthetics*. Bloosmbury Academics.

Bonard, C. (in preparation). *Super Semantics, Super Pragmatics, and Musical Meaning*.

Bonard, C., & Humbert-Droz, S. (2020). Art (entrée académique). In M. Kristanek (Ed.), *L'Encyclopédie philosophique*. http://encyclo-philo.fr/art-a/

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*(4), 542–573.

Borg, E. (2004). *Minimal Semantics*. Oxford University Press.

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation. In

Handbook of psychophysiology, 3rd ed (pp. 581–607). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396.025

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.

Brandom, R. (1983). Asserting. *Noûs*, 637–650.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. https://philpapers.org/rec/BRAIPA

Bratman, M. (1990). What is Intention? In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (MIT Press).

Bratman, M. (2018). *Planning, Time, and Self-Governance: Essays in Practical Rationality*. Oxford University Press.

Breznitz, S. (1984). *Cry Wolf: The Psychology of False Alarms*. Laurence Erlbaum Associates.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. University of California Press.

Buchanan, R. (2012). Meaning, Expression, and Evidence. *Thought: A Journal of Philosophy*, *1*(2), 152–157.

Bull, N. (1951). The attitude theory of emotion. *Archivio Di Psicologia, Neurologia e Psichiatria*, *12*(2), 108.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2017). *Handbook of Psychophysiology*. Cambridge University Press.

Camp, E. (2013). Slurring Perspectives. *Analytic Philosophy*, *54*(3), 330–349. https://doi.org/10.1111/phib.12022

Camp, E. (2018a). Slurs as dual-act expressions. In D. Sosa (Ed.), *Bad Words*. Oxford University Press.

Camp, E. (2018b). Why maps are not propositional. In A. Grzankowski & M. Montague (Eds.), *Non-propositional intentionality* (pp. 19–45). Oxford University Press.

Camp, E. (2018c). Insinuation, Common Ground, and the Conversational Record. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New Work on Speech Acts*. Oxford University Press.

Campbell, J. (2005). Joint Attention and Common Knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology* (pp. 287–297). Clarendon Press.

Cappelen, H., & Lepore, E. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Blackwell.

Carruthers, P. (2004). Phenomenal Concepts and Higher-Order Experiences. *Philosophy and Phenomenological Research*, *68*(2), 316–336. https://doi.org/10.1111/j.1933-1592.2004.tb00343.x

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Blackwell.

Cepollaro, B., & Stojanovic, I. (2016). Hybrid Evaluatives: In Defense of a Presuppositional Account. *Grazer Philosophische Studien*, *93*(3), 458–488. https://doi.org/10.1163/18756735-09303007

Cepollaro, B., & Thommen, T. (2019). What's wrong with truth-conditional accounts of slurs. *Linguistics and Philosophy*, *42*(4), 333–347. https://doi.org/10.1007/s10988-018-9249-8

Chernilovskaya, A., Condoravdi, C., & Lauer, S. (2012). On the Discourse Effects of wh-Exclamatives. *Proceedings of the 30th West Coast Conference on Formal Linguistics*, 109–119.

Chisholm, R. M. (1967). He could have done otherwise. *The Journal of Philosophy*, *64*(13), 409–417.

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Clark, B. (2013). *Relevance Theory*. Cambridge University Press.

Cleveland, A., & Striano, T. (2007). The effects of joint attention on object processing in 4- and 9-month-old infants. *Infant Behavior and Development*, *30*(3), 499–504. https://doi.org/10.1016/j.infbeh.2006.10.009

Colombetti, G., & Thompson, E. (2007). The feeling body: Toward an enactive approach to emotion. In *Developmental perspectives on embodiment and consciousness* (pp. 61–84). Psychology Press.

Coolidge, F., & Wynn, T. (2006). The effects of the tree-to-ground sleep transition in the evolution of cognition in early Homo. *Before Farming*, *2006*(4), 1–18.

Copp, D. (2009). Realist-Expressivism and Conventional Implicature. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 4, pp. 167–202). Oxford University Press.

Coppock, E., & Champollion, L. (2020). *Invitation to Formal Semantics*. Manuscript, August 2020 version, available at http://eecoppock.info/teaching.html.

Coyne, J. A. (2010). *Why Evolution is True*. Oxford University Press.

Croom, A. M. (2011). Slurs. *Language Sciences*, *33*(3), 343–358. https://doi.org/10.1016/j.langsci.2010.11.005

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153.

Curran, W., McKeown, G. J., Rychlowska, M., André, E., Wagner, J., & Lingenfelser, F. (2018). Social context disambiguates the interpretation of laughter. *Frontiers in Psychology*, *8*, 2342.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085.

Damasio, A. R. (1994). *Descartes' error*. Putnam.

Dänzer, L. (2020). The explanatory project of Gricean pragmatics. *Mind & Language*, *n/a*(n/a). https://doi.org/10.1111/mila.12295

D'Arms, J., & Jacobson, D. (2000). The Moralistic Fallacy: On the "Appropriateness" of Emotions. *Philosophy and Phenomenological Research*, *61*(1), 65–90. JSTOR. https://doi.org/10.2307/2653403

Davidson, D. (1979). Moods and Performances. In A. Margalit (Ed.), *Meaning and Use: Papers Presented at the Second Jerusalem Philosophical Encounter April 1976* (pp. 9–20). Springer Netherlands. https://doi.org/10.1007/978-1-4020-4104-4_2

Davidson, D. (1982). Rational Animals. *Dialectica*, *36*(4), 317–327.

Davidson, D. (2001). *Essays on Actions and Events: Philosophical Essays Volume 1: Philosophical Essays* (Vol. 1). Oxford University Press.

Davis, W. A. (1992). Speaker meaning. *Linguistics and Philosophy*, *15*(3), 223–253. https://doi.org/10.1007/BF00627678

Davis, W. A. (2003). *Meaning, Expression and Thought*. Cambridge University Press.

Davis, W. A. (2013). Meaning, Expression, and Indication: Reply to Buchanan. *Thought: A Journal of Philosophy*, *2*(1), 62–66.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or

manipulation. *Behavioural Ecology: An Evolutionary Approach*, *2*, 282–309.

De Gelder, B., Morris, J. S., & Dolan, R. J. (2005). Unconscious fear influences emotional awareness of faces and voices. *Proceedings of the National Academy of Sciences*, *102*(51), 18682–18687.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

de Saussure, F. (1916). *Cours de Linguistique Générale*. Payot.

De Sousa, R. (1987). *The rationality of emotion*. MIT Press.

Deigh, J. (1994). Cognitivism in the theory of emotions. *Ethics*, *104*(4), 824–854.

Denkel, A. (1992). Natural meaning. *Australasian Journal of Philosophy*, *70*(3), 296–306.

Deonna, J. (2006). Emotion, perception and perspective. *Dialectica*, *60*(1), 29–46.

Deonna, J. (2007). The structure of empathy. *Journal of Moral Philosophy*, *4*(1), 99–116.

Deonna, J., & Teroni, F. (2008). *Qu'est-ce qu'une émotion?* Vrin.

Deonna, J., & Teroni, F. (2012). *The emotions: A philosophical introduction*. Routledge.

Deonna, J., & Teroni, F. (2014). In what sense are emotions evaluations? In C. Todd & S. Roeser (Eds.), *Emotion and value* (pp. 15–31). Oxford University Press.

Deonna, J., & Teroni, F. (2015). Emotions as attitudes. *Dialectica*, *69*(3), 293–311.

Dezecache, G., Mercier, H., & Scott-Phillips, T. C. (2013). An evolutionary

approach to emotional communication. *Journal of Pragmatics*, *59*, 221–233.

Diaz-Leon, E. (2020). Pejorative Terms and the Semantic Strategy. *Acta Analytica*, *35*(1), 23–34. https://doi.org/10.1007/s12136-019-00392-2

Döring, S. A. (2007). Seeing what to do: Affective perception and rational motivation. *Dialectica*, *61*(3), 363–394.

Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, Content, and Function* (pp. 17–36). Oxford University Press.

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

Dretske, F. (1995). *Naturalizing the Mind—The Jean Nicod Lectures-1994*. MIT Press.

Dretske, F. (2006). Minimal Rationality. In S. Hurley & M. Nudds (eds) (Eds.), *Rational Animals?* (1st ed.). Oxford University Press, USA. http://gen.lib.rus.ec/book/index.php?md5=9a24312c0b7d06a17752be117dba211c

Dretske, F. (2008). Epistemology and information. In P. Adriaans & J. van Benthem (Eds.), *Philosophy of information*. Elsevier.

Dummett, M. (1973). *Frege: Philosophy of Language*. Harvard University Press.

Dunbar, R. (1996). *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.

Dunbar, R. I., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*(5843), 1344–1347.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384.

Ekman, P. (1997). Expression or communication about emotion. In N. L. Segal, G. E. Weisfeld, & C. C. Weisfeld (Eds.), *Uniting psychology and biology: Integrative perspectives on human development.* (pp. 315–338). American Psychological Association. https://doi.org/10.1037/10242-008

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124.

Elgin, C. Z. (2008). Emotion and Understanding. In G. Brun, U. Dogluoglu, & D. Kuenzle (Eds.), *Epistemology and Emotions*. Ashgate.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, Scherer, Klaus R., & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford University Press.

Evans, D. (2001). *Emotion: The Science of Sentiment*. Oxford University Press.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Fischer, J. M., & Ravizza, M. (2000). Précis of Responsibility and Control: A Theory of Moral Responsibility. *Philosophy and Phenomenological Research*, *61*(2), 441–445.

Fish, S. E. (1972). *Self-consuming Artifacts: The Experience of Seventeenth-century Literature*. University of California Press.

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (pp. xiii, 171). MIT Press.

Fogal, D., Harris, D. W., & Moss, M. (Eds.). (2018). *New Work on Speech*

Acts. Oxford University Press.

Foolen, A. (2012). The relevance of emotion for language and linguistics. In A. Foolen, U. M. Lüdtke, T. Racine, & J. Zlatev (Eds.), *Moving ourselves, moving others: Motion and emotion in intersubjectivity, consciousness and language* (pp. 349–369). John Benjamins Publishing.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy*, *66*(23), 829–839.

Frege, G. (1956). The thought: A logical inquiry. *Mind*, *65*(259), 289–311.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion*. Routledge. https://doi.org/10.4324/9781315086071

Frost, K. (2014). On the very idea of direction of fit. *Philosophical Review*, *123*(4), 429–484.

García-Carpintero, M. (2004). Assertion and the Semantics of Force-Markers. In C. Bianchi (Ed.), *The Semantics/Pragmatics Distinction* (pp. 133–166). CSLI Publications.

García-Carpintero, M. (2015). Contexts as shared commitments. *Frontiers in Psychology*, *6*, 1932.

García-Carpintero, M. (2017). Pejoratives, Contexts and Presuppositions. In P. Brézillon, R. Turner, & C. Penco (Eds.), *Modeling and Using Context* (Vol. 10257, pp. 15–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-57837-8_2

García-Carpintero, M. (2020). On the Nature of Presupposition: A Normative Speech Act Account. *Erkenntnis*, *85*, 269–293. https://doi.org/10.1007/s10670-018-0027-3

Geach, P. T. (1965). Assertion. *The Philosophical Review*, *74*(4), 449–465.

Gendron, M., & Feldman Barrett, L. (2009). Reconstructing the Past: A

Century of Ideas About Emotion in Psychology. *Emotion Review*, *1*(4), 316–339. https://doi.org/10.1177/1754073909338877

Gergely, G., & Király, I. (2019). Natural pedagogy of social emotions. *Foundations of Affective Social Learning: Conceptualizing the Social Transmission of Value*, 87.

Gert, J. (2018). Neo-pragmatism, Representationalism and the Emotions. *Philosophy and Phenomenological Research*, *97*(2), 454–478.

Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology*, *80*(4), 395–430.

Gibson, J. J. (1977). The theory of affordances. In J. B. Robert E Shaw (Ed.), *Perceiving, acting, and knowing: Toward an ecological psychology*. Lawrence Erlbaum Associates.

Ginzburg, J., Breitholtz, E., Cooper, R., Hough, J., & Tian, Y. (2015). *Understanding laughter*.

Godfrey-Smith, P. (1991). Signal, decision, action. *The Journal of Philosophy*, *88*(12), 709–722.

Godfrey-Smith, P. (1992). Indication and adaptation. *Synthese*, *92*(2), 283–312.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge University Press.

Goldie, P. (2000). *The emotions: A philosophical exploration*. Oxford University Press.

Goldie, P. (2002). Emotions, feelings and intentionality. *Phenomenology and the Cognitive Sciences*, *1*(3), 235–254.

Gómez, J.-C. (2007). Pointing Behaviors in Apes and Human Infants: A Balanced Interpretation. *Child Development*, *78*(3), 729–734.

https://doi.org/10.1111/j.1467-8624.2007.01027.x

Gorzelak, M. A., Asay, A. K., Pickles, B. J., & Simard, S. W. (2015). Inter-plant communication through mycorrhizal networks mediates complex adaptive behaviour in plant communities. *AoB Plants*, *7*.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146.

Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, *17*(2), 484–495.

Green, M. (2007). *Self-Expression*. Oxford University Press.

Green, M. (2009). Speech Acts, the Handicap Principle and the Expression of Psychological States. *Mind & Language*, *24*(2), 139–163. https://doi.org/10.1111/j.1468-0017.2008.01357.x

Green, M. (2019a). Assertion, implicature, and speaker meaning. *Rivista Italiana Di Filosofia Del Linguaggio*, *13*(1).

Green, M. (2019b). Organic Meaning: An Approach to Communication with Minimal Appeal to Minds. In A. Capone, M. Carapezza, & F. Lo Piparo (Eds.), *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications* (pp. 211–228). Springer International Publishing. https://doi.org/10.1007/978-3-030-00973-1_12

Green, M. (2010). Perceiving Emotions. *Aristotelian Society Supplementary Volume*, *84*(1), 45–61.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Grice, H. P. (1968). Utterer's meaning, sentence-meaning, and word-meaning. In *Philosophy, Language, and Artificial Intelligence* (pp. 49–66).

Springer.

Grice, H. P. (1969). Utterer's meaning and intentions. *The Philosophical Review*, *78*(2), 147–177.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Grice, H. P. (1982). Meaning revisited. In N. V. Smith (Ed.), *Mutual knowledge* (pp. 222–243). Academic Press.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Grzankowski, A., & Montague, M. (2018). *Non-propositional intentionality*. Oxford University Press.

Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2013). The Deliberate Duchenne Smile: Individual Differences in Expressive Control. *Journal of Nonverbal Behavior*, *37*(1), 29–41. https://doi.org/10.1007/s10919-012-0139-4

Hamm, A. O., Weike, A. I., Schupp, H. T., Treig, T., Dressel, A., & Kessler, C. (2003). Affective blindsight: Intact fear conditioning to a visual cue in a cortically blind patient. *Brain*, *126*(2), 267–275. https://doi.org/10.1093/brain/awg037

Hauser, M. D. (1996). *The Evolution of Communication*. MIT Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579.

Hedger, J. A. (2012). The semantics of racial slurs: Using Kaplan's framework to provide a theory of the meaning of derogatory epithets. *Linguistic and Philosophical Investigations*, *11*, 74–84.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Wiley.

Helm, B. W. (2009). Emotions as Evaluative Feelings. *Emotion Review*,

1(3), 248–255. https://doi.org/10.1177/1754073909103593

Hom, C. (2008). The semantics of racial epithets. *The Journal of Philosophy*, *105*(8), 416–440.

Hom, C. (2010). Pejoratives. *Philosophy Compass*, *5*(2), 164–185. https://doi.org/10.1111/j.1747-9991.2009.00274.x

Hom, C., & May, R. (2013). Moral and Semantic Innocence. *Analytic Philosophy*, *54*(3), 293–313. https://doi.org/10.1111/phib.12020

Hom, C., & May, R. (2018). Pejoratives as fiction. In D. Sosa (Ed.), *Bad words* (pp. 108–131). Oxford University Press.

Hopkins, W. D., Russell, J., McIntyre, J., & Leavens, D. A. (2013). Are chimpanzees really so poor at understanding imperative pointing? Some new data and an alternative view of canine and ape social cognition. *PLoS One*, *8*(11), e79338.

Hoque, M., & Picard, R. W. (2011). Acted vs. natural frustration and delight: Many people smile in natural frustration. *Face and Gesture 2011*, 354–359. https://doi.org/10.1109/FG.2011.5771425

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. *Meaning, Form, and Use in Context: Linguistic Applications*, *11*.

Hufendiek, R. (2018). Explaining Embodied Emotions – with and Without Representations. *Philosophical Explorations*, *21*(2), 319–331. https://doi.org/10.1080/13869795.2018.1477985

Hutto, D. D. (2012). Truly enactive emotion. *Emotion Review*, *4*(2), 176–181.

Hyslop, A. (1977). Grice without an audience. *Analysis*, *37*(2), 67–69.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, *100*(1), 68–90.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*(2), 211–225.

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522.

James, W. (1884). What is an emotion? *Mind*, *9*(34), 188–205.

Jankovic, M. (2014). Communication and shared information. *Philosophical Studies*, *169*(3), 489–508.

Jeshion, R. (2013). Slurs and Stereotypes. *Analytic Philosophy*, *54*(3), 314–329. https://doi.org/10.1111/phib.12021

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770.

Kant, I. (1798). *Anthropology From a Pragmatic Point of View*. Cambridge University Press.

Kaplan, D. (1999). The meaning of ouch and oops: Explorations in the theory of meaning as use. *Manuscript, UCLA*.

Kasher, A. (1982). Gricean Inference Revisited. *Philosophica*, *29*(1), 25–44.

Klein, C. (2007). An imperative theory of pain. *The Journal of Philosophy*, *104*(10), 517–532.

Kolodny, N., & Brunero, J. (2020). Instrumental Rationality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2020/entries/rationality-instrumental/

Korta, K., & Perry, J. (2007). Radical Minimalism, Moderate

Contextualism. In G. Preyer (Ed.), *Context Sensitivity and Semantic Minimalism* (pp. 94–111). Oxford University Press.

Korta, K., & Perry, J. (2020). Pragmatics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2020/entries/pragmatics/

Kriegel, U. (2014). Towards a New Feeling Theory of Emotion. *European Journal of Philosophy*, *22*(3), 420–442. https://doi.org/10.1111/j.1468-0378.2011.00493.x

Krumhuber, E. G., & Manstead, A. S. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion*, *9*(6), 807.

Lambie, J. A., & Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review*, *109*(2), 219.

Langton, R., Haslanger, S., & Anderson, L. (2012). Language and Race. In G. Russell & D. G. Fara (Eds.), *The Routledge Companion to Philosophy of Language*. Routledge.

Lavan, N., Rankin, G., Lorking, N., Scott, S., & McGettigan, C. (2017). Neural correlates of the affective properties of spontaneous and volitional laughter types. *Neuropsychologia*, *95*, 30–39.

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (2005). Understanding the point of chimpanzee pointing: Epigenesis and ecological validity. *Current Directions in Psychological Science*, *14*(4), 185–189.

Lepore, E., & Stone, M. (2015). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford University Press.

Leventhal, H., & Scherer, K. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion*, *1*(1), 3–28.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press.

Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press.

Lewis, D. (1970). General Semantics. *Synthese*, *22*, 18–67.

Lewis, D. (1979a). Attitudes De Dicto and De Se. *The Philosophical Review*, *88*(4), 513–543. JSTOR. https://doi.org/10.2307/2184843

Lewis, D. (1979b). Scorekeeping in a Language Game. *Journal of Philosophical Logic*, *8*(1), 339–359. https://doi.org/10.1007/BF00258436

LoBue, V., & Adolph, K. E. (2019). Fear in infancy: Lessons from snakes, spiders, heights, and strangers. *Developmental Psychology*, *55*(9), 1889.

Lycan, W. G. (1996). *Consciousness and Experience*. MIT Press.

Lycan, W. G. (2015). Slurs and lexical presumption. *Language Sciences*, *52*, 3–11. https://doi.org/10.1016/j.langsci.2015.05.001

Lyre, H. (2018). Socially Extended Cognition and Shared Intentionality. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00831

MacFarlane, J. (2011). What Is Assertion. In J. Brown & H. Cappelen (Eds.), *Assertion*. Oxford University Press.

Malcolm, N. (2001). *Ludwig Wittgenstein: A Memoir*. Oxford University Press.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, *50*(3), 629–658. https://doi.org/10.1111/nous.12089

Marques, T., & García-Carpintero, M. (2020). Really Expressive Presuppositions and How to Block Them. *Grazer Philosophische Studien*, *97*, 10–1163.

Martínez, M. (2011). Imperative content and the painfulness of pain. *Phenomenology and the Cognitive Sciences*, *10*(1), 67–90.

Martínez, M. (2015). Pains as reasons. *Philosophical Studies*, *172*(9), 2261–2274.

Martínez, M., & Klein, C. (2016). Pain signals are predominantly imperative. *Biology & Philosophy*, *31*(2), 283–298.

Marty, A. (1875). *Über den Ursprung der Sprache*. Stuber.

Matsumoto, D., & Ekman, P. (2009). Basic emotions. *The Oxford Companion to Emotion and the Affective Sciences*, 69–72.

Maynard Smith, J., & Harper, D. (2003). *Animal Signals*. Oxford University Press.

McCready, E. (2010). Varieties of conventional implicature. *Semantics and Pragmatics*, *3*. https://doi.org/10.3765/sp.3.8

McDougall, W. (1923). *An Outline of Psychology*. Methuen and Co.

McDowell, J. (1980). Meaning, Communication, and Knowledge. In Z. V. Straaten (Ed.), *Philosophical Subjects*. Oxford University Press.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2015). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex*, *25*(1), 246–257.

Mele, A. R. (1992). *Springs of Action: Understanding Intentional Behavior*. Oxford University Press.

Mele, A. R., & Moser, P. K. (1994). Intentional action. *Nous*, *28*(1), 39–68.

Mendelovici, A. (2014). Pure intentionalism about moods and emotions. *Current Controversies in Philosophy of Mind*, 135–157.

Menzel, R., & Giurfa, M. (2001). Cognitive architecture of a mini-brain: The

honeybee. *Trends in Cognitive Sciences*, *5*(2), 62–71. https://doi.org/10.1016/S1364-6613(00)01601-6

Mercier, H., & Sperber, D. (2017). *The Enigma of Reason*. Harvard University Press.

Messinger, D. S., Fogel, A., & Dickson, K. L. (2001). All smiles are positive, but some smiles are more positive than others. *Developmental Psychology*, *37*(5), 642.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, *86*(6), 281–297.

Millikan, R. G. (1995). Pushmi-pullyu representations. *Philosophical Perspectives*, *9*, 185–200.

Millikan, R. G. (2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge University Press.

Millikan, R. G. (2004). *Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT press.

Montague, M. (2007). Against propositionalism. *Noûs*, *41*(3), 503–518.

Moore, R. (2017). Gricean communication and cognitive development. *The Philosophical Quarterly*, *67*(267), 303–326.

Moore, R. (2018). Gricean communication, language development, and animal minds. *Philosophy Compass*, *13*(12), e12550.

Moors, A. (2017). Integration of Two Skeptical Emotion Theories: Dimensional Appraisal Theory and Russell's Psychological Construction Theory. *Psychological Inquiry*, *28*(1), 1–19. https://doi.org/10.1080/1047840X.2017.1235900

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal

theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124.

Moors, A., & Scherer, K. R. (2013). The role of appraisal in emotion. In M. Robinson, E. R. Watkins, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 135–155). The Guilford Press.

Mumenthaler, C., & Sander, D. (2015). Automatic integration of social information in emotion recognition. *Journal of Experimental Psychology: General*, *144*(2), 392.

Mumenthaler, C., & Sander, D. (2019). Socio-affective inferential mechanisms involved in emotion. In D. Dukes & F. Clément (Eds.), *Foundations of Affective Social Learning: Conceptualizing the Social Transmission of Value*. Cambridge University Press.

Murray, R. J., Brosch, T., & Sander, D. (2014). The functional profile of the human amygdala in affective processing: Insights from intracranial recordings. *Cortex*, *60*, 10–33. https://doi.org/10.1016/j.cortex.2014.06.010

Nanay, B. (2010). A Modal Theory of Function. *Journal of Philosophy*, *107*(8), 412–431. https://doi.org/10.5840/jphil2010107834

Nanay, B. (2013). *Between Perception and Action*. Oxford University Press.

Nanay, B. (2014). Teleosemantics without etiology. *Philosophy of Science*, *81*(5), 798–810.

Nanay, B. (2015). Perceptual Representation / Perceptual Content. In M. Matthen (Ed.), *Oxford Handbook for the Philosophy of Perception* (pp. 153–167). Oxford University Press.

Nanay, B. (2017). All actions are emotional actions. *Emotion Review*, *9*, 350–352.

Neale, S. (1992). Paul Grice and the philosophy of language. *Linguistics and Philosophy*, *15*(5), 509–559.

Neale, S. (2004). This, That, and the Other. In A. Bezuidenhout & M. Reimer (Eds.), *Descriptions and Beyond* (pp. 68–182). Oxford: Oxford University Press.

Neale, S. (2016). Implicit and aphonic reference. In G. Ostertag (Ed.), *Meanings and Other Things: Themes from the work of Stephen Schiffer* (pp. 229–344). Oxford University Press.

Neander, K. (1991). The teleological notion of 'function.' *Australasian Journal of Philosophy*, *69*(4), 454–468. https://doi.org/10.1080/00048409112344881

Neander, K. (2018). Teleological Theories of Mental Content. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2018/entries/content-teleological/

Nesse, R. M. (2019). *Good Reasons for Bad Feelings: Insights from the Frontier of Evolutionary Psychiatry*. Penguin.

Nummenmaa, L., & Saarimäki, H. (2019). Emotions as discrete patterns of systemic activity. *Neuroscience Letters*, *693*, 3–8. https://doi.org/10.1016/j.neulet.2017.07.012

Nussbaum, M. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge University Press.

Nussbaum, M. (2004). Emotions as judgements of value and importance. In R. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 183–199). Oxford University Press.

O'Connor, T., & Franklin, C. (2020). Free Will. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2020/entries/freewill/

Öhman, A., & Soares, J. J. F. (1994). "Unconscious anxiety": Phobic responses to masked stimuli. *Journal of Abnormal Psychology*, *103*(2), 231–240. https://doi.org/10.1037/0021-843X.103.2.231

Origgi, G., & Sperber, D. (2000). Evolution, Communication and the Proper Function of Language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Language, Modularity and Social Cognition* (pp. 140–169). Cambridge University Press.

O'Shaughnessy, B. (2008). *The Will: Volume 2, a Dual Aspect Theory*. Cambridge University Press.

Owren, M. J., & Bachorowski, J.-A. (2003). Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior*, *27*(3), 183–200.

Pacherie, E. (2006). Toward a Dynamic Theory of Intentions. In S. Pockett (Ed.), *Does Consciousness Cause Behaviour?* MIT Press.

Pacherie, E., & Haggard, P. (2010). What are intentions? In L. Nadel & W. Sinnott-Armstrong (Eds.), *Conscious Will and Responsability* (p. 16). Oxford University Press.

Panksepp, J., & Burgdorf, J. (2003). "Laughing" rats and the evolutionary antecedents of human joy? *Physiology & Behavior*, *79*(3), 533–547.

Papineau, D. (1984). Representation and Explanation. *Philosophy of Science*, *51*(4), 550–572. https://doi.org/10.1086/289205

Papineau, D. (1993). *Philosophical Naturalism*. Blackwell. http://gen.lib.rus.ec/book/index.php?md5=b8ea2a294704214883ffac9d0e9a404c

Pater, W. (1873). *The Renaissance*. Boni and Liveright.

Paternotte, C. (2011). Being realistic about common knowledge: A Lewisian approach. *Synthese*, *183*(2), 249–276. https://doi.org/10.1007/s11229-010-9770-y

Peacocke, C. (2005). Joint Attention: Its Nature, Reflexivity, and Relation to Common Knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler

(Eds.), *Joint Attention: Communication and Other Minds* (p. 298). Clarendon Press.

Peláez, I., Martínez-Iñigo, D., Barjola, P., Cardoso, S., & Mercado, F. (2016). Decreased pain perception by unconscious emotional pictures. *Frontiers in Psychology*, *7*, 1636.

Perry, S., & Manson, J. H. (2009). *Manipulative Monkeys*. Harvard University Press.

Portner, P. (2007). Imperatives and modals. *Natural Language Semantics*, *15*(4), 351–383. https://doi.org/10.1007/s11050-007-9022-y

Potts, C. (2007). The Expressive Dimension. *Theoretical Linguistics*, *33*(2), 165–198.

Poyatos, F. (2002). *Nonverbal Communication across Disciplines: Volume 2: Paralanguage, kinesics, silence, personal and environmental interaction*. John Benjamins Publishing.

Preuschoft, S. (1992). "Laughter" and "smile" in Barbary macaques (Macaca sylvanus). *Ethology*, *91*(3), 220–236.

Price, T., Ndiaye, O., Hammerschmidt, K., & Fischer, J. (2014). Limited geographic variation in the acoustic structure of and responses to adult male alarm barks of African green monkeys. *Behavioral Ecology and Sociobiology*, *68*(5), 815–825. https://doi.org/10.1007/s00265-014-1694-y

Prinz, J. (2004). *Gut reactions: A perceptual theory of the emotions*. Oxford University Press.

Prinz, J. (2007). *The emotional construction of morals*. Oxford University Press.

Prinz, J. (2011). Is Attention Necessary and Sufficient for Consciousness? In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and Psychological Essays* (pp. 174–204). Oxford University Press.

Provine, R. R. (2001). *Laughter: A Scientific Investigation*. Penguin.

Provine, R. R. (2004). Laughing, tickling, and the evolution of speech and self. *Current Directions in Psychological Science*, *13*(6), 215–218.

Provine, R. R. (2017). Philosopher's disease and its antidote: Perspectives from prenatal behavior and contagious yawning and laughing. *Behav. Brain Sci*, *40*, e399.

Quilty-Dunn, J., & Mandelbaum, E. (2018). Against dispositionalism: Belief in cognitive science. *Philosophical Studies*, *175*(9), 2353–2372. https://doi.org/10.1007/s11098-017-0962-x

Quine, W. V. (1960). *Word and Object*. M.I.T. Press.

Ratcliffe, M. (2005). The feeling of being. *Journal of Consciousness Studies*, *12*(8–9), 43–60.

Reboul, A. (2017). *Cognition and Communication in the Evolution of Language*. Oxford University Press.

Recanati, F. (1986). On Defining Communicative Intentions. *Mind & Language*, *1*(3), 213–241.

Recanati, F. (2004). *Literal Meaning*. Cambridge University Press.

Recanati, F. (2007). *Perspectival Thought: A Plea for Moderate Relativism*. Oxford University Press.

Recanati, F. (2008). Pragmatics and Semantics. In *The Handbook of Pragmatics* (pp. 442–462). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756959.ch20

Recanati, F. (2010). *Truth-Conditional Pragmatics*. Oxford University Press.

Reisenzein, R. (2006). Arnold's theory of emotion in historical perspective. *Cognition and Emotion*, *20*(7), 920–951.

Reisenzein, R. (2009). Emotions as metarepresentational states of mind: Naturalizing the belief–desire theory of emotion. *Cognitive Systems Research*, *10*(1), 6–20.

Rescorla, M. (2012). Millikan on Honeybee Navigation and Communication. In D. Ryder, J. Kingsbury, & K. Williford (Eds.), *Millikan and Her Critics* (pp. 87–106). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118328118.ch4

Richard, M. (2008). *When Truth Gives Out*. Oxford University Press.

Richards, I. A. (1926). *Poetries and Sciences* (Vol. 2). Routledge & Kegan Paul, Limited.

Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution* (pp. ix, 332). University of Chicago Press.

Riley, J. R., Greggers, U., Smith, A. D., Reynolds, D. R., & Menzel, R. (2005). The flight paths of honeybees recruited by the waggle dance. *Nature*, *435*(7039), 205–207. https://doi.org/10.1038/nature03526

Rivas, E. (2005). Recent use of signs by chimpanzees (Pan troglodytes) in interactions with humans. *Journal of Comparative Psychology*, *119*(4), 404–417. https://doi.org/10.1037/0735-7036.119.4.404

Roberts, C. (1996). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 6–1.

Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology*. Cambridge University Press.

Ronan, P. (2015). Categorizing expressive speech acts in the pragmatically annotated SPICE Ireland corpus. *ICAME Journal*, *39*(1), 25–45.

Rosenthal, D. M. (1986). Two Concepts of Consciousness. *Philosophical Studies*, *49*(May), 329–359. https://doi.org/10.1007/BF00355521

Rossi, M., & Tappolet, C. (2019). What kind of evaluative states are

emotions? The attitudinal theory vs. the perceptual theory of emotions. *Canadian Journal of Philosophy*, *49*(4), 544–563.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145.

Sander, D. (2013). Models of emotion. In J. Armony & P. Vuilleumier (Eds.), *The Cambridge handbook of human affective neuroscience* (pp. 5–54). Cambridge University Press. https://doi.org/10.1017/CBO9780511843716.003

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18*(4), 317–352.

Sander, D., Grandjean, D., & Scherer, K. R. (2018). An appraisal-driven componential approach to the emotional brain. *Emotion Review*, *10*(3), 219–231. https://doi.org/10.1177/1754073918765653

Sauerland, U. (2007). Beyond unpluggability. *Theoretical Linguistics*, *33*(2), 231–236.

Saul, J. (2018). Dogwhistles, Political Manipulation, and Philosophy of Language. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New Work on Speech Acts*. Oxford University Press.

Saussure, L. de, & Wharton, T. (2020). Relevance, effects and affect. *International Review of Pragmatics*, *12*(2), 183–205. https://doi.org/10.1163/18773109-01202001

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2015). *Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations: Correction.*

Scarantino, A. (2010). Insights and blindspots of the cognitivist theory of

emotions. *The British Journal for the Philosophy of Science*, *61*(4), 729–768.

Scarantino, A. (2013). Animal communication as information-mediated influence. In U. Stegmann (Ed.), *Animal communication theory: Information and influence* (pp. 63–88). Cambridge University Press.

Scarantino, A. (2014). The motivational theory of emotions. In J. D'Arms & D. Jacobson (Eds.), *Moral psychology and human agency* (pp. 156–185). Oxford University Press.

Scarantino, A. (2015a). Basic emotions, psychological construction, and the problem of variability. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 334–376). Guilford Press.

Scarantino, A. (2015b). Information as a probabilistic difference maker. *Australasian Journal of Philosophy*, *93*(3), 419–443.

Scarantino, A. (2017). How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry*, *28*(2–3), 165–185.

Scarantino, A., & De Sousa, R. (2018). Emotion. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2018/entries/emotion/

Scarantino, A., & Piccinini, G. (2010). Information without truth. *Metaphilosophy*, *41*(3), 313–330.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal Processes in Emotion: Theory, Methods, Research*, *92*(120), 57.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256.

Scherer, K. R., & Meuleman, B. (2013). Human Emotion Experiences Can Be Predicted on Theoretical Grounds: Evidence from Verbal Labeling. *PLOS ONE*, *8*(3), e58166. https://doi.org/10.1371/journal.pone.0058166

Scherer, K. R., & Moors, A. (2019). The Emotion Process: Event Appraisal and Component Differentiation. *Annual Review of Psychology*, *70*(1), 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Schiffer, S. (1972). *Meaning*. Clarendon Press.

Schiffer, S. (1982). Intention-based semantics. *Notre Dame Journal of Formal Logic*, *23*(2), 119–156.

Schlenker, P. (2007). Expressive presuppositions. *Theoretical Linguistics*, *33*(2), 237–245.

Schlenker, P. (2016). The semantics-pragmatics interface. In M. Aloni & P. Dekker (Eds.), *The Cambridge Handbook of Formal Semantics* (pp. 664–727). Cambridge University Press.

Schlenker, P. (2018). What is Super Semantics? *Philosophical Perspectives*, *32*(1), 365–453.

Schlenker, P. (manuscript). *Triggering Presuppositions*. https://ling.auf.net/lingbuzz/004696

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Cäsar, C., & Keenan, S. (2016). Formal monkey linguistics. *Theoretical Linguistics*, *42*(1–2), 1–90.

Schroeder, M. (2016a). Frege–Geach problem. In *Routledge Encyclopedia of Philosophy* (1st ed.). Routledge. https://doi.org/10.4324/9780415249126-L149-1

Schroeder, M. (2016b). Value Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2016/entries/value-theory/

Schroeter, L., Schroeter, F., & Jones, K. (2015). Do Emotions Represent Values? *Dialectica*, *69*(3), 357–380.

Schwitzgebel, E. (2019). Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2019/entries/belief/

Scott-Phillips, T. (2008). Defining biological communication. *Journal of Evolutionary Biology*, *21*(2), 387–395.

Scott-Phillips, T. (2015). *Speaking Our Minds: Why Human Communication is Different, and how Language Evolved to Make it Special*. Macmillan International Higher Education.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Searle, J. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge University Press.

Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Searle, J. (1995). *The Construction of Social Reality*. Free Press.

Seemann, A. (Ed.). (2011). *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*. MIT Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Shargel, D. (2015). Emotions Without Objects. *Biology and Philosophy*, *30*(6), 831–844. https://doi.org/10.1007/s10539-014-9473-8

Shargel, D., & Prinz, J. (2018). An Enactivist Theory of Emotional Content. In H. Naar & F. Teroni (Eds.), *The Ontology of Emotions*. Cambridge University Press.

Shea, N. (2007). Consumers need information: Supplementing

teleosemantics with an input condition. *Philosophy and Phenomenological Research*, *75*(2), 404–435.

Shea, N., Godfrey-Smith, P., & Cao, R. (2018). Content in Simple Signalling Systems. *The British Journal for the Philosophy of Science*, *69*(4), 1009–1035. https://doi.org/10.1093/bjps/axw036

Shi, Y., Zheng, X., & Li, T. (2018). Unconscious emotion recognition based on multi-scale sample entropy. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1221–1226.

Sievers, C., & Gruber, T. (2016). Reference in human and non-human primate communication: What does it take to refer? *Animal Cognition*, *19*(4), 759–768.

Skerry, A. E., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Current Biology*, *25*(15), 1945–1954.

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge University Press.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press.

Smith, R., & Lane, R. D. (2015). The neural basis of one's own conscious and unconscious emotional states. *Neuroscience & Biobehavioral Reviews*, *57*, 1–29.

Smith, R., & Lane, R. D. (2016). Unconscious emotion: A cognitive neuroscientific perspective. *Neuroscience & Biobehavioral Reviews*, *69*, 216–238.

Solomon, R. (1977). The logic of emotion. *Noûs*, *11*(1), 41–49. https://doi.org/10.2307/2214329

Solomon, R. (1993). *The passions: Emotions and the meaning of life*.

Hackett Publishing.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. *Metarepresentations: A Multidisciplinary Perspective*, 117–137.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Blackwell.

Sperber, D., & Wilson, D. (2015). Beyond speaker's meaning. *Croatian Journal of Philosophy*, *15*(2 (44)), 117–149.

Stalnaker, R. (1976). Possible Worlds. *Noûs*, *10*(1), 65–75. https://doi.org/10.2307/2214477

Stalnaker, R. (1978). Assertion. *Syntax and Semantics (New York Academic Press)*, *9*, 315–332.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*(5/6), 701–721.

Stalnaker, R. (2014). *Context*. Oxford University Press.

Stegmann, U. E. (2015). Prospects for probabilistic theories of natural information. *Erkenntnis*, *80*(4), 869–893.

Sterelny, K. (1990). *The Representational Theory of Mind: An Introduction* (1st ed.). Wiley-Blackwell. http://gen.lib.rus.ec/book/index.php?md5=557044a63accdecd3a8d3c8ba4a6d2b4

Sterelny, K. (2006). The evolution and evolvability of culture. *Mind & Language*, *21*(2), 137–165.

Sterelny, K. (2012). *The Evolved Apprentice*. MIT press.

Sterelny, K. (2017). From code to speaker meaning. *Biology & Philosophy*, *32*(6), 819–838. https://doi.org/10.1007/s10539-017-9597-8

Strawson, P. F. (1964a). Intention and convention in speech acts. *The*

Philosophical Review, *73*(4), 439–460.

Strawson, P. F. (1964b). Intention and convention in speech acts. *The Philosophical Review*, *73*(4), 439–460.

Suppes, P. (1983). Probaility and information. *Behavioral and Brain Sciences*, *6*(1), 81–82.

Szameitat, D. P., Alter, K., Szameitat, A. J., Wildgruber, D., Sterr, A., & Darwin, C. J. (2009). Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America*, *126*(1), 354–366.

Tanaka, H., & Campbell, N. (2014). Classification of social laughter in natural conversational speech. *Computer Speech & Language*, *28*(1), 314–325.

Tanaka, H., Kashioka, H., & Campbell, N. (2011). Laughter as a gesture accompanying speech–towards the creation of a tool for the support of children on the autistic dimension. *Proc. GESPIN, Bielefeld, September*.

Tappolet, C. (2000). *Émotions et valeurs*. Presses universitaires de France.

Tappolet, C. (2016). *Emotions, value, and agency*. Oxford University Press.

Todt, D., & Vettin, J. (2005). Human laughter, social play, and play vocalizations of non-human primates: An evolutionary approach. *Behaviour*, *142*(2), 217–240.

Tomasello, M. (2008). *Origins of Human Communication*. MIT press.

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., Goncalves, I. B., Burkart, J. M., Flower, T., Gaunet, F., Glock, H. J., Gruber, T., Jansen, D. A. W. A. M., Liebal, K., Linke, A., Miklósi, Á., Moore, R., Schaik, C. P. van, Stoll, S., … Manser, M. B. (2017). Exorcising Grice's ghost: An empirical approach to studying intentional communication in animals. *Biological Reviews*, *92*(3), 1427–1433.

https://doi.org/10.1111/brv.12289

Turpin, G. (1986). Effects of Stimulus Intensity on Autonomic Responding: The Problem of Differentiating Orienting and Defense Reflexes. *Psychophysiology*, *23*(1), 1–14. https://doi.org/10.1111/j.1469-8986.1986.tb00583.x

Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press.

Tye, M. (2008). The experience of emotion: An intentionalist theory. *Revue Internationale de Philosophie*, *1*, 25–50.

Tye, M. (2016). *Tense Bees and Shell-shocked Crabs: Are Animals Conscious?* Oxford University Press.

Van der Goot, M. H., Tomasello, M., & Liszkowski, U. (2014). Differences in the nonverbal requests of great apes and human infants. *Child Development*, *85*(2), 444–455.

Vetter, P., Badde, S., Phelps, E. A., & Carrasco, M. (2019). Emotional faces guide the eyes in the absence of awareness. *ELife*, *8*. https://doi.org/10.7554/eLife.43467

Vlach, F. (1981). Speaker's meaning. *Linguistics and Philosophy*, *4*(3), 359–391.

Welby, V. (1903). *What is Meaning?: Studies in the Development of Significance*. John Benjamins Publishing.

Wharton, T. (2009). *Pragmatics and Non-verbal Communication*. Cambridge University Press.

Wharton, T. (2016). That bloody so-and-so has retired: Expressives revisited. *Lingua*, *175*, 20–35.

Whiting, D. (2011). The Feeling Theory of Emotion and the Object-Directed Emotions. *European Journal of Philosophy*, *19*(2), 281–303. https://doi.org/10.1111/j.1468-0378.2009.00384.x

Whiting, D. J. (2008). Conservatives and Racists: Inferential Role Semantics and Pejoratives. *Philosophia*, *36*(3), 375–388. https://doi.org/10.1007/s11406-007-9109-1

Wild, B., Rodden, F. A., Grodd, W., & Ruch, W. (2003). Neural correlates of laughter and humour. *Brain*, *126*(10), 2121–2138.

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press. http://gen.lib.rus.ec/book/index.php?md5=be599a443a810443947d8ed79cb222e8

Williamson, T. (2009). Reference, Inference, and the Semantics of Pejoratives. In J. Almog & P. Leonardi (Eds.), *The Philosophy of David Kaplan*. Oxford University Press.

Wilson, D., & Sperber, D. (2006). Relevance theory. In L. Horn (Ed.), *The Handbook of pragmatics*. Blackwell.

Winkielman, P., & Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science*, *13*(3), 120–123.

Wittgenstein, L. (1953). *Philosophical Investigations* (G. E. M. Anscombe, Trans.). Basil Blackwell.

Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology*, *53*(1), 205–214.

Zeman, D. (2014). Meaning, Expression and Extremely Strong Evidence: A Reinforced Critique of Davis' Account of Speaker Meaning. *Thought: A Journal of Philosophy*, *3*(3), 218–224. https://doi.org/10.1002/tht3.137

Zuberbühler, K. (2018). Intentional communication in primates. *Revue Tranel*, *68*, 69–75.