Art and Learning

A Predictive Processing Proposal

Jacopo Frascaroli

PhD

University of York

Philosophy

August 2022

Abstract

This work investigates one of the most widespread yet elusive ideas about our experience of art: the idea that there is something cognitively valuable in engaging with great artworks, or, in other words, that we *learn* from them. This claim and the age-old controversy that surrounds it are reconsidered in light of the psychological and neuroscientific literature on learning, in one of the first systematic efforts to bridge the gap between philosophical and scientific inquiries on the topic. The work has five chapters. Chapter 1 lays down its conceptual bases: it explains what learning is taken to be in the current philosophical debate and it points out how Bayesian cognitive science (particularly in its predictive processing formulations) might be well-suited to capture the kind of learning involved in our engagement with the arts. The following chapters test this latter hypothesis with respect to particular art forms, namely literature and literary language (Chapter 2), narrative (Chapter 3), and visual art, music and motor activities (Chapter 4). The fine-grained discussions conducted in each of these areas will enable us to see that the relationship between art and learning is indeed fundamental and pervasive. The final chapter (Chapter 5) examines the consequences of this fact for our understanding of the role of art in our epistemic practices, its ultimate usefulness and value, and its place in the interdisciplinary study of the human mind. The upshot is a novel and wide-ranging picture, both philosophically informed and empirically sound, that bypasses many of the problems and dead ends of the current philosophical debate on the topic and captures the deep sense in which art and learning are interrelated.

List of Contents

List of Figures

Acknowledgements

I am very grateful to the following mentors, friends and colleagues for their help, guidance and support: Gregory Currie, Lena Wimmer, Heather Ferguson, Stacie Friend, Kayleigh Green, Jamie Cawthra, Celia Coll, Samuel Dickson, Eleanor Byrne, Daniel Kim, Peter Lamarque, James Clarke, David Worsley, Janet Eldred, Julie Kay, Andrea Blomkvist, Deborah Marber, Sander Van de Cruys, Kara Rancourt, Yu-Wen Lee.

I also wish to thank all the speakers and participants of the “Art and Affect in the Predictive Mind” conference I organised in June-July 2021, as well as all the participants of the conferences I attended in the past four years who provided useful feedback for my work.

Additional thanks to the Leverhulme Trust and the Department of Philosophy at the University of York for their financial support.

Declaration

I declare that this thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Introduction

Few impressions are more deeply rooted in our experience as readers, beholders and listeners and yet more elusive and difficult to capture than the impression that we gain something by being exposed to instances of great art. While engaging with a masterful novel, a great painting or a compelling piece of music, we feel captivated and intrigued and we find the experience rewarding and worth pursuing further – an impression so widespread and familiar that we hardly see it as something deserving of explanation. Yet, try to imagine how bizarre our behaviour in this respect must appear to one of our fellow living beings that is unaware of the subtle joys of art. Imagine it seeing one of us following the row of black characters on a page with the same excitement of an animal searching the ground for food, or inspecting a coloured canvas with the concentration of a prey looking for predators amidst the tall grass, or being wakened and alerted by a piece of music as if it were the call of a mate or a cry of alarm. The spectacle of this animal, frozen in place, engrossed and fascinated, tense or joyful, his mind and senses captured for minutes or hours by something apparently so devoid of the slightest biological significance would no doubt come as something of a mystery. But we feel moved and touched by these strange stimuli; we feel that there is something important in them, something we should care about; we would swear that they are affecting us in a way that is repaying and encouraging our serious and sustained attention. And even if one cannot say exactly what has changed when one puts down the book or leaves the theatre, when the painting is behind us or the music is over, one feels enriched in some way, sometimes shaken up quite devastatingly, at other times maybe just aware that the colouring with which one sees the world has changed slightly. When art works at its best, the contact with it seems to assume the traits of a transformative experience that leaves us not quite the same as we were before.

It seems quite natural, therefore, and close to our intuitions as readers, beholders and listeners in those circumstances, to claim that we *learn* something from our encounter with the work. The way the work affects us, the modifications that it prompts in us, have the appearance of a cognitive improvement. It is also quite natural to consider this improvement essential to great art. If after our encounter with an artwork we do not feel changed and enriched, if we are left cold and unshaken, then arguably the artwork in question is not a great artwork – it has failed to produce what great art characteristically produces. Whether or not one agrees on this latter point, it is true that, at least since Aristotle and throughout the history of aesthetics, the pleasure that we get from the arts has been related to knowledge acquisition and has been seen as fundamentally epistemic in nature. According to this line of thought, as we shall see, acquiring knowledge (learning, understanding, comprehending) is pleasurable, and art affords this pleasure to a characteristically high degree. The claim that art affords learning is therefore not only somewhat close to our intuitions as art consumers but also the object of a philosophical attention (and controversy) that runs throughout the history of philosophy, from Plato and Aristotle to the present-day opposition between cognitivists and anti-cognitivists in the analytic philosophy of art.

This work is meant as a contribution to the age-old debate about whether we learn from art. With respect to the existing philosophical discussion on the topic, however, it represents a decisive shift in focus and concerns. To illustrate the character and the reasons of this shift, it might be helpful to say a few words about my personal trajectory in exploring these issues over the past four years.

In 2018 I started my PhD in York as part of a Leverhulme-funded project entitled "Learning from Fiction: A Philosophical and Psychological Study", under the supervision of Prof. Gregory Currie. The project was, as the title suggests, eminently interdisciplinary in character: as part of a team of philosophers and psychologists from three universities in the UK, I was tasked with bridging the gap between the philosophical and psychological debates on the different kinds of learning that fictional stories might promote. An enterprise that, even back then, struck me as both fascinating and very difficult to accomplish. Having studied mainly literature up to that moment, I knew very little about the philosophical debate around learning from fiction, and even less about the psychology of learning. It was not, therefore, a matter of filling in a few gaps in my knowledge, but rather of building my understanding of the problems involved almost from scratch. As neither a philosopher approaching psychology nor a psychologist approaching philosophy, I had both the need and the opportunity to aim from the start at a more comprehensive picture, a picture able to integrate without preconceptions philosophical insights and empirical acquisitions. To this end, I audited both philosophy and psychology modules, gave talks at (and organised) conferences at the intersection of the two disciplines, learned what I could about cognitive science and neuroscience. This rich interdisciplinary engagement led me to discover many intriguing strands of research that I would not have encountered otherwise, and provided me with a vivid picture of what research looks like when it works at its best. However, the more I tried to draw meaningful connections between philosophy and psychology on these topics and combine them into a unitary picture, the more I became aware of the profound differences between the two disciplines in terms of backgrounds, interests and aims. The kind of notions that most contemporary philosophers in the learning from fiction debate use, the kind of questions they ask, just did not seem to match well with the notions and questions used and asked by psychologists, cognitive scientists and neuroscientists.

Take the notion of "learning", which is so central for our purposes. As with most contemporary epistemology in the analytic tradition, the project of the philosopher when it comes to defining learning appears to be largely a project of conceptual analysis. The point is not to discover how empirical subjects learn about the world, but to define, if possible by means of necessary and sufficient conditions, what it is to learn something. Truth, belief, justification, reliability are the notions most commonly invoked. Thus, one might say for example that learning is the acquisition of justified true (or closer to the truth) beliefs. Once a definition is given, counterexamples and imaginary scenarios might be devised that tease our intuitions about its soundness and might lead us to revise or refine it in various ways. All this can be done (and is normally done) without any reference to the psychology and neuroscience of learning. Whether or not subjects display behavioural, psychological or neurophysiological changes that reflect "learning" in this philosophical sense (i.e. whether this "learning" is learning for the subjects and their brains as well as for the epistemologist) is a different question, perhaps interesting, but certainly not decisive for establishing what learning is.

The project of the psychologist, the cognitive scientist or the neuroscientist, on the other hand, is eminently descriptive. What they are trying to do is to define what learning amounts to in and for the subject, in terms of psychological, behavioural and neurophysiological changes. Psychologists are not normally concerned with truth, beliefs, justification, reliability; instead, they talk about Pavlovian conditioning, "Aha!" experiences, Hebbian plasticity, ERP responses, and all the phenomena that seem to accompany or constitute learning from an empirical standpoint. From the study of these phenomena, theories of learning are developed that can account for empirical evidence. One might argue, for example, that based on what we know about brain plasticity and electrophysiology, learning can be seen as the process by which the brain updates its probabilistic model of environmental contingencies. From a psychological perspective, considerations about whether the beliefs that we acquire are true, justified, or reliably obtained do not play any role in defining learning – unless of course it is proved that these facts make some difference for the psychology, behaviour or neurophysiology of the learning subject. If for example no psychological, behavioural or neurophysiological event marks the acquisition of a belief "closer to the truth", then arguably this notion should have no place in a psychological theory of learning.

These two approaches, as we shall see, lead to very different conceptions of the sort of stimuli that are conducive to learning. To the philosopher, a stimulus will be conducive to (at least one kind of) learning if it can lead to the acquisition of many justified true beliefs; to the psychologist, instead, a stimulus will be conducive to learning if it can prompt to a significant degree the sort of changes in psychology, behaviour and neurophysiology that are indicative of learning. By the same token, these two approaches also lead to very different treatments of the question of whether we learn from fiction or art. The philosopher will try to find out whether our engagement with fiction or art leads to the acquisition of justified true beliefs to a significant degree. She will ask questions such as: what is the pathway from fiction to belief? Is this pathway apt to provide beliefs that are true and justified? Are the beliefs we acquire from fiction reliably acquired? Is the author trustworthy and is the process by which we trust her rational? The psychologist, instead, will look at whether fiction or art are apt to produce those psychological, behavioural and neurophysiological changes that are characteristic of learning. She will ask questions such as: do artistic stimuli tend to promote a subjective feeling of insight? Do they produce to a significantly high degree the electrophysiological responses or the changes in connectivity that are indicative of learning? Do they make us feel or behave differently, and, if so, how do they manage to do so?

One of the key points that I will make in this work is that if we want to capture that soul-altering force, that feeling of insight that characterises our encounter with great art we should put aside the standard philosophical characterisations of learning and take more seriously the psychological perspective. This is because, as I shall argue, the various characterisations of learning that dominate the current philosophical debate are inadequate for (and are in fact preventing us from) capturing what art does to us, how it changes us and how it manages to produce those epiphanies and revelations that we so consistently attribute to it. This inadequacy pertains not only to the notion of learning as the acquisition of justified true beliefs (what philosophers call "knowledge-that"), but also to all other notions most commonly invoked in the philosophical debate (knowing-how, knowing-what-it-is-like, and so on). In trying to assess whether we learn from art in these philosophical senses of learning, I will argue, the philosopher might be missing the fundamental and pervasive relationship between art and learning – a relationship that, meanwhile, a growing stream of psychological and neuroscientific research is starting to unveil.

This does not mean that these philosophical characterisations of learning should be dismissed as irrelevant, or that we should not try to assess whether we learn from fiction or art in these philosophical senses of learning. Trying to assess whether we acquire new justified true beliefs, new skills or empathic capacities from our encounter with fiction or art might well be a worthwhile enterprise (although I will suggest that it might be less interesting than it seems). The point is rather that even if we were able to establish that, in general, art is conducive to these kinds of learning, we would not have yet said anything about that powerful impression of learning and insight that we get while engaging with great art. And that impression is, arguably, what prompted the question of whether we learn from art in the first place. As I will show, one thing is to establish whether we acquire new true beliefs, skills or capacities from our engagement with art, and another thing is to discover the roots of its power to persuade, fascinate and transform, to touch us and matter to us the way it does. As a result, even if we were to settle the first question, we would not have even begun to address the latter.

With its descriptive ambitions, psychological and neuroscientific research seems to be well-positioned to tackle this latter task. At this point, psychology and neuroscience have in fact developed quite sophisticated tools to capture and describe in detail what learning looks like in the phenomenal experience of the art consumer, and in her brain. The flourishing psychological research on art engagement has by now produced an abundance of constructs (captivation, transportation, absorption, immersion, flow, affection, being moved, and so on) meant to capture this very experience. In the meantime, the neural and physiological correlates of our engagement with art are being examined with increasing precision, providing further indications about the ways artworks affect us. This strand of work is not only giving us an increasingly clearer picture of the general relationship between art and learning, but it is also unveiling the more minute details of this relationship. It is allowing us, for example, to follow the fast-paced dynamics of information gain during our aesthetic encounters, or to understand the strategies that artists use to win and maintain our interest through time – thus complementing in important ways the intuitive insights of artists and the theoretical acquisitions of aestheticians and art historians. To overlook this wealth of interdisciplinary insights would be unwise. Whether or not the reader will agree with particular proposals put forward in this work, I hope it will be apparent, by the end, that the philosophical discussion on whether and how we learn from art simply can no longer ignore the acquisitions of disciplines where the study of learning is a paramount concern: (developmental) psychology, cognitive science, neuroscience, artificial intelligence, machine learning and other related fields.

This is, therefore, the shift in focus that the present work tries to accomplish: bringing more decisively the psychological and neuroscientific understanding of perception, cognition and learning into the philosophical discussion about the cognitive value of art. The points from the psychological and neuroscientific literature that I will bring to bear on our discussion are for the most part rather uncontroversial, reflecting the current general understanding of perception, cognition and learning in these fields. In various circumstances, however, I will refer to a particular framework in order to put forward more precise and detailed proposals about the way we learn from art. This framework is predictive processing (PP), an increasingly important framework in cognitive science and neuroscience, with very wide-ranging explanatory ambitions. Originally conceived as a general theory of brain function, PP has rapidly been extended to explain "perception, action, reason, attention, emotion, experience, and learning" (Clark, 2015, p. 9) and even, according to the advocates of its most general formulation, life and sentience as such (Friston, 2013a). Apart from its ostensibly vast explanatory reach, the rationale for adopting PP in particular as a framework for my proposal is twofold. On the one hand, as we shall see, PP allows us to condense and operationalise many of the key insights about perception, cognition and learning commonly held in contemporary psychology and cognitive science. These include the Helmholtzian view of perception as a process of hypothesis-testing, the idea of the mind/brain as a probabilistic modeler of environmental contingencies, the importance of affect in perception and cognition: all aspects that, as we shall see, are crucial for our understanding of the relationship between art and learning and that within the PP framework seem to hold together in a particularly convincing way. On the other hand, the interest in the PP framework for our discussion resides in the fact that, while I write, this framework is being increasingly applied to illuminate issues about art and aesthetics. In fact, tentative PP accounts of several art forms have already been developed, and an intriguing convergence of interests and results is emerging in this area between philosophers and art historians on the one hand, and neuroscientists and cognitive scientists on the other. We shall return to this convergence throughout this work. For the moment, let us just note that PP seems to offer the ideal tools for articulating a discussion about art and learning both precise in its details and wide-ranging in its implications. The PP picture of learning, how this differs from the philosophical notions of learning, and what this difference entails for the question of whether we learn from art will be the object of the first chapter of this work.

As we shall see, one of the advantages of the PP story about art and learning is that it seems to apply equally well to all the arts. Whether we happen to be captured by a great novel, a painting or a piece of music, the feeling of epistemic gain that we experience can be conceived, in light of PP, as the product of the same underlying dynamics, active in different ways in different cases. This fact will allow us to expand considerably the scope of the current philosophical debate on art and learning, which has so far been focused mainly on literature and narrative. It will also allow us to remedy another traditional limitation of this philosophical debate, namely the stress put on propositions, propositional attitudes and propositional knowledge when assessing the cognitive value of art. Learning, we are often told, implies belief change, and beliefs are propositionally individuated. Even those who insist that fiction and art provide us with non-propositional kinds of knowledge define these other kinds mainly by contrast with propositional knowledge. To the philosopher, the stress on propositions might well be second nature, courtesy of the Fregean identification of thought and proposition and the legacy of the philosophy of language in shaping traditional debates in the philosophy of mind. But to the psychologist, the idea that a relevant part of our mental life happens in a propositional format seems increasingly strange and unlikely. In most contemporary psychology and cognitive science, the explanatory role played by propositional "beliefs" in traditional philosophy of mind is now played by states and processes whose best expression is non-sentential: increasingly, probability distributions, or parameters and hypotheses of probabilistic models. As we shall see, once we abandon the focus on propositions and propositional knowledge and embrace these other ways of characterising our mental life, the differences between traditional philosophical notions such as knowing-that, knowing-how and knowing-what-it-is-like start to dissolve. More importantly for our aims, abandoning the stress on propositions will allow us to see the profound similarities between the arts with respect to their capacity to elicit learning. "Propositional" works such as novels will appear to stand on the same ground as works that do not have a propositional content, or works that do not have any representational content at all. Literature, visual art, music, and even sensorimotor activities such as sport, dance and skilful action will be shown to function, in this respect, in a very similar way. For this reason, while I will devote considerable space to literature and narrative (Chapters 2 and 3) in line with the aims of the "Learning from Fiction" project, I have also chosen to illustrate how the same PP story might be applied across the spectrum aesthetic activities (Chapter 4), and how this has wide-ranging consequences for our understanding of the place of art in our epistemic practices (Chapter 5).

The upshot, I think, is a proposal with a pleasing generality, able to overcome many of the difficulties and dead ends of the philosophical debate on the topic and capture the profound sense in which art and learning are interrelated. It is a proposal that, while finding guidance and inspiration in disciplines outside philosophy, by no means abandons the fundamental preoccupations at the hearth of the philosophical debate on art and learning. The dialogue with psychology, cognitive science and neuroscience will instead help us identify the real philosophical points at stake. As a result of the contact with these disciplines, some of the questions in the current philosophical debate will appear to be marginal or misguided, and some others will be brought to the foreground. The resulting picture will be one that vindicates both scientific acquisitions and old and venerable philosophical intuitions about the cognitive power of art. What is more, once a clear link between art and learning is established, it can be pursued in both directions. If the scientific understanding of perception, cognition and learning can inform our study of art and aesthetics, then the study of art and aesthetics can contribute to our understanding of perception, cognition and learning. In this sense, the present work can be seen as a first step towards the establishment of a framework in which aesthetics and cognitive science might bring each other into proper focus, avoiding sectarian or colonialist attitudes and collaborating as equal partners in illuminating crucial aspects of the human mind.

1. Kinds of Learning

1.1 The Feeling of Having Learnt

In starting to think about art and learning, we should probably ask ourselves why the problem of their relationship arises in the first place. In other words, why has the question of whether we learn from art been considered interesting and worth of investigation by a long line of humanists, scientists and laypeople? Quite evidently, for the question to arise at all, the claim that we learn from art must have some intuitive plausibility but be at the same time not easy to establish.

What makes it plausible or intuitive that we learn from art is primarily, I will argue, our awareness of how great art makes us feel when we engage with it. When we encounter an effective artwork, we have a distinctive feeling of cognitive progress, an awareness that the work is being informative and enlightening; once we finish engaging with it, we have the impression that we do not stand where we were before. If one follows the centuries-old debate on art and learning, one gets the impression that it is this *feeling of having learnt*, more than any rational argument or empirical evidence, that grounds the widespread conviction that we learn from art – fuelling the debate about whether we actually do.[1] When advocates of the cognitive value of art insist that we learn from it, their conviction might be based primarily, I think, on their clear feeling that art affects them in beneficial ways. Even those who (from Plato onwards)[2] are sceptical about the cognitive value of art do not normally deny its power to generate this feeling of learning and insight, but warn instead that this power might be harmful, and this feeling misleading. Before we embark on the task of establishing whether and in what sense we learn from art, it might therefore be useful to try to describe this feeling in some detail, relying, for the time being, mainly on our shared experience as art consumers and on the characterisations of this experience that have been produced in the philosophical and psychological literature.[3]

One of the most noticeable features of successful engagement with art is its *self-sustaining* and *self-perpetuating* character. When an artwork is effective, we are compelled to devote to it our serious and sustained attention. The artwork in question feels worth attending to: we feel that if we keep reading on, if we keep examining the pictorial surface or listening to the piece of music as it unfolds, our attention will be rewarded. This feeling seems to be distinctively

[1] On this debate, as well as on the lack of rigorous arguments or empirical evidence in favour of the humanistic claims about the cognitive value of fiction and art, see Currie (2020). Currie points out that “the belief that reading quality fiction does us some good is almost an article of faith with liberal, educated people” (p. 4).

[2] The *locus classicus* is of course *Republic*, books II, III and X.

[3] To be clear, I do not take this feeling of having learnt to be always (or even typically) present in our engagement with art. In many cases, our encounter with an artwork leaves us cold and unshaken, and part of our task will be to understand why this happens. What I wish to claim, instead, is that when we do have this feeling of having learnt in relation to an artwork, then our experience tends to assume, with varying degrees of intensity, the traits that I am about to describe.

epistemic in character: it is the impression that the work is continuously disclosing – or might still disclose – something new, that there are still things that it has not told us yet and might tell us if we keep attending to it. As a result, we are compelled to prolong our engagement, and, in the best cases, to revisit the work many times, as with an inexhaustible source of insight. Kant captured this phenomenon quite aptly when he noted, in the third *Critique*, that

> [Pleasure in the beautiful] does have a causality inherent to it, namely that of preserving the state of contemplation itself and keeping the cognitive powers engaged without any further aim. We linger in our contemplation of the beautiful, because this contemplation strengthens and reproduces itself. (1987 [1790], §12)

This state of engrossment and fascination, familiar enough to any art consumer, has been an object of study for some time now in the fast-growing psychological literature on the experience of art. Here it is captured under a variety of constructs and definitions: captivation, transportation, absorption, immersion, flow, affection, being moved (see Schindler et al., 2017 and Menninghaus et al., 2019 for a review). We shall return to some of these constructs later; for now, let us simply notice that this ability of successful artworks to "keep our cognitive powers engaged" seems to be an important component of the feeling that we learn from them.

The second thing to notice about successful encounters with art is that they are often perceived and characterised as *transformative experiences*: great art, it is said, does not merely confirm the view of the world that we already possess; instead, it challenges and questions it, prompting us to revise or expand it in significant ways. As a result, after the experience we feel that the artwork has affected us, that it has left traces in us, that it has contributed to making us somewhat different from what we were before.[4] These changes might be more or less dramatic, of course, and more or less lasting. In the best cases, one might even come to see the encounter with the work as a memorable turning point, an important shift in one's understanding of oneself and the world. In many other cases, the changes brought about by the work will be peripheral and short-lived. But that there be *some* change seems to be one of the requisites for the success of an artwork. If the work does not change us, if it has no influence, conscious or unconscious, on the people we end up being, then arguably it has failed to provide what great art should provide. Or, in any event, it has failed to provide the felt experience of learning we are concerned with here.

Successful encounters with art however do not simply change us, but do so *for the better* – or so it seems to us. The shifts in worldview that great artworks bring about have often the taste of an improvement. We might experience a feeling of mental enlargement, of having now access to a range of thoughts to which we were before strangers; we may have the impression of new worlds being opened up, of deep, powerful truths being conveyed in some way; we may sense that we have made progress toward the understanding of puzzling aspects of our world or our condition, our mind and its workings, our feelings and impulses, and that, thanks to the artwork,

[4] See Pelowski and Akiba (2011) for a brief history of the line of thought that stresses disruption and transformation as essential aspects of an aesthetic experience.

we have achieved some decisive clarification. Accordingly, we often credit the great artist with especially penetrative powers and with an uncommon sensitivity to aspects of the real world, especially the human world of motivation, thought and feeling. There is perhaps no need to trace in any detail the history of such an attribution, which is so common and obstinate, and perhaps as old as art itself. Suffice it to remember that Homer, Hesiod and other poets were regarded in ancient Greece as paradigmatic givers of wisdom, and used as evidential sources in disputes about the most diverse matters; similarly, the great Greek tragedians are regarded still today as masters of in-depth psychological thinking, capable of profound and exact portrayals of human instincts and motivations; Shakespeare was also surrounded by a similar reputation, being considered by Samuel Johnson as able to unveil the human condition as such, and being deemed "inconceivably wise" by Emerson (1987 [1850], p. 121); and similar enthusiastic judgements are frequently made about the landmark novelists of the last two centuries, to which many observers in both literary and philosophical circles are apt to attribute competences and capacities that exceed those of the psychologist and the systematic thinker.[5] Nor are such claims confined to verbal art alone, if Heidegger could discover, thanks to Van Gogh's *A Pair of Shoes*, "what shoes are in truth" (1971, p. 35), and Elgin can claim that Bach's *B-Minor Mass* "conveys, more adequately than any theology text, the utter incomprehensibility of the divine forgiveness" (2002, p. 10). These kinds of epiphanies and revelations are again something that we seem to naturally associate with great art. As noted by the art historian David Freedberg (1989, p. 60), people are not only "stirred" by works of art: they "expect to be elevated by them… they have always responded in these ways; they still do."

A last trait of the experience we want to consider here is that it is *pleasurable*. The feeling of insight and discovery that characterises our engagement with great art is normally accompanied by positive affect. Granted, art might arouse in us all sorts of emotion, pleasant and unpleasant. But the idea here is rather that, insofar as an artwork affords discoveries or insights, these insights are accompanied by a pleasurable affective reaction, a kind of joy in having made cognitive progress. Indeed, according to a venerable philosophical tradition, the pleasure elicited by art is epistemic in nature, and has to do with knowledge acquisition. Art would be pleasurable *because* it affords learning. This line of thought runs throughout the history of aesthetics. Aristotle had already linked the pleasure elicited by the imitative arts with their ability to promote learning (see *Poetics*, 1448b 13-19 and *Rhetoric*, 1371b 4-10);[6] a very similar idea underlies Baumgarten's claim that aesthetics is concerned with the "perfection of sensitive cognition" (1936 [1750], §14), as well as Kant's (1987 [1790]) view that the experience of the beautiful is linked with the intuition of "a formal purposiveness" in the manifold of experience; Dewey (2005 [1934]) and more recently Johnson (2018) have similarly considered the arts as "intensified,

[5] See for example Nussbaum (1990) and Mack (2012).

[6] In *Poetics* 1448b 13-19 Aristotle notes: "it is natural for all to delight in works of imitation... though the objects themselves may be painful to see, we delight to view the most realistic representations of them in art... to be learning something is the greatest of pleasures not only to the philosopher but also to the rest of mankind... the reason of the delight in seeing the picture is that one is at the same time learning - gathering the meaning of things, e.g. that the man there is so and so."

nuanced and complex realizations of the processes of meaning in everyday life" (Johnson, 2018, p. 25). Nor is this view just held by philosophers: as we shall see, in recent psychological and neuroscientific research, perceptual pleasure is increasingly seen as an effect or marker of successful assimilation of sensory information.[7] The common theme of an ancient and very much alive strand of research is in sum that art provides exemplary instances of comprehension, insight or understanding, and owes to this very fact its pleasure and its allure.

The above characterisation of our engagement with successful artworks is no doubt sketchy, but it is indicative, I think, of the kind of experiences that fuel our intuition that we learn from art. We are left with the picture of a pleasurable, self-sustaining transformative experience that has the appearance of a cognitive gain. On this description, the layperson, the scientist and the humanist seem to agree. It is close to the experience of readers, listeners and beholders, it is captured by a wealth of psychological constructs and an increasing number of empirical studies in the psychology and neuroscience of art, and it is supported by an ancient philosophical tradition that links art and knowledge acquisition. The issue becomes, then, how we are to account for this phenomenon. What is the notion of learning that can best capture this persistent feeling of having learnt? What kind of learning, if any, is involved in our engagement with the arts? In the rest of this chapter, we shall try to establish whether current philosophical or psychological research might provide an answer to these questions.

1.2 Learning According to the Philosopher

1.2.1 Learning-That, Learning-How, Learning-What-It-Is-Like

When (analytic) philosophers approach the question of what should count as "learning", they tend to do so against a very specific background: the work done in epistemology roughly since the middle of the twentieth century. A major strand of this work revolves, understandably, around the notion of knowledge, and the understanding of this notion is in turn influenced by the very specific theoretical purposes philosophers put it to.[8] In this tradition, the most important form of knowledge is knowledge of facts. This is also known as "propositional knowledge", or "knowledge-that", on the assumption that a fact corresponds to a true proposition *p*, and therefore knowing a fact is knowing *that p*. The task of the epistemologist becomes then to identify the set of conditions that must obtain for a subject S to know that *p*.[9] According to the most influential analysis of this kind, there are three necessary and sufficient conditions for S to

[7] See for example Biederman and Vessel (2006), Schmidhuber (2010), Schoeller and Perlovsky (2016); Sarasso et al. (2020). The developing PP story about aesthetics is also, as we shall see, perfectly in line with this hypothesis.

[8] See Jackson (2011) for a discussion on this point.

[9] This can be understood as a task of conceptual analysis or as a metaphysical inquiry into the nature of knowledge. Quite often, however, epistemologists engaging in the project of analysing knowledge do not clarify whether their claims are to be intended as conceptual or metaphysical ones. See Ichikawa and Steup (2017) on this ambiguity.

know that *p*: S needs to *believe* that *p* (you can't "know" propositions you do not believe), *p* needs to be *true* (you can't "know" false propositions), and the subject needs to be *justified* in believing that *p* (you can't "know" a proposition if you arrived at it by sheer luck). According to this analysis, therefore, knowledge of facts is the possession of *justified true beliefs*.

Intuitive as it might seem, this analysis faces well-known problems. It is unclear, for example, how exactly we should understand these three conditions, whether the conditions are sufficient for knowledge (Gettier, 1963), or whether the notion of knowledge is analysable at all (Williamson, 2002). More radically – and more importantly for our purposes – it is anything but clear that we should think of beliefs as being propositionally individuated. As we noted in the Introduction, in most contemporary psychology and cognitive science the idea that our mental life takes place in a propositional format is considered strange, and the role of propositional beliefs is played by other states or processes whose best expression is non-sentential. In the impossibility of settling definitively any of these questions, a general understanding remains in contemporary epistemology that "knowledge-that" is an important form of knowledge, and that it can be analysed at least on first approximation in terms of the possession of propositionally individuated justified true beliefs.

Now if we start from this background, we arrive quite naturally at the first notion of learning that is discussed in the philosophical literature. If learning is the acquisition of knowledge, and if factual knowledge (knowledge-that) is the possession of justified true beliefs, then there ought to be a kind of learning that amounts to the acquisition of justified true beliefs: a *learning-that*. This, we are told by the philosopher, can take various forms. For example, it can be the shift from the absence of belief about a certain state of affairs to a true belief about that state of affairs. If I was previously ignorant of the capital of Suriname, I learn something when I find out that it is Paramaribo. Now I know more about Central America than I did yesterday. But learning can also take the form of a shift from a false belief to a true belief: If I thought that the capital of Suriname was Cayenne, I learn something when I find out that it is in fact Paramaribo. Finally, it seems reasonable, as Currie (2016) urges us to do, to admit that learning takes place also when there is a shift from a false belief to a belief that is not exactly true, but at least "closer to the truth" than the belief it replaces. Thus, it seems intuitively right that "converts from the Aristotelian-Ptolemaic view to Newtonian mechanics and gravitation learned something, though it was not a transition from falsehood to truth. Their learning consisted in epistemic improvement, but not in knowledge" (Currie 2016, p. 407).[10]

Whatever the specific case, what is crucial in this characterization of the learning process is the relation of the newly acquired belief to truth: you only learn something if your newly acquired belief is true, or at least closer to the truth than the belief it replaces. This is said to be reflected also in our linguistic use: "knowing" and "learning" are factive verbs, that is, verbs that presuppose the truthfulness of the proposition to which they apply. I shall refer to this conception as the *epistemic notion of learning*. Learning in this epistemic sense will designate

[10] Even if, as the author notes, "no one seems to have come up with a satisfactory account of what it takes for one proposition to be closer to the truth than another" (p. 417).

therefore the kind of cognitive progress one makes when one gets closer to the truth than one was before – when one gets closer to "getting things right".

If we adopt the epistemic notion of learning, we would have proved that we gain factual knowledge from art if we can prove that our encounter with it reliably leads to the acquisition of new justified true – or closer to the truth – beliefs. This, however, seems an arduous task. After all, it is far from clear why we should regard art as a privileged source of factual knowledge. Even setting aside for the moment the difficult question of how truths can be conveyed in visual art or music, what are the guarantees that the representation of the world or the human mind that we find in great literary works is true or truthlike? Why should we attribute to the writer a hold on such matters as human nature that even the bolder psychologist would never reasonably claim to know? On the face of it, these claims face the many reasonable objections that have been raised against the cognitive value of art in the last decades of debate. It has been repeatedly argued that fiction rarely conveys much new true information, and when it does, the information is either trivial or more reliably provided by the kind of systematic, empirically grounded inquiry of the sciences.[11] It has been pointed out that artists, unlike scientists or historians, do not have an obligation towards truth, nor are they subject of the same institutional constraints aimed at avoiding error or deception.[12] In art, instead, errors and deformations are often not sanctioned, but encouraged. It is indeed widely held, among artists and critics too, that faithful reproduction is not the goal, and not even a desirable outcome of art. Many philosophers also insist that we do not value works of art because they are offering us truths.[13] Where truthfulness is invoked, in criticism as well as in everyday discourse, it often seems to mark the achievement of some vivid effect obtained by the artist, with no obvious connection to the idea of being right about something.

To be sure, fiction seems to affect beliefs. There are now many empirical studies pointing to such an influence. In one of these, psychologists Green and Brock (2000) had subjects read a short story – *Murder in the Mall* – about the murder of a girl by an unrestrained psychiatric patient. After the reading, even subjects that were made aware that the story was fictional were more prone to agree with propositions like “The likelihood of a death by stabbing in a shopping mall is quite high”, or “Psychiatric patients who live in an institution should not be allowed out in the community during the day”. These conclusions are not arrived at in an epistemically rigorous way, and are not warranted by any examination of what is the case.

This is to point out that there is nothing in art or fiction as general categories that can make us think of them as particularly reliable sources of true beliefs. Instead, there are many reasons to suspect that they are not. Fiction and art do affect belief, but can spread ignorance and prejudice as well as truth, leading us to a worse belief-state than the one we previously possessed.[14] The question of whether we learn from art in the epistemic sense (i.e. whether we

[11] See for example Stolnitz (1992).

[12] See Currie (2014). On the problems of reliability and testimony in fiction, see also Ichino and Currie (2017).

[13] See for example the “no-truth theory” developed by Lamarque and Olsen (1994).

[14] See also Currie (2020) and Goffin and Friend (2022) for further reasons to think that this might be the case.

acquire true or at least better beliefs from it) becomes then, it seems to me, an empirical one, and one with no general answer: it would reduce itself to the assessment, done on a case-by-case basis, of whether the encounter of a certain person with a certain artwork has put her in a better belief-state or not.

But to frame learning in terms of acquisition of better beliefs is often seen as a particularly implausible move if we want to understand the kind of learning offered by art, and as such it is not frequently adopted by advocates of the positive view that we learn from it.[15] What advocates of this view tend to stress, instead, is that art can afford other kinds of learning that do not involve acquisition of better beliefs. What we would gain from great art are not or not primarily true propositions, but rather skills, abilities, sensitivities; art would enhance or exercise our moral or interpersonal capacities, make us better empathisers or better in theory of mind. All those changes should certainly count as forms of cognitive improvement, but they are not reducible, at least according to many, to the acquisition of factual beliefs about something. Using a standard distinction due to Ryle,[16] art would provide not (or not primarily) a learning-that, but a *learning-how*.

But art is also said to lead to *learning-what-it-is-like*, that is, to afford and encourage projections into another person's situation, acquainting us with alien perspectives. This is sometimes taken to be a third kind of learning, different from both learning-that and learning-how. In reading *Anna Karenina*, we are not learning what happened to a Russian married socialite in the second half of the nineteenth century, but (perhaps) we are learning what it is like for a woman in that condition to be torn between a devouring passion and the weight of family ties, social institutions and faith. Reading other stories, we might become acquainted with war, tyranny, chivalric love, the grief over the death of a child and a multitude of other experiences we would not have otherwise encountered. In engaging in such projections, we are perhaps enlarging our grasp of the possibilities of the human mind and behaviour, our awareness of what can possibly be felt and experienced.[17]

The philosopher has therefore other choices apart from the acquisition of better beliefs when it comes to assessing whether we learn from art. Learning-how and learning-what-it-is-like are two prominent ones, but there are others in the literature.[18] There is no principled reason to exclude that we undergo any of these different kinds of learning when we are exposed to a great work of fiction or art, and it is very much possible that we undergo many of them at the same time. The point is that, as with the acquisition of better beliefs, these kinds of learning are also *epistemic* in the sense we have defined, and, as such, they need to be susceptible to some sort of verification. There has to be some condition to gauge success. We cannot simply say that art

[15] For an appraisal of the limits of the notion of propositional knowledge for characterising the cognitive value of art, see Elgin (2002).

[16] See Ryle (1949), especially pp. 40-47. Ryle's distinction is however contested, and some argue that knowing-how is just a species of knowing-that (see e.g. Stanley and Williamson, 2001).

[17] See Gibson (2008) for suggestions in this direction.

[18] Some also talk about "understanding" or "wisdom". We shall deal with the notion of understanding shortly. For the notion of wisdom, as well as for a discussion on all these various kinds of learning, see Currie (2020), Chapter 5.

provides some kind of non-propositional learning like learning-how or learning what-it-is-like and leave it at that. The next natural step is to verify if this kind of learning *really* takes place. In the case of learning-how, one needs to establish in some way whether the encounter with the artwork has really enhanced our capacities (moral, interpersonal, empathic, or otherwise). In the case of learning what-it-is-like, one needs to verify in some way that the imaginative projections afforded by the artwork are accurate, that they *really* give us a faithful rendering of what a certain experience is like and not just a misleading impression that this is the case. For the concerns of the epistemologist, learning-how, learning-what-it-is-like and any other alleged kinds of learning are not different from learning-that, in that there should be also for them such a thing as getting it right, and one would have proven that genuine learning takes place only if one can show that the reader has got it right. Verification is, I believe, the only rigorous way to assess if learning in the epistemic sense takes place. In this sense, Currie (2020, p. 7) is right in not being content with the "enthusiastic and somewhat poorly evidenced" commitments to the positive value of literature and in asking instead for some form of empirical assessment.[19]

As with learning-that, therefore, the question becomes whether art is capable of reliably generating learning-how, learning what-it-is-like, or any other kind of learning epistemically understood. This is an empirical question that requires empirical inquiry. But now we should ask what a systematic empirical inquiry of this sort is likely to show. I suspect that those who are expecting a neat, yes or no answer would be disappointed. Since there are no principled reasons to think that artworks *qua* artworks provide any sort of knowledge, one should not expect from a systematic empirical testing a general answer to questions such as "Does art sharpen our moral sensitivity?" or "Can we learn from art or fiction how we would feel in situations we have not encountered?". What a systematic empirical inquiry is likely to show, is, at best, that certain artworks improve the epistemic condition of certain agents in certain situations. Nothing with the generality or the consistency of our attribution of insight to great art. The question of whether we learn from art would once again receive a rather uninteresting answer, but the only reasonable one to such a generic question: it depends; do your testing and see whether the cognitive change brought about by this specific artwork in this specific subject does more good or harm in each specific situation.[20]

[19] Currie (2020, p. 183) writes: "it cannot count as the generation of genuine insight merely that people have the feeling that insight has been generated... we must offer some standards, no doubt fallible, by which to tell when people have arrived at whatever form of enlightenment is at issue... confirmation that the learner has gotten it right." On the need to take seriously the empirical challenge to the claim that we learn from art, see also Nannicelli (2020) and Vidmar Jovanović (2021).

[20] Currie (2020, pp. 3-4) seems to make similar predictions about the likely outcomes of empirically assessing whether we learn from fiction: "I don't say we cannot or do not learn from fictions. That will never, I am sure, be a reasonable conclusion to draw. At best we will be able to say, with some confidence, that learning from fiction is apt/not apt to take place for some people, in some circumstances, for some works of fiction." Reid (1964, p. 321) similarly notes: "No one who approaches aesthetics from some knowledge of the several arts (and not a priori) and who has long reflected on these questions, could possibly expect a simple general answer to the question of the relation between art and reality."

1.2.2 Two Different Explanatory Targets

At this point, however, we might wonder whether there are not really two different issues at stake in our discussion, and if, in following the concerns of the epistemologist, we have lost track of the issue from which we started. One thing, one might say, is to understand whether art changes us *for the better*, that is, whether it provides learning in the epistemic sense. Another thing is to understand how and why art changes us at all, how and why it manages to generate a feeling of having learnt. These two questions are too often treated as if they were one, as if discovering whether we really learn from art coincided with understanding the roots of its suggestive power, and vice versa. But that is not the case. An answer to the first question does not constitute an answer to the second. If we want to understand anything at all about the relationship between art and learning, we need to disentangle these two problems first.

Whether art changes us for the better is, I have suggested, an empirical question: it requires us to demonstrate that art makes some positive difference for the people we become. That art changes us, instead, seems indisputable: the challenge is just to capture in the most accurate way what this change amounts to. Whether art changes us for the better is also, I have suggested, a question that is likely to dissolve into a myriad of particular questions about specific artworks in specific circumstances for specific subjects – something that threatens to make the issue of learning from art in its generality meaningless in the first place. The attempt to capture how and why art changes us, instead, can be conducted in its generality, insofar as we seem to recognise to art *qua art* the power to affect us in specific (beneficial) ways. Finally, in our attempt to establish whether art changes us for the better, we should be wary about the "enthusiastic and somewhat poorly evidenced" impressions of readers, beholders and listeners that the artwork has affected them in beneficial ways. In the attempt to explain how and why art changes us, instead, these impressions become our starting point and our explanatory target.

After all, it is not difficult to see that the two are separate matters: a text can be full of true propositions, or capable of enhancing certain abilities without being deemed particularly valuable or insightful. One can acquire many true beliefs from a history textbook, and lots of know-hows from a cookbook, without experiencing a particular feeling of cognitive gain while reading and without feeling particularly enriched after the experience. The first step, then, is to recognise that the "learning" that the philosopher is interested in verifying is not the "learning" experienced by the subject. We can learn in the epistemic sense of the philosopher while not feeling a subjective sense of enrichment, and we can have a subjective feeling of enrichment when, from the perspective of the philosopher, we are not learning at all. Many of us can perhaps sympathise with the observation of the American writer Charles Baxter when he notes, in an essay significantly entitled *Against Epiphanies* (1997, p. 63): "In retrospect, I can say with some certainty that most of my own large-scale insights have turned out to be completely false. They have arrived with a powerful, soul-altering force; and they have all been dead wrong." But the fact that some (or even most) of our insights may be false does not exempt us from explaining the powerful, soul-altering force with which they arrive. If anything, it makes that explanation even more urgent.

Which of the two questions shall we pursue, then? Shall we try to assess if and in what circumstances specific artworks may turn out to be epistemically advantageous? Or shall we try to understand the roots of the power that art has to attract, transform and convey a feeling of insight? My feeling is that we should lean towards the second option. Not only because it might be the only one that preserves the meaningfulness of such a general question as whether we learn from art, but also because the first direction of inquiry seems to miss the point. I suspect that behind the curiosity of the humanist, the scientist and the layperson about art and learning lies a concern that cannot be addressed by a case-by-case examination of what true beliefs or capacities the reader has effectively acquired from being exposed to the artwork. When most people wonder whether we learn from art, what they seem to care about is finding an explanation for this consistent, widespread impression that, by being exposed to a great artwork, they have undergone an enriching experience. They are concerned with the phenomenal aspects of this experience that they can so clearly perceive: surprise, rumination, progress, achievement. They are trying to justify (perhaps to themselves more than to others) their undeniable impression that the artwork does repay their serious, sustained and repeated attention, that there is a gain in attending to it, that it is a worthwhile experience to have. This is why they worry so rarely about finding empirical evidence for their claims that we learn from art: the learning is just there, so vividly felt. And this is also why the question is so often formulated with this blunt generality (whether we learn from "art" or "fiction", or "literature" as if it were clear that the entities in such categories are uniformly related to truth). We consistently feel that it is great art (or fiction, or literature) *qua* great art (fiction, literature) that affords this enriching experience, and that this is in fact one of its essential characteristics. And this is also why an empirical, case-by-case inquiry, although it is arguably the only rigorous way to assess if a "learning" in the sense of the epistemologist takes place, runs the risk of missing the point.[21] Even if one had assessed empirically that exposure to art reliably provides better beliefs or improves a certain capacity, one would not have yet said anything about the feeling of having learnt, which is arguably what prompted the question of learning from art in the first place.

In what follows, therefore, our focus will be on the feeling of having learnt. We will try to understand what art does to us, how it changes us, and why these changes appear to us in the form of discoveries and revelations. To do so, we need a shift in how we conceptualise learning. In particular, we need a notion of learning that is non-epistemic, our aim being to explain why certain sentences, narratives, paintings, pieces of music touch us and move us more than others, irrespective of the epistemic advantages they might ultimately confer. And we need a notion of learning that has something to say about the phenomenal experience of the learning subject: her feeling of progress, success, insight and discovery. In this respect, the notions of learning that are most frequently discussed in the philosophical debate do not seem to be particularly helpful. Not

[21] On why an empirical assessment of what we learn from art might miss the point, see also what literary scholar Derek Attridge says in the passage quoted in Chapter 2, section 2.1 below.

only do they have all a primarily epistemic concern, but, due to the metaphysical and conceptual task they were created for, they are also quite disconnected from the phenomenology of learning. They were not meant to describe our experience and behaviours as cognitive agents. For these reasons, we must leave the terrain of the epistemologist and look at disciplines that have primarily a descriptive concern and place learning centre-stage: (developmental) psychology, cognitive science, neuroscience. This shift in focus will allow us to see that the philosopher, by approaching the issue from its epistemological standpoint, might have so far missed a fundamental relationship between art and learning that from other disciplinary perspectives is instead increasingly clear, almost self-evident.

1.2.3 Understanding: A Better Candidate?

Before we turn to psychology, cognitive science and neuroscience, however, let us examine one last notion that is often discussed in the philosophical literature: the notion of "understanding". Some philosophers have defended the view that understanding is a valuable epistemic notion that differs from knowledge.[22] The idea that seems to emerge from the literature on the topic is that while knowledge has to do with the possession of individual pieces of information, understanding has to do with our grasp of how pieces of information are connected with each other in more comprehensive bodies of information. In this view, understanding a fact or proposition would be not merely entertaining it as a stand-alone item of knowledge, but grasping the significance (the consequences, the relationships, the implications) of that fact or proposition for a larger body of knowledge. Thus, for example, "I understand that Athens defeated Persia in the battle of Marathon, because I grasp how the proposition stating that fact fits into, contributes to, and is justified by reference to a more comprehensive understanding that embeds it" (Elgin 2007, p. 35), an understanding that might include, for example, other relevant facts about the history, culture and geography of ancient Greece.

Understanding would therefore be a type of cognitive progress that, while related in complex ways to the acquisition of factual knowledge, does not coincide with it. One might acquire knowledge without advancing in one's understanding. Being completely ignorant of the history, culture or geography of Ancient Greece, I might hear from an authoritative source that Athens defeated Persia in the battle of Marathon, and come to believe it; based on the traditional analysis of knowledge, I will therefore know that Athens defeated Persia in the battle of Marathon. But if this fact does not have consequences for other facts I know, if it remains an atomised piece of information completely disconnected from the rest of my knowledge, then arguably I do not understand it. And if I acquire many such atomised facts, I might accumulate a great deal of knowledge with little or no advancement in my understanding. Moreover, being concerned with how much an item of knowledge reverberates within a whole, the notion of understanding seems to admit of degrees in a way that knowledge does not. While I either know

[22] See for example Zagzebski (2001), Elgin (2007), Kvanvig (2003, 2009), Pritchard (2014) and Hills (2016).

or do not know that *p*, my understanding that *p* would depend on the quantity of connections that I am able to draw between *p* and other items in my web of beliefs.[23] There is no particular number of such connections one is required to make in order to understand; arguably, one has to be able to make some, and more will count (in complex ways) as possessing a better understanding. Thus, while a fresher might have some understanding of the Athenian victory based on a very sketchy background of relevant pieces of information, a professor of ancient history will have a deeper understanding of the same fact, being able to embed that fact in a broader web of beliefs.[24]

Whether or not we think of understanding as a true epistemic capacity in its own right, the use of this notion seems to bring to the forefront an important aspect of cognitive progress that the notion of learning-that (the acquisition of factual knowledge, or justified true beliefs) might have led us to overlook. Conceiving learning-that as the acquisition of justified true beliefs seems to suggest a picture in which information comes in discrete bits and knowledge grows cumulatively, progressing potentially with no disruptions or restructurings. A person learns a new fact and smoothly incorporates it into the stock of facts that she already knows. In this picture, humans would gather information "in the way that squirrels gather nuts" (Elgin 2002, p.1), amassing data bit by bit and storing them away against future need. This image of cognitive progress seems unlikely. We should recall here Quine's (1951, p. 35) dictum that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body." Our knowledge does not look like a pile or stock of amassed facts, but more like a constantly changing web of relations and mutual consistencies. Incorporating new information in this web of relations means therefore making changes to the whole web – and not all new information is likely to behave equally in this respect. Certain information is more surprising, and may cause major reassessments of what we already believe; other information is more predictable, and fits more neatly into what we already believe. But to the extent that we understand the new information at all, some restructuring has to take place. If I am told that there is not a pink elephant floating in my room right now, or that sodium potassium aluminium silicate has a moderate capacity to bioaccumulate (both facts that happen to be true), I won't have the feeling of having made a significant leap in my understanding of the world, precisely because both these facts, for different reasons, leave my previous knowledge almost unchanged. Information that changes us most, it seems is the one that disrupts and rearticulates our previous knowledge of the world to some extent (and the more it changes it, the more enlightening it may seem). Insofar as the notion of understanding captures this process of rearticulation, it has the merit to point to an important fact: cognitive progress (at least as is felt by the subject) might be

[23] See Elgin (2007) and Hills (2016) for characterisations of understanding as something that comes in degrees.

[24] A worry might remain that, in fact, instances of understanding reduce to instances of knowing (see for example Silwa, 2015 for one such reductionistic account). Perhaps understanding is just knowledge-that plus the capacity to put that knowledge into use; perhaps it is just a way of acknowledging the non-linear process by which factual knowledge accumulates. Whether understanding is reducible to knowledge is an issue we do not need to take a stance on. The crucial question for our purposes is, as we shall see, whether understanding is factive.

less concerned with the quantity of justified true beliefs we acquire and more with how much the new information reverberates on and changes our pre-existing knowledge of the world.

At this point, we might wonder whether understanding is a better candidate for the kind of experience that great art characteristically affords.[25] After all, as we saw, when people extoll the cognitive benefits of art, they do not normally characterise them in terms of circumscribed additions to their stock of known facts; they stress instead how the artwork has altered and transformed (more or less radically) their view of the world. Reading *Anna Karenina*, I might well come to learn that nineteenth-century Russian aristocracy often spoke French in social situations, as well as many other facts of this sort. But when asked what I learned from the *Anna Karenina*, I am more likely to point to how the novel rearticulated my understanding of important aspects of the human condition: social constraints, passion, love, jealousy, desperation, faith or the search for the value of life.[26] The notion of understanding, with its stress on how information reverberates in our body of pre-existing knowledge, seems quite apt to capture this transformation. Moreover, insofar as understanding means seeing something as part of a broader pattern, the notion seems also to better capture our epistemic behaviour when we engage with an artwork. Reading a novel, inspecting a painted canvas or listening to a piece of music seems to involve in an important way grasping the significance that individual elements (facts, shapes and colours, musical notes) have within the overall pattern (the narrative, the painted figure, the melody) that the work outlines. Understanding Anna Karenina's unfortunate parable seems not so much a matter of learning *that* she did a certain action, then another one, and so on, but a matter of grasping how her actions make sense in the organic whole of her fictive personality and within the organic whole of the novel. In many respects, therefore, the notion of understanding seems to point in the right direction.

The important point, however, is whether we consider understanding to be factive, like "knowing" and "learning" according to the epistemologist. If we do, we think that a subject S understands that *p* if and only If *p* is true and the connections that S draws between *p* and other items in her pre-existing knowledge are valid. Thus, I understand that Athens defeated Persia in the battle of Marathon only if this fact is true; additionally, I might also think that the Athenian victory was due in part to the geographic features of the Marathon plain, or to the intervention of allied forces coming from Plataea, or to the military manoeuvres of the Greek commander Callimachus; I will understand that Athens defeated Persia in the battle of Marathon only if these facts are also (mostly)[27] true and if they really contributed to the Athenian victory. If we do not

[25] See Elgin (2002) and Gibson (2003) for proposals in this direction.

[26] Even a rather peripheral piece of information such as that Russian elites tended to speak French seems to contribute to our cognitive progress not so much as a stand-alone piece of information acquired in isolation, but mainly insofar as it prompts rearticulations in relevant areas of our previous knowledge (for example, if we are able to see it as a symptom of the affectation of the members or Russian aristocracy at that time, of their striving towards a model of "Europeanness", or their attempt to distance themselves from their traditionally rural roots).

[27] See Kvanvig (2003, pp. 201-2) for the reasons for this qualification: "it is hard to resist the view that understanding may be correctly ascribed even in the presence of some false beliefs concerning the subject matter… When the falsehoods are peripheral, we can ascribe understanding based on the rest of the information grasped that is true

think that understanding is factive, on the other hand, these constraints do not apply: I might gain understanding even by drawing invalid connections between propositions that are in themselves unwarranted or false. The wildest phantasies of the poet will count as instances of understanding as much as the accurate reconstructions of the historian, as long as they are able to provide the same inner click. The jury is out on this issue: according to some (Kvanvig, 2003, 2006; Hills, 2016), understanding is certainly factive, according to others (Zagzebski, 2001; Elgin, 2007, 2009; Riggs, 2009) it need not be. We do not need to enter the debate, but we should note what each of these alternatives entails for the aim that we are pursuing. If we think that understanding is factive, then the notion faces the same problem encountered by the other epistemic notions of learning: it becomes useless for explaining the phenomenon of insights that arrive with soul-altering force but that are, as it were, deadly wrong. If we think that understanding is not factive, then we might have a useful tool to describe this feeling of having learnt (illusory or not it does not matter) that great art seems to provide.

Given the nature of my enterprise, I favour the second alternative. In this way, the notion of understanding will turn out to be useful (especially in Chapter 2 and 3) and to map well with psychological and neuroscientific characterisations of the learning process. For the moment, let us simply treat understanding as an interesting notion, a notion that, if conceived non-factively, can perhaps capture some interesting aspects of our engagement with art.

1.3 Learning According to the Brain

1.3.1 Grasping a Gestalt: Learning, Problem-Solving and Insight

It is time to move from how learning is conceived in contemporary philosophical discussions to how it is conceived in psychology, cognitive science and neuroscience. In approaching these disciplines we encounter a very different set of concerns, and we discover that notions crucial to the philosopher play here little or no explanatory role. The stress is no longer on truth, reliability, justifications or propositionally individuated beliefs, but instead on how agents (humans as well as animals) adapt to their environment, modifying their behaviour in response to practical challenges. Knowledge and learning are here not notions to be defined conceptually or metaphysically, but part of an explanatory project aimed at discovering how experience changes the way we perceive, think, plan and act.

The study of learning as the capacity to adapt to environmental contingencies owes much to the school of Gestalt psychology that flourished in the first half of the twentieth century. A landmark work in this tradition is Wolfgang Koehler's *The Mentality of the Apes* (1927 [1921]).

and contains no falsehoods. In such a case, the false beliefs are not a part of the understanding the person has, even though they concern the very material regarding which the person has understanding. So in this way, the factive character of understanding can be preserved without having to say that a person with false beliefs about a subject matter can have no understanding of it."

Forced to a prolonged stay in Tenerife by the outbreak of World War I, Koehler conducted in those years several experiments with the chimpanzees of the local Anthropoid Station. His focus was on the problem-solving abilities of those primates. Thanks to his work, he was able to challenge the behaviourist view that animals learn only blindly via trial-and-error, showing that chimpanzees were able to solve problems intelligently by means of "insight" (*Einsicht*). Koehler's best-known experiments involved the use of tools to gain access to food: typically, a chimp would have to reach some fruit outside its cage using objects inside the cage made available by the experimenter. If the chimp was given a stick, after an attempt to grab the fruit directly, it could be seen pausing and scanning its surroundings; in most cases, it would then grab the stick and retrieve securely the food with it. But then Koehler would complicate the experimental situation: the chimp was now given two sticks, one narrow and the other thicker and hollow, none of which was long enough to allow the chimp to push the food towards itself. Most of the chimps then, after various attempts to reach for the food with either of the sticks, would exhibit increasing frustration and finally give up on the task. Koehler's star chimp Sultan, however, after some failed attempts of the same kind, paused and concentrated on the two sticks. It worked on them for over an hour. When they finally fitted together, Sultan immediately used the new tool to retrieve the food. From this and several other experiments, Koehler concluded that the chimps were displaying a goal-directedness and an awareness of the "problem space" which was incompatible with simple trial-and-error learning.

Since Koehler's pioneering observations, insightful problem-solving has been studied in humans under several paradigms. One of the best known is Duncker's "candle problem". In this experimental setting, subjects are given a candle, some matches and a box of tacks and are asked to find a way to attach the candle to the wall in order to light up the room. The solution requires the participants to empty the box of tacks, set the candle inside the box, tack the box to the wall, and light the candle with the matches, as shown in Figure 1 below. The difficulty of the task lies in the fact that it requires subjects to use objects in a way they are not accustomed to. The optimal solution to the candle problem is achieved only if the subject is able to overcome his tendency to see the box as merely a container for the tacks and succeed in reconceptualising it in a way that meets the demand of the particular context. Interestingly, if the task is presented with the tacks piled next to the box (rather than inside it), virtually all participants achieve the optimal solution.[28]

[28] See Duncker (1945). Many participants in the experiment explored other creative, but less efficient, methods to achieve the goal. Some tried to tack the candle to the wall without using the thumbtack box, and others attempted to melt some of the candle's wax and use it as an adhesive to stick the candle to the wall. Neither method really worked.

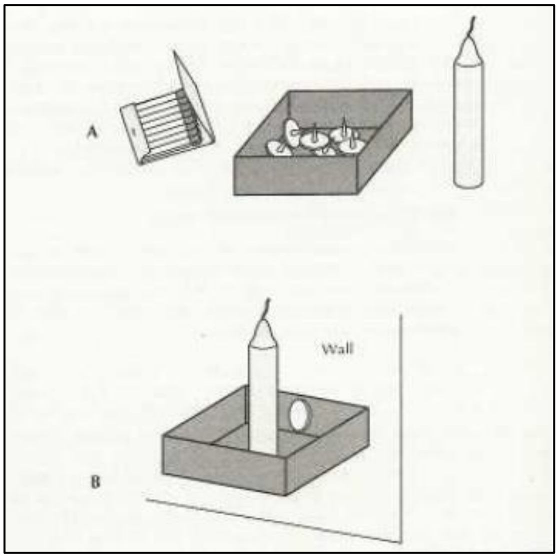

Figure 1: Duncker’s “candle problem” (Duncker, 1945).

Even from the two examples above, it should be clear that insightful problem-solving requires two fundamental conditions. The first, quite evidently, is *that there be a problem to solve*. The subject must find herself at odds with something: her set of expectations and behavioural routines does not fit the demands of the current situation (the stick that the chimp has used to reach for the food in the past does not serve the purpose this time; the way in which the subject conceptualise the box of tacks is not helping her solve the candle problem). In other words, the subject pays the price for what psychologists call “functional fixedness”, or “*Einstellung* effect”: the tendency to attribute to an object its most conventional function, being thus unable to grasp its relevance to the situation at hand.[29] But then, and this is the second condition for insightful problem-solving, *there must be a solution to the problem*. The subject must then be capable of overcoming her functional fixedness, renegotiating the value of the objects in a way that optimally fits the demands of the current situation. Here is where learning takes place. The terms of the problem, once unrelated, become parts of a meaningful functional whole (the stick becomes a part of a two-component tool, the box of tacks becomes a holder for the candle). Some describe the operation as involving the “breaking of a mental set”, or a “mental restructuring”.[30] Interestingly for our purposes, the solution to the problem is usually accompanied by a pleasurable feeling of cognitive gain which is referred to in the literature as a “Aha!” experience.[31]

[29] On the notion of functional fixedness, see again Duncker (1945). For some classic works on the *Einstellung* effect, see Luchins (1942) and Luchins and Luchins (1959).

[30] See Wertheimer (1959) and Öllinger, Jones and Knoblich (2008).

[31] On the “Aha!” experience, see Gick and Lockhart (1995), Topolinski and Reber (2010) and Shen et al. (2016). Interestingly, most research in this area stresses that experience involves not just positive affect, but also an

From these rather simple observations, we may draw a lesson that will become a running theme of this work: problem-solving and the pleasurable feeling of insight that comes with it can only take place if 1. the subject finds herself at odds with a certain situation (there is a problem to solve); 2. the subject is able to restructure her understanding and meet the demands of the situation (there is a solution to the problem). If the solution is too obvious or too difficult to be achieved, then insightful problem-solving can't occur. Of course, among the extremes of absolute triviality and excessive complexity, all kinds of intermediate cases are likely to occur. There will be simple problems that require simple solutions, and complex problems that just allow for less-than-optimal solutions. Moreover, the line of what we consider a problem to solve is likely to be a movable one: once we solve the candle problem the first time, solving it the second time requires far less effort (and affords far less satisfaction). This is in fact what learning is from a psychological viewpoint: to accumulate an understanding of how environmental problems are to be solved in a way that allows the agent not to approach every situation as for the first time.

In examining insightful problem-solving, we might have already taken some steps towards comprehending the roots of the feeling of having learnt afforded by art. Insightful problem-solving certainly seems to have some of the ingredients of our paradigmatic contact with the artwork: the "mental restructuring" characteristic of a transformative experience, the impression of having attained clarification, the pleasurable affective response. Moreover, insofar as it involves grasping the role that disparate elements may play as parts of a meaningful whole, insightful problem-solving is also quite close to the notion of understanding as discussed in the philosophical literature.

Things become even more interesting once we recognise that insightful problem-solving has a broader character than that suggested by the previous examples. It does not pertain just to the solution of practical tasks: it also features in the solution of verbal problems such as the interpretation of jokes and metaphors, and in scientific and mathematical discoveries (see Shen et al. 2016 for a review). More generally, it seems to be a feature of all perceptual experience. To understand this, consider the well-known image below (taken from Gregory, 1997, p. 12):

"intuitive sense of success" (Gick and Lockhart, 1995, p. 215) and a high level of confidence in the correctness of the found solution (Topolinski and Reber, 2010).

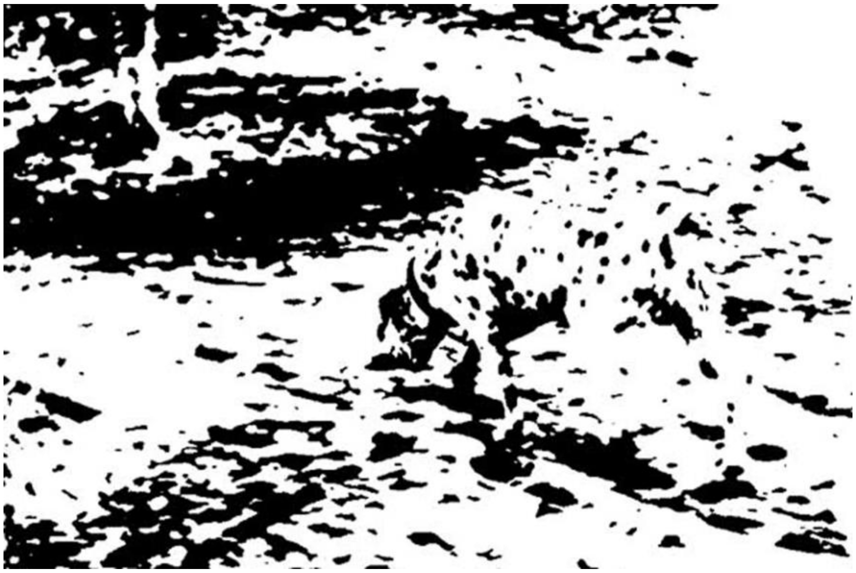

Figure 2: A hidden Gestalt (from Gregory, 1997, p. 17).

Confronted for the first time with this image, the observer is likely to perceive only a disorderly group of black and white patches. But things change dramatically once she is able to recognise the Dalmatian dog hidden in it (cue: the head is near the centre of the image, and the dog is sniffing the ground). We can readily see the analogy of this case with the two examples of problem-solving illustrated above. At first, the observer of this image is just in the same position as the chimpanzee that is given the two sticks or the subject that is presented with the objects of the candle problem: what she perceives, the black and white patches, are just a group of disparate visual objects, with no clear relationship to one another. But then suddenly, just as the two sticks for the chimp or the objects in the candle problem, the patches start to make sense as parts of a newly grasped whole (a Gestalt), and the perceptual system settles down on an interpretation of the visual input that allows the agent to make the most out of what it perceives. Together with the solution to the visual problem comes a pleasurable feeling of discovery, an "Aha!" experience that is identical in nature to that experienced by the solver of the candle problem.

Ambiguous images of this kind make clear, in sum, that perception is a problem-solving activity. Taking vision as a paradigmatic case, we must acknowledge that, far from being a mere bottom-up rendering of what is out there, every act of seeing involves an intelligent disambiguation of a number of aspects that are left underspecified by our retinal image. Here Gestalt psychology makes contact with a long tradition of perceptual research – with Helmholtz (2013 [1867]) and Gregory (1997) as some of its major proponents – that understands perception as a process of hypothesis-testing, whereby the agent actively tries to infer the most probable cause of the stimulations impinging on its sensory organs. In this view, to perceive the Dalmatian would be to entertain a Dalmatian hypothesis as the one that happens to best explain the pattern of stimulations impinging on your retina. Finding a Gestalt means therefore both solving a perceptual problem and finding a viable hypothesis about the causes of sensory stimulations.

Of course, in most cases we are not aware of solving problems or testing hypotheses while perceiving. Perception normally appears to us to simply mirror already-structured external objects. But the story we have told so far has the tools to explain why this is the case. Perception might well be always a problem-solving activity, but not all percepts are equally problematic. Many percepts, like many situations, are relatively unambiguous: their organisation happens quickly and almost entirely at a subpersonal level (this is why Helmholtz (2013 [1867]) characterised perception as involving "*unconscious* inferences"). Other percepts might be too ambiguous to allow for a solution. It is just percepts with the right level of ambiguity that allow for the problem-solving character of perception to come to the forefront. To appreciate this, consider the three images below (Figure 3):

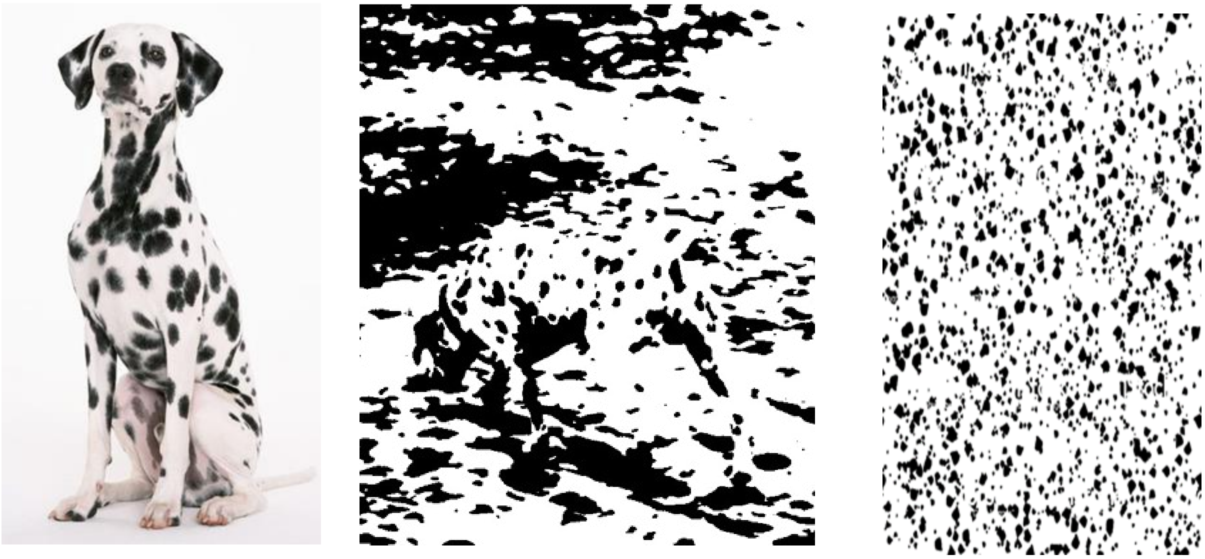

Figure 3: An optimal level of ambiguity.

If prompted to find a Dalmatian in each of these three images, the observer will actively try to construct a Dalmatian out of the ambiguous sensory data. But the task is not equally demanding or feasible in the three cases. In the image on the left, the task is too simple; in the image on the right, it is impossible. It is only with images with the right level of ambiguity that the active and inferential character of perception, as well as its pleasurable affective correlates, can more forcefully emerge.[32] This speaks again of the substantial analogy between perception and the solution of practical tasks: as practical tasks, percepts too vary in the degree in which they require

[32] Again, the "right level" of ambiguity is of course constantly shifting, as the observer gets gradually habituated to the percept in question (the camouflaged Dalmatian seen for the tenth time does not pose the same perceptual problem that it posed the first time and does not generate the same feeling of discovery).

(and allow for) insightful problem-solving and the pleasurable affective experience that accompanies it.

An important thing to notice about the kind of learning in question in insightful problem-solving is that it is not "epistemic" in the sense we have defined above: it does not require the subject to get "closer to the truth". The solution that we find to the perceptual (or conceptual, or practical) problem could in fact be wrong or suboptimal. The pattern that we find might just not be there. Pareidolias and visual illusions bear witness to this fact. From the perspective of the learning subject, however, this does not prevent the experience from counting as learning and being lived as such. The discovery of a face in a cloud can afford insight as much as the recognition of a true face (and in fact even more, if the face in the cloud is somewhat ambiguous and difficult to detect). The observer has succeeded in finding a pattern in scattered sensory data, and this has prompted changes in her experience and behaviour, and that is all that is required for learning. Whether or not that pattern ultimately reflects a true state of affairs is an altogether different question. Our perceptual and cognitive habits are just a reflection of what has worked in most cases given the particular set of experiences that we have encountered up to that moment. Whether future experience would disprove these habits is something that the learning subject – as well as the psychologist – is never able to say.[33] In the impossibility to tell what would ultimately be epistemically advantageous, it would be unreasonable to say that no change that occurs in response to experience deserves the name of "learning". It is more reasonable to describe all these changes as learning, suspending the judgement over whether they will ultimately turn out to be advantageous. This points to a substantial difference in the way philosophers on the one hand and psychologists and neuroscientists on the other usually talk about learning. While the first tend to concentrate on truth as a condition for learning, the latter are more concerned with capturing the agent's changes in physiology and behaviour in response to environmental contingencies, whatever the ultimate epistemic status of those changes may be.

All things considered, therefore, the study of insightful problem-solving might have provided the beginning of a promising story. If we manage to generalise this story across sensory modalities and the perception/cognition divide, we will have a principled explanation of what kind of percepts provide the pleasurable, non-truth-directed learning experience we are interested in. A hypothesis would then become readily available that artworks tend in fact to be percepts of just this kind. Let us see if we can articulate this story a bit more.

1.3.2 Neuroplasticity and the Brain as a Probabilistic Model

We have seen that there is a way of casting perception as a problem-solving activity in which scattered and ambiguous sensory data are assembled and organised into patterns (Gestalts) or viable hypotheses about their distal causes (as per Helmholtz and Gregory). We should now ask

[33] This is, quite evidently, Hume's problem of induction.

how these patterns come to be internalised by the subject in such a way that allows her to recognise them more readily in subsequent experiences. What is it that makes it far easier and more immediate to see the Dalmatian in Figure 2 the second time? Why don't we have to start anew every time we look at the image? To answer these questions means to go at the core of what learning Is taken to be in psychology and neuroscience.

From a neuroscientific perspective, the internalisation of a pattern of perception, thought or action ought to correspond to a modification of some kind in the subject's nervous system. In other words, changes in the way in which a subject organises a percept, thinks or carries out a practical task should correspond to changes in the structure and activity of her nervous system. These changes are made possible by neuroplasticity, the remarkable capacity of the nervous system to rearrange itself in response to experience. Neuroplasticity is really an umbrella term that covers a variety of mechanisms, including neurogenesis (the growth of new neurons), synaptic pruning (the elimination of unused synaptic connections), synaptic plasticity (changes in the strength of synaptic connections), changes in white or grey matter density and cortical remappings.[34] These changes happen at different timescales (from milliseconds to years), and are more or less long-lasting. They are more radical during childhood and development but continue throughout life. Thanks to neuroplasticity, we are constantly growing our brain by using it, and we grow it in particular ways by using it in particular ways. Neuronal circuits are being assembled and rearranged continuously while we explore our environment, play a musical instrument, acquire a new language.[35] This leads us to perform the task in question with increased effectiveness and decreased energy consumption. This is in fact what learning amounts to from a neuroscientific perspective: "the rewiring by experience of a plastic brain so as to make the operation of that brain better suited to the environment in which it finds itself" (Gallistel and King, 2010, p. 187).[36]

Incidentally, this is also what learning is taken to be in the vast strand of artificial intelligence and machine learning that makes use of artificial neural networks, computational devices that

[34] See Costandi (2016) for a handy introduction to these various kinds of neuroplasticity.

[35] See Oakes (2017) for a summary of research on neuroplasticity induced by various kinds of tasks and experiences. Classic experiments in this area include: Greenough, Black, and Wallace (1987), who compared rats raised in enriched environments with rats raised in captivity and found that the first showed better learning and effects on brain development including heavier and thicker cortices, more dendrites per neuron and more spines per dendrite; Maguire et al. (2000), who found that experienced London taxi drivers had larger brain regions related to spatial memory (posterior hippocampi) compared to age-matched controls; Elbert et al., (1995), who examined cortical representations in the sensorimotor cortex of violin players and found larger-than-average representations for the digits of the left hand (the one used on the fingerboard).

[36] One should notice that the process in question is one of reciprocal causation: the experience of the environment changes the agent's brain, but the agent selects which parts of the environment to experience. As Sporns (2007, p. 179) puts it: "The architecture of the brain… and the statistics of the environment [are] not fixed. Rather, brain-connectivity is subject to a broad spectrum of input-, experience-, and activity-dependent processes which shape and structure its patterning and strengths… These changes, in turn, result in altered interactions with the environment, exerting causal influences on what is experienced and sensed in the future." See section 1.3.4 below for some implications of this fact.

mimic the functioning of biological brains. Here a multi-layered structure of interconnected nodes is fed data from a data set and changes the weights of its connections as training progresses, coming to represent properties of the dataset in question. As with biological brains, learning in artificial neural networks is the adaptation of the network to better handle a task by considering sample observations.

Is the kind of learning that brains carry out and artificial neural networks try to simulate epistemic in our sense? In other words, does it require that the agent end up "closer to the truth"? Again: no. In talking about learning, the neuroscientist is interested in capturing how the brain adapts in response to experience, not in whether these adaptations are ultimately epistemically good. Granted: part of the interest of the neuroscientist is to explain *effective* behaviour, and effective behaviour seems to have a lot to do with getting things right. But getting things right locally might mean becoming less able to get things right overall. If a kitten is raised wearing a mask that exposes its eyes just to vertical or horizontal stripes, the receptive fields of neurons in its primary visual cortex will tend to grow vertically or horizontally to an abnormal degree, making the animal better able to detect vertical or horizontal lines (Hirsch and Spinelli, 1970). This counts as learning in the neuroscientific sense: the cat has reorganised its cortex based on experience. But this learning did not bring the cat "closer to the truth" of the visual world: due to this cortical reorganisation, the cat is in fact visually impaired. In the same vein, if an artificial neural network used for face recognition is trained with a dataset containing mostly pictures of white men, it will become very accurate in recognising faces of white men, but far less accurate in recognising faces of black women, and thus not very accurate in recognising faces overall (Zou and Schiebinger, 2018). It is therefore perfectly consistent with the neuroscientific notion of learning that an agent might at the same time learn and worsen its general epistemic position. Learning in neuroscience (and in artificial neural networks) might be adaptive, but it is not truth-directed.

Our observations about neuroplasticity and learning have been so far rather uncontroversial, reflecting the mainstream understanding of learning and brain function in neuroscience. The picture that we obtained is that of the brain as a plastic organ, constantly rearranging its structure and dynamics in response to experience. We shall now try to move towards a more specific interpretation of these phenomena, an interpretation that reflects the increasing popularity in cognitive science and neuroscience of probability theory as a tool for describing human cognition.

According to the developing story we are about to consider, the fact that the brain is altered by experience indicates that it comes to embody certain implicit beliefs about the shape of its environment.[37] These beliefs might be rather abstract (for example, based on previous

[37] These "beliefs" are effectively "hypotheses", "best guesses" or "predictions" (I will use the terms interchangeably throughout this work) about the shape of the external world. In other words, the brain might be seen as hypothesising, guessing or predicting its inputs by its very structure and dynamics. Note however that these "beliefs" differ from the propositionally individuated beliefs of the philosopher in at least two important respects. First, their best expression is probabilistic and not sentential (see below). Second, they might be (and for the most part are) held on a subpersonal level. In this context, therefore, to say "I believe" may not be very different from saying "my visual cortex believes".

encounters with dogs, I have acquired general beliefs about what a dog should look like), or they might be more concrete (after seeing the Dalmatian in Figure 2, I have acquired beliefs about how this particular configuration of stimuli should be organised – here is the head, here is a pawn, here is the tail, etc.). All these beliefs might be conceived *probabilistically*, as probability distributions defining which world structures are seen as most likely at any given time. Plasticity itself would be a means for the brain to encode these probabilistic beliefs at multiple timescales. In this picture, the brain would therefore embody, in its structures and dynamics, a probabilistic model of the world.

If the brain is seen as embodying a probabilistic model of the world, then to learn means to update this model in light of new evidence coming from the senses. According to an increasingly popular view, this process might be described using Bayes' theorem, a formalism that offers a rational way to go from what we already know to what we should believe next based on what we are currently observing.[38] In other words, Bayes' theorem allows us to combine a *prior* probability distribution (an estimate of how probable a certain hypothesis is prior to the encounter with new evidence) and a *likelihood* (an estimate of how probable the new evidence is assuming the hypothesis in question) to obtain a *posterior* probability distribution (an estimate of how probable our hypothesis is in light of the new evidence).[39] The process of going from a prior to a posterior probability distribution is called belief updating, and the divergence between these two probability distributions indicates the degree of belief updating: the bigger the divergence, the bigger the update. This divergence is also known as Bayesian surprise and, being a measure of the difference between the subject's beliefs before and after encountering the new evidence, it is effectively a measure of how much the subject has learnt as a result of the observation. This process of belief updating is normally continuous and unfolds in cycles: the posterior probability distribution thus obtained becomes the prior distribution for a new cycle of belief updating, and so on until the subject has reached a sufficiently stable interpretation of the portion of the world she is considering.

All this might sound rather obscure, but a graphic rendition might help. Consider again the experience of finding the Dalmatian dog in Figure 2 above. When I start looking at the image, I am not likely to have a precise idea of what it might contain. In Bayesian terms, I will have competing *prior* beliefs, all judged to be similarly (un)likely. These beliefs might be represented as a set of Gaussian probability distributions each specified by two parameters: a mean (where the curve peaks), indicating how probable I judge that belief to be, and a variance (how spread out the curve is), indicating how confident I am about this judgement. My prior belief that the image contains a Dalmatian dog will therefore be in the initial set of competing beliefs and might be represented as a Gaussian probability distribution such as the dotted curve in Figure 4 below.

[38] For introductions to Bayesian approaches to brain function, see Knill and Richards (1996), Knill and Pouget (2004) and Griffiths, Kemp and Tenenbaum (2008).

[39] Bayes theorem states that the conditional probability of a hypothesis H given an event E is equal to the probability of E given H multiplied by the probability of H and divided by the probability of E. In other words: $P(H|E) = [P(E|H) \times P(H)] / P(E)$. $P(H|E)$ is the posterior probability, $P(E|H)$ is the likelihood and $P(H)$ is the prior probability. $P(E)$ is a constant scaling factor that allows the posterior to be between zero and one.

This probability distribution has a low mean (indicating that I find the Dalmatian dog hypothesis to be quite unlikely) and a high variance (indicating that I am not very confident about this estimate). But then suppose that I continue inspecting the image and find something that looks like a dog's head, towards the centre of the image. This is significant new information: it makes sense in light of one of the hypotheses that I am entertaining (namely the hypothesis that there is a Dalmatian dog in the picture). The dashed curve indicates the relevance of this new evidence: it represents the *likelihood*, the probability of observing the data I am observing (a dog's head) given the Dalmatian hypothesis. This probability distribution has a high means (indicating that I find it highly probable to see a dog's head if the image contains a dog), and a low variance (indicating a high confidence that the presence of a dog in the picture is a good explanation for the presence of a dog's head). Using Bayes' rule, we can now combine these two probability distributions to obtain a *posterior* probability, represented by the solid curve. This represents the probability of the Dalmatian hypothesis in light of the new evidence.

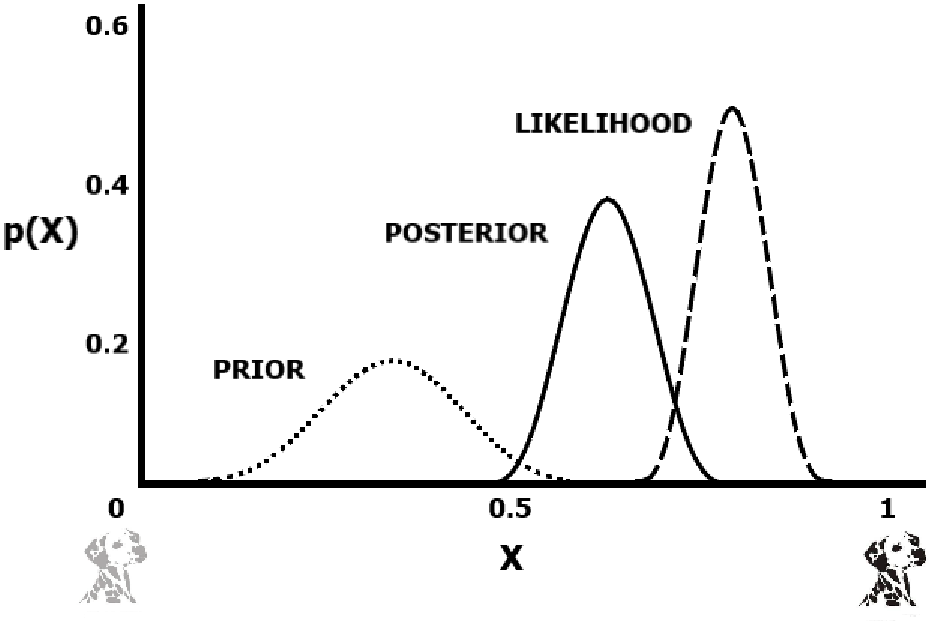

Figure 4: An example of Bayesian belief updating with Gaussian probability distributions (adapted from Seth, 2021, p. 105).

The shift from the prior to the posterior probability constitutes the belief updating. It captures how much my probabilistic model of the world has changed as a result of experience (that is, how much I have *learnt*). From the brain's perspective, this shift represents the amount of change in connectivity brought about by the experience; phenomenally, it indicates the fact that the subject has reduced the uncertainty about how to organise her sensorium, as one hypothesis among many has become more likely. This process, as we said, proceeds in an iterative fashion:

the posterior probability thus obtained becomes the new prior probability in the next cycle of belief updating. The Dalmatian hypothesis becomes thus the basis for new observations that can either weaken it or strengthen it even more. If all goes well, after a few cycles of belief-updating (which might last a few milliseconds) the Dalmatian hypothesis becomes likely enough and the subject end up seeing the Dalmatian as a relatively stable object.

This image of belief updating is no doubt quite simplistic, and we will need to refine it as we proceed, seeing how it might work in detail and in particular instances. But it has also various merits. First, it allows us to formalise the idea of perception as insightful problem-solving and give to it neural substance: finding the solution to perceptual problems may be now understood in terms of the updating of probabilistic beliefs about the causes of our sensory stimulations, in line with the widespread idea that brains distil causal regularities in the sensorium and embody them in models of their world. Second, this view allows us to obtain a measure of learning, and an idea of what percepts are likely to promote it. Learning becomes quantifiable in terms of the divergence between a prior and posterior probability distribution, which captures how much our brain as a probabilistic model of the world has changed as a result of an experience. The bigger the divergence, the bigger the learning. What remains to be seen is how exactly the brain might be implementing this kind of Bayesian belief updating. How are these complex statistical estimates carried out flexibly and continuously by biological brains? To be able to see this and further finesse our story, we need to consider a particular proposal within the broader family of Bayesian approaches to cognition, a proposal that in the last few years has grown to become one of the major frameworks in contemporary cognitive science. This proposal is predictive processing.

1.3.3 Predictive Processing: Finessing the Story

Predictive processing (henceforth: PP) allows us to bring together all the observations we made so far about learning in psychology and neuroscience. It develops the interpretations of perception as hypothesis-testing and the brain as a probabilistic model of the world and offers a more detailed picture of how the rolling process of belief updating might be carried out by the brain. For these reasons – and others that we shall discuss – it is worth expanding on, as we build the premises for our inquiry on art and learning.[40]

PP shares with other Bayesian approaches to brain function the idea that agents like us maintain a grip on their environment by embodying and constantly revising a probabilistic model of the world. This model, instantiated by our brain structures and dynamics, defines the kind of world that the agent expects to encounter at any given time. In other words, the model generates the predictions (or beliefs, or hypotheses) that are tested against the incoming sensory stimulations. According to PP – and this is a crucial addition to the story – these predictions unfold

[40] The literature on PP is extremely fast-growing and often quite technical. Here I shall only discuss, in an accessible way, the aspects of the framework that are relevant to my aims. For handy philosophical introductions to PP, see Clark (2013) and Hohwy (2020). For more extensive treatments, see Hohwy (2013), Clark (2015) and Seth (2021).

in a hierarchical fashion across many spatial and temporal scales, following the hierarchical organisation of cortical sensory areas. A high-level prediction to see a dog, for example, may give rise to "lower-level predictions about limbs, eyes, ears and fur, which then cascade further down in terms of predictions about colours, textures and edges, and finally into anticipated variations of brightness across the visual field" (Seth 2021, p. 108). At each level of the hierarchy, predictions are compared with the sensory stimulation coming from the level below, and to the extent that there is a mismatch between the two, a "prediction error" is generated. This prediction error is propagated up in the hierarchy and used to recruit new and better predictions; these are then compared again with the incoming sensory stimulation, and so on in an iterative fashion. This reciprocal exchange of bottom-up prediction errors and top-down predictions proceeds until at all levels prediction error is minimised and predictions are optimised. If this picture is right, then, the brain approximates the kind of Bayesian belief updating described in the previous section. Reducing prediction error at each level of the hierarchy, it minimises the uncertainty about the causes of its sensory inputs, coming up with the best conditional explanation of the data it is observing. By constantly updating a hierarchical probabilistic model of this kind, the agent can therefore make contact with a structured world full of objects, people and places and keep track of their changes over time.

In line with the views we have explored so far, PP conceives perception not as a passive extraction of sensory information, but as a proactive process, driven by the attempt to predict the incoming stream of sensory stimulations. In this picture, the bottom-up sensory flow acts in fact more as a feedback signal on how well the agent is doing in predicting the environment. Correct predictions are a signal that the agent is coping well with the environment. Prediction errors signal that the model embodied by the agent needs some revision; they "carry the news", highlighting surprising aspects of the incoming stimulations that still need to be accounted for by the model. Overall, this process points to a fundamental continuity between perception and learning, which are both seen, in PP, as part of the process by which the brain restructures itself to best account for the incoming sensory stimulation. The difference between perception and learning is only in their timescale, with perception reducing prediction error by fast changes in neural connectivity (those that allow you to recognise the Dalmatian in Figure 2 the first time) and learning being realised by slower processes of brain plasticity (those that allow you to recognise the Dalmatian faster the second time). Indeed, one of the major ambitions of PP is to unify perception, learning and decision-making under a common explanatory principle, seeing all of them as "processes that resolve uncertainty about the world" (Friston et al. 2017, p. 2634).[41]

All these processes of uncertainty reduction (or prediction error minimisation) are said to happen mostly at the subpersonal level, below our conscious awareness. There is indeed an

[41] See also Friston (2018, pp. 1019-1021): "[One can] explain both fast neuronal dynamics that underwrite perceptual synthesis and the slow fluctuations in synaptic efficacy that mediate perceptual learning with just one imperative: to minimize prediction error… perceptual inference (i.e., neurodynamics) and learning (i.e., neuroplasticity) are in the game of optimizing the same thing; namely, model evidence or its variational equivalent (i.e., free energy)." Thus, "the brain has gracefully integrated perception and learning within the same computational anatomy… learning and perception are two sides of the same coin."

ongoing debate as to whether PP, as a theory of subpersonal processes, can illuminate aspects of our conscious, person-level experience.[42] Given what we have said so far, however, it seems plausible that, at least in some cases, these dynamics might rise above the threshold of consciousness and impact our phenomenal and affective experience. We have seen for example that the solution of complex-enough perceptual problems tends to be accompanied by a pleasurable affective response (an "Aha!" experience). In experiences of this sort, the positive affective response seems strictly linked with the cognitive success in resolving the uncertainty about the stimulus in question. From a PP perspective, it makes sense for the dynamics of uncertainty reduction to be affectively charged. In this framework, the agent is seen as a "self-evidencing" creature (Hohwy, 2016), an embodied model of the world constantly searching for evidence for its own validity. To minimise prediction error means therefore to maximise the evidence for the agent's own existence. In other words, what becomes more or less likely as we test hypotheses on how to organise our sensorium is effectively our existence as viable models of our world. Given these underlying existential implications, one should expect our perceptual processes, with their ups and downs of uncertainty, to be marked (perhaps intrinsically) by affect. This is indeed the hypothesis of a growing stream of research within the PP framework that is trying to model affective valence on the dynamics of prediction error minimisation (see Joffily and Coricelli, 2013; Van de Cruys, 2017; Kiverstein, Miller and Rietveld, 2019; Hesp et al, 2021). According to this developing view, increased or irreducible prediction error will be felt as unpleasant, and, by contrast, a reduction in prediction error will be marked by positive affect. As Van de Cruys (2017, p. 22) puts it,

> affective valence is determined by the change in… prediction error over time, with positive valence linked to active reduction of prediction errors, and negative to increasing prediction errors. This makes sense because these temporal dynamics signal whether the organism is making progress (or regress) in predicting its environment, which in the long run translates in proper functioning of the processes of life (fitness).

PP seems therefore to confirm the role of affect in insightful problem-solving and to generalise it to all experience. If this story is on track, all perceptual experience is in fact accompanied by an evolving affective profile that signals, with varying degrees of intensity, how well we are doing in our attempt to remain viable models of our world. In turn, the joy of finding structures in our sensorium takes on the traits of an existential conquest. We shall keep this in mind, as it will be important later in our discussion about the pleasure afforded by art.

PP has a complex story to tell about how exactly this process of prediction error minimisation is implemented in the brain, a story that we do not need to follow here in detail.[43] Suffice it to say that research on the neural plausibility of PP is still very much in its infancy, and while the proposal seems to map well with many known facts about neuroanatomy and electrophysiology,

[42] See for example Block (2018), Clark (2018) and Ramstead et al. (forth.).

[43] But see Friston (2010) for a detailed account.

some aspect of it remains to be gauged by empirical research.[44] But even putting aside the specific PP story, that the brain tracks the statistical regularities of its environment and the surprisingness of its sensory states remains quite clear. Neuroplasticity, as we saw, is standardly seen as the brain's way to encode environmental regularities at multiple timescales. In addition, a wealth of data on electrophysiological responses and many well-known phenomena such as repetition suppression, mismatch negativity and the N400 and P300 in electroencephalography suggest that the brain is evaluating the predictability or surprisingness of sensory stimulations as they come along. Surprising words, sounds or visual percepts for example are known to elicit a variety of event-related brain responses whose amplitude seems to be proportional to the surprisingness of the stimulus in question. We shall return to some of these responses in later chapters. Here let us simply note that there is by now compelling evidence that "neural responses are shaped by expectations and that these expectations are hierarchically organized" (Heilbron and Chait, 2018, p. 54). As such, even if many details of the specific PP proposal about neural implementation will turn out to be wrong or imprecise (as is likely to be the case), this does not compromise the idea of the brain as a hierarchical probabilistic model of the world relentlessly trying to predict the incoming flow of sensory stimulations at multiple levels.

This idea is in fact most of what we are going to need in what follows. From it, we can distil an image of learning (i.e., an image of how experience changes us) that is in line with mainstream psychology and neuroscience and precise enough to be applied in interesting ways to the arts. This image depicts learning as the updating of a hierarchical probabilistic model of the world accompanied by corresponding changes in the brain and by characteristic (electro)physiological responses. This model-updating may in turn be conceived as a Bayes-optimal shift from a prior to a posterior probability distribution (indicating respectively the state of the learner's beliefs before and after the experience) or, in PP terms, as the minimisation of prediction error – i.e., the minimisation of the mismatch between the predictions of the model and the evidence coming from the senses.

Given this characterisation of learning, we also get a clear idea of what kind of stimuli are most conducive to it. If learning amounts to the minimisation of prediction error, it becomes clear that, for it to take place, 1. there ought to be some prediction error to reduce (the subject's probabilistic model of the world should fail to predict a certain aspect of the incoming sensory stimulation); and 2. the prediction error ought to be reduced (it should be possible for the subject to optimise her model of the world in light of the incoming sensory stimulation). This means that percepts that maximise learning will tend to be neither too predictable nor too unpredictable given the learner's current model of the world: they will occupy a region with an optimum of unpredictability that maximally affords reducible prediction errors. Conversely, learning will tend to decrease with percepts that are either too predictable or too unpredictable, as these both afford, for opposite reasons, less reducible prediction error – less margin for improvement.

[44] For discussion on the neurophysiological evidence for PP, see Friston (2005), Clark (2013), Heilbron and Chait (2018), Walsh et al. (2020).

Here our conclusion mirrors once again our discussion of insightful problem-solving. Learning in PP terms, as insight for the psychologist, appears to be stronger in regions of the input space with an optimal level of ambiguity.[45] If we consider once more the case of a perceiver primed to see a Dalmatian dog in each of the three images in Figure 3, we can say, in PP terms, that the image on the left produces little prediction error (since it conforms more to our predictions about what a Dalmatian dog should look like); the image on the right produces an irreducible prediction error (since it deviates too much from our predictions about how a Dalmatian dog should look like). The image in the middle, instead, deviates from our predictions enough to cause a prediction error, but not enough to make it impossible to reduce it. As such, it produces the highest reduction in prediction error, driving more learning and plasticity and causing a pleasurable affective response. Reformulating the phenomenon of insight in PP terms, however, allows for two important additions to the story: we are now able to follow the temporal dynamics of perceptual problem-solving with a finer level of detail, and to explain the pleasure we get from the solution to the problem in terms of an existential conquest.

In wrapping up our discussion of PP, it is worth noting how the image of learning we have arrived at through psychology and neuroscience differs from that of the philosopher. A few differences immediately stand out. Instead of a taxonomy including various kinds of learning (learning-that, learning-how, learning-what-it-is-like, understanding, etc.) we get a picture with just one overarching process optimising perception, reasoning, and action. Instead of an important distinction between propositional and non-propositional kinds of learning, we get a picture in which the notion of proposition plays no apparent role, and our mental life is better described in non-sentential, probabilistic terms. More importantly, instead of an epistemic notion of learning as a change that brings us "closer to the truth", we get a notion that captures how and how much we are changed by experience, irrespectively of the ultimate epistemic status of these changes. Learning is no longer defined by the relation between the subject's model of the world and truth, but by the relation between the subject's model of the world and the stimulus in question. We do not learn by acquiring true information, but by being exposed to surprising information that prompts cognitive rearrangements. As a result, we also get a notion of learning that is better equipped to capture what learning means for the subject and her brain and body. If the psychological and neuroscientific notion of learning captures (to the extent that it is successful) the dynamics of belief updating in the brain, it is clear, on the other hand, that the epistemic notion of learning does not describe what our brain is responsive to. There is nothing in our brain that marks the accumulation of true beliefs, or a shift from a false to a true belief. The way our brain assesses its cognitive progress has little or nothing to do with the acquisition of justified true belief, and much more with how predictable/unpredictable the sensory inputs are judged to be in relation to our current model of the world. This is, *pace* the

[45] See also Friston (2013b, p. 1329): "Ambiguity (or perhaps its resolution) gets to the heart of perceptual inference: there is no point—or pleasure—in making statistical inferences about sure bets. The raison d'être for inference is to disambiguate among plausible and competing hypotheses."

epistemologist, what seems to matter for the brain and the phenomenal and affective experience of the learning subject.

1.3.4 Tuned for Learning

We have therefore arrived at a picture of learning that seems to capture important aspects of our phenomenal, affective and bodily experience as learning subjects. We have seen that we tend to learn most from stimuli with an optimal level of unpredictability. This latter consideration is not just well-supported by theoretical and empirical work in psychology and neuroscience; it is, I take, almost trivial. What else could learning be if not something we carry out in relation to what we do not already know but can still assimilate? Quite evidently, if you get exactly what you predict, you are not learning anything new; nor are you learning if you are not getting better at predicting something. Learning must consist in the successful assimilation of new information, and this, as we saw, might be conceived in terms of the updating of a multi-layered probabilistic model of the world. If this picture of learning is accurate, however, we should also expect it to capture something about our behaviour as epistemic agents. If for example there really is an optimum of unpredictability in learning, we should expect this fact to make some difference in the ways we go about acquiring information. This is therefore the last aspect of our picture of learning that we shall consider before moving to art: whether learning as we described it has any role to play in explaining how we behave as information-seekers.

It turns out that it has. Mounting evidence in developmental psychology seems to point to the fact that stimuli with an optimum of unpredictability not only maximise learning, but also tend to be the ones we actively seek and try to bring about, the ones that most attract our attention and arouse our curiosity. As information seekers, we seem to display an appetite for information that can change us the most, and this appetite seems to guide our development and behaviour as cognitive agents from the very beginning. Seven- and eight-month-old infants, for example, have been shown to prefer and concentrate on stimuli with an intermediate level of unpredictability as compared with either highly predictable, or highly unpredictable stimuli – a phenomenon known as the "Goldilocks effect". This attentional strategy seems to hold for multiple types of visual displays (Kidd, Piantadosi and Aslin, 2012) and auditory stimuli (Kidd, Piantadosi and Aslin, 2014). It also seems to inform the way infants acquire language. 17-month-old infants for example attend longer to learnable linguistic grammars than unlearnable ones, displaying an implicit sense of where a valid generalization over linguistic stimuli can be made (Gerken, Balcomb and Minton, 2011). These results seem to suggest that "infants implicitly decide to direct attention to maintain intermediate rates of information absorption. This attentional strategy likely prevents them from wasting cognitive resources on overly predictable or overly complex events, therefore helping to maximize their learning potential" (Kidd and Hayden, 2015, p. 455). In other words, the growing infant quickly gets bored by things it already understands well, but also by those it does not understand at all, always searching for percepts exhibiting some yet unknown but learnable regularity. From the start, then, the way infants develop as cognitive agents and assess their cognitive progress seems not to be marked by an accumulation of true belief about the world, but by the systematic maximisation of learning in the sense we have defined.

Growing up, this appetite for learning continues to manifest itself in the way the child plays, manipulates objects and explores the environment. There is by now an extensive empirical literature that shows that children structure their play in a way that is sensitive to the rate of learning progress (i.e., the rate at which they can reduce uncertainty about the world).[46] Preschoolers have been shown to preferentially play with toys that violate their expectations, or whose underlying mechanisms are not yet understood (Bonawitz et al., 2012). They have also been shown to prefer toys that yield promise to solve ambiguities about their causal workings ("which of the two levers produces the effect that I am observing?") as compared with completely novel toys. (Schulz and Bonawitz, 2007). When there are ambiguities about the toy's causal workings, children appear to formulate hypotheses and then act efficiently to minimise uncertainty over such hypotheses (Cook, Goldman and Schulz, 2011). In displaying this sensitivity towards regions of the world that allow for optimal reduction of uncertainty, children are effectively promoting their own developmental trajectory, creating "progress niches"[47] that maximise their chance to learn new causal structures.

Of course, the child is not left alone in this attempt to grow and learn optimally. Adults decisively intervene in the child's environment to scaffold her learning, presenting her with stimuli and situations that are well-fitted to her evolving level of expertise. In doing so, the parents make sure the child stays in what Vygotsky (1978) called the "zone of proximal development", the space of problems and activities just at the edge of the child's current capacities. But, to a good and increasing extent, the child displays a propensity to craft her own developmental path. Quite often, the toys and activities that the parents think would elicit her delight are simply disregarded in favour of an object that suddenly and unpredictably attracts her attention – their father's keys, their mother's necklace, a wooden spoon or a frying pan. The child plays with it for a while and then abandons it in favour of the next object, in a way that is apparently capricious, but probably serves quite well the imperative of optimal learning.[48] The capacity of the child to autonomously craft her own learning trajectory is in fact so remarkable that it has inspired a strand of work in developmental robotics (to which we shall return) aimed

[46] For a summary of research in this area, see Kidd and Hayden (2015).

[47] I take this notion from Oudeyer, Kaplan and Hafner (2007). The authors describe progress niches as follows (p. 282): "the progress drive pushes the agent to discover and focus on situations which lead to maximal learning progress. These situations, neither too predictable nor too difficult to predict, are 'progress niches'. Progress niches are not intrinsic properties of the environment. They result from a relation between a particular environment, a particular embodiment… and a particular time in the developmental history of the agent. Once discovered, progress niches progressively disappear as they become more predictable." See also Smith et al. (2018) on how children create these optimal learning environments for themselves.

[48] In this autonomous selection of regions of the input space that maximise learning might lie, as we shall see, the germ of aesthetic sensitivity.

at constructing artificial agents that can mimic the information-seeking behaviour of children (Oudeyer and Kaplan, 2007; Oudeyer, Kaplan and Hafner, 2007; Schmidhuber, 2010).

Is this drive towards percepts that can maximise learning a feature of infancy and childhood only? It seems unlikely. As adults, we do not just stop looking for salient information, and we do not cease to concentrate our energies and attention on regions of the input space that seems to promise further learning. We shall return to this point throughout this work: it seems that beings like us are constantly in the game of sampling their environment so as to experience this ongoing improvement of our model of the world, avoiding wasting cognitive resources on overly simple or overly complex stimuli. In fact, if the PP story we have presented above is on track, this might even be an existential imperative. Remember that in PP the agent is seen as a "self-evidencing" creature, an embodied model of the world constantly trying to reaffirm its own validity. If this is true, then stimuli that can promote new learning are to be sought after, because they allow the agent to minimise uncertainty over the long run, thus preserving its viability. As we have seen, this intuition underlies the PP thesis that reductions of prediction errors are pleasurable, but it is also increasingly seen as the key to understanding human curiosity, happiness and wellbeing.[49] According to this developing view, humans are slope-chasers, beings always in search of regions of the input space "that, given our current mental models, hold the highest potential for – the best slopes of – prediction error (uncertainty) reduction" (Van de Cruys, Bervoets and Moors, forth).

We shall have more to say about this view at the end of our inquiry about art and learning. For now, let us notice that, quite apart from the individual progress niche that each of us might be able to carve out for ourselves, humans as a species are also undeniably "expert at deliberately manipulating our physical and social worlds so as they provide new and ever-more-challenging patterns that will drive new learning" (Clark, 2015, p.277). As humans, we construct artefacts (from the abacus to the telescope) that scaffold our thought processes and widen the reach of our senses. We introduce practices (mathematics, reading, writing, schooling, scientific inquiry, and so on) that exponentially increase the number of learnable causal structures we have access to. Our permeability to (and our appetite for) statistical regularities, if combined with such pattern-rich human environments, triggers a self-fuelling process that greatly expands the scope of our models of the world. Contrary to other animals endowed with similar neural machineries for prediction error minimisation, we are thus able to access opportunities for learning that bring our modelling activity far beyond the few environmental contingencies that an animal normally encounters, towards the recondite worlds of astrophysics, poetry, piano playing and software engineering. Human-built worlds are in sum progress niches in their own rights, not different in principle from the ones babies craft for themselves. They drive new learning and expand the range of patterns that we can come to embody. By structuring and restructuring these worlds,

[49] See Kiverstein, Miller and Rietveld (2019), Miller, Kiverstein and Rietveld (2022), Andersen et al. (forth.), Van de Cruys, Bervoets and Moors (forth.)

we structure and restructure our own minds, in repeated cycles of reciprocal influence.[50] The role of these forms of material and sociocultural scaffolding in our cognitive progress cannot be overlooked if we want to understand whether and how we learn from art.

1.4 The Guiding Hypothesis

Let us wrap up our preliminary discussion on art and learning by pointing out what we have discovered so far and where we are headed now. We started by identifying a distinctive feeling of having learnt that seems to accompany our experience of art and fuel our intuitions about its cognitive value. Consequently, we have tried to identify the notion of learning that might be best suited to capture the experience that art seems to provide. We have seen that the notions of learning most often discussed in the philosophical debate, due to their concern with truth and a certain disconnect from the phenomenology of learning, are not particularly helpful in this respect. In search of better options, we turned to psychology, neuroscience and cognitive science. Here we found a notion of learning that seems to capture important aspects of our phenomenal experience as learners and reflect known facts about brain function. To further define this notion of learning and make it modellable and quantifiable, we turned to recent Bayesian approaches to cognition, particularly PP. This allowed us to define learning as the updating of a hierarchical probabilistic model of the world embodied in our brain structures and dynamics. We saw that this updating can be conceived in terms of the minimisation of prediction error, i.e. the minimisation of the mismatch between the model predictions and the evidence coming from the senses. This allowed us to identify the kind of percepts that maximise learning, namely percepts with an optimal level of (un)predictability. Finally, in the previous section, we also saw that this notion of learning seems to capture important aspects of our behaviour as epistemic agents: from the very beginning, we seem to concentrate our attention on portions of the world that allow for the optimal reduction in prediction error.

Given these premises, the question of whether we learn from art becomes quite tractable. Now that we have an idea of what learning amounts to and what stimuli are most conducive to it, it will be a matter of assessing whether artworks are among those stimuli that tend to maximise learning in the sense we have defined. If that will turn out to be the case, then we will have to admit that artworks are also “progress niches” in their own right, prominent instances of

[50] See Clark (2013, p. 15) for an apt description of this phenomenon: “we thus self-construct a kind of rolling ‘cognitive niche’ able to induce the acquisition of generative models whose reach and depth far exceeds their apparent base in simple forms of sensory contact with the world. The combination of ‘iterated cognitive niche construction’ and profound neural permeability by the statistical structures of the training environment is both potent and self-fueling. When these two forces interact, repeatedly reconfigured agents are enabled to operate in repeatedly reconfigured worlds, and the human mind becomes a constantly moving target. The full potential of the prediction-error minimization model of how cortical processing fundamentally operates will emerge only (I submit) when that model is paired with an appreciation of what immersion in all those socio-cultural designer environments can do.”

those "deliberate manipulations of our physical worlds designed to provide new and ever-more-challenging patterns that will drive new learning", as Clark puts it. We would discover that artists somehow manage to carefully exploit the dynamics of belief updating we have delineated above, designing percepts that offer this experience to an enhanced degree. Art would then be just an expression of the more general drive towards learning that characterises our behaviour as epistemic agents. This will be the hypothesis that will guide the rest of our discussion.

To the psychologist of art, the fact that art might afford this kind of experiences would probably seem something more than just a tentative hypothesis. That there could be an optimal level of unpredictability (or surprisingness, novelty, uncertainty, ambiguity, complexity, arousal, etc.)[51] in the experience of art is something that was hypothesised by Wundt in the early days of modern psychology (see Wundt, 1874). Wundt's intuition was then famously formalised by the British-Canadian psychologist Daniel Berlyne (1970, 1971), who hypothesised that the relation between aesthetic liking and some "collative variables" (such as complexity, novelty, uncertainty, surprisingness and ambiguity) could be captured by an inverted U-shaped function (the so-called "Wundt curve"). In this model, aesthetic liking would increase with the increase of these collative variables but only up to an optimum, after which it decreases in a symmetrical way. With the introduction of information theory in the late nineteen-fifties, the same basic idea has then been reformulated in terms of an optimum between redundancy (order) and entropy (chaos). The hypothesis of the optimum has been investigated empirically by many studies since then, and has become one of the most enduring acquisitions in the field (see Van Geert and Wagemans, 2020 for a summary). In fact, Gombrich (1984, p. 9) had already laid out quite a while ago what he called "the most basic fact of aesthetic experience… the fact that delight lies somewhere between boredom and confusion." To complete the picture, in the last few years a PP version of this same story is developing and is being applied to art forms as diverse as visual art, music, cinema and games.[52] According to this story,

> individuals are likely to have an optimal amount of unpredictability that they most appreciate. Too much prediction error is unpleasant or even disturbing; none or too little is boring… An optimum of mild violation of predictions will be experienced as most pleasant, because the experiencer manages to return to a familiar mental schema. (Van de Cruys and Wagemans, 2011, pp. 1045-46)

This can hardly be a coincidence. On the one hand, we have a consensus in psychology and neuroscience that stimuli with an optimum of unpredictability (or novelty, ambiguity, complexity, etc.) maximise learning. On the other hand, we have a long and lively tradition in the psychology

[51] The constructs used in this literature are many, and one of the problems of this strand of work is that it is unclear if and how they overlap. As we shall see, PP might remedy this confusion by reconceptualising all these variables in terms of the unpredictability of the stimulus in question, i.e. the amount of prediction error that it produces.

[52] On PP and visual art, see e.g. Van de Cruys and Wagemans (2011), Kesner (2014), Seth (2019). On PP and music, see e.g. Vuust and Witek (2014), Koelsch, Vuust and Friston (2019), Mencke et al. (2019). On PP and cinema, see Miller et al. (forth.), and on PP and games Deterding et al. (2022) and Andersen et al. (forth.). We shall return to this literature in Chapter 4.

of art that tells us that stimuli with an optimum of unpredictability (or novelty, ambiguity, complexity, etc.) are the ones we tend to accord our aesthetic preference to. This seems to point to a fundamental relationship between art and learning that deserves further attention.

Unfortunately, this connection and this lively psychological literature are mostly ignored in the contemporary philosophical discussion on art and learning,[53] to the effect that the arts have yet to be systematically analysed from this perspective, and the broader philosophical consequences of such an analysis are still all to be drawn. In what follows, we will try to fill this gap. In the next three chapters, we will embark on a detailed examination of the ways in which artworks can elicit learning in the sense just defined, with all the rich phenomenology associated with it. We shall concentrate in particular on literary language (Chapter 2) and narrative (Chapter 3), as both are particularly underexplored topics from this perspective. In Chapter 4, we shall expand our discussion to visual art, music and sensorimotor activities in order to show that our conclusions are quite general and can be applied across the spectrum of aesthetic endeavours. This careful examination of a variety of art forms will allow us to see that there is indeed a fundamental and pervasive relationship between art and learning, and that this relationship can tell us a great deal about why we engage with artworks, what we expect from them, how they tend to be structured, and how they make us feel.

[53] A notable exception is Cochrane (2021). See especially pp. 31-49.

2. Literature, Language and Insight

2.1 Insightful Language?

The concerns of this chapter can be usefully introduced by means of an example. Consider the following two sentences:

> (1) Spain – a great whale stranded on the shores of Europe.
>
> (1a) Spain – this country whose cartographic representation has an enlarged form and is located on the borders of Europe.

(1) is attributed to Edmund Burke and was included by Melville among the epigraphs of *Moby Dick*. (1a) is a paraphrase of (1) provided by a renowned group of semioticians and literary scholars.

Most readers will probably agree that there is something suggestive in (1) that is missing in its paraphrastic counterpart. In its conciseness, Burke's sentence seems to convey a dense array of subtle implications. There is probably an allusion to the shape and collocation of the country on the world map, as stressed by (1a); but there seems to be much more than that. The fact that Spain is "stranded on the shores of Europe" seems to suggest that the country is not part of the continent, that it lies at its borders like a foreign body, culturally and politically marginalised. And why is Spain a "whale"? The reference to a marine animal might suggest a naval vocation. And why is the whale referred to as "great"? Perhaps Burke alludes to an authentic domain of the sea, to an era of primacy. But then, why is the whale-Spain "stranded"? It could have another attitude, hostile, intimidating; it could be a Leviathan, or a kraken that threatens to devour Europe; or it could be just "asleep", a latent power. The allusion is evidently to a loss of primacy, to a time of crisis, which incidentally was that of eighteenth-century Spain, which had to surrender its ambitions of sea hegemony to the British Empire. But a stranded whale may also suggest the image of a mighty political power, made unable to react readily by the stateliness of its own bureaucratic or military apparatus… The conjectures of the reader can go on indefinitely, mobilising also her knowledge of the history of the period, of the position of the British Burke in relation to Spain, of other literary texts, and so on – in an open-ended, and characteristically subjective, fashion.

In reading (1a), instead, we get a completely different impression. If (1) seems informative, worth taking into consideration, perhaps even revealing, (1a) seems a rather plain and dull statement, not particularly worthy of attention or thought. Clearly, (1a) would never have impressed Melville enough to find its place among the epigraphs of *Moby Dick*.

But perhaps this is just a fault of this specific paraphrase; we might want to suggest other equivalents:

(1b) Spain – this country that is alien to the European political and cultural identity.

(1c) Spain – this naval power now in decline.

(1d) Spain – this country burdened by a big bureaucratic and military apparatus.

We must admit that we are not quite satisfied with those either. It seems that none of these paraphrases, not even a set of them has the same power of mobilising us in the way that (1) does. This is why so many paraphrases are accompanied by qualifiers like "roughly" and "and so on". On the one hand, the paraphrase seems to say too much, making too explicit and clear-cut that which the original conveys in a much more nuanced way; on the other hand, it seems to say too little, leaving out things that one would like to mention. In any case, it fails to give the same impression of informativeness and worthiness that the original gives. Max Black summarised this widespread intuition in his classic work on metaphor (1962, p. 46):

> Suppose we try to state the cognitive content of a... metaphor in ''plain language.'' Up to a point, we may succeed... But the set of literal statements so obtained will not have the same power to inform and enlighten as the original. For one thing, the implications, previously left for a suitable reader to educe for himself, with a nice feeling for their relative priorities and degrees of importance, are now presented explicitly as though having equal weight. The literal paraphrase inevitably says too much – and with the wrong emphasis. One of the points I most wish to stress is that the loss in such cases is a loss in cognitive content; the relevant weakness of the literal paraphrase is not that it may be tiresomely prolix or boringly explicit (or deficient in qualities of style); it fails to be a translation because it fails to give the insight that the metaphor did.

The point is not limited to metaphor, of course. When Shakespeare's Juliet, leaving her lover under the balcony, says that "Parting is such sweet sorrow", the phrase "sweet sorrow" is not a metaphor, but, with its oxymoronic quality, it can be just as suggestive as Burke's metaphor. We might have the impression that Shakespeare has perfectly captured that particular feeling of joyful melancholy that a lover feels when the absence of the beloved elicits in her mind fantasies about their next encounter, and memories about the encounter just concluded; or maybe Shakespeare wanted to point to a quite general feature of love, suggesting that it is lived more intensely when it is hindered, delayed.[1] Again, there is probably no way to define precisely what the "cognitive content" of that "sweet sorrow" is, or to give an exact definite equivalent. But that

[1] The seeming ability of great writers to describe (more accurately than the philosopher or the psychologist) the subtle nuances and complexities of our experience and inner life is often remarked upon. See for example Nussbaum (1990, p. 5): "Certain truths about human life can only be fittingly and accurately stated in the language and forms characteristic of the narrative artist. With respect to certain elements of human life, the terms of the novelist's art are alert winged creatures, perceiving where the blunt terms of ordinary speech, or of abstract theoretical discourse, are blind, acute where they are obtuse, winged where they are dull, and heavy." In a similar vein, Dewey (2005 [1934], p. 70) notes: "Poet and novelist have an immense advantage over even an expert psychologist in dealing with an emotion. For the former build up a concrete situation and permit it to evoke emotional response. Instead of a description of an emotion in intellectual and symbolic terms, the artist 'does the deed that breeds' the emotion."

expression has nonetheless provided a certain experience that we consider valuable, and that a paraphrase does not seem to provide.

In sum, there seem to be certain verbal expressions – of the kind that abound in great literary works – that are perceived as particularly informative, telling or insightful. They might give the impression of conveying with great aptness a certain feeling or situation, or they might prompt some thought that we judge worth having. These are the kind of passages that make us pause in reflection while reading, that we underline with a pencil and that we recall more vividly; these are the kind of passages that make us inclined to believe that what we are reading is being effective in expanding horizons and introducing new ways of thinking, feeling, and understanding the world.

Our task for this chapter will be to capture the roots of this feeling of insight that seems to accompany the encounter with linguistic expressions of the kind we normally find in literature. In other words, we will try to understand why certain linguistic expressions appear to be particularly suggestive, informative and insightful, while others do not. The question touches upon an old controversy that parallels the more general one about art and learning. It is the controversy about the cognitive value of literary language, a problem that has always elicited quite opposite reactions (see Camp, 2006 for a summary).

On the one hand, there is a tendency, particularly widespread in analytic philosophy, to be wary of claims of supposed ineffability or non-paraphrasability. What is it, one might reasonably ask, that is unique about such felicitous linguistic creations and such that no definite, finite paraphrase, however sophisticated and carefully worked out, could in principle do justice to them? What kind of insight disappears whenever one tries to make it more explicit? Is there really a special cognitive content attached to such expressions? According to many philosophers, if figurative utterances like (1) express any genuine content at all, then that content can in principle be paraphrased into literal terms. There is no such thing as a special cognitive content that resists paraphrase. Granted: a figurative utterance like (1), in addition to conveying a certain content, may prompt or inspire various reactions, in the form of an undefinable, undetermined and open-ended array of non-propositional "effects". These can include loose impressions, images, analogies, things that the utterances make us notice or think about. But all these fleeting and subjective effects are not a matter of cognitive content. Content is always determinable. In this view, therefore, felicitous literary expressions do not afford any particular insight and do not carry with them any distinctive kind of content.[2]

To some extent, this tendency to limit the domain of content only to what is definite and finite may be considered an undesirable heritage of the myth of a logically perfect language that characterised analytic philosophy of language in its beginnings, a myth according to which, as

[2] Davidson famously held this view about metaphors: "If a metaphor has a special cognitive content, why should it be so difficult or impossible to set it out?... Can't we, if we are clever enough, come as close as we please?'' (Davidson, 1978, p. 44). The same view has recently been articulated comprehensively by Lepore and Stone (2015). On the challenges posed by non-propositional "effects" to current theories of linguistic pragmatics, see Wilson and Carston (2019). For more on the problem of paraphrase in literature and poetry, see Currie and Frascaroli (2021).

Frege (1903, §56) puts it, an area that is not clearly delimited is not an area. But the hostility towards indeterminacy is probably also indicative of a more fundamental difficulty of truth-conditional semantics. If you conceive the meaning of a proposition as being the same as its truth conditions, you need to be able to state what these truth conditions are; you need, in other words, to be able to state exactly what you will observe in the world should the proposition be true. If this is not possible, the proposition is not provable or disprovable empirically and should therefore be dismissed as meaningless and metaphysically loaded. To the logical positivists, a sentence like (1) should have seemed just that kind of meaningless metaphysical proposition that the philosopher should avoid in her pursuit of truth. Now, if we accept this picture, it follows quite naturally that the golden standard of the "meaningful" proposition is to be found in the unambiguous statements usually associated with scientific practice, and that, by contrast, ambiguous, open-ended literary sentences like (1) are defective; their ambiguity is a fault that needs to be remedied if the proposition is to make sense at all. A paraphrase like (1a), then, being clearer, is much more suited to convey truth and, consequently, to contribute to knowledge.[3]

And yet, there is another tradition, well-represented in both analytic and continental philosophy and predominant among writers and literary scholars, that is more sympathetic towards the intuitions of readers and does maintain that felicitous literary utterances have a peculiar cognitive potential. According to this tradition, it is a "heresy" either to deny that what such utterances do is genuinely cognitive or to assume that it can be translated without loss into literal terms. In this perspective, the ambiguity of literary utterances is not a flaw that needs to be remedied, but precisely what allows a literary text to be so suggestive and to disclose an inexhaustible heuristic potential over time.[4] The extraordinary longevity of the great literary work in comparison to scientific works – the fact that, for example, whereas we still read Euripides, we no longer study Ptolemy – would be due precisely to its ability to remain open and challenging, to preserve a "supply of structural indeterminacy" (Lotman, 1990, p. 80), postponing indefinitely the moment when it will be completely clear.

[3] Even if not many nowadays support the strict verificationism of the logical positivists, the ideas that to know the meaning of a sentence means to know its truth conditions, and that ambiguity represents a problem when trying to establish the truth conditions of a sentence remain widely held (see Saka, 2007 for discussion). For the problems that this line of thought encounters with literature, see the discussion in Currie (2020, pp. 104-105). Here Currie takes as an example a sentence by Proust: "The whole art of living is to use the people who make us suffer simply as steps enabling us to obtain access to their divine form and thus joyfully to people our lives with divinities." While Currie grants that "this is not an obviously empirical claim; indeed it is far from clear what it means", he maintains that "the claim that the remark is worth taking account of in one's practical life can't be insulated from empirical inquiry… Not knowing how to proceed empirically does not license us to proceed non-empirically."

[4] See for example Merleau-Ponty (1964 [1960], p. 90): "What is hazardous in literary communication, or ambiguous and irreducible to a single theme in all the great works of art, is not a provisional weakness of literature which we could hope to overcome. It is the price we must pay to have a conquering language which, instead of limiting itself to pronouncing what we already know, introduces us to new experiences and to perspectives that can never be ours." Similar claims are also made by Richards (1936), Ricoeur (1975), Lotman (1977 [1971], 1990). Cavell (1976 [1969]) too, while maintaining that utterances can always be paraphrased, notes that "the overreading of metaphors so often complained of, no doubt justly, is a hazard they must run for their high interest" (p. 79).

This positive attitude towards ambiguity goes hand in hand with the deeply rooted distrust, particularly in literary criticism, towards those approaches that pretend to find the "true meaning" of a text. Especially after deconstruction, Paul Valéry's adage that "Il n'y a pas de vrai sens d'un texte" (1936, p. 74) has become almost common sense in literary circles. Added to this, there is a widespread concern, among both philosophers and literary critics, that seeing a literary text as carrying some sort of detachable lesson that needs to be extracted from it is to misunderstand the value that we attribute to the reading experience. As literary scholar Derek Attridge puts it in a recent interview,

> the point of reading, putting it very simply, is not to carry away some nugget of knowledge about the world or some moral truth about how real people should live their lives, some fact about the Napoleonic wars or whatever the subject of the work might be… a text becomes a work of literature when it is the experience of reading it, hearing it, or seeing it performed which constitutes the value or importance to the reader, listener or viewer. It is not anything that that person could put into words afterwards as the message or the truth that the work has conveyed, but rather something that has happened during the course of the experience that has left that individual different. It is to be hoped the change has been for the better, but you cannot say that it is always a good thing to have read a certain work of literature or watched a certain play.[5]

It should be noticed, however, that this debate around the cognitive value of literary language is rather inconclusive when it comes to accounting for the phenomenon that originated our interrogation in the first place: namely that certain utterances, like (1), display a particular suggestiveness and certain others, like (1a), do not. Deniers of the cognitive value of literary language tend to insist that what (1) conveys is either determinate or not genuinely cognitive, and advocates of the cognitive value of literary language tend to insist that it is both indeterminate and genuinely cognitive. But this still tells us nothing about the difference in suggestiveness between (1) and its paraphrases. To tackle this issue, we do not need to embark on long disquisitions about the nature of the cognitive content conveyed by (1). We can concede to the anti-cognitivist that a felicitous literary expression has no special cognitive content and merely produces "effects". Or we can grant instead to the cognitivist that the cognitive content of a literary utterance cannot be expressed in a paraphrase. In both cases, the fact remains that certain expressions are perceived as more insightful than others, and we would like to know why. This is, as I have argued in the previous chapter, the interesting question, and this is the question we shall address here. Whether literary language provides true insight or just a misleading appearance of it, we need to explain why it makes us feel the way it does.

[5] In Cools and Verheyen (2019, p. 11). In the same interview, philosopher Peter Lamarque agrees with Attridge and reiterates: "the idea that literature offers us some kind of nugget of knowledge or moral truth, a sort of take-away detachable proposition which gives the work its interest and its value, is very reductive and not the best way to think of literature. If you start with literature as art, as offering some kind of experience, you are less inclined to think of literature in terms of something that offers us nuggets of knowledge" (p.13).

Doing so requires, quite evidently, a shift from the current philosophical discussion on the topic. This latter has revolved so far around the question of whether there is some special "content" or "meaning" conveyed uniquely by the literary expression. But talking about content or meaning means focusing only on the final product of the interpretive activity and missing its dynamics, which, as we shall see, are all-important. To the reader, the text appears not as a set of propositions with truth values, but – as Attridge and Lamarque aptly stress – as an experience: a percept that unfolds before her eyes and that she needs to make sense of. To capture this experience, we shall rely heavily on the theoretical apparatus I have introduced in the previous chapter, and particularly on the notion of insightful problem-solving in Gestalt psychology and the picture of learning that emerges in Bayesian cognitive science in general and PP in particular. By the end of the chapter, we will hopefully have reached an image of the experience of reading literature that explains its revelatory power and its allure and agrees with the phenomenological intuitions of readers and the theoretical acquisitions of psychologists, neuroscientists and literary scholars.

2.2 Meaning-Making as Problem-Solving

In Chapter 1, we saw that humans (and at least chimps too) sometimes learn to perform a successful piece of behaviour by insightful problem-solving. This happens when an agent's habitual behavioural patterns do not meet the demands of a certain situation, resulting in a circumstance that the agent does not know how to handle or interpret. Solving the problem presented by the unexpected situation entails overcoming our tendency to conceive the elements of the situation in the habitual way and grasping the relevance that each element has for the situation at hand. In Ducker's candle problem, for example, the optimal solution is achieved only if the subject is able to overcome her tendency to see the box as merely a container for the tacks and succeed in reconceptualising it in a way that meets the demand of the particular task (i.e., as a holder for the candle). After the solution, the terms of the problem, once unrelated, become parts of a meaningful functional whole (a Gestalt) in which each of them plays the most contextually appropriate role. If the problem was challenging enough, the solution is also accompanied by a positively-valenced feeling of cognitive success (an "Aha!" experience), as the agent becomes increasingly confident about the correctness of the found solution.

We have also seen that, although this kind of process is likely to be involved in every sort of cognitive activity and is probably the basis of adaptive behaviour as such, not every situation lends itself to insightful problem-solving to the same extent, and, therefore, not every situation provides to the same degree the positive feeling of discovery associated with the solution to a problem. In many cases, the situation in which the agent finds itself is well taken care of by its habitual behavioural patterns, and as such does not require any strong "mental restructuring". Moreover, what was once genuinely a problem is no more a problem once the new appropriate behavioural pattern is acquired (going through the task of the candle problem for the tenth time in a row is to apply an acquired procedure more than to solve a problem). The result is that in most cases we solve problems automatically, by applying overlearned behavioural routines that we are not aware of applying unless something goes wrong. In many other cases, on the other hand, something does go wrong, but we are not able to do the required mental restructuring and to devise the optimal solution to the problem. To design an experience that allows for insightful problem-solving, then, one needs to arrange a situation that is ambiguous and novel enough for the agent to require some sort of restructuring (otherwise, there is no problem to solve) but not so ambiguous and novel as to render the restructuring impossible (otherwise, there is no solution to the problem).

What I want to suggest now is that the author of a suggestive utterance designs such an experience for her reader. In doing so, she provides the reader with the opportunity to experience the thrill of a cognitive success. To be able to see this, however, we first need to clarify in what sense the interpretation of a linguistic expression is akin to a problem-solving activity.

Let us start with a mundane case:

(2) The container held the apples.

In understanding this simple sentence, the reader has already done a considerable amount of interpretive work, even if she is normally not aware of it. For example, she is likely to have inferred that the “container” mentioned here is something like a basket, and not, say, something like a bottle. Why so? Because a basket is the most appropriate container to think of given the sentential context. In a different context, “container” could well have been interpreted as indicating something like a bottle. In fact, this is what Anderson and Ortony (1975) found in a now-classic psycholinguistic experiment. The study consisted of a cued recall task: subjects were exposed to a series of sentences like (2) that they had to recall after a period of time with the help of cues; (2) was recalled more easily if the cue was “basket” compared to “bottle”; the opposite was true, however, if the sentence was “The container held the cola”.[6]

But the reader of (2) is likely to have drawn many other inferences too. She might have imagined that the apples are green, red or yellow, but probably not blue or pink, that they are probably a dozen or so and not hundreds or thousands, that the container is perhaps on a table but probably not in outer space somewhere near Mars, and so on. None of this is explicitly stated by (2), and different people are likely to draw slightly different inferences. Try to ask different subjects to read (2) and you will see how different the response may be: one will imagine the “container” as a wicker basket, another as a crate of the kind you see in a market, yet another as the ceramic tray she has in her kitchen. Thus, even with this simple proposition one could argue at will about what its content is, what it says and what it merely implicates and what a paraphrase of it should look like. What interests us, however, is that all these inferences are produced during comprehension in the attempt to find out how the elements of the sentence can fit together in

[6] Similar experiments yielded similar results. For example, Barclay et al. (1974) in another cue recall task found that a sentence like “The man tuned the piano” was better recalled if the cue was "Something with a nice sound" compared to "Something heavy", but the opposite was true for the sentence "The man lifted the piano".

the most contextually appropriate way. That is, even if we are unable to say exactly what "container" means here, we can say with some certainty that, during comprehension, the reader tries to adjust the value of this word in relation to the other elements of the sentence. The "container" is something like a basket (or a crate, or a tray, etc.) in (2) because that is the most appropriate role that the word can play in that specific context; in another context, the same word will have a different role to play.

This is tantamount to pointing out the well-known fact that interpretation of even familiar and apparently unambiguous words varies with their context of occurrence. Consider for example this series (taken from Anderson and Ortony, 1975, p. 169):

(3a) John ate the soup.

(3b) John ate the steak.

(3c) John ate the apple.

(3d) The executive ate the steak.

(3e) The baby ate the steak.

(3f) The dog ate the steak.

(3g) Lord Raleigh ate the soup.

(3h) The tramp ate the soup.

We can notice that by virtue of the context in which it appears, the same word "ate", gives rise to very different suppositions about location, circumstance, manner, instrumentality, and antecedent and consequent actions. Again, it will be difficult to state exactly what "ate" means in each sentence, but what is sure is that we interpret it as suggesting different things according to the context in which it occurs.

But to say that words assume different values in different contexts is to say that we are endowed with the ability to devise the most appropriate value for them in each context.[7] And the process whereby one devises the most appropriate value for an object in a context is, as we saw, problem-solving. In other words, linguistic comprehension, insofar as it entails figuring out the most contextually appropriate meaning for the elements involved, is a problem-solving activity. We can recognise in it the same kind of practical intelligence that allows us to recognise what can be more useful for the task that we are trying to accomplish. The parallel with practical problem-solving is, I hope, self-evident: if, in solving a task like the candle problem, we renegotiate the value of the box of tacks to obtain the one that best serves the present purpose

[7] This remarkable ability to respond appropriately to a combination of words even if never encountered before is known to linguists and philosophers as "generativity" and corresponds to what Chomsky calls the "creative aspect of language use". Interestingly, for Chomsky, this is the "central fact to which any significant linguistic theory must address itself" (1964, p.7), even if it might always remain a "mystery" (1982, *passim*).

(i.e. a holder for the candle), in comprehending a sentence like (2) we renegotiate the value of its words to obtain the one that best fits the present linguistic and extralinguistic context. Linguistic comprehension seems thus part and parcel of the more general ability to act in a way that is appropriate to the situation at hand.

Now, even if linguistic comprehension seems to be always, to some extent, a problem-solving activity, not every linguistic expression lends itself to problem-solving to the same degree. In many utterances, words are employed in ways that are highly compatible with their habitual use (what we might call their vocabulary or "literal" value). As such, figuring out the way in which they can fit together in the most appropriate way is straightforward, and does not require a big restructuring. This is the case with (2), (3a-h) and most of our daily verbal exchanges, where the work of contextual adjustment is done so rapidly and effortlessly that we don't become aware of the problems we are solving. In other cases, words are employed in a way that is so remote from their habitual use that we cannot figure out how they can fit together into a unitary whole. This is the case every time we are not quite sure about what the author of the expression meant with it, and it happens paradigmatically with semantically anomalous or nonsensical sentences like "Colourless green ideas sleep furiously" and "Quadruplicity drinks temporalisation". Here words are perceived as unrelated objects that refuse to coalesce in a meaningful whole, like the objects in the candle problem before the solution is devised. Still in other cases, however, words are employed in a way that differs significantly from their habitual use, but the reader is able to adjust and grasp the value that they have in that particular case.

My suggestion is that the author of a suggestive utterance sets up this latter kind of experience. She shapes its utterance in such a way that its interpretation is challenging but at the same time possible. The suggestiveness and allure of literary language is due to the occasions it provides to exercise this inferential, problem-solving activity.

This idea should not sound too surprising, as it is in line with much of the theorising around literary language. If there is one thing that all literary scholars seem to agree about, it is that literary language displays a systematic tendency to deviate from "normal" ("literal", "ordinary") linguistic use. There has been talk of "abuse" (Valery), "violation" (Cohen), "scandal" (Barthes), "anomaly" (Todorov), "deviation" (Spitzer), "subversion" (Peytard), "infraction" (Thiry);[8] according to the linguist Charles Bally, "the first person who called a sailing vessel a sail made a mistake". In the Anglo-American area, Beardsley (1958, p. 141) speaks of metaphorical utterances as "self-contradictions", while some describe metaphor and other figures of speech in terms of Ryle's "category mistake" (Camp, 2004). These judgments seem reasonable. In fact, metaphors, metonymies, synecdoches, oxymorons and all the other instruments that crowd the toolbox of the poet and the prose writer lend themselves quite easily to be conceived in terms of the violation of some norm – semantic, syntactic, metric, prosodic, of tone, etc. A violation of any kind makes the verbal chain more improbable: it increases its level of ambiguity and makes

[8] The list, and the quote from Bally that follows, is taken from Groupe µ (1981 [1970], p. 9). See this book and Groupe µ (1977) for an early formulation of some of the ideas that follow in the jargon of Gestalt psychology and information theory.

it more difficult for the reader to see how the various pieces (phonemes, words, phrases) could fit together. Leave out these accidents and, it seems, the utterance becomes at the same time easier to comprehend and more underwhelming. It is well-known, for example, that once a figure of speech enters common usage and ceases to be perceived as a violation, it also ceases to cause the cognitive effects typically associated with figurative language. Nobody pauses in wonder to imagine a humanised table when someone mentions a “table leg”: the figure is lexicalised, dead. It seems quite clear, therefore, that to have a lively figure of speech, a figure that elicits thoughts and inferences, there must be a perceived violation of some kind of norm.

But to define the figure of speech as a mere violation is obviously not enough. If this were the case, there would be no way to distinguish a suggestive figure of speech from a simple nonsense or an error, and any strangeness inserted in a text would contribute in the same way to its literary quality. What most scholars agree about is that, for there to be a figure of speech, the violation should somehow be reabsorbed, it should receive some explanation. This is what distinguishes, for example, the devious but meaningful Shakespearian sentence “Juliet is the sun” from a simple nonsense like the Russellian “Quadruplicity drink temporalisation”. In the first case, we manage somehow to explain the anomaly; in the second case, on the other hand, this operation fails, and we remain baffled. The effective, lively figure of speech, therefore, is said to live on a tension between violation (without which it is not perceived) and motivation (without which it is not understood). Hence the slightly paradoxical formulas with which it is often designated: "deliberate mistake" (Valéry), "organised violence" (Jakobson), “significant self-contradiction (Beardsley), "coherent deformation" (Merleau-Ponty), "calculated error" (Ricoeur).[9] What these contradictory expressions point to is that for a felicitous figure of speech to constitute the kind of situation that demands – and allow for – insightful problem-solving: one that presents a challenge which is nevertheless solvable.[10]

In linking the interpretation of figurative utterances to insightful problem-solving, we are therefore providing a story that is consistent with an important strand of work in linguistics and literary theory and also, quite evidently, with the picture of the kinds of stimuli that maximise learning that emerged in Chapter 1.[11] As such, we can begin to answer the question of why we tend to attribute a particular suggestiveness to certain linguistic expressions. Those expressions are deemed suggestive that allow for insightful problem-solving. To return to our initial example, why Burke’s metaphorical utterance (1) is particularly suggestive to many readers? A preliminary response seems to be: because it presents these readers with a challenge of the right magnitude

[9] See Groupe μ (1981 [1970]), Jakobson (1923), Beardsley (1958), Merleau-Ponty (1964 [1960]), Ricoeur (1975) respectively.

[10] This as far as literary theory is concerned. For psychological evidence that there might be an optimum of unpredictability/complexity in our engagement with literary texts (and texts more generally), see Kammann (1966), Evans (1970) and Form (2019). Intriguingly, there also seem to be evidence that verbal expressions that lend themselves to insightful problem-solving are not only preferred, but also better recalled: see Auble, Franks and Soraci (1979).

[11] Going back to Figure 3 in Chapter 1 and trying to find an equivalent in the verbal domain, we might come up with something like: “Juliet is beautiful” (left); “Juliet is the sun” (centre); “Quadruplicity drinks temporalisation” (right).

– it is neither too simple to interpret, nor too complex. In particular, the unusual juxtaposition of "Spain" and "whale" presents the reader with a challenge that, once managed, assures the pleasure of an intellectual conquest. At the same time, we can also begin to answer the question raised by Black when he notices that the paraphrase of a suggestive utterance "fails to give the insight" that the original gives. "Insight" as defined by psychologists is the kind of positive affective response that accompanies the solution to a problem. To the extent that a paraphrase is an utterance that says things "more clearly", it is also an utterance that poses to the reader a simpler problem than the one posed by the original, a problem that, once solved, is not likely to provide the same inner click. When Black says that the paraphrase disappoints because it states explicitly what in the original is left for the reader to educe for himself, he is pointing to this very fact. Black is imprecise, however, when he claims that this is a loss in "cognitive content". Paraphrases do not fail because they do not convey all the content that the original conveys; this is at most an improper way of stating what happens. Paraphrases fail because they eliminate the challenge of grasping a content. If we, despite this, often value paraphrases and consider them legitimate and useful enterprises, it is because they can lead us to solve the problem generated by the original (much like the indication that there is a Dalmatian in Figure 2 facilitates its recognition). In other words, paraphrases are seen as informative only against the backdrop of the problem left open by the text and only if they help to solve it. They act as a cue given during a problem-solving task. But, quite evidently, they cannot replace the task itself.

2.3 Progress Without Truth

It is important to notice that the problem-solving activity described so far unfolds dynamically over time: readers do not wait until the end of a sentence to start interpreting it, but construct its meaning incrementally as reading progresses.[12] Using our previous example, it is not the case that the reader of "The container held the apples" first reads the entire sentence and then starts to figure out how all its elements can fit together in the most contextually-appropriate way. Instead, the reader starts interpreting the first word immediately, based on the available linguistic and extralinguistic context. This causes a revision of the context itself, which is updated in light of the word encountered (reading the phrase "The container…" I get new contextual expectations about what the sentence might mean). This new context then becomes the basis for the interpretation of the next word encountered and so on, in an incremental fashion.[13]

That linguistic meaning-making is incremental will be no news for those who, like the psycholinguist or the literary scholar, are accustomed to the study of linguistic comprehension in

[12] This is known among psycholinguists as the "principle of immediacy of interpretation". See Just and Carpenter (1980) for discussion.

[13] The continentally inclined person will recognise in this mutual dependency of the whole on the parts and of the part on the whole what is known as the "hermeneutic circle".

its unfolding, but it is hardly ever pointed out in the philosophical debate. As I noted above, the debate about the cognitive value of literary language seems to revolve around the question of whether there is a special “content” or “meaning” conveyed by a literary utterance. But to talk about content and meaning is to isolate the product and to overlook the process. The philosopher can abstract and say that a linguistic expression has such and such content. But for the reader the content of a sentence is nothing fixed or extracted once and for all, but a project being constantly revised as new linguistic material arrives.

Capturing this incremental process of meaning-making requires that we complement the Gestalt picture of problem-solving we used so far with what we learned in Chapter 1 about the dynamics of belief updating. In particular, we need to conceive the solution to the verbal problem (the Gestalt) as a hypothesis about the underlying structure of the linguistic expression (its “meaning”) that gains in probability while reading progresses. As we saw, this process of hypothesis-testing can be described in Bayesian terms. At any given time, one might say,[14] the reader entertains various competing hypotheses about the underlying structure of the linguistic expression. These hypotheses, each of which is held with different degrees of confidence, might be represented as a set of probability distributions. As reading unfolds, this set of probability distributions is updated (using Bayes’ rule) in light of the new linguistic evidence encountered, yielding a set of posterior probability distributions. The difference between the prior and the posterior probability distributions reflects how much the reader’s beliefs about the meaning of the linguistic expression have changed. Over each cycle of belief updating, the newly computed set of posterior probabilities (the new set of inferred hypotheses) becomes the set of prior probabilities for the next cycle, just before new input is encountered. The process is then repeated until one hypothesis in the set becomes sufficiently probable and the reader becomes reasonably sure about the meaning of the expression. As we saw in Chapter 1, this process is inherently *predictive*, because each new set of prior probabilities corresponds to a set of predictions about the structure of the sensory flow. The result is a model that captures the incremental process of linguistic interpretation, with its ups and downs in uncertainty as current interpretive hypotheses become more or less probable depending on the new linguistic material that reaches the reader.

Now, at each cycle of belief updating, the revision of our current hypotheses can be more or less dramatic, depending on how much the new element encountered along the verbal chain (e.g. the new word) confirms or disrupts them. Consider these two sentences:

(4) He was stung by a bee.

(5) He was stung by a mile.

In (4), the final word “bee” is in line with the interpretive hypotheses that most readers will entertain as most probable while reading the sentence up to that point. Another way of putting it is that the occurrence of “bee” is quite predictable, given the preceding context: if required to

[14] See Kuperberg and Jaeger (2016) for a useful discussion on these developing probabilistic models of linguistic comprehension. For a hierarchical PP model of linguistic comprehension, see Friston et al. (2018).

complete “He was stung by a…” in a plausible way, most readers will do so with “bee”. Being in line with the ongoing reconstruction of the reader, “bee” is therefore easily integrated into it, and the revision that it promotes is minimal. Its occurrence confirms the interpretation that the reader was building up, further reducing the uncertainty about the meaning of the sentence. Using a Gestaltic jargon, we can say that the occurrence of “bee” “stabilises” or “closes” the interpretation of (4): the sentence seems now complete and self-sufficient. With (5), instead, the opposite happens. Given the preceding context, the word “mile” is surprising: the reader could not see it coming. As such, its occurrence questions the hypotheses that the reader entertained as most probable at that point, and the uncertainty about the meaning of the sentence increases. In Gestaltic terms, the occurrence of “mile” “reopens” the interpretation of (5), introducing a “tension” that needs to be resolved. As it is, (5) does not “make sense”: further information is required.

Thus, it seems that not all words contribute in the same way to the incremental process of interpretation: some words, being more predictable, promote small revisions of our interpretive hypotheses and tend to stabilise interpretation; other words, being more surprising, promote stronger revisions of our interpretive hypotheses and reopen interpretation. In general, however, every new word encountered (to the extent that it is informative) causes some revision of our interpretation, even if in most cases the revision is so minute that we are not aware of it. As a rule, the more the interpretation of a sentence has stabilised at a given point, the more disruptive a new element is that does not confirm that interpretation, and the more aware the reader becomes that a revision is in order. Consider for instance the following two sentences:

(6) The horse raced past the barn fell.

(7) The woman painted by the artist fell.

Most people do a double take on (6); this is because, having encountered many sentences in the form subject-verb, readers initially tend to build an interpretation in which “raced” is the main verb of the sentence; later, however, they encounter “fell”, which contradicts this first interpretation and promotes a revision. This revision is felt by most readers: they are aware of having initially considered an interpretation and then having changed it in light of the new evidence. (7) has the same structure; therefore, here too most people initially build an interpretation in which “painted” is the main verb of the sentence (i.e. interpreting the woman as the one doing the painting), before revising their interpretation in light of the new evidence. But in this case most people are unaware of having done such revision. This difference is due to the fact that in (6) the interpretation of “raced” as the main verb is compatible with the phrase that follows, “past the barn”; as such, by the time the reader arrives at “fell”, this interpretive hypothesis has consolidated enough to make the occurrence of “fell” surprising. In (7), instead, the interpretation of “painted” as the main verb of the sentence is contradicted by “by” which comes immediately after, so it does not have the time to stabilise enough for us to be aware of having formulated and then revised it.

In sentences like (5) and (6) above, it is the last word that disproves our interpretive hypothesis and reopens it. However, it should be noted that, even if it is always possible to encounter such surprising elements in any position of the sentence, in general, the elements tend to become more predictable the further into the sentence they occur. This is because, as the sentence unfolds and interpretation builds up, the possibilities of how the sentence can continue are usually narrowed down. For the reader that begins to read (4), pretty much everything can follow "He…". Towards the end of the sentence, after reading "He was stung by a…", the reader has a reasonably precise idea of what to expect.[15]

An examination of all the subtle ways in which a sentence can confirm or violate our predictions is out of question here. But even from these unsystematic remarks I hope to have made clear that 1. the meaning of linguistic expressions is built up incrementally, by progressive revisions of the reader's interpretive hypotheses as she encounters new elements; and 2. this process is uneven and irregular in character, because some elements are integrated more easily into the current interpretation of the reader and tend to stabilise it, while others are more surprising and tend to question it more sharply. For these reasons, to talk about the "content" of a linguistic expression, or to say that that content is determinate or indeterminate, is to abstract and to miss the dynamic character of meaning-making. When the reader experiences a linguistic expression, there is no "determinate" or "indeterminate" content being extracted once and for all, but an interpretive attempt evolving over time, being now more indeterminate, now more determinate, stabilising and re-opening continuously. These dynamics are not captured by the notion of (propositional) content, but are felt by the reader, and are indeed a fundamental aspect of the experience of reading. They constitute what the Groupe μ (1977) called the "semantic rhythm" of a text: the impression, for example, that the reading is proceeding smoothly or irregularly, or that a certain sentence is creating an informational tension in a certain place, or that another is creating suspense with the insertion of a parenthetical, or that yet another has closed incisively, and so on. As we shall see, the skilful writer can manipulate these dynamics in the same way that she controls the unfolding of the plot of a story, shaping the experience that the reader will undergo.

That these dynamics are really part of the phenomenal experience of the reader is corroborated by the acquisitions of many psycholinguistic and neurolinguistic studies, which seem to paint the same picture of linguistic comprehension as an incremental and uneven process. The studies on eye movements during reading are an interesting case in point. By tracking the gaze of the reader as it progresses through the text and measuring the amount of time spent in each fixation, eye-tracking experiments make the progressive effort of meaning-making observable with some precision.[16] These experiments reveal that the interpretation of a written sentence does not proceed at a regular pace, but by way of arrests and accelerations, regressions and leaps forward. Readers do not pause the same amount of time on every word:

[15] The same must be true, on a different timescale, for phrases, and even for words. "A…" is a more open Gestalt than "Appl…".

[16] For a review of the literature on eye movements during reading, see Rayner (1998).

fixation on a word during reading reflects the predictability of the word given the preceding context. Words that are highly predictable given the preceding context are fixated for less time than words that are less predictable, and often they are skipped altogether. On the other hand, infrequent or surprising words are fixated for more time and sometimes cause regressive fixations to preceding portions of the text. Moreover, the gaze lingers more at the end of phrases and sentences, when interpretation tends to stabilise, than in the middle of phrases and sentences, when interpretation is more open and incomplete. The result is an uneven progress, faster when words are more predictable and slower when they are surprising, as shown in Figure 5 below, which reports the average fixation times for each word of a sample sentence.

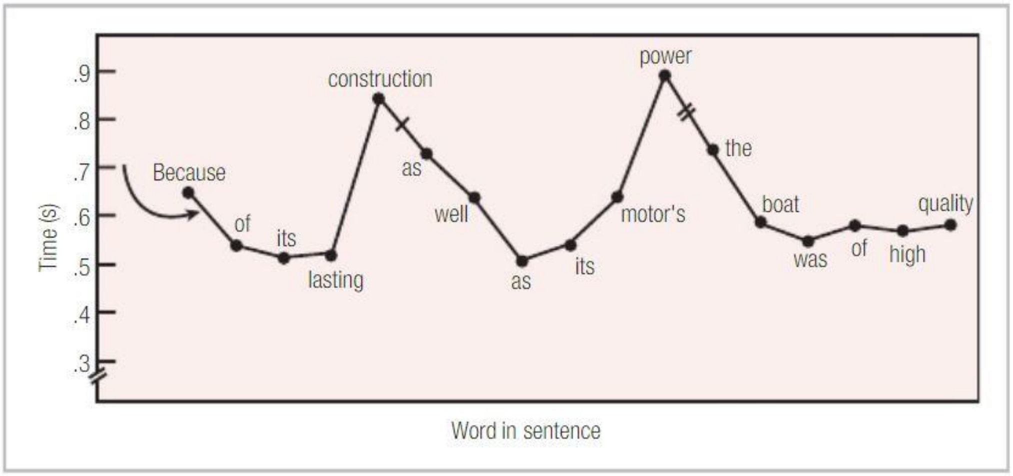

Figure 5: Average word fixation times for a sample sentence
(adapted from Aaronson and Scarborough, 1976).

Even more interesting are, in this context, the experiments examining brain activity during linguistic comprehension. In Chapter 1, we saw that the brain seems to be very sensitive to the degree of predictability of its sensory inputs. In particular, there seems to be a variety of electrophysiological responses that occur in response to novel or unpredicted stimuli, and whose amplitude is inversely correlated with the subjective probability of the stimulus in question (the less probable the stimulus, the larger the response). Now, one of these responses, a relative negativity peaking around 400 msec after stimulus onset (and thus called N400), has been found to occur reliably whenever readers or listeners encounter a semantic anomaly. Specifically, a strong N400 seems to be associated with a written, spoken or signed word that is incongruent with the context of occurrence.[17] Figure 6 below illustrates the phenomenon, showing the ERPs elicited by the last word in "He was stung by a bee/mile". The anomalous word "mile" elicits a

[17] The effect was described for the first time in Kutas and Hillyard (1980). Since then, the literature on the N400 has flourished and this ERP response is now considered a reliable measure of semantic processing. See Kutas and Federmeier (2011) for a review.

strong N400, while the expected word “bee” does not; an anomalous but semantically related word like “hive” elicits an N400 of intermediate amplitude.

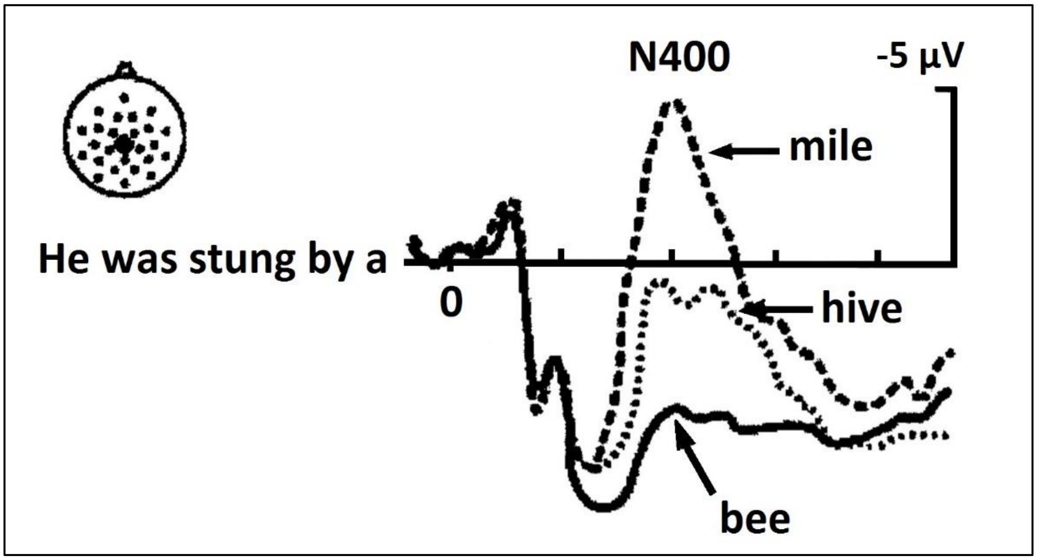

Figure 6: N400 electrophysiological responses to words with different degrees of semantic anomaly (from Kutas and Schmitt, 2003).

The N400 is particularly revealing for the study of the dynamics of linguistic comprehension because its amplitude seems to be exquisitely sensitive to changes in predictability along the verbal chain. With words presented in isolation, N400 amplitude is determined by word frequency (larger for low-frequency words). Within minimal context, such as in a word pair, the N400 to the second word is reduced by repeating the same word exactly or by a word that is semantically related. In a sentence context all words seem to elicit some N400 activity, with amplitude determined by how expected a word is and thus how readily it can be integrated within the current context at a semantic level. Interestingly, within a sentence, with all else held constant, the amplitude of the N400 to any content word tends to become smaller and smaller the further into the sentence the word occurs, probably reflecting the stabilisation of interpretation as one of the interpretive hypotheses becomes more and more probable. By contrast, studies using words in semantically anomalous sentences (meaningless but syntactically correct sentences, such as “Colourless green ideas sleep furiously”) show no such reduction in N400 (see Van Petten and Kutas, 1990 and Van Petten and Kutas, 1991 for details).

Thus, the brain seems to have its own signatures for the dynamics of linguistic meaning-making: the N400 seems to correlate with the probability of the word in question given the current set of interpretive hypotheses of the reader, and its variations in amplitude allow us to observe with some degree of precision the uneven process of belief updating. It should also be noted here that, within the PP framework, ERP responses like the N400 are considered learning signals: they indicate a mismatch between the predicted and the actual stimulus, and signal that the current model of the world embodied by the reader needs some revision.[18] If the revision is successful and the error is explained away, the agent has *learnt*. Linguistic comprehension is then effectively a learning process in the Bayesian/PP sense, and learning takes place with every new revision of the set of interpretive hypotheses of the reader (and the stronger the revision, the larger the learning). More generally, the N400 (as well as other language-related ERP responses) seems to lend further support to the image of the brain as a probabilistic model of the world relentlessly trying to predict the incoming flow of sensory stimulations.

Before we turn to the consequence that this picture of linguistic comprehension has for our understanding of the cognitive value of literary language, a last addition is in order. So far, we have talked only about interpretation advancing incrementally, by means of continuous revisions, within a sentence. But a text is usually made of many sentences, and interpretation usually proceeds beyond sentence boundaries. The same dynamics of Bayesian belief updating are then likely to take place *between sentences*. That is: once the reader's interpretation has stabilised at the end of a sentence, it is likely to be questioned, more or less sharply, by the next one. Suppose for example that the reader has just read "The container held the apples" and has come to imagine something like a basket, with a dozen of green apples in it. Then, she reads one of the following:

(8) It was a basket, with some apples in it.

(9) Ten thousand juicy red apples ready to be turned into jam.

Here, again, the first sentence is more in line with the interpretation that the reader has constructed up to that point; as such, it is likely to promote minimal revisions. (9), instead, questions the interpretation of the reader more sharply, introducing new uncertainty and promoting stronger revisions (for example, the reader would have to abandon the hypothesis that the "container" is something like a basket and start picturing, possibly, something like a big industrial container). Notice that, in this context, (8) does not seem very informative. Since it limits itself to reiterate what the reader has already hypothesised with a certain degree of confidence, it does not seem to add much. Overall, we get the impression of a description that is quite prolix and boringly explicit. (9), instead seems more informative. Overall, we get the impression of a denser description that leaves more for us to infer.

Thus, meaning-making proceeds incrementally not only within sentences, but also between them, with revisions being more or less dramatic depending on how much the new sentence contradicts the reader's current interpretive hypotheses. Certain sentences will follow more naturally from the preceding context, others will be more surprising. It is indeed possible to manipulate the degree of probability of a sentence given the preceding one. This has been done in several psycholinguistics experiments that studied the effect of "causal relatedness" between

[18] See Friston (2005). Explicit interpretations of the N400 in a PP perspective have been suggested, among others, by Kuperberg and Jaeger (2016) and Fitz and Chang (2019). See also Van Petten and Luka (2012) for related discussion.

sentences on comprehension. Classically, subjects are presented with various versions of a two-sentence paragraph, the second sentence of which is kept constant, while the first is manipulated to display a decreasing level of causal relatedness with the second. For example:

> (10a) Joey's big brother punched him again and again. The next day, his body was covered with bruises.
>
> (10b) Racing down the hill, Joey fell off his bike. The next day, his body was covered with bruises.
>
> (10c) Joey's mother became furiously angry with him. The next day, his body was covered with bruises.
>
> (10d) Joey went to a neighbour's house to play. The next day, his body was covered with bruises.[19]

What has been found in these studies is that reading times for the second sentence steadily increase as causal relatedness decreases: it takes more time to read "The next day, his body was covered with bruises" in the context (10d) than in the context of (10a).[20] This is usually attributed to the increasing difficulty to integrate information from the new sentence as causal relatedness decreases. Understanding why Joey's body is covered with bruises in the context of (10a) is straightforward: it requires only the recruitment of the simple piece of real-world knowledge "A punch can produce a bruise". However, when statements are less causally related, more complex inferences are necessary to make sense of them as a whole. In (10d), we need to imagine a more complex story to explain why Joey finds himself covered with bruises after having been in a neighbour's house. This can give the impression of a statement that is richer in implicit meaning and that says much more than what it states explicitly. At the same time, the statement will also be more vague, open-ended and idiosyncratic: each reader is likely to construct a (slightly) different story. In some cases, the sentences might be so unrelated to one another that the reader simply cannot conjure up any viable hypothesis about their relationship. In such cases, unless some more information is provided further on in the text, interpretation remains too open for the reader to understand. There is, then, presumably, an optimum of "causal relatedness" between the sentences (and words) of a text that allow for a maximum of inferential suggestiveness: if a writer manages to create sequences open enough to engage our inferential activity, but not enough to confound us, her prose will seem dense in implicit meanings, agile, and witty.[21] Intriguingly, pairs of sentences with intermediate level of causal relatedness have

[19] The example is taken from Keenan, Baillet and Brown (1984).

[20] Brain activity also seems to vary systematically with causal relatedness. In a study with similar material, Kuperberg Paczynski and Ditman (2011) found that a critical word like "bruises" in the second sentence elicited a smaller N400 if the preceding sentence was explicitly supportive (as in (10a)) than if the preceding sentence was poorly supportive (as in (10d)), even if the sentential context was the same in each case. See also Kuperberg et al. (2006).

[21] A very dubious legend has it that Hemingway was once challenged to write a short story in only six words. Its reply is said to have been: "For sale: baby shoes. Never worn." Whether the anecdote is true or not, this minimal story

also been shown to be recalled best compared to pairs that are either more or less causally related (see Keenan, Baillet and Brown, 1984, and compare with results reported in footnote 10 above).

From what has been said so far, it will be apparent that linguistic comprehension happens between two extremes (with all possible intermediate cases): 1. the new element in the verbal chain (word, sentence) confirms completely the predictions of the reader and, as such, promotes no revision of her interpretive hypotheses (in PP terms, there is no prediction error); 2. the new element in the verbal chain (word, sentence) is completely unpredictable, and, as such, does not allow for any interpretive hypotheses to be stably entertained (in PP terms, there is irreducible prediction error). In both cases, for opposite reasons, the reader makes no progress in her effort to understand the text in question. If one lines up many such completely predictable or unpredictable elements, one obtains a text where nothing adds up, nothing is building up as reading progresses. This means that merely adding elements to a verbal chain is not adding information. To be informative, the new element should be unpredictable enough to demand a revision of our interpretive hypotheses, but not so much as to render any hypothesis impossible. In this latter case, the reader experiences a subjective feeling of progress being made, as an interpretive hypothesis gains in probability over the others, revision after revision. Once again, we come to the conclusion that there ought to be an optimum of unpredictability for linguistic stimuli, and that this optimum maximises learning and cognitive gain.

To some extent, the above might sound obvious. It is obvious, for example, that texts consisting of the same word or sentence repeated indefinitely, or of unrelated words or sentences are not informative. But pointing to such obvious facts gives us hints on the kind of communicative exchanges that we do consider worth having. There seems to be an expectation that a linguistic expression will structure itself in such a way that every new element will modify the interpretation that we have built up to that point. This necessarily requires the succession of elements to be neither repetitive, nor random, neither too predictable, nor too unpredictable. In Chapter 1, we saw that infant and adult learners tend to exhibit such a preference for objects and percepts that, being neither too familiar nor too novel, afford the best chances for learning. Our linguistic behaviour seems to lend further support to the idea that this might be a more general feature of humans as cognitive agents, a feature that defines the kind of percepts we find worth devoting our attention to. Information that we already possess or that does not connect in any plausible way with what we already know, is not worth considering. It does not take beyond what we already have: we do not learn from it.

It should be emphasised now that if the above picture is correct, what "feels" informative and worth processing for the reader has little to do with notions such as truth or falsity. Rather, it is a matter of how the flow of information is organised by the text and perceived by the reader, how often and how strongly the text surprises him, how much it allows for an ever-growing interpretive work. In other words, the reader will have the impression of a meaningful, enriching

seems to owe a good deal of its suggestiveness to a careful manipulation of the level of causal relatedness between its phrases, and analogous devices are likely to be exploited by writers quite frequently.

and worthwhile experience if the text that she is reading manages to consistently promote revisions of her interpretive hypotheses, irrespective of the truth or falsity – or even verifiability – of the propositions in the text. By contrast, a text full of true propositions will not feel informative or enriching if it fails to consistently promote revisions of the interpretive hypotheses of the reader. In talking about the notion of understanding in Chapter 1, we have seen that there are many instances of acquisition of true beliefs that are not lived as a cognitive progress. Now we have the means to explain why. Many true propositions merely state what is already manifest to us. As such, they do not produce a notable revision of our interpretive hypotheses and are not felt as particularly informative. If I tell you that it is raining outside and then I add that "The streets are wet", this second observation might well be true, and you might well come to believe it, but it will hardly come as a revelation. Other true propositions, instead, might be completely unrelated to our ongoing train of thoughts and expectations; if we do not manage to make them relevant and to grasp what bearing they might have on the context at hand, they won't sound as revelatory either. I might well inform you now that "5 March 1895 was a sunny day in Teheran", but you have every right to reply: "So what?". And if we assemble for a reader a collection of such true-but-irrelevant propositions, arranged in an unrelated fashion, and the reader comes to believe every one of them, we might well have a "learning" in the epistemologist's sense, but we won't have an experience that feels like learning, and that anyone would consider as particularly enriching or worth having. Again, to experience a subjective feeling of learning seems not a matter of acquiring true beliefs, but of being exposed to a flow of sensory stimulations that demands (and affords) continuous revisions of our interpretive attempt.[22]

This view goes a long way, I think, in helping us characterise the feeling of cognitive progress that we associate with certain linguistic expressions. The view evidently entails a change in focus away from traditional truth-conditional semantics, insofar as to communicate seems not a matter of conveying true propositions, but to produce revisions in the subject's model of the world. It is encouraging to see, therefore, that in the last few decades a similar change in focus has gained some traction also in the philosophy of language and in pragmatics, yielding views of linguistic communication that emphasise its dynamic and context-sensitive character.

One such view is relevance theory, one of the current dominant approaches in pragmatics and cognitive linguistics. According to relevance theorists, to communicate is to claim someone's attention, and hence to imply that the information communicated is relevant. Information is relevant if it manages to modify the "cognitive environment" of the addressee enough to repay her attention. The cognitive environment of the addressee is the set of assumptions available to her at a given time. As a discourse proceeds, this set of assumptions forms a gradually changing background against which the new information is processed. The new information, if it is to be

[22] Another hint that the brain assesses informativeness in a way that has little to do with truth comes once again from N400 studies: in an experiment using simple categorical statements which could be true or false and affirmative or negative (four combinations), Fischler et al. (1983) found that the amplitude of the N400 elicited by the final word of such sentences depended on the relationship between the subject and the object rather than the truth value of the sentence. Statements such as "A sparrow is (not) a bird" yield a smaller N400 than statements such as "A sparrow is (not) a vehicle", regardless of the overall truth or falsity of the statement.

relevant, should modify the set of assumptions of the addressee in some way, i.e. it should yield some "contextual effects". The act of interpretation, then, crucially involves working out the consequences of adding the new piece of information to the current set of assumptions that we are considering. But not every piece of new information manages to affect us with the same strength. As Sperber and Wilson (1995, p. 48) put it:

> some information is old: it is already present in the individual's representation of the world. Unless it is needed for the performance of a particular cognitive task, and is easier to access from the environment than from memory, such information is not worth processing at all. Other information is not only new but entirely unconnected with anything in the individual's representation of the world. It can only be added to its representation as isolated bits and pieces, and this usually means too much processing cost for too little benefit. Still other information is new but connected with old information. When these interconnected new and old items of information are used together as premises in an inference process, further new information can be derived: information which could not have been inferred without this combination of old and new premises. When the processing of new information gives rises to such a multiplication effect, we call it *relevant*. The greater the multiplication effect, the greater the relevance.[23]

According to Sperber and Wilson, people have intuitions of relevance: they can sense what information elicits more change in their cognitive environment.[24] And they expect their interlocutor to be relevant. They do not expect her to be truthful, however. For relevance theorists, no Gricean maxim of quality governs conversation. The relationship between truth and the interest of the addressee is, at most, mediated by relevance, in the sense that what is true tends to be more relevant.[25] But what matters for us in communication is that what is communicated changes us in some way.

Very similar points are made by a group of so-called "theories of dynamic semantics", which have been developed in the last thirty years in opposition to more static classical truth-conditional semantics. The most influential and representative of these theories is probably Discourse Representation Theory (DTR). Consistently with our Bayesian/PP picture, the tenet underlying DRT is that language interpretation proceeds by incremental updating of an ever-growing discourse model. Every discourse gives rise to a mental representation that is continually revised, as new sentences are incorporated. Each new sentence is interpreted in the light of the existing discourse representation, and at the same time modifies it. The communicative contribution of the sentence is then the change it brings to the current representation. In other words, "sentences have meaning only derivatively, *depending on their capacity to change the*

[23] See also Sperber and Wilson (1995, pp. 118-123).

[24] See Sperber and Wilson (1995, p. 119). This lines up with the hypothesis made in the PP literature that "people have an (imperfect) sense of where predictive progress can be made" (Van de Cruys, Bervoets and Moors, forth.).

[25] See Wilson and Sperber (2012, p. 83): "Yes, hearers expect to be provided with true information. But there is an infinite supply of true information which is not worth attending to. Actual expectations are of relevant information, which (because it is information) is also true".

existing discourse representation" (Brogaard, 2019, p. 382, my emphasis). Suffice it to quote another brief passage from some of the main advocates of this framework to see how much their position is in line with what we have seen so far:

> utterance meaning depends on context. Moreover, the interaction between utterance and context is reciprocal. Each utterance contributes (via the interpretation which it is given) to the context in which it is made. It modifies the context into a new context, in which this contribution is reflected: and it is this new context which then informs the interpretation of whatever utterance comes next.
>
> The focus on context dependence has led to an important shift in paradigm, away from the "classical" conception of formal semantics, which sees semantic theory as primarily concerned with reference and truth and toward a perspective in which the central concept is not that of truth but of information. In this perspective, the meaning of a sentence is not its truth conditions but its "information change potential" – its capacity for modifying given contexts or information states into new ones. (Kamp, Genabith and Reyle, 2011, p. 125)

In sum, we can recognise, beyond the terminological differences, the same recurring idea that what is more important for us as communicative and epistemic agents is not the acquisition of true beliefs, but of stimuli that can change us. As information seekers, we seem not to be geared towards true information, but towards information that has the largest impact (in terms of "belief updating", "contextual effects", "information change potential", etc.) on us. As such, if a text structures its element in a way that promotes continuous revisions of our hypotheses, the experience of reading it will feel enriching and worth having and will be lived as progress irrespective of the truth of its statements. In reaching this conclusion, we have laid out all we need, I think, to understand the cognitive value that we attribute to literary language.

2.4 Insightful Language Reconsidered

Let us wrap up our discussion in this chapter and see whether we managed to capture something of the feeling of insight that literary language seems to provide. We have seen that linguistic interpretation can be conceived in terms of a problem-solving activity in which previously unrelated elements (words, phrases, sentences) become part of a meaningful whole (a Gestalt) where each of them plays the most contextually appropriate role. As with any other kind of problem-solving, this can be more or less straightforward and more or less achievable depending on how difficult it is for the reader to structure the various elements in such a unitary whole. Elements that are either too easy or impossible to organise effectively are less likely to promote problem-solving; elements that present the right level of challenge as to the way they should be organised are more likely to promote problem-solving and the subjective feeling of insight that accompanies it. We then saw that there are at least preliminary reasons to think that successful literary expressions offer experiences of this latter kind.

To further capture the dynamic character of linguistic interpretation, we then reconceived linguistic problem-solving in Bayesian terms. We saw that trying to find the solution to a verbal problem can be seen as updating a set of probabilistic hypotheses about the underlying structure of a linguistic expression. These hypotheses can become more or less probable depending on whether the new elements encountered along the verbal chain confirm or disprove them. We saw that these ups can and downs in uncertainty can be captured with the formal apparatus of Bayesian belief updating. We also saw that they are felt by the reader and have precise behavioural and neurophysiological correlates. Finally, based on this picture, we defined what a linguistic expression ought to look like to maximise its informativeness. We saw that linguistic expressions that are either too predictable or too unpredictable are not informative, because they are redundant and meaningless respectively. It appears then, once more, that the good communicator ought to strike a balance between predictability and unpredictability, closure and openness while structuring her text. If she manages to do so consistently, her text will provide the reader with a constant experience of cognitive gain, quite independently from the truth value of the propositions that are being conveyed. Given this picture, we can return to the question of what is special about literary language and how it might differ from other kinds of linguistic uses (such as ordinary or scientific language) that are normally seen as less striking, enriching and enlightening.

The first thing to point out is that, as we saw, literary language tends to be more ambiguous, and this ambiguity is taken by some to be among its virtues and among the things that set it apart from ordinary or scientific language. From our perspective, ambiguity just means uncertainty over the best solution to the verbal problem (or the hypothesis that best captures the verbal expression’s underlying structure), something which is obtained by including unpredictable elements in the verbal chain.[26] In this sense, ambiguity is the prerequisite for insightful problem-solving and learning in the Bayesian sense: there is no problem-solving if there is no problem to solve, and there is no Bayesian belief updating if no hypothesis is becoming more probable as reading progresses. We can therefore see the reasons of those who value the ambiguity of literary language, insisting that it is not a flaw, but precisely what makes literary language cognitively valuable. The fact that the meaning of a literary expression is not clear or easy to pinpoint is precisely what allows the reader to live an experience of discovery while she works her way inferentially towards it. But again, an expression should be ambiguous up to a certain point. It should not reach the point of incomprehensibility. It should not be just a nonsense. The situation that is conducive to insight and learning is the one in which ambiguity is transient, reducible. We start not quite sure about the meaning of the expression, and then we dispel this uncertainty as reading progresses. This suggests, once more, that there must be a sweet spot of ambiguity that is most conducive to learning, a region that in the case of language is occupied by those expressions that are neither trivial nor random but rather look “like they are going to make

[26] Some other talk of the literary text as being markedly polysemous (see e.g. Groupe µ, 1977). From our perspective, polysemy is just the lexical aspect of ambiguity: it means uncertainty over which value a word ought to take in the context of the verbal problem at hand.

sense" (Cochrane 2021, p. 43).[27] The somewhat paradoxical characterisations of literary language that we discussed in 2.2 seem to point to the fact that literature tends to provide expressions of this kind. If this is the case, then our attraction towards literary language might be just part and parcel of the more general tendency to concentrate on regions of the input space that yield the greatest promise of new learnable regularities.

Another virtue that is frequently attributed to literary language is, as we saw, precision: literary language is said to be better than ordinary or scientific language in capturing the subtle nuances of our experience and our mental life (see footnote 1 above for some such claims). Literary scholar Cleanth Brooks, in commenting on a verse from Keats' famous *Ode on a Grecian Urn* ("Thou, silent form! dost tease us out of thought / As doth eternity") observes that

> the word "tease" is tremendously important. The metaphorical sense is complex. The poet uses the word "tease" to imply an attitude of mischievous mockery on the part of the urn, though he immediately qualifies the quality of this mockery by suggesting that it is of a kind which may be shared by eternity itself; and he qualifies it further by reminding us that it is the kind of mockery which is conveyed not by words but by silence. Keats here, like all other poets, is really building a more precise sort of language than the dictionaries contain, by playing off the connotations and denotations of words against each other so as to make a total statement of a great deal more accuracy than is ordinarily attained... We continually remark that poets are always remaking language. But the point is often forgotten. (Brooks, 1939, p. 16)

On the face of it, the purported precision of literary language seems to be in contrast with its purported ambiguity. How can a linguistic expression be vague and at the same time designate with great accuracy an attitude, feeling or impression? The contrast, however, is only apparent. From our perspective, this precision that an expression might display is just a reflection of the strengths of the contextual adjustments that are needed in order for its terms to coalesce into a unitary whole (i.e., how much ambiguity there is to be reduced). This means that ambiguity is not in contrast with precision but is instead its precondition: an utterance needs to be ambiguous to be seen as precise later on. In other words, that the elements of the utterance do not seem to coalesce is the precondition for discovering how fitting they are once the solution to the verbal problem is found. As we noticed in 2.2, many of our ordinary utterances use words in ways that are close to their dictionary meaning, and, as such, they do not demand from us any strong contextual adjustments. Literary expressions, instead, force us to more marked renegotiations of the value of their terms (they force us, as Brooks puts it, to "play off the connotations and denotations of words against each other"). The result is utterances whose terms seem more

[27] It should also be noted that the experience of something "making sense" in this way can be described in terms of the discovery of "formal perfection" that many see as characteristic of the experience of beautiful objects. See e.g. Cochrane (2021, p. 31): "It is widely acknowledged that beautiful things… display formal perfection. I understand form here in what has been called its 'descriptive' sense… That is, something has form when it has discernible parts that bear some kind of relation to each other (be it conceptual, material or functional). The object then shows formal perfection when these parts relate to each other in a definitely ordered way; a way that makes sense."

precise than usual and denser with subtle implications. In this sense, the author of a successful literary utterance (and her reader with her) is indeed "remaking language", as she is effectively renegotiating (and making her reader renegotiate) the value of words in light of the particular verbal problem that she is setting up (pretty much as the creator of the candle problem renegotiates – and makes the subject renegotiate – the value of the objects involved in the problem). In this sense, the creation and comprehension of a literary expression is also an achievement, just like the solution to a practical problem. Using words the way they are ordinarily used is not much of an achievement: it is like using a box full of tacks as a container for the tacks. But devising new ways in which elements can coalesce into meaningful wholes denotes imaginativeness and ingenuity, and that is an important part of what readers praise in a well-rounded literary expression (and something that a paraphrase cannot convey).

The achievement of remaking ordinary language is carried out, as Brooks aptly notes, in time. Keats first introduces the word "tease" that suggests "an attitude of mischievous mockery", then qualifies this attitude, and then qualifies it further. It is this dynamic aspect of linguistic meaning-making that we really need to consider if we want to understand the power of literary language to enlighten, and the Bayesian/PP apparatus we adopted help us do just that. It allows us to follow with some accuracy the ups and downs of uncertainty that an author sets up along the verbal chain to obtain an experience much more moved, striking and captivating than those normally afforded by less reasoned linguistic exchanges. As we saw, one of the most pervasive features of literary language seems to be the calculated violation of our predictions. When this strategy is employed systematically, the result is a text that promotes continuous revisions of our interpretive hypotheses. In contrast with our mundane verbal productions, where often interpretation stabilises quickly and more and more as the sentence proceeds, literary utterances are often designed to keep interpretation from stabilising and to reopen it continuously. As Iser (1980 [1976], p. 48) puts it: "in everyday speech, there is an increasing degree of redundance as the parts of speech become more and more predictable, but in literary speech the opposite is true." This means that, ideally, every word encountered in a literary text carries some news: it is neither too predictable, nor too unpredictable; in one word: it is relevant (in the sense of Sperber and Wilson, 1995). The reader has therefore the impression that what is reading has a point, that is going somewhere and building up to something. She has the impression of acquiring new information and changing her perspective as reading progresses; and she is, quite literally, changing alongside the text, because her probabilistic model of the world (which, ultimately, is embodied in her brain organisation and dynamics) is being constantly updated. Literary language, in other words, appears to be designed to offer close-to-optimal learning experiences.[28]

[28] This does not mean that a literary text always succeeds in promoting such an experience. But when it fails, it fails for one of the aforementioned reasons: by being either too predictable or too unpredictable to allow for the incremental work of interpretation to take place. Joyce once said of Proust that the reader ends his sentences before him (Joyce, 1961, p. 118). Maybe this is ungenerous in relation to Proust, but we can all think of unimaginative pieces of writing whose language unfolds predictably and without surprise, leading their readers to boredom. On the other hand, many readers of some of Joyce's work find it too chaotic to be enjoyable. Suffice it to recall what C.G. Jung wrote in his review of *Ulysses* (see Deming, 2005, p. 585): "Every sentence raises an expectation which is not fulfilled;

If we now return to our initial example, we should be able to say more as to why Burke's metaphorical utterance about Spain (1) seems more insightful and suggestive than its paraphrase (1a). It is not that (1) is truer than (1a), or that (1) conveys a content that (1a) does not convey; the difference is in the experience that the two sentences promote. The original sentence has a distinctive pattern of tensions and resolutions, openings and closures. "Spain", the first word of the sentence, set some expectations that are refuted soon after by "whale", which comes as a surprise. This causes the interpretive hypotheses of the reader to decrease in probability and ambiguity to rise. The following elements, "…stranded on the shores of Europe", are therefore examined by the reader with particular care, as they are seen as having a role to play in disambiguating the sentence and closing its interpretation. In its conciseness (1) has almost the concentrated dramatic arc of an entire story, with its beginning, its dramatic turn and its final reconciliation.[29] To the extent that a paraphrase – like (1a) – alters this structure of surprises and confirmations, of openings and closures, it alters what is responsible for our feeling of insight. Obviously, however immediate it may seem, (1a) too requires a considerable number of interpretive revisions as reading progresses. But these revisions are less marked and take place very quickly. While we don't expect Spain to be defined as a "whale", we expect Spain to be a "country". So, when the reader encounters the word "country", the occurrence of this term does not question her interpretation as strongly as the occurrence of "whale" in (1) does, but rather confirms and stabilise it. In this way, the attention that the reader must pay to the following words of the sentence decreases. They are no longer useful elements that can inform a tentative hypothesis, but redundant elements that do not add much to what has already been expressed.

The above analysis is no doubt very coarse-grained and imprecise. A more satisfactory analysis would have to capture the process of linguistic comprehension at a finer scale, and hopefully be supplemented by the kind of electrophysiological and behavioural evidence we discussed in 2.3 above, and perhaps also by computational models tracking uncertainty or surprise along the verbal chain (see Frank et al., 2013 for a discussion of some such developing models). Still, it is this kind of analysis, I believe, that we need to carry out if we want to understand the capacity of a text to inform, fascinate and enlighten by means of its language. One could even see in the above a few sketches of a "Bayesian rhetoric" to which cognitive scientists, literary scholars and writers could all contribute with their different insights into the dynamics of meaning-making. As far as our current and much more limited aims are concerned, I hope to have shown at least that, to understand the feeling of insight that we attribute to certain linguistic expressions

finally, out of sheer resignation, you come to expect nothing any longer. Then, bit by bit, again to your horror, it dawns upon you that in all truth you have hit the nail on the head. It is actual fact that nothing happens and nothing comes of it, and yet a secret expectation at war with hopeless resignation drags the reader from page to page… You read and read and read and you pretend to understand what you read. Occasionally you drop through an air pocket into another sentence, but when once the proper degree of resignation has been reached you accustom yourself to anything. So I, too, read to page one hundred and thirty-five with despair in my heart, falling asleep twice on the way… Nothing comes to meet the reader, everything turns away from him, leaving him gaping after it."

[29] See also Brooks (1970 [1949], *passim*, especially Chapter 8), who talks about a "dramatic structure" that poems seem to have and that no paraphrase can capture.

(paradigmatically literary ones), we need to take a close look at what happens when the reader experiences them in their unfolding. This requires a significant shift in the philosophical debate on the cognitive potential of literary language, which has revolved so far around whether there is some special "content" or "meaning" conveyed uniquely by the literary text. In talking about content or meaning, one misses the dynamics of linguistic interpretation, which, in this case, are all important. To examine these dynamics, I relied more on those disciplines that have been more sensitive to the subtleties of the reading process: literary theory on the one hand and the psychology and neuroscience of reading on the other hand. This examination has allowed us to recognise that the literary text provides “insight” and “learning” in a sense that is coherent with what we know about the psychology and neuroscience of insight and learning, even if this sense might be in tension with the epistemically-charged understanding of learning current in philosophy. The philosophical consequences of this fact are, however, considerable: we shall elaborate on these in Chapter 5. Now, however, we need to say more about how a literary text may give us the impression of insight, extending our analysis from individual linguistic expressions to the entire plot of a literary work.

3. Learning from Narratives

3.1 Narrative Understanding

In the previous chapter, we looked at the ways in which a literary work can generate an experience of cognitive gain by means of its language. We tried to shed light on the minute, fast-paced dynamics involved in the interpretation of a sentence or in the shift from a sentence to the following one, and we discovered how a text can prove cognitively rewarding even in the short space of a few surprising words. But literary works surely are more than a collection of striking sentences. When we read a novel, a theatrical piece, a short story or a tale, we have first and foremost the impression of following a *narrative*, a sequence of actions or events carried out by or happening to characters, people with goals and motivations. It is time, then, to adopt a more comprehensive view, examining how literary texts engage their readers not only word by word, but event by event, along their whole narrative progression. In this chapter, we will try to clarify how and why a literary work can convey a feeling of cognitive gain by means of its narrative.

That narratives, *qua* narratives, convey some sort of insight or understanding is almost common sense.[1] It is often remarked that stories do not merely present events, but organise them in a way that makes them intelligible. In a story, we are told, events do not merely follow one another but explain one another, finding their place in a unitary whole. The events so organised, then, seem to disclose some overarching significance or sense, a "moral" or "message" that the reader can grasp. This renders narratives perhaps unavoidably tendentious, but justifies at the same time the enormous fortune they have always encountered, in every human culture, as instruments of teaching and learning. From etiological myths to bedtime stories, from religious parables to the daily news, narratives are a privileged way in which human communities come to articulate, negotiate and transmit their understanding of the world, their values and identities, the sense of their past and their present. Their function in this regard is indisputable.

Thus, when the philosopher asks whether we learn from narratives, her question may sound rather idle. Of course we learn from narratives! The whole history of human culture bears witness to that! The fact is that we must distinguish, once more, the epistemic sense of learning, with its interest in truth, from the kind of learning that goes on when we read a story and gain an understanding of its events and their implications. As I have argued in Chapter 1, whether we learn from narratives in the epistemic sense will have to be ascertained empirically, on a case-by-case basis, in an inquiry that is not likely to yield any general answer. If you want to know

[1] See for example Velleman (2003, p. 18): "the understanding provided by narrative should be attributable to the nature of narrative itself – to that in virtue of which a recounting of events qualifies as a story". Brooks (1984, p. 10) similarly notes: "Plot, let us say in preliminary definition, is the logic and dynamic of narrative, and narrative itself a form of understanding and explanation." Both Velleman and Brooks trace this idea back to the Aristotelian notion of *mythos* (see *Poetics*, *passim*).

whether someone has learned something from reading *Macbeth* or *Anna Karenina*, go check out whether the cognitive change brought about by this specific work in this specific reader does more good or harm in specific situations. This might be an interesting inquiry in its own right, but in pursuing it we probably won't even begin to touch on the question of why narratives make us feel like we are learning something, and what is this kind of learning that *all* narratives, *qua* narratives, seem to provide.

It is these latter questions that we will try to illuminate here. It will be a matter of finding out what features of a narrative text are responsible for its capacity to convey a feeling of having learnt. And this latter issue is far less clear, and much more debated. Even a quick survey of the recent philosophical literature on the topic reveals that, beyond the almost general consensus that narratives convey understanding, there is large disagreement on how they manage to do so. It has been proposed that narratives give some kind of (illusory) explanation (Velleman, 2003), that they pose and answer questions (Carroll, 2007), or that they contain implicit arguments (Schultz, 1979; Plumer, 2015; Olmos, 2017). As we shall see, there is something true in all these formulations; but to see this, we will have to delve into the details of how narratives organise their material and structure the experience of their readers. In doing so, we shall notice that what we say about narratives in this chapter mirrors closely what we said about sentences in the previous one: the same underlying dynamics will be found to be active, at different timescales, in both cases. If there we saw that a sentence can be conceived as a little story, with its own compressed dramatic arc, here we will see that a narrative is similar, in many respects, to a long and complex sentence. The conclusions that we will reach will hopefully do justice to both our common-sense intuition that we learn from narratives and what psychology and neuroscience tell us about learning. They will also help us add another piece to the Bayesian/PP story about art and learning that we are progressively articulating, further establishing its explanatory reach and potential.

3.2 What is a Narrative?

If we want to understand how narratives convey understanding, we had better understand what a narrative is. Indeed, if narratives *qua* narratives convey understanding as it is often claimed, one should expect the answer to the question "What is a narrative?" to be also an answer to the question "How do narratives convey understanding?". We must then ask ourselves what are the conditions that a text must fulfil to be identified as a narrative (or "story", or "plot": I will use these terms interchangeably here).

The literature on the definition of narrative and the conditions for narrativity spans from Aristotle's *Poetics* to a very lively contemporary debate, and is so vast and varied to be practically unsurveyable.[2] There is however one point on which there seems to be general agreement: in

[2] But see Ryan (2007) for a useful summary of recent approaches.

order for a text to qualify as a narrative, there must be among the events that it recounts some kind of connection, some sort of relationship that the reader can grasp.[3] E. M. Forster (2022 [1927], p. 52) famously remarked that "The king died and then the queen died" is not a plot, whereas "The king died and then the queen died of grief" is, the reason being that the first sentence merely reports two events which could be completely unrelated, whereas the second makes it clear that there is a causal relation between them. But, as we saw in the previous chapter, when a reader is confronted with two sentences in succession, she tries her best to infer the relation between them, especially when that relation is not already clear. So even if "The king died and then the queen died" does not specify any connection between the two events, the reader is likely to infer that the two deaths are connected (and even if she does not, in most cases she will at least assume that the "queen" the text is talking about is the dead king's wife, and not some other unrelated queen from another unrelated kingdom).[4] Readers, it seems, are always trying their best to see patterns in events, and that they be able to see them seems to be one of the necessary conditions for them to perceive something as narrative.

But it is of course possible to create texts that report events so unrelated that the reader is simply unable to establish any kind of underlying pattern. Lotman (1990, p. 223) exemplifies this condition with a passage from Evgeny Shvarts' play *The Dragon*, where a character is pretending to be schizophrenic (but any other sequence of unrelated sentences will do):

> (1) Millers have just got in a new delivery of cheese. Nothing goes better on a girl than modesty and a dress you can see through. At sunset wild ducks flew over the cradle. They are waiting for you at the council meeting, Sir Lancelot.

This will not appear as a narrative to most readers. It seems just a collection of random considerations, none of which justifies or explains the others in any way. So one must preliminarily agree at least on this point: in order to have a narrative we need to have a group of events bearing some relationship to one another for some reader. Notice also that this requisite for narrativity is a requisite for comprehension as such. In other words, it is not just that if we do not grasp the relationships between the events reported in the text, we are not inclined to consider it a narrative: it is also that we do not understand what, if anything, the text wants to say.

So we seem to have found a necessary condition for having a narrative. Is it also a sufficient one? It seems that it is not. Eco (1979a) points out that not just any sequence of related events might deserve the name of narrative. A narrative, he says, must contain events that are not only

[3] On the nature of this relationship there is, however, an open debate. Carroll (2001, 2007), for example, defends the view that they are causal relations. For Velleman (2003) instead, the events of a story need not be related causally but should articulate a kind of "emotional cadence". See also Currie (2006) and Feagin (2007) for discussion.

[4] Are these relationships in the text or in the reader? For example, is the relationship between the king dying and the queen dying in the text itself, or is it imposed, imagined or hallucinated by the reader? This is, I believe, the textual version of the problem of induction, and, as with its more general counterpart, not much progress has been done on it since Hume. I'll let the reader decide if there are narratives out there or if they are only in our minds, or even if the question is unanswerable in principle. For more on this point, however, see Chapter 5, section 5.3 below.

related to one another, but also unexpected or non-obvious in some regard. It could be that the actions of a character are difficult to accomplish, or that she faces problems or moral dilemmas with no apparent solution. Whatever the source of difficulty, it seems that for a sequence of events to have a narrative feel there must be in it some non-obviousness and uncertainty, some indication that the events could take or have taken another turn from the one that they eventually took. This is why, in Eco's view, one should not regard as narrative a text like the following (Eco, 1979a, p. 108, my translation):

> (2) Yesterday I left home to take the 8.30 train to Turin. I took a taxi to the station, then I bought a ticket and went to the right platform; at 8.20 I got on the train, which left on time and took me to Turin.

Confronted with a text like this, says Eco, we are left wondering why the narrator dwelled at length on such an uninformative episode. It seems that there is very little in this sequence of events that the reader might not have guessed by herself or that makes the whole episode narratively interesting and worth sharing. This might be true, but the text (just like Forster's little parable about the king and the queen) still retains a certain narrative feel. After all this little trip could well have gone otherwise, and sometimes to catch a train on time and to reach one's destination is not such an obvious deed, as anyone that has ever travelled by train in Italy can testify. But what about texts like the following:

> (3) Every day, peasant Ivan would work in the fields all morning, and then tend his master's flock in the afternoon. That day, Ivan worked in the fields all morning, and then tended his master's flock in the afternoon. The next day, Ivan worked in the fields all morning, and then tended his master's flock in the afternoon. The day after that, Ivan worked…

And so on, indefinitely. Is this a narrative? In this case, we are even less inclined to say so. To be sure, there is an orderly series of events, here. Yet we do not feel like anything is really happening or progressing, because the events follow one another along the same, predetermined pattern. Although there are events following one another, the text seems to portray a state, not a progression. At most, (3) sounds like the setting of the stage for a narrative that is yet to start, and that will really begin only when Ivan does something different, whether deciding to abandon his rural life or starting to conceive a plan to murder his master. Narrative seems to be about variation, change, openness to many possible outcomes, and not about what is certain and unfolds deterministically given certain premises.

It seems therefore that to have a narrative consequentiality is not enough. The events must indeed be related, but in a non-obvious way, a way that we cannot predict in advance and that we need to discover progressively while reading. This seems to place upon narratives the two somewhat contradictory demands of surprisingness and consequentiality that Aristotle had already placed upon the plot of the tragedy in his *Poetics*:

> Tragedy, however, is an imitation not only of a complete action, but also of events arousing pity and fear. Such events have the very greatest effect on the mind when they occur *unexpectedly and at the same time in consequence of one another;* there is more of the marvellous in them then than if they happened of themselves or by mere chance. (1452a 1-6, my emphasis).

What Aristotle says of tragedy might well be said, I think, of narratives in general. There seems to be a demand on the part of readers and spectators that the facts of a narrative follow one another unexpectedly (παρὰ τὴν δόξαν, or "contrary to the common opinion", as Aristotle puts it) but at the same time in a highly consequential way (κατὰ τὸ εἰκὸς, or "according to what is probable"). If by contrast the facts follow one another by a logic that is either self-evident or impossible to grasp, we simply do not experience that movement forward that we attribute to narratives. We might have an accumulation of facts, like in (1) or (3) above, but not a story that progresses.

If this is the case, however, we might just have reached an explanation of how narrative and understanding are interrelated, an explanation that is perfectly in line with the broader story we have been articulating so far. We have seen that learning and understanding might be conceived in terms of insightful problem-solving: a process by means of which previously unrelated elements (the objects in the candle problem, the black and white patches in the image of the Dalmatian, the words in a sentence, etc.) find their place in a unitary whole. Insight will be maximal, we said, when such integration of elements is not immediate (i.e., there is a problem to solve) but still possible (i.e., there is a solution to the problem). Now, the same line of reasoning can be applied to narratives, with the difference that this time the elements that need to find a place in a unitary whole are higher-order units that we might call "facts" (e.g., the fact that the queen died, or that Millers have just got in a new delivery of cheese, or that Ivan is tending his master's flock). Integrating facts in a unitary whole can be more or less easy or feasible depending on the sequence of facts in question. In (1) for example such integration is impossible, whereas in (3) it is very easy, and becomes increasingly easy every time Ivan starts a new day of work. If we do not have a narrative in either of the two cases, then quite evidently the conditions for narrativity are also the conditions for insightful problem-solving. In particular, the condition that there be a relation between the facts of the narrative corresponds to the condition that there be a solution to the problem, and the condition that this relation be non-obvious corresponds to the condition that there be a problem to solve. This means, then, that good narratives promote insightful problem-solving and the pleasant feeling of discovery associated with it.

As we know by now, the same process could be given a probabilistic/Bayesian interpretation. Finding the solution to the narrative problem might be conceived as the process by which a hypothesis about the underlying causal structure of the narrative (i.e., about how its facts are related to one another) becomes more and more probable. In fact, there is perhaps no other way to make sense of Aristotle's point mentioned above. Since a sequence of facts cannot be at the same time highly improbable and highly probable, Aristotle's point must amount to the recommendation that there be in the course of a good narrative an appreciable shift from improbable to probable, as one of the hypotheses about the causal structure of the narrative becomes more and more likely. In other words, the uncertainty about the causal structure of the narrative ought to decrease as the narrative progresses. The problem with (1) and (3) is then that they afford such a decrease in uncertainty (or increase in probability) only to a very limited degree. In (1) the underlying causal structure remains unclear, and no hypothesis about it becomes more probable; in (3) the underlying causal structure is already clear, and our hypothesis about it is already held with a high probability. This means that in both cases we are not learning any new causal pattern. Good narratives, on the other hand, would be those that allow us to learn new causal patterns in an optimal fashion (or, which is the same, we would perceive as more narrative the material that allows us to learn new causal patterns in an optimal fashion).[5]

In sum, whether we characterise narratives by means of the Gestaltic notion of problem-solving or in Bayesian terms, a hypothesis becomes available that is in line with our story so far: narratives might be devices that maximise learning. They might do so by offering sequences of facts whose underlying causal structure is not clear but is still discoverable (that is, sequences of facts that look "like they are going to make sense"; see again Cochrane 2021, p. 43).[6] If this is true, then our attraction towards narratives can be seen as part of the more general human attraction towards "progress niches", regions of the input space that allow for optimal assimilation of new causal patterns. In turn, our dislike for "bad" or imperfect" narratives like (1), (2), and (3) above would be part of our general dislike for regions of the input space that do not promise to disclose new causal patterns, either because their pattern is already clear or because they do not seem to have one.

What we have said so far should have made it clear, then, that narrative and understanding are so intimately related to be indistinguishable. The conditions for having a narrative are also the condition for understanding the facts of the narrative. There is narrative progression in a sequence of facts only if we as readers are able to work our way to the causal structure that underlies them. Indeed, it could be said that a narrative is not a thing, a collection of facts with such and such properties and relations, but the process we engage in when we understand those facts, integrating them in a unitary whole. If this is true, narrative dynamics – the ways in which

[5] I say "more narrative" because the story on offer here suggests that we should regard narrative as a gradational rather than categorical notion. In other words, something can be more or less narrative (or, if you prefer, higher or lower in narrativity) depending on how much it promotes insightful problem-solving (or learning in the Bayesian sense). Moreover, since each subject will have her own probabilistic model of the world, one should expect that what appears to be high in narrativity to a subject might not appear in the same way to another subject, or even to the same subject upon a second encounter with the same sequence of facts. For a similarly gradational notion of narrative, see Currie (2006).

[6] See also Mink (1970, p. 545): "in following a story, as in being a spectator at a [cricket] match, there must be a quickly established sense of a promised although unpredictable outcome: the county team will win, lose, or draw, the separated lovers will be reunited or will not. Surprises and contingencies are the stuff of stories, as of games, yet by virtue of the promised yet open outcome we are enabled to follow a series of events across their contingent relations and to understand them as leading to an as yet unrevealed conclusion without however necessitating that conclusion." Mink is here summarising the view put forth by Gallie (1964).

a story accelerates or slows down, comes to a close or opens up again – are really the dynamics of our act of comprehension.

Once we grasp the fundamental relationship between narrative and understanding, we can begin to notice that many of the principles for creating a "good story" are also principles for enhancing understanding. This is true not just of Aristotle's suggestion discussed above: on close examination, many of the rules and techniques that have been theorised, discussed and stably applied throughout the history and practice of (literary) storytelling reveal themselves as ways to preserve, promote and maximise the reader's sense of cognitive progress. We will discuss some of these rules and techniques in the next section. For now, let us just mention another of Aristotle's narratological principles that came out of the *Poetics* and became very influential: the principle of the unities of time, place and action. This is the principle according to which a story should ideally deal with only one complete action carried forward by the same character without discontinuities in time or space. These prescriptions are clearly designed to spare us a nuisance that is aesthetic and cognitive at the same time. To see this, it suffices to consider cases in which they are not respected. Propp (1968 [1928]) notices that if we are reading a tale where the hero leaves home in search of a horse and he returns with a princess, the tale seems unsatisfactory. It is as if the narrator lost track, while narrating, of the point she was trying to make. Similarly, Carroll (2007) notices that if a character that seemed important at the beginning of the story gets dropped in favour of some other character, or if a subplot is set in motion but then dropped without explanation, we may feel "a kind of intellectual discomfort" (p. 6), "a species of dissatisfaction" (p.7). In light of what we have said so far, this dissatisfaction is perfectly understandable. If a narrative is such only to the extent that it leads to the establishment of a viable hypothesis about the facts' underlying causal structure, then narratives of the kind that Propp and Carroll describe are imperfect, since they do not allow for such a hypothesis to be established (it is difficult to say what keeps together causally the search for a horse and the return with a princess, or two unrelated subplots). It is interesting to notice once more that defects in a narrative are also, at the same time, things that get in the way of understanding. This further suggests that the principle of the unity of time, space and action, as many other principles of narrative organisation, is not merely a matter of good taste, but reflects some underlying cognitive imperative.[7] After all, an author of fiction could well organise her material in the wildest ways, disregarding any rule. But there are certain limitations that authors almost invariably

[7] Which is also why certain prescriptions about the way of crafting a good story appear to be surprisingly stable across times and cultures. The reader might find it instructive to compare, for example, the recommendations contained in Aristotle's *Poetics* and the rules for writing a good mystery story suggested some twenty-two centuries later by classic mystery writers such as Ronald Knox, S. S. Van Dine, Gilbert K. Chesterton, Dorothy Leigh Sayers and Raymond Chandler. The similarities are striking. Chesterton even wrote a *Defence of Dramatic Unities* (1923), where he recommended the unities of time, place and action. Sayers, on her part, made it clear in a conference significantly entitled *Aristotle on Detective Fiction* that Aristotle managed to conceive "a theory of detective fiction so shrewd, all-embracing and practical that the *Poetics* remains the finest guide to the writing of such fiction that could be put, at this day, into the hands of an aspiring author", and that "any writer who tries to make a detective story a work of art at all will do well if he writes it in such a way that Aristotle could have enjoyed and approved it" (1936, pp. 23-24, 35).

impose themselves, certain combinations that are preferred over others. And these preferences are quite often revelatory. Apparently marginal aesthetic discussions such that on how plots are structured, might therefore acquire a more general significance and help us shed light on our human cognitive needs and capacities. Let us pursue this line of thought further.

3.3 Narrative Dynamics

We have seen therefore in what sense narratives provide learning experiences: they offer us the possibility to grasp new causal patterns in the facts that they present. The idea is intuitive enough, and allows us to capture the often-stated idea that a narrative does not merely present facts, but organise them in a way that makes them intelligible. But if we give this idea a Bayesian/PP formulation, we can say a lot more about how narratives generate these learning experiences. In particular, we can follow in some detail the dynamic process by which the reader tries to guess the causal structure of the narrative.

Here we can put to work the same apparatus we used in the last chapter with sentences. We saw there that interpretation of a sentence proceeds incrementally as reading progresses and new elements (words, in that case) are encountered along the verbal chain. As we saw, this process might be captured by supposing that at any given time the reader entertains various hypotheses about the underlying structure of the sentence (i.e., about how the words of the sentence are related to one another). Each of these hypotheses gains or loses in probability with each new element encountered, in a process that might be cast in terms of Bayesian belief updating. This produces characteristic ups and downs in uncertainty that, as we saw, have psychological, behavioural and physiological correlates. If all goes well, after a few cycles of belief updating a hypothesis gains in probability over the others, and the reader experiences a subjective feeling of cognitive gain as she becomes increasingly confident about the underlying structure of the sentence. Now, the same can be said just as aptly about narratives. The only difference being, again, that the new elements encountered as reading progresses this time are not words, but facts. Here, again, one might suppose that at any given time the reader entertains various hypotheses about the causal structure of the narrative (i.e., about how the facts of the narrative are related to one another). These hypotheses gain or lose in probability as reading progresses and new facts are encountered, until hopefully one of them wins over the others and the reader becomes increasingly confident about the causal structure of the narrative. When this happens, the facts of the narrative begin to “make sense”, that is, to fit into a unitary whole. This means, however, that the reader will feel that she is progressing in her understanding only if a hypothesis about the causal structure of the narrative is becoming more probable (or, which is the same, if the uncertainty over the causal structure of the narrative is being reduced). If the new facts encountered along the sequence just confirm a hypothesis that is already very probable or if they do not allow any hypothesis to be stably entertained, then the reader will not have the impression of gaining an understanding of the facts in question. She will have facts accumulating without a narrative progressing.

Maybe an example will help clarify what I mean. Say that we are reading a story about Joe, head of a big company based in London. That day, Joe decides to have lunch at a restaurant near his office. At lunchtime he leaves work, goes to the restaurant, and orders something; after a while, the meal arrives at his table; he eats it and then asks for the bill; when the waiter comes with the bill, Joe gives a look at it, pays and leaves the place.

We can notice that this little narrative looks and feels very much like (2) or (3) above: consequential, but rather uninformative. From our perspective, this is because the reader has already a clear hypothesis about how the facts of this sequence should be organised, a hypothesis that the sequence does very little to put into question. Using two terms that narratology has been borrowing from cognitive science for some time now, we might say that the facts in this little sequence are all part of a "frame" or "script" that the average reader knows very well, a frame or script that we might call "lunch at the restaurant" and that involves precisely the kind of actors and actions that Joe's little sequence portrays.[8] We all have had experience of restaurants: we know that first you enter the place and sit at a table, then you order some food, then the food is brought to you, and so on. Since Joe's sequence adheres strictly to this script, it does very little to challenge our evolving understanding of the narrative situation. Each new fact of the sequence follows naturally from the previous ones, according to a logic that is already well understood. As such, the question of the significance of each fact, of the place that it should take in the general picture that we are articulating, does not have the time to arise. We do not wonder why, say, Joe pays the bill, because we do not need to find an underlying causal structure that accounts for this fact: we know it already. As a result, we do not seem to gain from reading Joe's sequence any deep understanding of the facts mentioned in it. They remain brute facts, dumb and self-evident like all the facts whose significance we no longer perceive for having encountered them all too often.

But now imagine that, when the waiter comes with the bill, Joe gives a look at it and then violently slaps the waiter in the face. This fact surely challenges our evolving understanding of the narrative situation! After encountering it, we no longer know how all the facts of the sequence fit together. As such, the question of their significance, of the place they should take in the general picture we are articulating, has a way of arising explicitly. We do wonder why Joe slapped the waiter, because we do need to find an underlying causal structure that accounts for this fact. But if we do not find a such causal structure, the fact remains unexplained. We will have to accept as just a brute fact that Joe slapped the waiter. This time, however, the fact is "brute"

[8] For the notion of frame in cognitive science, the *locus classicus* is Minsky (1975). For a classic application of these notions to the interpretation of narrative texts, see Eco (1979a), pp. 79-84. An interesting point made by Eco is that narrative texts in most cases do not mobilise only "common scripts" (i.e., scripts that the reader has acquired from her experience with the world, such as "lunch at the restaurant"), but also "intertextual scripts" (i.e., scripts that the reader has acquired from her experience with other texts). These include genre plots ("the mystery story"), as well as what literary criticism has traditionally called "motives" or *topoi* ("the rescued princess"), and, more generally, every kind of codified, stereotypical narrative situation ("the western gun duel").

not because it is self-evident, but for the very opposite reason that it is obscure. If the fact that Joe pays the bill fits into a pattern all too clearly, the fact that Joe slaps the waiter does not seem to fit into a pattern at all. The result, however, is the same: we do not get the impression of having gained a deep understanding of the facts of the sequence. If the story goes on presenting many other such unexplainable facts, it will fall apart as a narrative.

But now suppose that the narrator, after having told us that Joe slaps the waiter, reveals that there was a very offensive, very personal insult written in pen on the bill. This new fact will indeed come as informative. The reader has now the opportunity to hypothesise a new causal structure underlying the whole sequence ("the waiter wrote the insult on the bill, Joe read the insult, Joe slapped the waiter"). As this new hypothesis becomes more probable, the reader gets the impression of having understood the facts of the sequence: each fact now fits a newly discovered unitary whole. Of course, this does not solve all the difficulties. There is now the new problem of understanding why the waiter wrote the insult on the bill. As we shall see, stories are often made in such a way that finding a new causal pattern opens the problem of how this pattern should be integrated at a higher, more comprehensive level. The narrator might then inform us, for example, that the waiter was that employee that Joe fired unjustly some months before. This will enable us to hypothesise a new and more comprehensive causal structure ("Joe fired the now-waiter, the waiter was angry with Joe, the waiter wrote the insult on the bill, Joe read the insult, Joe slapped the waiter"), which in turn will pose new problems of integration (has Joe recognised the man? How will the man react to the slap?). If the narrator manages to carry on in this fashion, we as readers will keep finding new causal structures, and we will therefore have a sequence of facts that do not merely follow one another but explain one another, advancing our understanding. We will have a narrative.[9]

Our discussion so far should have suggested that the Bayesian/PP story we are articulating might have a lot to say about the affective phenomenology involved in following a narrative, as well as the ways in which effective narratives are structured. To the extent that the Bayesian/PP apparatus can capture the dynamics of belief updating, it opens a window on the experience of tension, release, suspense and closure that narratives offer, and on how narratives structure their material to promote such experiences. Let me reiterate and expand upon what we have said so far to make this explicit. When following a narrative, the reader entertains and constantly updates hypotheses about the underlying causal structure of the narrative (that is, how the facts of the narrative "make sense", fit in a coherent interpretive whole). Based on her current hypotheses, she comes to expect certain things more than others. If the reader is told that Joe goes to the restaurant, eats his lunch and gets his bill, she will expect Joe to pay the bill more than she expects Joe to slap the waiter. In other words, it is just part and parcel of the interpretive process that the reader constantly entertains predictions about what will happen next. The new

[9] This is consistent with both Carroll's (2007) proposal that narratives raise and answer questions, and Velleman's (2003) proposal that narratives convey understanding by providing explanations. As both Carroll and Velleman notice (quoting Forster, 2022 [1927]), "if it is in the narrative, we ask why". But in order for such a question to arise at all, the answer must not be self-evident. And in order for the narrative to provide explanations, an answer should be given. If there is either no question or no answer, quite evidently, there cannot be any insight.

fact in the sequence can confirm or violate these predictions, and can do so more or less strongly. If the new fact confirms the predictions, the reader's grasp on the causal structure of the narrative is strengthened, and the narrative appears to "stabilise" or "close", at least momentarily.[10] On the other hand, if the new fact disproves the predictions, the reader's grasp on the causal structure of the narrative is put into question and the narrative "reopens" (a "tension" has been introduced that demands a resolution). In general, if the narrator does not want to lose her readers along the way, she must try to keep this predictive game alive throughout the work, giving the reader the impression that there are still causal relationships among the facts of the narrative that need to (and can) be grasped. If the narrator closes the narrative too much, with every new fact confirming ever more the pre-existing picture, then the reader is left with little or nothing to guess; the text will soon become boring and uninformative as in (2) or (3) above. On the other hand, if the narrative opens too much, with every new fact undermining ever more the reader's interpretive hypotheses, then the reader is soon left with no hypothesis to test; the text becomes too ambiguous to mean anything, as in (1) above, and equally uninformative and boring to the reader. To keep the reader glued to the narrative and hungry for more, to keep providing her with a constant feeling of advancement and movement forward, the narrator must then carefully balance moments in which predictions are confirmed and moments in which they are violated, openings and closings, tensions and releases.

The interpretive process is therefore uneven and "bumpy" by necessity, with certain new elements that stabilise interpretation more or less strongly and certain others that reopen it. In the previous chapter we saw that this unevenness is responsible, at the level of the sentence, for what we called "semantic rhythm". We saw that the reader can feel whether the interpretation of a sentence is stabilising or opening, and that these feelings have physiological and behavioural correlates. We also saw how a writer might exploit these dynamics by carefully shaping her sentences. Now we should add that the same unevenness of the interpretive process is responsible, at the level of the story, for the story's *narrative rhythm*. A story might give the impression of accelerating, slowing down, lingering, changing abruptly, heading towards a close, and so on. All these effects are a reflection of how the story plays with our expectations, regimenting our interpretive act. The expert storyteller manipulates these dynamics. By careful dosing of elements that are now more, now less surprising, she gives her story a certain rhythmic profile that the reader might feel in terms of tensions and releases, openings and closings.

It is in the possibilities offered by these manipulations that lies all the complexity, variety and charm of narrative constructions. For, even if every story, if it is to be a story, must offer us the experience of cognitive progress that we have been describing, every story will realise this progress in its own way – the ways in which a story can play with our expectation being virtually infinite. If every story, *qua* story, ultimately enacts the successful struggle of the mind of the reader in coping with the text, each story offers its own particular struggle. Let us see then what

[10] As in the previous chapter, the Gestaltic jargon of "stabilisation" and "closure" comes in handy here. As we shall see, it is also remarkably in agreement with the proposals of those who speak, like Eco (1979b) or Carroll (2007), of "closed works", or "narrative closure".

are some of the ways in which the narrator might articulate that struggle, starting with a few techniques that are fairly general and applied across the board and moving then to others that are typical of certain particular narrative genres.

One of the main ways in which narrators manipulate narrative dynamics to maintain our interest alive is by presenting the facts in a different order from the one in which they happened (fictionally or not). In other words, in any but the simplest stories there is a difference between what narratologists call "fabula" (the facts of the story as they happened, in their chronological succession) and the "syuzhet" (the facts of the story as they are narrated).[11] Facts that happened earlier in the story might be revealed to the reader only later on and vice versa, originating temporal alterations that are quite often rather complex, and whose most familiar manifestations are flashbacks and flashforwards. Now, if one reflects on the logic underlying these chronological alterations, it will be clear that in almost all cases they are made in order to promote and enhance the reader's understanding. Facts that, if presented in their normal order, would make for a rather dull or incomprehensible story are altered to yield a succession that has the right kind of tensions and resolutions to make us feel that we are making cognitive progress.

Take our previous example of Joe and the waiter. The facts of that little story in their chronological order are: Joe fires an employee unjustly; the employee finds a job as a waiter in a nearby restaurant; Joe goes to that restaurant to have lunch; at the end of the lunch, Joe asks for the bill; the waiter writes an insult on the bill and gives it to Joe; Joe slaps the waiter. Recounted in this order, the facts are much more predictable and consequential, the narrative much less eventful. But the same facts, if told in the order in which we encountered them the first time, acquire a whole other poignancy. If we are told first that Joe fires an employee unjustly, then that, after some time, he goes to a restaurant, eats, asks for the bill, and then slaps the waiter, this latter fact comes much more as a surprise, because the narrator has omitted the antecedents that would have made it predictable; the fact now creates an explanatory tension that asks for a resolution; if then the author reveals that there is an insult written on the bill, this tension is partially resolved, but a new tension arises, because we do not know how to interpret the fact that there was an insult on the bill; if then the narrator reveals that the waiter was the man that Joe fired moths before, then not only is this second tension resolved, but a fact happened months before is also charged retrospectively with a new significance. The reader gets the impression that every event has found its place.

Thus, by simply altering the order of the facts, we have enhanced the reader's understanding and we have produced a story that is much more dynamic. Every skilful narrator has an implicit knowledge of these dynamics. She knows what order will be more effective, and what will spoil the fun. Many stories owe to such temporal manipulations a great deal of their fascination, but certain stories are even unthinkable without them. Mystery stories are a case in point: in most cases, the first event that happened chronologically (the culprit) is the last to be reported in the story; doing otherwise would spoil the mystery altogether. But the same is true, for example, of

[11] The distinction is due to the Russian formalists, and particularly to Tomashevsky (1999 [1925]). For a classic systematization of these notions, see Genette (1972).

many ancient Greek dramas that end with what Aristotle in the *Poetics* calls "recognition" (ἀναγνώρισις), as well as of the majority of stories based on a *coup de théatre*.

Another pervasive way in which a narrator displays her awareness and control of narrative dynamics is by playing in the difference in duration between the time of the fabula and the time of the syuzhet. If the events (fictional or not) of the story have a certain duration, the story has the liberty of narrating them at its own pace: events that in the world of the story took years to unfold might be narrated in a few lines or omitted altogether; events that in the world of the story took a few seconds might be described in pages and pages.[12] Virtually every narrative contains differences between the duration of the events and the duration of the story. But, again, it is interesting to notice what is the logic that seems to govern these alterations. It seems that in almost every case where there is a difference in duration between the events and the story, this difference goes in the direction of creating a more eventful and dynamic story; a story, that is, that can maximise the reader's sense of cognitive gain. Thus, facts that follow one another in a predictable fashion and without surprises, such as those in (2) or (3) above, are normally summed up or omitted altogether. The narrator "cuts to the chase", comes directly to the point of narrative interest: when the speaker in (2) has already arrived in Turin, or when peasant Ivan decides to leave his village to seek his fortune. In other cases, on the other hand, the narrator extends the duration of the story to report crucial series of events (think of Mrs. Ramsay in Woolf's *To the Lighthouse*, ladling out the soup for lines and lines, while realising that she has had an empty life), or stops the narration altogether to make important points or insert relevant descriptions. The descriptions too are usually geared towards the promotion of cognitive gain. They give information that is neither obvious, nor unconnected with what preceded, and that, as such, might advance the reader's interpretive project. When Tolstoy first introduces Anna Karenina, he does not tell us that she has two arms and two legs; nor does he care to say how many moles she has on her neck; but he does pause to notice that she looked sadder and more thoughtful than the people around her in hearing of a man run over by a train. All these narratorial choices might seem obvious and natural, but on closer inspection they reveal a great deal about what we expect from a narrator and what we take narratives to be. Once again, it seems that stories are crafted to enhance the reader's understanding.

Apart from these quite general ways of creating narrative dynamics, there are other techniques that are characteristic of narratives of certain epochs, genres or styles. An Aesopian fable, a Russian folktale, a chivalric poem, a picaresque novel, a serial novel, a mystery story, etc. have all very different and very distinctive ways of playing with our expectations and managing our attention. They all have very different dynamic profiles. Something which testifies to the infinite richness and variety that narrative forms might reach within the same constraints for

[12] A classic treatment of the topic is in Genette (1972). Genette distinguishes four cases of difference between the duration of the events and the duration of the story: "ellipsis" when events that happen in the narrative world are omitted in the story (time of the fabula t, time of the syuzhet 0); "summary" when events that happen in the narrative world are summed up in the story (time of the fabula > time of the syuzhet); "scene" when events that happen in the narrative world are punctually described in the story (time of the fabula = time of the syuzhet); "pause" when the story stops to make room for descriptions or reflections (time of the fabula 0, time of the syuzhet t).

narrativity that we have defined. This is certainly not the place for a systematic examination of this variety, but let me gesture towards it by giving some examples.

In medieval chivalric poems we find already quite sophisticated ways of modulating narrative rhythm. The most notable one is a technique known as *entrelacement*, which consists in interlacing several simultaneous stories of different characters in one larger narrative. This allows the narrator to generate suspense merely by abandoning a certain character in a crucial moment and starting following another one; each time that the narrative shifts, there is both tension and release at the same time: the reader is left with the desire to know what will happen to the character she leaves, but at the same time relieved by finding out what has happened in the meantime to the character she is now presented with. And this movement propels the reader forward. In addition, the narrator would often bring the characters together at some point in the story, so that sequences of facts that seemed disparate stand revealed in the end as parts of the same underlying causal structure.

Quite often, the dynamic profile assumed by a narrative is influenced by the ways in which in that historical moment narratives of that kind are consumed. Most serial novels of the XIX century (Dickens, Dumas, etc.) for example tend to display an episodic character due to their being published in monthly or weekly instalments, in periodicals and *feuilletons*. Typically, the narrator would introduce a surprising turn of the event right at the end of the instalment to reopen the narrative and entice the readers to follow it further. This confers to the narrative an intermittent profile similar to that of today's TV series, which is not necessarily as successful when the novel is read in book form.

Another clear example of a narrative genre with its distinctive dynamic profile is the mystery story. A typical mystery story has a highly polarised structure: there is a very open beginning, where a mystery is built by introducing in the narrative many surprising facts; and there is a very closed ending, where all the facts find their place in the glorious reconstruction of the detective. Characteristically, the passage from a pole to the other should be abrupt: facts are not to be explained one at a time, with the explanatory tension waning gradually, but all together, at the very end; the more sudden is the passage and the more compelling is the solution, the more successful is the story. The ability of the mystery writer lies then in laying down all the premises for the solution of the mystery in front of the reader while at the same time keeping her interpretive project open until the very last moment.

Even with these few considerations, I hope to have given a sense of how narrators manage to provide us with a rewarding experience of continuous cognitive progress, and of the variety of the means at their disposal. Again, the experience that results from their efforts feels like learning to the reader, and is learning, at least from the perspective of psychology and neuroscience. But it is certainly not "learning" in the sense of the epistemologist. That which compels a reader to read further is certainly not a curiosity of a referential kind, a desire to acquire new true propositions: it is an appetite for richly patterned sensory steams, a desire for cognitive progress of the kind we have been describing. Engaging narratives provide us with a richly structured stream of information, and this, more than truth, is what we seem to ask for, and what we seem to find rewarding.

3.4 Ending Well

To have a lasting impact on the reader, however, a narrative must not only captivate her in its unfolding, affording her the chance of continuous cognitive progress. It is also crucial that it ends in the right way. The end is a particularly delicate place, because it is there that the narrative takes its final shape and exposes itself to our comprehensive evaluation. While we are still following the narrative in its ups and downs of uncertainty, we give it the benefit of the doubt: perhaps these inconsistencies will be explained later on, perhaps this boring part will serve as a background for a surprising turn of events. But when the story ends, there is no longer any doubt. In the end, after having strived to leave the narrative open and uncertain enough to propel the reader forward, the narrator must bring the inferential game to some sort of conclusion;[13] and the reader might like that conclusion or not, quite independently of what she has thought of the narrative before. A narrative that we have enjoyed thoroughly might be ruined by a bad ending, while a rather dull narrative might be revived by a successful one. The end, in sum, is the place where the narrative is called to take its final form and to disclose its ultimate significance, the "moral" or "message" that it implies. Ending well is thus critical for the ability of a narrative to convey insight. We shall therefore try to clarify how narratives manage to convey, in ending, a sense of their worthiness and importance, or, in other words, what in the end of a narrative determines whether a reader finds it enlightening.

First, let us notice that for a narrative to end is not necessarily for a narrative to close. The end of a narrative is just where the text stops. Narrative closure instead has to do with the way in which the text manages our inferential activity. As we saw, a narrative "closes" (or "stabilises") when a hypothesis of the reader about its underlying causal structure gains in probability. If a hypothesis becomes probable enough, the structure of the narrative will become apparent to the reader and all the facts of the narrative will appear to "make sense" (i.e. to find a place in a unitary whole). A narrative that closes is therefore a narrative whose facts make sense in such a way, and the "moral" of a narrative is simply the sense that its facts are making. It is the relationship, which the reader has now inferred, among the facts of the story. When a narrative closes, the reader can feel it. It is one of those facts about narrative rhythm that we discussed above. A closed narrative conveys a "phenomenological feeling of finality" (Carroll, 2007, p. 1), just like a well-rounded sentence or a musical passage after a perfect cadence. The reader can feel that, if there was uncertainty about how the story will develop, now that uncertainty has been dispelled, and no further addition is needed.[14]

[13] As Eco (1979b, p. 32) puts it, "the end of the text… authenticates or inauthenticates the whole system of long-distanced hypotheses hazarded by the reader about the final state of the fabula."

[14] See Carroll (2007, p. 7): "the narrator wraps the story up then when she has answered all the questions that have stoked the audience's curiosity. Those questions, needless to say, do not come from nowhere. They have been planted by the author in a way that makes them practically unavoidable for the intended audience. These questions hold onto our attention as the story moves forward. Closure then transpires when all of the questions that have been saliently posed by the narrative get answered. It is the point after which the audience may assume, for example, that the couple lived happily ever after and leave it at that."

Since ending and closing are different things, they may also not coincide. Certain narratives will close before they end: their causal structure will become apparent to the reader well before she arrives at the end of the text. As we said, in general this will have to be avoided by the narrator if she wants to keep the reader engaged throughout the narrative. Other narratives, on the other hand, will end without closing. Indeed, it is perhaps easier to understand what it means for a narrative to close if we consider a case where the narrative does not close. In his essay *The Storyteller*, Walter Benjamin discusses an example that is worth considering. It is the story of the Egyptian king Psammenitus, told by Herodotus in the fourteenth chapter of the third book of his *Histories*. As Benjamin (1969, p. 89) summarises it:

> When the Egyptian king Psammenitus had been beaten and captured by the Persian king Cambyses, Cambyses was bent on humbling his prisoner. He gave orders to place Psammenitus on the road along which the Persian triumphal procession was to pass. And he further arranged that the prisoner should see his daughter pass by as a maid going to the well with her pitcher. While all the Egyptians were lamenting and bewailing this spectacle, Psammenitus stood alone, mute and motionless, his eyes fixed on the ground; and when presently he saw his son, who was being taken along in the procession to be executed, he likewise remained unmoved. But when afterwards he recognized one of his servants, an old impoverished man, in the ranks of the prisoners, he beat his fists against his head and gave all the signs of deepest mourning.

The story of Psammenitus, quite evidently, does not close. To be sure, there is the beginning of a narrative progression, here: against our expectations, something unusual happens; the king is not moved by the fate of his daughter and son, but mourns deeply the fate of his servant. But the explanatory tension raised by these facts finds no solution. The text provides no answer as to what the relationship between these facts is. As such, we are not sure about what value we should attribute to the facts, and about the moral that they disclose. The same facts could be interpreted in many ways and convey different morals.[15] The narrative is left open.

But, as Benjamin notes, the persistent fascination of this story is due precisely to its being left open. Since we don’t find an answer in the text, we try to find it by ourselves, as the inferential game of the narrative leaves the page to continue, so to speak, in our minds, in our attempts to make sense of what we read after we read it. Thus, a narrative that does not close is not necessarily a bad narrative. On the contrary: it might be that we find it informative, profound and

[15] See Benjamin (1969, p. 89): “From this story it may be seen what the nature of true storytelling is. A story ... does not expend itself. It preserves and concentrates its strength and is capable of releasing it even after a long time. Thus Montaigne referred to this Egyptian king and asked himself why he mourned only when he caught sight of his servant. Montaigne answers: ‘Since was already overfull of grief, it took only the smallest increase for it burst through its dams.’ Thus Montaigne. But one could also say: The king is not moved by the fate of those of royal blood, for it is his own fate. Or: We are moved by much on the stage that does not move us in real life – to the king, this servant is only an actor. Or: Great grief is pent up and breaks forth only with relaxation. Seeing this servant was the relaxation. Herodotus offers no explanations. His report is the driest. That is why this story from ancient Egypt is still capable after thousands of years arousing astonishment and thoughtfulness.”

telling precisely because we cannot point out immediately what it means. As Benjamin (1969, p. 88) puts it, it might be "half the art of storytelling to keep a story free from explanation".

We can't help but notice that this latter conclusion parallels once again what we said about sentences in the previous chapter. There we saw that, if with most sentences our interpretation stabilises rapidly, there are other sentences (paradigmatically literary ones) whose meaning is much less clear. Recall Burke's metaphorical statement: "Spain – a great whale stranded on the shores of Europe". Like the story of Psammenitus, this sentence is ambiguous: although we might propose various interpretations, none of them seem to exhaust what the sentence expresses. Hence its suggestiveness. The problem faced by the reader of Burke's sentence and the problem faced by the reader of the story of Psammenitus is, in hindsight, the same: finding out how the elements of the text – words in the first case, facts in the second – fit into a unitary whole. Something that suggests that, at a fundamental level, a story means as a sentence means; and that stories, just like sentences, can be more or less ambiguous, more or less open.

Of course, open as they might seem, both Burke's sentence and the story of Psammenitus are not *completely* open. A completely open sentence would look more like the Russellian "Quadruplicity drinks temporalisation", and a completely open narrative would be more like (1) above. A sentence or a story that leaves interpretation completely open does not mean many things: it does not mean anything. As Velleman (2003, p. 18) notices in commenting on the same story of Psammenitus,

> not just any telling of an unexplained event would have left us with the sense of a story awaiting completion. Any number of unexplained developments might have ensued upon Psammenitus' attendance at the triumphal procession, and he might have done or said any number of unexplained things, most of which, if placed at the close of Herodotus' tale, would have turned it into a surreal fragment of prose rather than a protean story. ("Immediately following Psammenitus' son in the procession came a man walking on his hands. Psammenitus turned to Cambyses and remarked, 'You have helmet-hair.' And so on).

Ending a story well, then, just like recounting it effectively, seems to be a matter of striking a good balance between openness and closure. To have that lingering evocativeness that we attribute to great art, a narrative should achieve in ending a condition in which the connection among its events (the moral of the story) is not self-evident, but is still discoverable. To put it crudely, the dilemma that the narrator faces is between the banality of the hero who rescues the princess held captive in a high tower "and they lived happily thereafter" and the unintelligibility of the hero who climbs up the tower, punches the princess in the face and jumps off the tower with no reason whatsoever. If the moral is evident, then chances are high that the narrative will sound pedantic, excessively constraining, or pointing to something we already knew and did not need to be told.[16] But if there is no moral at all, if there is no way of integrating the facts of the story into a coherent whole, then the story teaches us nothing either. As when composing the

[16] French fabulist Jean de La Fontaine pointed to this fact when he observed that "Une morale nue apporte de l'ennui: / Le conte fait passer le précepte avec lui. / En ces sortes de feinte il faut instruire et plaire'' (*Fables*, VI, 1).

story more generally, therefore, in ending it the narrator is called to walk a thin line of fertile ambiguity, avoiding boredom on the one hand, and confusion on the other. How much openness she leaves to her narrative decides whether we ultimately find the story meaningless (too open), trivial (too closed) or profound and rich in implications still to be discovered.

Of course, whether a story ends with the right level of openness will be a matter of individual taste. A surrealist writer and a five-year-old child might have different views about it. At one extreme of the spectrum, we will have the advocates of extreme openness. It is indeed a trait of modern literature to have lost faith in the capacity of narratives to convey true understanding. To the modern writer, the world and life have often appeared so chaotic and absurd that any attempt to establish narrative cohesion sounds like a fiction and a lie. “Life does not conclude”, says Pirandello in *One, No One, and One Hundred Thousand*; so why should a narrative do so? And so many of the finest works of modern literature put the reader in front of facts that do not seem to disclose any underlying causal structure. A man can be processed and condemned for a crime unknown to him (Kafka, *The Trial*). A man can kill another man for no reason (Camus, *The Stranger*). These, however, are very consequential choices. If you think that life is “a tale told by an idiot… signifying nothing”, then, the only way to be faithful to life is by crafting a narrative signifying nothing. For the suitably predisposed reader, then, it may make sense that facts do not make sense. The absence of a graspable causal structure in the narrative may become a sign in itself that strengthens a certain hypothesis about the world and human life, thus giving the reader, despite everything, a sense of cognitive progress.

3.5 A Bayesian Narratology

Let us wrap up our discussion in this chapter by pointing out what we have achieved and by suggesting a few directions for further research that could be fruitfully pursued. We started from the common-sense intuition that narratives *qua* narratives convey some sort of insight or understanding. Narratives, we said, do not merely recount events, but present them in a way that makes them intelligible. An examination of this property of narratives using the Gestaltic notion of insightful problem-solving and the Bayesian/PP apparatus of belief updating has persuaded us that narrative and understanding are indeed intimately related. A sequence of facts gives us a sense of narrative progression only if we are able to reduce uncertainty about the causal structure underlying the facts of the sequence as reading progresses. Good narratives (or texts high in narrativity) are those that afford such a reduction of uncertainty to a higher degree. This means that stories are devices for optimal learning in the sense that is current in psychology and neuroscience: they allow us to grasp new causal patterns in an optimal fashion. We showed how this hypothesis not only aligns with our discussion on literary language in the previous chapter but also allows us to explain aspects of the phenomenology of narrative (the felt rhythms and dynamics, the sense of closure) as well as the ways in which narratives tend to be structured.

If the above is accurate, then it ought to have some consequence for our study of narratives in general. Here we have a formal apparatus (that of Bayesian belief updating) that is linked with an increasingly important strand of research in the sciences of mind and that can capture important aspects of what narratives are and how they are constructed. This fact should pave the way, I think, to analyses on the lines of those sketched in this chapter that would unveil how individual narratives or narratives belonging to various genres and using various techniques manage to captivate and persuade us and regiment our understanding of sequences of facts. Such an analysis would be the object of a "Bayesian narratology" that would parallel and dialogue with the "Bayesian rhetoric" envisaged in the previous chapter. Thanks to the clear link between cognitive dynamics and narrative principles that we were able to establish, this Bayesian narratology could be an eminently interdisciplinary enterprise, to which literary scholars, philosophers, writers, psychologists, cognitive scientists and neuroscientists could all contribute with their own tools. The knowledge of how narrative works that such an enterprise would generate could have repercussions not just for art and aesthetics, but also for history, education, testimonial and legal practices and all the multifarious contexts in which narrative is used as a means to make sense of human experience.

An interesting explanatory target for such Bayesian narratology – one that we did not have the space to explore here – would be character. After all, narratives are not simply about events; they are about events as lived and experienced by characters, people acting according to goals, desires and emotions. A great part of our engagement with narratives is occupied by characters and their psychological vicissitudes, and it is perhaps one of the most distinctive features of modern literature that the drama tends to move from the observable world actions and facts to the inner world of the character's mind. Even when we follow sequences of external events objectively reported, we often see them in relation to a human project that gives them structure and meaning, or as signs pointing to a character's intangible goals and desires. A full Bayesian/PP story about narrative of the kind we are envisaging, then, will have to say something about this crucial element in our engagement with narratives.

The issue of character has direct bearing on the problem of art and learning that concerns us here. Quite often, in fact, the insights we feel we have gained from a great narrative work have to do with characters and their psychology. We often feel that the great narrator has given us complex and lively renditions of the mental lives of some individuals (fictive or not), and that by attending to these renditions we are learning a great deal about the inner workings of human beings. Consequently, we often praise the great authors as masters of psychological depiction, gifted diviners of the most minute movements of the human mind. From our perspective, the study of this aspect of great narratives is no different from any other issue we have confronted so far. It will be a matter of understanding how a narrative can make (fictive) personalities intelligible to us, supporting and enabling a rich body of inferences about their motivations, interests and desires. Presumably, then, this inferential activity can be described with the Bayesian/PP apparatus we have been using so far: while reading, the reader would make probabilistic hypotheses about a character's psychology starting from what the author tells her about the character's actions, reactions and feelings. This inferential activity, then, will be subject to the same dynamics we have come to know: certain psychologies will be easier to divine than others, depending on how predictable the character's actions, reactions and feelings are. If they are either too predictable (think about the hero climbing up the tower to rescue the princess) or too unpredictable (think about the hero punching the princess and jumping off the tower for no apparent reason) we won't have the feeling of having penetrated deep into a character's psychology or having unveiled some of the mysteries of the human mind. Again, then, the condition in which the text will generate the impression of psychological insight will be the one where the character's actions, reactions and feelings will be unpredictable enough to stimulate our guesses about her psychology but not so much as to make every guess impossible.

Of course, this story about character will have to be spelled out in more detail and will have to be supported by analyses of concrete cases. But it seems to have already some intuitive appeal, as it maps well with existing narratological distinctions. Here E. M. Forster may come in handy again. In his *Aspects of the novel,* he makes a well-known distinction between "flat" and "round" characters. Flat characters are human types, constructed out of a few stereotypical traits; they "are easily recognized whenever they come in", behave predictably and consistently throughout the story and can be "expressed in one sentence" (Forster, 2022 [1927], p. 40). While they have their use in literature, they are not the kind of characters that give us the impression of having gained important psychological insights. Round characters, instead, are complex individuals, endowed with diverse and often discordant personal traits; they act in surprising but convincing ways and evolve in the course of the narrative.[17] They cannot be easily summed up or captured with a convenient formula: the whole of their fictive personality remains a moving target that the reader is constantly called to guess. For this reason, they do give us the impression of having gained from them important psychological insights. But push the complexity of the round character a little beyond a certain limit and you start to get what Lotman (2009 [1992]) calls the "madman", a character "whose actions reveal no causal connection" (*ivi*, p. 82) and whose psychology remains unfathomable. Many works of modern literature are populated by characters of this kind, as they enact the dissolution of a character's psychological identity. In extreme cases, the mind of the character becomes inaccessible and the only psychological insight we are left with is that the other is just inscrutable.

There is therefore in narratology an awareness of the conditions for psychological access in the minds of characters, as well as a terminology to describe them. What remains to be done is to formulate the same intuitions about psychological insight in narrative in Bayesian/PP terms, seeing the character's inner world as an underlying causal structure that the reader tries to infer, with more or less ease and success, as the narrative unfolds. Such a reformulation would give us a formal apparatus to follow the dynamics of empathy and psychological inference at a finer level

[17] See Forster (2022 [1927], p. 46): "the test of a round character is whether it is capable of surprising in a convincing way. If it never surprises, it is flat. If it does not convince, it is a flat pretending to be round. It has the incalculability of life about it—life within the pages of a book." Notice how close Forster's suggestion that round characters "surprise in a convincing way" is to Aristotle's suggestion that in a good tragedy events occur "unexpectedly and at the same time in consequence of one another".

of detail, allowing us to see when and by what structural means an author shuts down or reopens the access to a character’s psychology, managing our empathic response over time. The usefulness and feasibility of such an approach are all to be assessed, but the problems surrounding character can usefully be seen as an example of the kind of questions a Bayesian narratology could have a lot to say about.

As for our more immediate concerns, however, the picture is sufficiently clear. Narratives seem to be designed to enhance our understanding, and this fact tells us a lot about how we experience them and how they tend to be structured. In our developing Bayesian/PP story, they stand revealed as part of a broader class of stimuli (like ambiguous images or felicitous literary utterances) that facilitate perception, cognition and learning, insofar as all these processes might be cast in terms of probabilistic inference aimed at finding new patterns in data. Now that we have tested the explanatory potential of this idea with literature and narrative, it is time to see whether it might have a broader appeal and tell us something fundamental about art in general.

4. Broadening the Picture

4.1 A General Story?

In the two preceding chapters, we have built and tested a certain picture of our engagement with the literary text. In this picture, the reader is seen as embodying a hierarchical probabilistic model of the world and constantly testing the predictions of this model against the incoming sensory stimulations, at multiple temporal scales. The literary text, on the other hand, is seen as providing a particularly well-structured sensory stream that disrupts and confirms the reader's predictions strategically, to maximise her experience of cognitive gain. It is now time to consider how the picture we have developed so far scores with respect to arts other than literature. After all, as we noticed at the beginning of our inquiry, the lived sense of cognitive gain that we may experience while reading a good literary work is by no means limited to literature and narrative: it seems to generalise to all instances of great art. A painting or a piece of music can prove just as conducive to insight, transformation and powerful epiphanies as a good work of literature, and can elicit and sustain our serious and devoted attention in a very similar way. In all cases, we have the same feeling that the contact with the work is cognitively rewarding. If the source of this feeling was different in the case of literature and in the case of visual art or music, this would be strange indeed. This chapter will therefore consider how we can understand visual art, music, and other stimuli and activities as providing the same experience of cognitive gain that we have described in the verbal and narrative domain. The reason for broadening the picture in such a way is threefold.

First, discussing other arts will allow us to link the work we have done with literature and narrative with the acquisitions of a growing stream of research on PP and the arts. As we noted in Chapter 1, a developing PP story about the arts and aesthetics has been gaining traction in recent years, and it has been applied so far mainly to arts different from literature: visual art, music, and, increasingly, cinema and games. All these applications seem to converge invariably around the same idea: effective artworks are clever ways of playing with our predictions to provide us with a rewarding experience of cognitive gain. Discussing this growing stream of research will therefore provide additional support for the view developed in the previous chapters and will place it firmly into a broader and fast-developing research agenda.

Secondly, broadening the picture in this way will allow us to recognise that the proposal that we are advancing has deep roots in the psychology of art and in general psychology. As we anticipated in Chapter 1, the experience of cognitive gain that we are trying to capture has been described before, in intuitive terms, by psychologists of art such as Ernst Gombrich and Rudolf Arnheim, as well as by general psychologists under different headings ("Aha!" experiences, "flow states", etc.). More recently the same experience has been at the centre of an interesting convergence of results between developmental psychologists and researchers working on artificial intelligence and machine learning. Examining this literature will allow us to appreciate that our engagement with the arts might be just part and parcel of a more general human tendency to direct attention towards regions of our sensorimotor space that make us grow and flourish.

This will allow us, thirdly, to understand how deep and generalised the relationship between art and learning is. At the end of this chapter, we will be able to see that learning – in the epistemically-neutral sense that we have by now defined – can be said to underlie all aesthetic experiences and to be essential to our engagement of art *qua* art. This rather bold conclusion will set up the framework for the final discussion about art and learning that will occupy us in the next and last chapter.

But now let us start by saying something about the ways we learn from pictures.

4.2 Learning from Pictures

In Chapter 1, we have seen that PP and other strands of research in cognitive science have embraced and formalised in Bayesian terms a view of perception as an inferential process. This view maintains that perception is not a passive reception of sensory inputs, but an active process of hypothesis-testing in which the agent tries to guess the hidden causes of the stimulations that impinge on its sensory organs. For creatures like us, these "guesses" take the form of the predictions of a hierarchical probabilistic model embodied in our brain's structures and dynamics. These predictions, as we saw, unfold in a hierarchical fashion. In the case of vision, for example, a high-level prediction that I will see a dog gives rise to "lower-level predictions about limbs, eyes, ears and fur, which then cascade further down in terms of predictions about colours, textures and edges, and finally into anticipated variations of brightness across the visual field" (Seth, 2021, pp. 107-8). At each level of the perceptual hierarchy, predictions are updated in light of the evidence coming from the level below, until the perceptual system settles down to what it considers to be the best interpretation of the sensory data. "Learning", in this picture, indicates precisely the process through which hypotheses throughout the hierarchy are updated and revised in light of the evidence coming from the senses, until a satisfactory explanation is reached.

An important consequence of this view is that different percepts afford learning to different degrees. Going back once more to our Dalmatian example (Figure 3 in Chapter 1), we may recall that in the case of the left picture, perceptual organisation is fast and effortless: the picture conforms more to our hypotheses and, as such, triggers few cycles of belief updating (i.e., a small learning). In the case of the picture on the right, perceptual organisation is impossible: the picture does not allow for any hypothesis to gain in probability, and, as such, triggers few cycles of belief updating (we have again a small learning, for opposite reasons). It is with the central picture that we get the highest rate of learning: in this case, perceptual organisation is difficult, but still possible. Perception needs to proceed more gradually and tentatively, but can still proceed. It is when confronted with ambiguous percepts of this kind that the active and inferential character of perception, as well as its pleasurable affective correlates, are made more manifest.

Keeping this framework in mind, we can now start our exploration of how we learn from pictures by noticing that many paintings exploit precisely the kind of low-level perceptual ambiguity exemplified by the image of the Dalmatian. They hinder the recognition of objects, restoring to perception the character of a problem to solve. Take for example impressionist paintings, such as the Monet reproduced below:

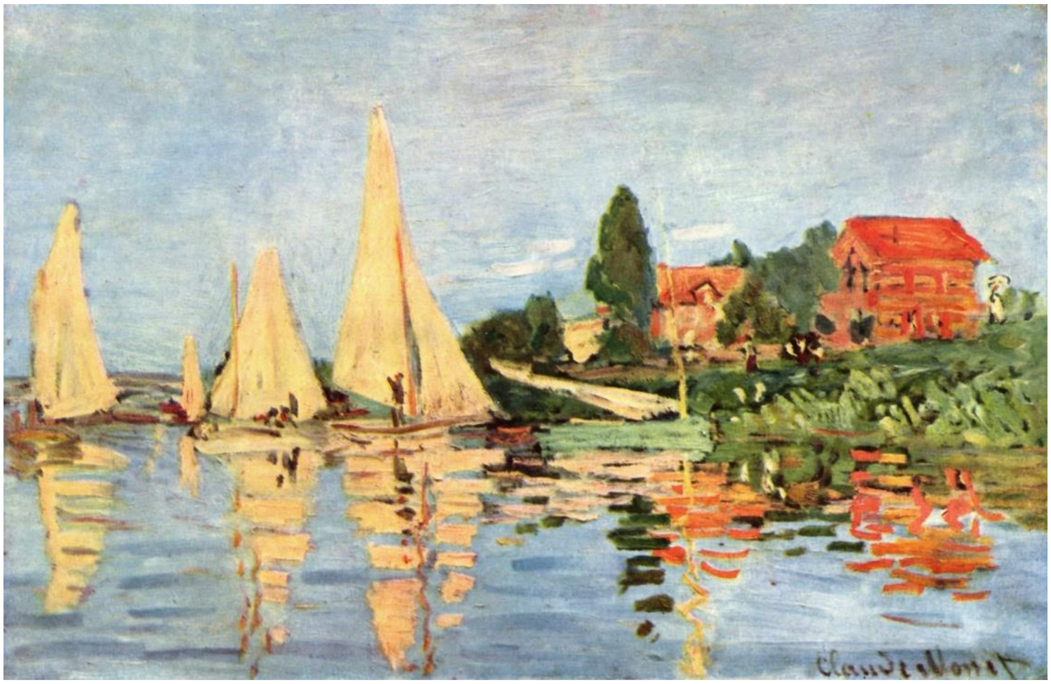

Figure 7: Claude Monet, *Regatta at Argenteuil*.

Perceptual organisation in a case like this is not immediate. Especially in the lower portion of the painting, the horizontal brushstrokes create a broken pattern that is difficult to decipher. But precisely because organisation is not immediate, once we find it, the depicted scene may appear much more vivid than it would if depicted in a naturalistic style. This is because, insofar as what we were confronting did not look like an ordinary landscape, we had to work our way inferentially towards it. This makes the imaginative contribution of the perceiver (what Gombrich would call the "beholder's share") more apparent. As Gombrich puts it, "when we say that the blots and brushstrokes of the Impressionist canvas 'suddenly come to life', we mean we have been led to project a landscape into these dabs of pigment" (1960, p. 170). In other words, in much the same way as the picture of the Dalmatian above, impressionist paintings "leave space for the observer's visual system to perform its interpretive work" (Seth, 2021, p. 118). They portray a pre-structured perceptual world whose organisation is not obvious and need to be discovered.[1] This discovery is felt as such, and rightly so: it has all the features of a learning experience from a neuroscientific and psychological perspective.

Something similar could be said of cubist art, which, according to Seth (2021, p. 128), "draws the observer into imaginatively creating perceptual objects out of a jumble of possibilities". Take for example Braque's *Violin and Candlestick*:

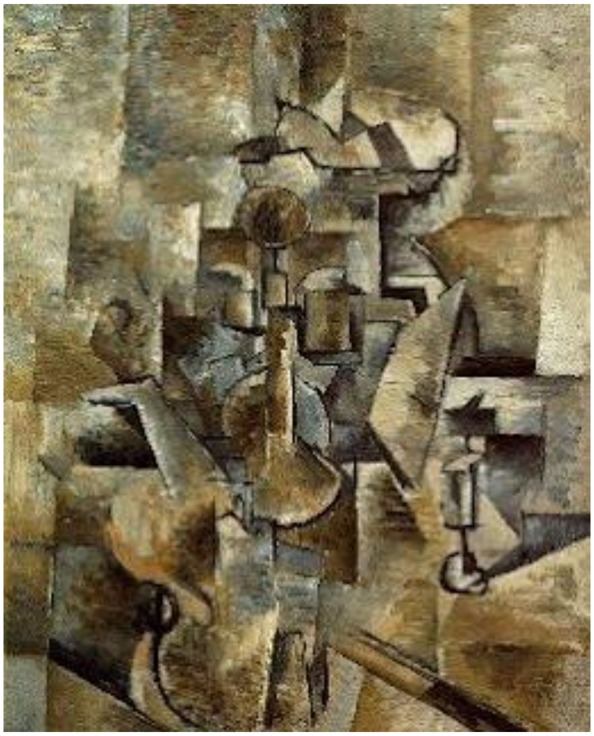

Figure 8: Georges Braque, *Violin and Candlestick*.

Here too the first impression is of a general perceptual disorder. But (encouraged also by the expectations installed in us by the title) we gradually come to find evidence of the two objects depicted. The process is somewhat different from the one that leads us to discover a body of water in Monet's painting. If there we were confronted with the task to combine more atomized brushstrokes, here we need to put together temporal and spatial slices of more clearly outlined objects. With this decomposition, the painter complicates and makes manifest the rather

[1] See also Seth (2019) for an interesting discussion along the same lines. Here Seth observes that impressionist art "explores the idea that the painted image provides, not a detailed pictorial representation of some external situation, but the raw material to ignite perceptual and associative representations… Impressionist painting can therefore be understood as a series of experiments into the inferential operations of the visual system and – more broadly – into the nature of the subjective experiences entailed by these operations. These artistic 'experiments' complement contemporary neuroscientific attempts to reveal how top-down perceptual predictions underpin visual experience within the framework of P[rediction]E[rror]M[inimisation]" (p. 385).

automatic inferential process through which, in ordinary perception, we put together different views of the same object to perceive it as one. Despite these differences, however, the same general process is at stake in both cases: both paintings provide the right quantity of visual cues to start us off in the process of recognition without however delivering an immediate solution. They look "like they are going to make sense". The result is a dynamic process in which what initially does not look like a known object ends up looking much more familiar. Picasso's famous reply to those who observed that Gertrude Stein did not look like the portrait he drew of her ("That does not make any difference, she will")[2] seems to point to the same phenomenon. Picasso's painting evokes the true Stein more powerfully precisely because it does not represent her perfectly, allowing us to discover her in the painting, and the painting in her.

The play with our hypotheses that the artist can generate even at the relatively low level of perceptual recognition is however much more complex and varied than what we may suspect from our discussion so far. Artists do not only have control on whether to delay the recognition of patterns in the pictorial space or not, but also on *how* to do so: recognition might be made faster or slower, more or less stable, irreversible or shifting continuously between parallel hypotheses. There is, in this respect, an infinite array of possibilities, and charting this territory in more detail can hold many insights into how visual art channels and promotes our inferential activity. Interesting work in this direction has been done by Muth and Carbon (2016), who, starting from a PP perspective, sketched a useful preliminary typology of ambiguity resolution in visual art. Following them, we may start distinguishing a few differences in the dynamics of visual belief updating. Images such as the hidden Dalmatian for example, as well as many impressionist paintings, contain "transient ambiguities": their perceptual organisation is not immediate, but once we find it, the picture tends to coalesce rapidly into a relatively stable interpretation ("here is a dog", "here is a body of water", "here is a boat", and so on).

In other cases, however, which Muth and Carbon put under the heading "indeterminacy", the picture is so ambiguous that it barely suggests viable interpretive hypotheses, none of which stabilises enough to allow for precise recognition. Many cubist paintings are of this kind: their work of perspectival deconstruction is carried to such extremes that every perceptual hypothesis about their organisation is fleeting and precarious. A series of paintings by the contemporary artist Robert Pepperell exploit similar mechanisms (see e.g. *Succulus*, Figure 9 below): depicted in a style vaguely reminiscent of neoclassical art, these paintings are more than simple chaos; they are designed to systematically encourage the observer's search for meaningful objects, without however never satisfying it. The observer is thus led to scan the pictorial surface repeatedly, and constantly finds hints of objects ("here is a head", "here is a piece of clothing") that are nevertheless too vague to allow for full recognition.

In other cases still, the pictorial surface allows for the development of two (sometimes more) mutually exclusive interpretations, each of which might reach relative stability before is overridden by another. An example of this switch between determinate interpretations or "bistability" is offered by Dalí's *Slave Market with Disappearing Bust of Voltaire* (Figure 10), nicely

[2] The anecdote is reported in Stein (2019 [1933], p. 9).

analysed by Carroll, Moore and Seeley (2012). In cases like this, focusing on one or the other part of the picture can make one or the other interpretation more salient and prompt a global restructuring that might extend, more or less readily, to all neighbouring areas of the image. Whether we are able to consciously control the switch from one hypothesis to the other will depend on how much each of them is perceived as being clear and compelling: the less compelling a hypothesis is, the more conscious effort is needed to work towards its stabilisation.

In still other cases, which Muth and Carbon call "dichotomy", different perceptual interpretations might coexist in the same picture in a locally consistent, but globally contradictory way. In these cases, a region of the pictorial space suggests a perfectly consistent interpretation that needs to be abandoned once the observer moves her gaze to another portion of the painting suggesting another perfectly consistent interpretation incompatible with the first. Escher's depictions of impossible architectural spaces (Figure 11) are a clear case in point, but subtler effects of the same kind are also possible. Arnheim (1974 [1954], pp. 299-301) for example notices how in de Chirico's "metaphysical" landscapes the architectural elements often suggest more than one contradictory linear perspective, creating an unsettling effect whose source the viewer is often unable to pinpoint.

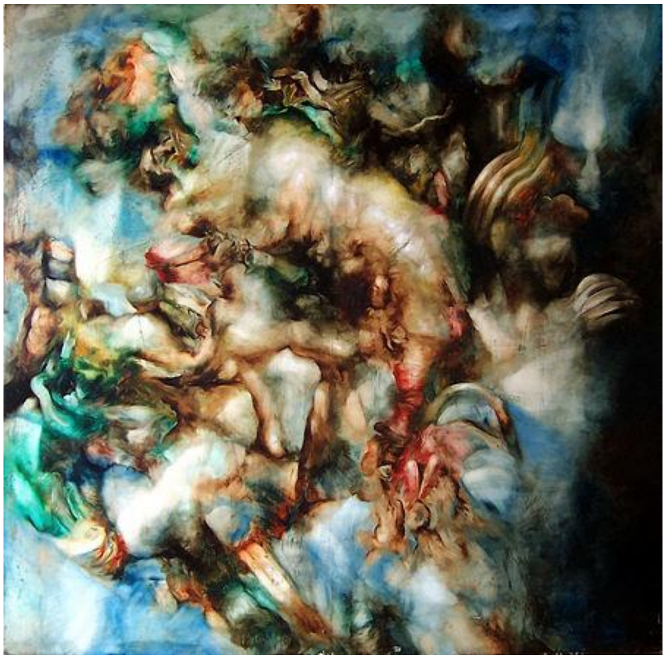

Figure 9: Robert Pepperell, *Succulus*.

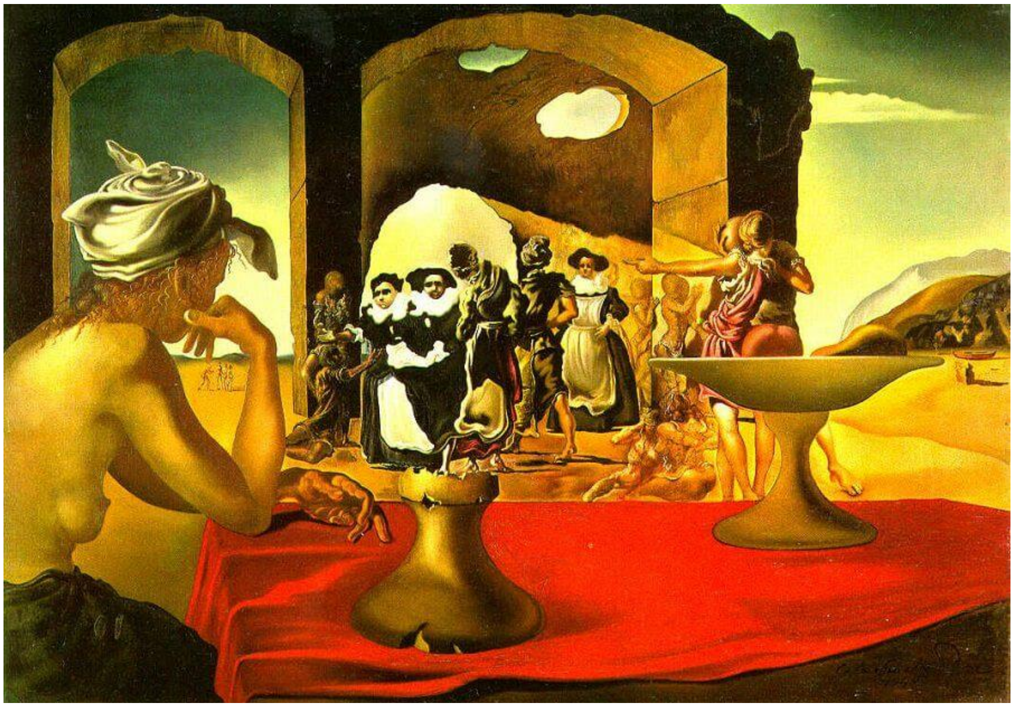

Figure 10: Salvador Dalí, *Slave Market with Disappearing Bust of Voltaire*.

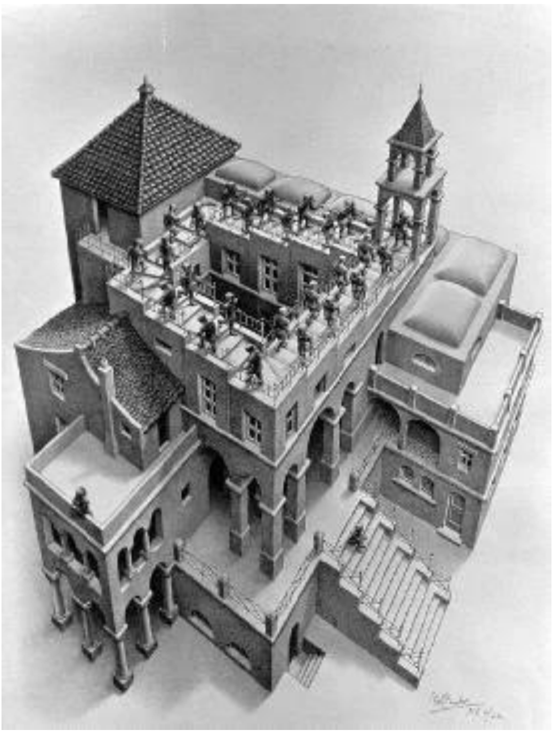

Figure 11: Maurits Cornelis Escher, *Ascending and Descending*.

This typology is no doubt partial and coarse-grained, but it gives us an idea of the multitude of ways in which artists can exploit the hypothesis-testing character of perception, effectively channelling and controlling our inferential activity to promote a prolonged exploration of the pictorial surface. This exploration (or belief updating in Bayesian terms) is "learning" in the perfectly legitimate sense we defined in Chapter 1, and, moreover, it feels like learning: it is by means such as these that paintings manage to sustain our prolonged attention and appear to us as constant sources of new information.

Given the examples above, the reader might have come to think that the view of visual art and learning I am articulating here applies only to a rather restricted group of experimental works – the kind of works that leverage and thematise the inferential character of perception. But one could object that many other works are far more straightforward in their perceptual interpretation and do not contain any marked ambiguity of the above kinds. We will address this worry shortly, when we will notice that visual art can leverage ambiguity at many levels of the perceptual-conceptual hierarchy. But for now, let us notice that the capacity to exploit low-level perceptual ambiguities is by no means relegated to a few experimental paintings or styles. The most realistic painting often contains very sophisticated devices of the very same nature, designed to spur and scaffold our inferential activity.

In fact, Carroll, Moore and Seeley (2012) – supported in this by Gombrich (1960) and Livingstone (2002) – suggest that Leonardo's *Mona Lisa* might owe a good part of its fascination to a very similar kind of ambiguity. Leonardo's use of *sfumato* around the corners of Mona Lisa's mouth and eyes makes these features, so crucial for the recognition of facial expression, much more ambiguous than they would otherwise be. As a result, the viewers are forced "to use their imagination to interpret an expression that cannot ever be discretely resolved" (p. 51). Moreover, this ambiguity is enhanced by "spatial inconsistencies in the background landscape, that alter the way viewers perceive the posture and relative size of the figure as they scan the painting" (pp. 51-52). Semir Zeki (2004) makes analogous remarks about Vermeer's portrait *Girl with a Pearl Earring*. Here too, according to Zeki, the subject's facial expression lends itself to multiple interpretations: "at once inviting, yet distant, erotically charged but chaste, resentful and yet pleased" (Zeki, 2004, p. 189). The observer oscillates between these interpretations, and the more she looks for cues that would confirm one of them, the more she finds cues that seem to lead her towards another. In this way, the painting approaches that inexhaustibility that many sees as characteristic of great art.

More generally, we should notice once again that the PP story that we are articulating does not offer a view of art that privileges strangeness over normality or deviation over norm, but one that does justice to the dialectics of strangeness and normality, deviation and norm, that is characteristic of artistic engagement and perception more generally. After all, if the PP story is on track, finding an object in a painting means experiencing a certain hypothesis about the structure of our sensorium as increasingly probable, or, if you prefer, being able to see how fitting those particular patches of colour are to represent the particular object we are currently hypothesising. This means that we can take delight in a perfectly realistic painting if this involves keeping discovering how close its patches of colour are to how things ought to look based on our current perceptual best guesses. In this sense, realism is not in contrast with ambiguity but rather requires it, since for recognition to take place an initially less probable hypothesis must become more and more probable.[3] At the same time, this means that we might dislike pictures with ambiguities and deviations that we are not able to explain. In these cases, we would just be inclined to say that we are confronted with a bad rendition of a dog, a boat, a body of water, and so on. Our story can therefore account for the tendency towards what is proper, fitting and adequate just as well as it accounts for the tendency towards deviation. This parallels our discussion of literary language and narrative and constitutes yet another confirmation of Gombrich's (1984, p. 9) observation about "the most basic fact of aesthetic experience", namely "that delight lies somewhere between boredom and confusion."

So far, however, we have mainly talked about rather low-level processes of perceptual recognition. But of course visual art can prompt much more sophisticated experiences of cognitive gain than those involved in merely discovering objects in a painting. We must remember that the observer's work of hypothesis-testing unfolds hierarchically, across multiple spatial and temporal scales and along a continuum that does not seem to admit any clear-cut distinction between perception and cognition.[4] This means that artists do not necessarily need to exploit low-level perceptual ambiguities to generate cognitive gain but can – and normally do – engage our inferential abilities in their full spectrum and scope. Here, again, the possibilities are infinite. A painting can be, for example, very straightforward at a relatively low perceptual level but very ambiguous iconographically, thematically or stylistically. Take this painting by Vincent Desiderio, nicely analysed from a PP perspective by Kesner (2014):

[3] We already mentioned that the pleasure of the imitative arts is explained by Aristotle in very similar terms. "The reason of the delight in seeing the picture", he says, "is that one is at the same time learning - gathering the meaning of things, e.g. that the man there is so and so" (see Chapter 1, footnote 6). Plutarch (1969, p. 93) echoes the same idea when he observes: "when we see a lizard or an ape or the face of Thersites in a picture, we are pleased with it and admire it, not as a beautiful thing, but as a likeness. For by its essential nature what is ugly cannot become beautiful; but the imitation, be it concerned with what is base or with what is good, if only it attains to the likeness, is commended. If, on the other hand, it produces a beautiful picture of an ugly body, it fails to give what propriety and probability require… What we commend is not the action which is the subject of the imitation, but the art, in case the subject in hand has been properly imitated." Interestingly, Plutarch in this passage (as Aristotle throughout the *Poetics*) talks about "τὸ εἰκὸς" (that which is "probable"), giving to his explanation a probabilistic undertone that perfectly fits the story we are articulating.

[4] The issue of what PP tells us about the perception/cognition divide is complex and debated. However, there is large agreement that, in principle, the hierarchical structure of PP generative models admits different levels of abstraction without requiring any sharp line separating perception from cognition. See e.g. Hohwy (2013, p. 138), according to whom "low-level perceptual inference and high-level conceptual judgement are not different in kind, but merely inferential processes at different levels of the perceptual hierarchy."

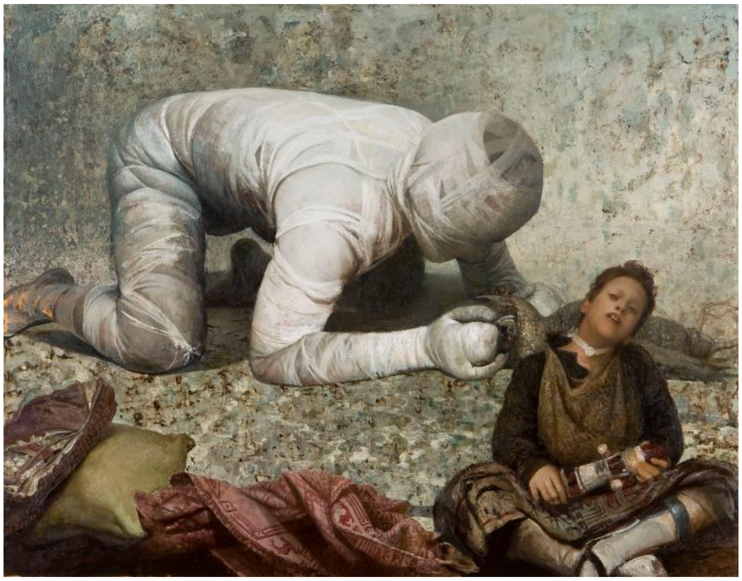

Figure 12: Vincent Desiderio, *Spiegel Im Spiegel*.

Since the subjects and objects in the painting are represented quite clearly, the observer is likely to settle quickly on a reasonably stable and definite low-level interpretation of the pictorial space. We can easily identify an adult man on his hands and knees wrapped in bandages, a child sitting cross-legged with a toy in his hand, some pillows and a blanket in the foreground. But now that we have (partially) reduced ambiguity at a relatively low perceptual level, the problem has evolved: it is no more what to make with the edges, shapes and colours that constitute the surface of the painting, but what to make with the elements that we have now (provisionally) identified at this relatively low level. The strange objects, subjects, poses and attitudes of the paintings are elements that ask for an explanation in a way not different in principle from how the scattered patches of the image of the hidden Dalmatian asked for an explanation. The problem becomes then what is the hypothesis that might account for these.[5]

As observers, we are then prompted to scan the pictorial surface in search of an underlying explanation for what we see. The more we look, however, the more we are likely to encounter other puzzling details (what is the adult holding in his hand? Is the child sleeping, is he ill or comatose? What is the role of the objects in the foreground?). If we don't find any viable explanation to test, our attention starts to wane; we feel that the painting is ceasing to be informative. After a while, if this puzzlement is protracted, we cease to engage with the painting altogether.

[5] As Kesner (2014, pp. 4-8) rightly points out, "despite the relative ease with which most individual objects in a pictorial space can be identified, the beholder is left puzzled as to what is transpiring in the depicted scene… Reducing the uncertainty related to the representational content of the painting by minimizing prediction error, far from 'explaining the painting away', decisively changes the generative model of the motivated viewer, thus triggering a new productive cycle of perceptual, cognitive, and affective inferences."

But now suppose that we know by previous encounters with Desiderio's body of work (or that we are informed by someone) that Desiderio has in fact a son who is severely physically and mentally handicapped, and that he has been constantly caring for him and has portrayed him in various paintings. Equipped with these prior expectations, we may start to hypothesise that the child in the painting is in fact Desiderio's son – something which may justify his comatose appearance. But if this is the case, then, perhaps the adult leaning over the child is the painter himself, and the bandages that wrap him are a means to indicate his helplessness in the face of his son's condition. The more we find details that confirm our hypothesis, the more what the artist has offered us will appear as an accurate and revealing depiction of whatever condition we are hypothesising the painting to express.

Moreover, while we mobilise our attention to test these higher-level hypotheses, we are led to examine more carefully portions of the painting that before we might have overlooked.[6] We might thus come to notice, for example, that the strange object around the child's neck is in fact the kind of neckband that holds a breathing tube in place. In turn, these newly discovered details can change our higher-level hypotheses, prompting new explorations of the pictorial surface in search of new details that may confirm them. And so on, in a cycle of explorations that, in the best cases, glues the observer to the picture, keeping her curious and inquisitive, confident that the painting still holds possibilities for insights. In the best cases, therefore, the interpretive process of visual art acquires the same sort of temporal dept and directionality that in Chapter 3 we described as characteristic of an engaging narrative. The observer can feel that her interpretive activity is going somewhere, that it is progressing.

This latter example has made clear that, if a PP story about visual art and learning is to be of any use, it should recognise the variety of factors that influence our predictive processes. These span from expectations about our immediate perceptual context (neighbouring edges, shapes, objects) to expectations about a specific work, author or style, up to evolutionarily-acquired expectations about the statistical properties of our environment. If we take this stance, we can interpret a variety of art-critical concepts (such as norms, genres, styles, "categories of art") as referring to predictions operating at different levels of the hierarchy. This vastly expands the kinds of predictions visual artists can play with, and, consequently, the learning experiences that they can provide. It suggests that when we engage with a painting we are not merely making and updating probabilistic best guesses about what the painting represent and what its elements mean, but we are also making and updating best guesses about the style of the author, the genre of the painting, the artistic movement it belongs to, the historical period in which it was produced, etc., and our guesses at one level inform and are informed by our guesses at another

[6] This is not dissimilar from informing a subject who is confronted with the image of the hidden Dalmatian that there is in fact a Dalmatian there. When we get this information, we know what kind of expectations we need to apply to the image, and where to direct our attention if we want to confirm them. In fact, every shift in these higher-order hypotheses is accompanied by a shift in the portions of the painting that become salient to us, and towards which we tend to direct our attention. As such, I see my proposal here as compatible with recent suggestions that artworks are "attentional engines" (see e.g. Carroll and Seeley, 2013; Seeley, 2020).

level.[7] At a very high level of abstraction, we may find artworks whose functioning as artworks depends on their ability to challenge and rearticulate very general predictions about what an "artwork" should look like. This is why Duchamp could make art by signing a urinal, and Warhol by gathering together pieces of commercial packaging. A significant portion of contemporary art has this character: it asks its observer to stretch and rework her hypothesis about what constitutes art. The game in these situations becomes: "this is art: see it as such". As strange as it may initially seem, this operation may not be different in principle from presenting someone with a confused group of black and white patches and asking her to see it as a dog. If PP is on the right track, the two are both inferential processes happening at different levels in the same perceptual-conceptual hierarchy.

In the PP view that we are articulating, higher-level predictions about artworlds, epochs, genres, schools, individual styles, etc. have therefore paramount importance in our experience of an artwork and constitute an integral part of the artist's toolkit. The importance of these predictions is always clear, but is even more manifest in contemporary art, where the rejection of representational conventions and the fragmentation of artistic languages in highly idiosyncratic idiolects have deprived the observer of much of her interpretive common grounds. Contemporary non-figurative art for example chooses to do without a whole set of largely shared expectations having to do with how the things depicted normally look. In approaching a figurative painting, even if you don't know anything about the artist, style, epoch, you can still leverage a common set of predictions to get your inferential activity going ("here is a man; and here is a child; they are wearing such and such clothes, their faces look in such and such way…"). In non-figurative art, this kind of visual *lingua franca* disappears. The well-documented coldness and aversion that contemporary art often elicits in the observers is due to a large extent, I think, to their inability to identify what kind of expectations they should bring to bear on what they are seeing, what hypotheses they are supposed to entertain.[8]

But if the rejection of a common set of expectations is not to lead to a complete paralysis of the inferential activity, the artist should be able to provide alternative expectations, often in the form of the constraints of her own personal poetics. In other words, the ability of contemporary non-figurative artworks to still function as inferentially rich stimuli depends crucially on the artist's ability to institute her own frame of reference and make this clear enough to her public. Hence the greater importance, in contemporary art, of paratextual information – titles, descriptions, information about the author, the author's writings, her correspondence,

[7] As Seth (2019, p. 401) points out, we can generalise "P[rediction] E[rror] M[inimization] to inference about the historical causes of current sensory inputs: in other words, one's perceptual experience of an artwork becomes partly constituted by predictions about the artist's motives and methods. … The PEM framework thus encompasses and operationalizes the 'period eye', at least as to the extent that it concerns influences of high-level cultural or social predictions about the causes of sensory signals."

[8] See Csikszentmihalyi and Robinson (1990, p. 103): "It is likely that the inability to have an aesthetic response is often the result of a lack of goals in the aesthetic encounter. Most people, when confronted with a work of art, simply do not know what to do. Without a goal, a problem to solve, they remain on the outside, unable to interact with the work."

manifestos, anything that can install in the observer a set of expectations that she can bring to bear on what she is seeing. (Think of how much the title influences interpretation and appreciation in works like Braque's *Violin and Candlestick* or Picasso's *Bull's Head*, or how much poorer the interpretation of Kandinsky's works would be without an awareness of his theoretical writings). This opens interesting additional spaces for artistic initiative, because insofar as authors control the discourse around their work, they have yet another set of predictions to play with. They are able to provide constraints to our inferential activity without intervening in the work itself.[9] This poses interesting questions about where the "boundaries" of the artwork lie, and if we should consider the discourse around the work (and more generally all our relevant expectations) as being part of the work itself. The postmodernist dictum that "Il n'y a pas de hors-texte" (Derrida, 1976, p. 158) may find here a new convincing formulation.

More generally, this highlights the importance of background knowledge for aesthetic appreciation, something philosophers should already be familiar with (see e.g. Walton, 1970), but that, in light of the Bayesian/PP picture of cognition, acquires new interesting implications. A work can resonate with us and invite us to a pleasurable inferential journey only if we can bring to it an adequate set of expectations that the work then shapes and modifies. In saying this, we are clearly echoing our conclusions about literary language and narratives, and our story is beginning to acquire some general plausibility. Let us see If we can expand our discussion further.

4.3 Learning from Music

Among all the arts, music is perhaps the one that lends itself most obviously to the Bayesian/PP story about art and learning that we are articulating. Not only the role of expectations, predictions and probabilities has been recognised for a long time in music research, but music is also, among the arts, the one that has been studied more extensively in PP terms in the last few years. As such, we shall devote comparatively less space to it, delineating the main acquisitions of research in this area and showing how these resonate with our conclusions in other artistic domains. As with the other arts, the problem will be to understand if and how music can produce a flow of stimulations that is conducive to learning in the psychological and neuroscientific sense.

PP, as we know, sees perception as an active inferential process in which the predictions of a hierarchical probabilistic model are tested against the incoming sensory stimulations. We saw how verbal narratives and visual art might be plausibly understood as providing particularly rich sensory streams that disrupt and confirm the subject's predictions strategically, providing her with a constant experience of cognitive gain. In articulating a PP story about our engagement with music, one would expect to find something of the same kind. In fact, it has long been

[9] Artists can also modulate the strengths with which they install predictions in their viewers (and, consequently, the level of ambiguity in their artworks) by choosing different means to express their authorial intentions. A title for example will be a clearer indication than a mere hint given during an interview.

recognised, in philosophy, musicology and psychology, that expectations affect almost every aspect of musical experience.[10] They are taken to play a major role in the perception of all major dimensions of music perception (melody, harmony, rhythm, volume and timbre) and to be related in fundamental ways to learning, emotional arousal, feelings of tension and resolution, and aesthetic appreciation.[11] In line with the PP story we have been telling so far, it has also been recognised that expectations in music play out across temporal scales and levels of abstraction. There are very general expectations generated by the musical culture one has been exposed to (e.g. the Western tonal system); there are also more precise expectations acquired through exposure to certain musical genres, styles or forms (e.g. chamber music, or the sonata form); there are also expectations prompted by the structure of a particular piece (e.g. "since the first chorus was thus and thus, the next one should be similar"); and there are also short-term expectations prompted by the immediately preceding musical context (e.g. those about the next note in a musical phrase, or the next chord in a progression).

From the emergence of information theory halfway through the last century, it has also been increasingly clear to music scholars that expectations at all these levels could be conceptualised in terms of a complex system of probabilities that composers, performers and listeners progressively learn and internalise (Meyer, 2008 [1956], 1957; Cohen, 1962).[12] Once more, it should be noted that the process whereby a perceiver comes to embody and internalise patterns of statistical regularities in her environment is what learning amounts to according to mainstream neuroscience and cognitive science. This means that, effectively, we are learning from music all the time, insofar as we are detecting and internalising its statistical regularities at multiple levels. It also means that music theory, to the extent that it reflects the internalised probabilistic models of listeners and composers, is a form of psychological knowledge (indeed maybe "the most formally developed example of a folk psychology currently extant," according to Pearce and Wiggins, 2012, p. 645). Composers and listeners master this knowledge, to a greater or lesser extent and mostly intuitively, while composing or engaging with music.

[10] See Judge and Nanay (2021, p. 997): "Almost every facet of the experience of musical listening—from pitch, to rhythm, to the experience of emotion—is thought to be shaped by the meeting and thwarting of expectations." In their paper, Judge and Nanay also point to "the need for increased collaboration between music psychologists and philosophers in order to arrive at a more detailed characterization of conscious musical experience and the role of expectations therein than has previously been offered". As we shall see, PP might be a very productive framework to do just that.

[11] See Meyer (2008 [1956]) and Huron (2006) for influential accounts of the role of expectations in all these facets of music listening. See also Pearce and Wiggins (2012) for a summary of the psychological, neuroscientific and musicological evidence on the importance of musical expectations.

[12] See e.g. Meyer (1957, p. 414): "Once a musical style has become part of the habit responses of composers, performers, and practiced listeners it may be regarded as a complex system of probabilities. That musical styles are internalized probability systems is demonstrated by the rules of musical grammar and syntax found in textbooks of harmony, counterpoint, and theory in general. The rule given is such books are almost invariably stated in terms of probability. For example, we are told that in the tonal harmonic system of Western music by the dominant, frequently by the subdominant, sometimes by the submediant, and so forth."

The intuition that music listeners command a probabilistic model of the statistical regularities of the sensory input has led in recent years to ever more refined work in computational modelling of the expectations of music listeners (see Temperley, 2007; Pearce and Wiggins, 2012).[13] In the last few years, computational models have been developed that can simulate reasonably well listeners' probabilistic expectations (i.e. the perceived predictability or surprisingness) of musical events. The most effective model to date, the Informational Dynamics of Music (IDyOM) model, has at its core a melodic pitch predictor that incorporates both a long-term and a short-term model. The long-term model is trained on all stimuli in a representative corpus of musical compositions, simulating the listener's whole prior musical experience; the short-term model instead is trained incrementally only on the current melody, simulating the listener's learning of statistical regularities specific to the current musical piece. Each model produces a distribution predicting each note as the melody proceeds, and the two distributions may be combined to give a final output that maps reasonably well with the listeners' own judgement about the subjective surprisingness of the musical event in question (Pearce and Wiggins, 2012; Hansen and Pearce, 2014).

Another interesting confirmation of the fact that music listeners are constantly tracking – and are sensitive to – the statistical properties of unfolding musical pieces comes from work in neuroscience. It has been known for some time that music tends to evoke electrophysiological responses in the brain whose amplitude seems to correlate with the surprisingness of musical events. These ERPs (the ERAN, the P3a, the P3b, etc.) parallels the N400 response to semantic deviations that we discussed in Chapter 2, and are normally interpreted as "learning signals", signs that the listener is indeed updating her probabilistic model of the cause of sensory inputs (see Koelsch, Vuust and Friston, 2019 for discussion). Moreover, recent fMRI findings (Cheung et al., 2019) show that uncertainty and surprise about musical events (as assessed by the IDyOM model) seem to modulate activity in certain brain regions (the amygdala, the hippocampus and the auditory cortex), another sign that the brain is sensitive to statistical properties of the musical inputs.

In sum, all the evidence from musicology, psychology, computational modelling and neuroscience seems to concur in confirming a fundamental role for predictions in music listening. These predictions seem to take the form of a multi-layered probabilistic model internalised by listeners and composers and put to work in the interpretation and creation of the unfolding musical piece. The developing PP story about music is the last in this long lineage of approaches that embrace these assumptions.[14] As compared to its predecessors, it stresses even more the exploratory component of music listening: in this perspective, engaged music listeners are avid

[13] Temperley (2007) makes explicit use of a Bayesian framework very similar to PP, on the three assumptions that "1. Perception is an inferential, multileveled, uncertain process"; "2. Our knowledge of probabilities comes, in large part, from regularities in the environment"; and "3. Producers of communication are sensitive to, and affected by, its probabilistic nature" (p. 3).

[14] See e.g. Koelsh, Vuust and Friston (2019), Cheug et al. (2019), Mencke et al. (2019), Vuust and Witek (2014).

information seekers, constantly generating and testing hypotheses about the structure of the unfolding musical narrative. In other words,

> when listening to music, we entertain a number of predictions, or hypotheses, about future musical events (e.g., in terms of meter, rhythm, melody, and harmony), which are resolved in the near future, usually within the next few tones. Music thus provides the opportunity to continuously resolve uncertainty over such hypotheses (cf. perception as hypothesis testing)… we derive pleasure from musical prediction errors, even if we know a musical piece, because they invariably resolve uncertainty about what we might have heard. (Koelsch, Vuust and Friston, 2019, p.73)[15]

As this passage makes clear, however, the PP story about music makes further claims, and it is on these further claims that we should concentrate, as they are the ones that are crucial for our present concerns. The idea is not just that a predictive process is involved in fundamental ways in the perception of music, or that music affords learning (in terms of the update of a probabilistic model of the causes of its sensory input). After all, if the general PP story about perception is on track, we are learning in that sense all the time. Instead, the thought is that "we derive *pleasure* from musical prediction error", or that, in other words, "music may… *elicit pleasure* by encouraging the listener to continuously generate and resolve expectations as the piece unfolds in time" (Cheung et al., 2019, p. 4087, my emphasis). To prove this latter claim, we need to show not only that engaging music affords learning, but that it is particularly well-suited to maximise learning as compared to other kinds of stimuli, and that we find it more aesthetically appealing and rewarding *for that reason*. Only then we would have a PP story about music and learning that parallels closely our proposal on literary language, narrative and visual art. Is there any evidence for that?

It turns out that there is. In Chapter 1 we mentioned the Wundt-Berlyne hypothesis according to which there is an inverted-U-shaped relationship between aesthetic appreciation and some "collative variables" (such as complexity, novelty, uncertainty, surprisingness and ambiguity). According to this hypothesis, as we said, an increase in these collative variables correlates positively with aesthetic appreciation up to an optimal point, after which a further increase reverses the effect. We noticed back then that in a PP perspective these collative variables can usefully be conceptualised in terms of subjective unpredictability (i.e., the amount of prediction error that the stimulus can produce). Once we do so, it becomes apparent that the idea of an optimum of unpredictability in engaging artworks is completely consistent with the idea that

[15] Compare with Temperley (2007, p. 3): "In listening to a piece of music, we hear a pattern of notes and we draw conclusions about the underlying structures that gave rise to those notes; structures of tonality, meter, and other things. These judgements are often somewhat uncertain… In the development section of a sonata movement, for example, we may be uncertain as to what key we are really in – and this ambiguity is an important part of musical experience. The probabilistic nature of music perception applies not only to these underlying structures, but to the note pattern itself; Certain note patterns are probable, others are not; and our mental representation of these probabilities accounts for important musical phenomena such as surprise, tension, expectation, error detection, and pitch identification."

engaging artworks maximise learning, because learning is maximal with stimuli that are neither too predictable, nor too unpredictable. In the case of music, for example, one should expect this subjective optimum to be somewhere between the opposite extremes of noise (where no statistical regularities can be detected) and complete repetition (where there is no new statistical regularity to be learnt), in a region where the statistical regularities of the musical piece are not obvious, but still learnable.[16] If this is the case, we may have found the relationship between music, learning, and aesthetic pleasure that we are looking for.

The inverted-U hypothesis has been tested extensively in music for several years, and, in general and despite some methodological difficulties, it seems to be a solid acquisition of music research. As one would expect, what constitutes the "optimal level" of unpredictability appears to be variable depending on expertise, personality, musical context and listening preferences (see Witek et al., 2014), but the overall pattern is robust. A recent meta-analysis (Chmiel and Schubert, 2017) conducted on fifty-seven studies examining the relationship between musical preference and unpredictability (and related constructs) found that fifty of them were compatible with the inverted-U hypothesis, five were mixed, and just two were completely incompatible with the hypothesis. Recent studies conducted explicitly from a PP perspective have also confirmed the preference for stimuli of intermediate unpredictability (Delplanque et al., 2019; Gold et al., 2019). Gold et al. (2019) explained the result by explicit reference to a "reward for learning". Overall, then, the picture on music and learning is clear, and consistent with our discussion of literary language, narrative and visual art. The music that engages us, the one that we prefer, it seems, is the one that can produce particularly rich sensory streams that disrupt and confirm our predictions strategically, providing us with a constant experience of cognitive gain. Effective, engaging art and learning appear once more to be connected by a fundamental relationship.

The story, of course, could be made more fine-grained. It would be interesting, for example, to go beyond just proving *that* music maximises learning to examine *how* it does so. Doing this would require a close analysis on the lines of those that we conducted for literary language, narrative and – to a lesser extent – visual art: an analysis that dwells more on the various tricks in the bag of the music composer, the ways she plays with our predictions to keep us engaged

[16] Talking from an information-theoretic perspective, Coons and Kraehenbuehl (1958, p.148) had already observed more than sixty years ago that "if a composition is to be effective, its pattern must be one that, first of all, attracts and holds the attention of the listener and, secondly, rewards the listener for his attention. It is evident that only something which is informative will attract the listener's attention. This means that the pattern must be as high-informed [i.e., unpredictable] as possible. On the other hand, the listener is rewarded by confirmations of his predictions. He would become frustrated and ultimately inattentive if a composition consistently nonconfirmed his predictions. This means that the pattern must be as low-informed [i.e., predictable] as possible. The paradox of artistic creation could be no more bluntly stated. It is necessary to make a pattern which is simultaneously both as high-informed and as low-informed as possible in order to accomplish the contradictory requirements for keeping a listener." In a similar vein, Huron (2006) notes that: "expectations that prove to be correct represent successful mental functioning. Successful predictions are rewarded by the brain" (p. 361); on the other hand, "although expected events are generally preferred, highly predictable environments can lead to reduced attention and lowered arousal – often leading to sleepiness" (p. 362).

throughout the musical piece.[17] Such an analysis could uncover the specific dynamics of prediction error minimisation in specific artworks, genres or styles, or the finer interplays between predictions at different levels of the perceptual-cognitive hierarchy. This more fine-grained analysis could also alert us to similarities between the ways in which the various arts engage our attention and shape our experience, opening new parallelisms that we were not able to see before. Here a host of interesting questions come to mind: can we get a better grasp of narrative tension by looking at the same phenomenon in music, and vice versa? Can the study of visual ambiguities elucidate certain dynamics of music perception? Is atonal music similar to non-figurative art in its refusal to adopt deeply engrained predictive routines? Such a line of inquiry would benefit decisively from the participation of both cognitive scientists and neuroscientists providing the required formal framework and humanists (aestheticians, art historians, narrators, painters, composers, etc.) offering their implicit knowledge and their wealth of phenomenological insights. In this interdisciplinary enterprise music could have a primary role, since, as we saw, it can boast both a large set of relatively well-defined expectations represented by music theory (a well-developed folk psychology, as we have noticed) and a set of well-known electrophysiological responses that seem to correlate with the statistical properties of musical events.

4.4 Exploring the Sensorimotor Space

So far, our developing PP story about art and learning has been focused on *perceptual learning*, that is, the kind of learning triggered by exteroceptive stimulation. I read a novel, see a picture, listen to a piece of music, and my brain is busy updating its multi-layered model of the underlying causes of such changing sensory inputs. On the other hand, we haven't said much – or anything at all – about *motor learning*, that is, the kind of learning that leads to successful and skillful action and that is concerned primarily with proprioceptive stimulation (information coming from our limbs and muscles). One of the considerable advantages of PP however is that it seems to blur the distinction between perception and action, allowing us to expand the story developed so far to activities normally not included within the purview of art and aesthetics, and to perceive a deep continuity between our engagement with art and sensorimotor exploration in general. This is one of the most promising and underdeveloped aspects of the PP approach to aesthetics and is the one we shall now discuss.

In the familiar "sandwich model of the mind" adopted by a good portion of work in classical cognitive science,[18] perception and action are clearly distinct and separated from each other by the inner filling of cognition. The idea is, roughly, that perception involves a bottom-up reconstruction of the features of the external world and its transmission to the cognitive system,

[17] For some suggestions in this regard, see Temperley (2014, 2019).

[18] See Hurley (1998) for an early critique of this model.

whereas action involves the generation of motor commands in the cognitive system and its top-down transmission to the motor plant. We have already seen how PP usefully departs from this story with respect to perception: perception is not passive feature detection driven from the bottom up, but rather an active, multi-layered process of prediction of the incoming stream of sensory stimulations. Now, it turns out that, in PP, action can be treated with the same predictive machinery. In this perspective, actions are not motor commands but rather predictions about expected proprioceptive sensations. In other words, performing an action is not executing a motor command, but rather predicting a certain pattern of proprioceptive stimulations, and then fulfilling these predictions by moving your limbs in the appropriate way. As with perceptual predictions, these motor predictions are organised hierarchically,

> with abstract goals and intentions at the top to be unpacked into policies (action sequences) and further into specific motor programs represented as the exteroceptive and proprioceptive outcomes that they are expected to realize. At the lowest, peripheral levels, proprioceptive predictions are confronted with the current state (before any movement) of muscle and tendon receptors to form prediction errors… here prediction errors trigger classical reflex arcs executing the movement. (Van de Cruys, Bervoets and Moors, forth.)[19]

The parallels between perception and action in the PP story are evident. In vision, according to PP, a high-level prediction that I will see, say, a dog gives rise to "lower-level predictions about limbs, eyes, ears and fur, which then cascade further down in terms of predictions about colours, textures and edges, and finally into anticipated variations of brightness across the visual field" (Seth, 2021, pp. 107-8). In the case of action, a high-level intention to, say, take a sip from the mug in front of me gives rise to lower-level predictions about specific motor programs (extend the harm, grasp, raise, bring to the mouth, etc.), which finally cascade down in terms of anticipated responses from muscle and tendon receptors; predictions are then updated in a recurring fashion throughout the hierarchy until the divergence between what I expect and what I am getting proprioceptively from muscle and tendons receptors is minimised and I feel that my limbs are behaving the way they should.

This blurring of perception and action seems to be confirmed by neuroanatomical evidence. The traditional sandwich model would predict anatomical asymmetry between perceptual and motor pathways. In fact, however, "the descending projections from motor cortex share many features with top-down or backward connections in visual cortex" (Adam, Shipp and Friston, 2013, p. 611), a fact that seems to suggest that

> the primary motor cortex is no more or less a motor cortical area than striate (visual) cortex. The only difference between the motor cortex and visual cortex is that one predicts retinotopic input while the other predicts proprioceptive input from the motor plant. (Friston, Mattout and Kilner, 2011, p. 138)

[19] See also Friston et al. (2010) for a detailed presentation.

As such,

> The perceptual and motor systems should not be regarded as separate but instead as a single active inference machine that tries to predict its sensory input in all domains: visual, auditory, somatosensory, interceptive and, in the case of the motor system, *proprioceptive*. (Adams, Shipp, and Friston, 2013, p. 614).

The overlap and interplay between perception and action are made stronger and more complex by the fact that, of course, each needs the other to function properly. We are constantly using exteroceptive predictions to guide action and proprioceptive predictions to guide perception. If I want to grasp this mug, I need to pay attention not just to the flow of proprioceptive stimulations but also to the flow of visual and somatosensory stimulations that tells me if my movement is appropriate. On the other hand, if I want to successfully predict patterns of stimulation in my visual or somatosensory field, I need to take into account the (predictable) changes that my own movements produce on the flow of stimulations. The result is that each act of sensorimotor exploration involves complex interactions between perceptual and motor systems predicting their input in the respective domains and working in tandem under the common general imperative of prediction error minimisation.

In light of the above account of action and motor control, a story about motor *learning* becomes available that parallels the story about perceptual learning we have been exploring so far. In particular, it becomes clear that motor learning consists in acquiring and updating a probabilistic model that accurately predicts the proprioceptive and exteroceptive consequences of your own movements so as to bring these consequences about ever more accurately. It also becomes clear that, just as not all percepts are equal with respect to the perceptual learning that they afford, not actions are equal with respect to the motor learning that they afford. Motor tasks that are too easy or too difficult to accomplish do not afford maximal learning. As with any other kind of learning, motor learning will be maximal when prediction errors are present but reducible.

Evidence suggests that it is precisely by concentrating on these kinds of actions that babies learn to master increasingly complex motor behaviours. A newborn infant does not just have to pick up and internalise causal regularities in her environment. She needs (perhaps more urgently) to pick up and internalise the causal regularities occurring in her own body, and between her body and the environment. The “motor babbling” displayed by infants may effectively be seen as a means to achieve just that. It is a form of active experimentation by means of which the infant acquires an increasing awareness of the exteroceptive and proprioceptive consequences of her own movements (“what sensory feedback do I get if I move my eyes or my fingers or my tongue just like that?” see Schmidhuber, 2010). The movements are therefore quite random at the beginning, but then through repetition and Hebbian learning the child gradually acquires a forward model that associates action plans and their expected sensory consequences.[20] If the child is to acquire ever complex motor behaviours, however, she needs to move in a particular way. Performing always the same movement would not afford further learning, because the proprioceptive and exteroceptive consequences of the movement will be always very similar and will soon become predictable. At the same time, the child is normally unwilling to embark in motor tasks that do not yield promise of an approachable solution, as this too would slow down their learning (for example, children do not try to stand before having learnt how to roll over, crawl and sit). If actions are experiments, or hypotheses about the occurrence of certain sensations, it makes sense for the child to work where the scientist would: at the edge of current understanding. What we noticed in Chapter 1 about the information-seeking behaviour of children seems to be true also in the motor domain: children tend to stay in what Vygotsky called the "zone of proximal development", trying to perform actions that are just above their current capacities. By staying in this space, they manage to learn at an optimal rate and to master increasingly complex actions at a rapid pace.

The intuition that infants are optimal learners lies at the basis of an interesting strand of work in developmental robotics. Here the attempt is to build artificial agents that can simulate the exploratory behaviour of children and autonomously develop an internal model of their body and their environment in an optimal fashion (Oudeyer and Kaplan, 2007; Oudeyer, Kaplan and Hafner, 2007; Schmidhuber, 2010). Quite in line with the PP picture, robots in this kind of research are normally equipped with an architecture comprising two modules (see Oudeyer, Kaplan and Hafner, 2007): a module "M" that learns to predict the sensorimotor consequences of the robot's actions, and another module "metaM" that learns to predict the error that M makes in its predictions and uses this error as a measure of the potential interestingness of the situation, steering the robot towards regions of its sensorimotor space that afford further learning. Interestingly enough, robots equipped with this kind of architecture display an "artificial curiosity" and an "intrinsic motivation" to explore their environment that parallel in many striking ways the behaviour of children.

Equipped with this picture of motor learning, we might finally move towards the promised extension of our PP story about art and learning. To do so, we should try to answer two key questions: 1. Is that sweet spot of unpredictability that maximises learning pleasurable and rewarding in motor activities as it seems to be in the case of perception? And 2. Are there activities explicitly designed to maximise motor learning, in the same way as paintings and pieces of music seem to be designed to maximise perceptual learning? If so, our engagement with the arts might seem less peculiar than what we initially thought and might be conceived as part of a more general tendency towards learning and exploration of our sensorimotor space. The answer to both questions seems to be yes.

To see this, we might usefully refer to the study of what the psychologist Mihaly Csikszentmihalyi called "flow states". Flow states are rewarding psychological experiences that

[20] See Piaget (1952) for a classic account of this process and Clark (2013, p. 134) for how this might work from a PP perspective.

tend to occur when subjects are engaged in activities just above their current skill level. When engaged in such activities, subjects report a higher involvement and concentration, a sense of fulfilment and an intrinsic motivation to continue the activity in question (see Csikszentmihalyi, 1990). This rewarding condition is maximal, Csikszentmihalyi insists, when the challenge offered by the situation is not too easy but also not too difficult. This is because, according to him, these are the situations that afford maximal growth and learning.[21] Thus, for Csikszentmihalyi, exploratory behaviour can be explained by a motivation for reaching situations that represent a learning challenge. Subjects are normally aware of whether their current experience is optimal or suboptimal in this respect, and they will be motivated, in the execution of motor tasks, to seek situations of the right level of complexity to remain in the flow state. Csikszentmihalyi exemplifies this careful dosing of the complexity of the situation with the case of a tennis player, represented in the figure below (1990, p. 74).

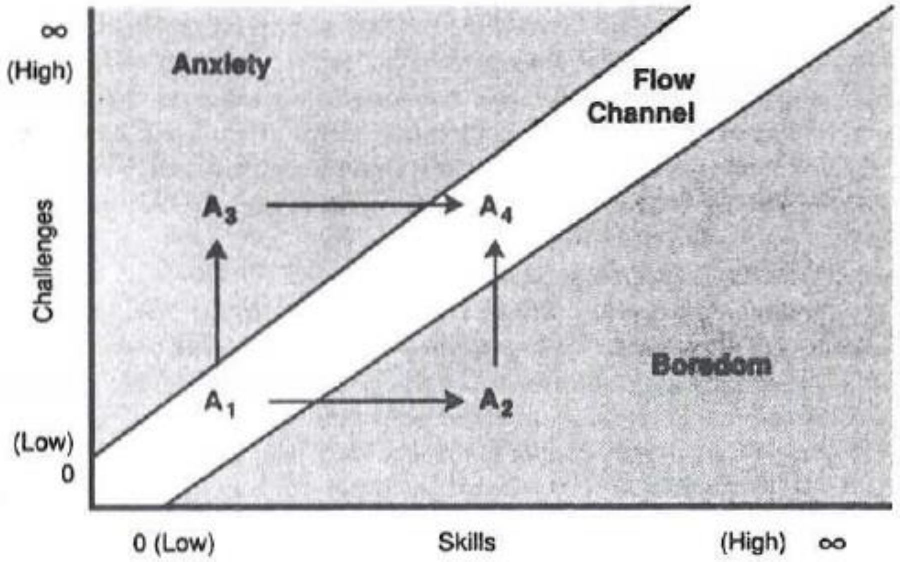

Figure 13: The flow channel (from Csikszentmihalyi, 1990, p. 74).

To the novice that just started playing (A_1 in the schema), even rudimentary actions such as hitting the ball over the net will appear challenging. In PP terms, the subject does not yet command a good probabilistic model of the sensory consequences of her actions. As the subject keeps practicing, one of two things happen: either she finds that the situation is becoming less

[21] See Csikszentmihaly (1990, p. 74): "In our studies, we found that every flow activity, whether it involved competition, chance or any other dimension of experience, had this in common: It provided a sense of discovery, a creative feeling of transporting the person into a new reality. It pushed the person to higher levels of performance, and led to previously undreamed-of states of consciousness. In short, it transformed the self by making it more complex. In this growth of the self lies the key to flow activities." One might wish to compare this description with the cognitivist's claims, mentioned in Chapter 1, about the cognitive benefits that fiction or art would provide.

and less challenging (the sensory consequences of her actions become more and more predictable) in which case she will get increasingly bored (A_2); or she finds that the challenge is too demanding for her current level of skill (the sensory consequences of her actions keep being surprising) in which case she will get increasingly anxious and frustrated (A_3). At this point, if she does not choose to stop playing altogether, she will be pushed to respectively increase or decrease the level of challenge (making the flow of sensory stimulations more or less unpredictable) to remain in the flow channel (she might do so, for instance, by changing her opponent, or focusing her practice on an easier/more difficult set of movements). If she does so, she will soon reach a point A_4, which, compared to her initial state A_1 is characterised by an increase in the complexity of the sensorimotor contingencies that she can master. In other words, she has learned something. If all goes well, A_4 will then become the starting point for further learning, growth and discovery. In this way, pretty much like the baby selecting the toys to play with, the tennis player scaffolds her own developmental parable, displaying a taste for situations that allow for optimal development.

Of course this picture is supposed to hold true for all motor tasks, not just sports: whether it is rock climbing, playing an instrument, driving, or riding a bicycle, studies and qualitative reports in the flow literature seem to suggest that an activity is pleasurable and rewarding when it affords optimal learning. This seems to be in line, I think, with our ordinary intuitions about what makes an action rewarding. Anyone who has tried to learn how to play an instrument or how to drive a car will be familiar with the frustration of not being able to line up your intentions (your predictions about how your body should behave) with the actual proprioceptive or exteroceptive feedback that we get from sensory receptors (the out-of-time stroke on the snare drum, or the roar of the engine when the clutch pedal has not been raised enough). Anyone will also be familiar with the inner click (the equivalent of an "Aha!" experience in the motor sphere) that takes place when a complex motor task is finally accomplished, and then mastered with increasing accuracy – until the moment in which it ceases to be noticeable, and to give us pleasure. We seem to get the peak of involvement and pleasure when we are progressing, when we are learning optimally, in the sense that we have been defining so far.

With their phenomenological plausibility, therefore, the studies of flow activities complement nicely the story we have been telling so far about motor learning, but they also add a critical piece to it. They suggest that we not only seek, from the very beginning, conditions in which motor learning is optimal, but also that we *like* these conditions, that we see them as rewarding and entailing a pleasant phenomenology of involvement, fulfilment, growth and discovery. Quite evidently, the conclusions of the flow literature are compatible with the general PP story about learning and well-being – the idea that we search for good slopes of prediction error reduction (see Chapter 1, section 1.3). For our present discussion, it should be noted how strikingly this story about motor learning and its pleasures matches what we saw in the perceptual domain. The idea that there is, in motor tasks, an optimal condition between boredom and anxiety constitutes an almost literal reinstatement in the motor domain of Gombrich's principle that aesthetic pleasure lies "between boredom and confusion". It parallels not only our discussion of visual and auditory learning, but also our conclusions about how successful literary texts and narratives engage their readers. This, I think, makes typical occasions for skilful actions (sports, instrument playing, etc.) the motor analogous of novels, paintings and music, and confers to our whole story a pleasing generality.

Despite its intuitiveness and phenomenological plausibility, we might however demand some more empirical confirmation for this pleasingly general story in the motor domain. Apart from the literature on flow states, empirical research on the pleasure or motor activities is still very much in its infancy, but other useful indications seem to come from the psychology and neuroscience of dance. Dance is particularly interesting in this respect: as a systematic and purposeful organisation of movements, one would expect it to be to ordinary motion what music is to random environmental sounds and literary language is to everyday communication: it should provide to our probabilistic models a particularly patterned and rich flow of proprioceptive stimulations to be predicted. Our hypothesis about motor learning and pleasure would suggest that dance movements of an intermediate level of complexity (relative to the dancer's current ability) should provide the most pleasurable and engrossing experience.

This seems to be true, according to some recent empirical studies that have investigated the relationship between rhythmic complexity, desire to move and pleasure in music (Witek et al., 2014; Matthews et al., 2019, Matthews et al., 2020; Vander Elst, Vuust and Kringelbach, 2021). The main finding in this case is that the desire to move in reaction to music and the pleasure we get from it is maximal with music with intermediate levels of rhythmic complexity. Very simple, highly predictable rhythms with no syncopations or very complex, highly unpredictable rhythms with many syncopations are both less likely to induce body movements or pleasure than rhythms with intermediate levels of syncopation. These latter are also associated with enhanced activity in brain regions linked to reward, as well as prefrontal and parietal regions implicated in generating and updating stimuli-based expectations (Matthews et al., 2020). These results are of course in line with the story we have been telling so far. From our perspective, it is to be expected that rhythms with intermediate complexity (relative to the subject's current abilities) elicit a stronger pleasure and urge to move since they are the ones that afford optimal learning, both in terms of motor learning strictly conceived (learning to predict proprioceptive signals) and in terms of sensorimotor coupling (learning to integrate exteroceptive and proprioceptive signals). Further studies will have to address the question of whether the pleasure we get in these circumstances is due to the optimal update of proprioceptive predictions (as our story about motor learning would suggest) or just to the optimal update of exteroceptive predictions about the unfolding musical piece. For now, however, these studies on dance seem to provide another useful indication of the relationship between motor learning and pleasure.

Let us add a last piece to the puzzle before wrapping up. So far, we have talked about the pleasure and cognitive gain involved in *performing* actions of the right level of complexity. But what about *perceiving* the actions of others? Another interesting aspect of the PP story we are considering is that it seems to account for both the execution and the comprehension of actions within the same explanatory apparatus (see Kilner, Friston and Frith, 2007, and Friston, Mattout and Kilner, 2011). In this perspective, once you acquire a model that associates your motor plans with some expected sensory consequences, you can leverage this model to guess the motor plans of others based on the actions you are currently observing. In other words, within this scheme (which seems to be compatible with what we know about the mirror neuron system), in interpreting the motor plans of others, we are effectively trying to guess "the most likely cause of an observed action… by minimizing the prediction error at all levels of the cortical hierarchy that are engaged during action observation" (Kilner, Friston and Frith, 2007, p. 165). If this is true, then many interesting hypotheses become readily available based on what we have said so far. In particular, it becomes clear that certain observed sequences of actions will be more unpredictable than others, and that the predictability of each sequence will vary over time, creating the same ups and downs of uncertainty, the same dynamics of stabilisation and reopening that we encountered when discussing literary language and narrative.[22] It also becomes clear that there might be an optimum of unpredictability in the action sequences of others that allows for optimal learning (see Cross et al., 2011 and Leder, Bär and Topolinski, 2012 for preliminary evidence that points in this direction). It becomes then reasonable to assume that skilful and dextrous actions carried out by others elicit our pleasure and appreciation because they maximise our learning. More generally, in this perspective the movements of others (and our own too) immediately acquire an evaluative, aesthetic dimension, since in interpreting them we are constantly evaluating how people *ought to behave* given the goals that we are currently attributing to them.[23]

These latter hypotheses, as promising as they might be, have for the moment no higher status than that of vague suggestions. What is sufficiently clear, however, is that the story we have told on sensorimotor exploration lines up very well with our conclusions about literature, narrative, visual art and music. The upshot is a picture with a pleasing generality. If we consider art as providing paradigmatically good instances of growth and learning, we must admit that our treatment of motor learning expands considerably the scope of the "aesthetic", unveiling similarities between apparently distant phenomena (like playing tennis and reading a novel) that might turn out to be engrossing and gratifying for the same underlying reason. The consequences of this general picture are what we shall now try to assess, in the next and last chapter of this work.

[22] Evidence that we might be sensitive to the dynamics of uncertainty in observed actions as we are in language and narrative might come from film editing. It is well-known for example that cuts made when the subject in the frame is carrying out a body movement tend to be perceived as more or less abrupt depending on when exactly they are made (cuts made in the middle of a movement are more abrupt than cuts made when the movement is completed). This seems to suggest that we do sense when our inferential attempts to understand bodily movements open up (i.e., uncertainty increases) or stabilise (uncertainty decreases).

[23] Expanding our story to include inferences about motor intentions could have repercussion for our understanding of the other arts too. Marks and brushstrokes on a canvas, deformations in the metal or stone of a sculpture or certain acoustic features of instrumental and vocal music become as many signs of the physical action that was required to produce them, thus encouraging a vast array of inferential processes. This would bring our story close to those accounts that stress the importance of "embodied simulation" in aesthetic experience (see e.g. Freedberg and Gallese, 2007; Currie, 2011; Kirsch, Urgesi and Cross, 2016; Gallese, 2017).

5. Art and Learning

5.1 A Fundamental Relationship

After having examined various arts from a Bayesian/PP perspective, it is now time to bring our inquiry into art and learning to a close, seeing whether what we said in the previous chapters warrants or suggests some general conclusions.

One of the points that I think emerges more clearly from our discussion – and quite independently from the details of particular proposals – is that our fascination with art could be part of a fundamental drive towards pattern-finding, or “learning” in a psychological and neuroscientific sense. We have seen that there are several reasons to think that humans tend to concentrate and actively search for regions of their input space that allow for optimal assimilation of new causal patterns. This tendency seems to orient our behaviour as information-seekers from the very beginning, influencing what stimuli infants tend to focus on, how they approach motor tasks, and how they manipulate objects during play. As adults, as we have seen, this drive continues to define what sort of stimuli and situations we find rewarding and are intrinsically motivated to pursue. This suggests that humans must be sensitive to the dynamics of their learning progress: they must be able to assess whether and in what measure what they are perceiving is disclosing – or might disclose – new causal patterns. In other words, we must be endowed with a sense of the worthiness of what we are perceiving that steers us towards regions of the world that allow for the fullest exercise of our cognitive capacities (insofar as perception, cognition and learning are processes of probabilistic modelling of our world).

Now if artworks are regions of the world of this kind, if they are “progress niches” as we have tried to show throughout this work, then our attraction towards them might be less strange than it seemed at first. Our tendency to be engrossed and fascinated in the contemplation of effective artworks is just part and parcel of how we function as good information seekers – agents that seek out new learnable causal patterns and avoid spending time both where causal patterns are already learnt and where they are unlearnable. As artists and art consumers, our behaviour would be in continuity with the exploratory drive of children, which directs them towards parts of the world that can foster their flourishing. The sense of worthiness that we have while engaging with art would be due to our intuitions about what is fuelling our learning trajectory.

This conclusion, as we know, vindicates a line of thought that runs throughout the history of aesthetics and is very much alive in contemporary empirical research, a line of thought that sees the pleasure of art as fundamentally linked with learning and knowledge acquisition. The intuitions of Aristotle, Baumgarten, Kant and Dewey mentioned in Chapter 1 may find now, I think, a convincing reformulation. Aristotle, as we saw, attributes the pleasure of the imitative arts to the discovery that they afford that their objects have been properly represented. Kant suggests the experience of the beautiful is linked with the intuition of “a formal purposiveness” in the manifold of experience; Baumgarten and Dewey conceive aesthetic experiences as being in continuity with the processes by means of which we give structure to our sensations. In continuity with these views, our story suggests that we take pleasure in discovering how stimuli ought to be organised, i.e. what is the causal pattern that underlies them. Art happens to afford this pleasure to a higher degree. The pleasure afforded by art is therefore the pleasure of making sense of what lies before us: it's the pleasure of successful cognition.[1]

If this story is on track, then, art is indeed fundamentally related to learning and knowledge acquisition: a relationship we were initially unable to see because of the way in which knowledge and learning are characterised in the contemporary philosophical debate on the cognitive value of art. If knowledge is characterised as the possession of justified true belief and learning as the acquisition of such beliefs, it is indeed difficult to see in what sense art affords learning and knowledge acquisition to an enhanced degree. Once we start conceiving learning in terms of the modification of a probabilistic model of the world embodied by our brain structures and dynamics (in line with a major strand of work in contemporary cognitive science), the relationships between art, learning and related phenomena (curiosity, exploration, information-seeking, motivation) become evident. This shift in perspective should influence the way in which the question of the cognitive value of art is normally approached in the philosophical debate. The stress on truth and justification, the reliance on the conceptual apparatus of the epistemologist and the disregard for the empirical and phenomenal aspects of learning should give way to the dialogue with disciplines that have a lot to say about learning, perception and cognition. From psychology, neuroscience and cognitive science, the philosopher can draw important suggestions on the cognitive gain provided by art and why we value it at all, suggestions that can in turn clarify what we expect from artworks, how they tend to be structured, and how they make us feel.

An interesting consequence of the relationship between art and learning that we have traced is that it can also be examined in the other direction. If the pleasure and allure of art are due to its capacity to provide enhanced learning experiences, then enhanced learning experiences have also in them something of the artistic and the aesthetic. If, that is, the pleasure and allure of art are due to the discovery of lawful structures in our sensorium, then all perception and cognition, insofar as they involve such a discovery, take on the traits of an aesthetic experience. Here the Bayesian/PP story we have articulated makes contact with the vexed question of the scope of the "aesthetic", and raises a potential objection: if the dynamics of ambiguity reduction that make for a pleasurable aesthetic experience are the same dynamics that make possible experience as such, then there is no principled distinction between something called "aesthetic experience" and some other forms of experience. Experience, it seems, is always aesthetic.[2] But if this is the case, then our story has not explained what it aimed to explain: it has not told

[1] More on what I mean by "successful cognition" in sections 5.2 and 5.3 below.

[2] This echoes once again Dewey's picture of (aesthetic) experience. See Dewey (2005 [1934]). Following Dewey, Johnson has recently argued to the same effect that "aesthetics… extends broadly to encompass all the processes by which we enact meaning through perception… In other words, all meaningful experience is aesthetic experience" (2018, p. 2). He adds: "the arts are therefore exemplary modes of meaning-making, because they give us intensified, nuanced and complex realizations of the processes of meaning in everyday life" (2018, p. 25).

anything about the fascination, allure and fulfilment that artworks *qua* artworks seem to provide.[3]

Our discussion in the previous chapters should have already provided us with an answer to this objection. If our Bayesian/ PP story suggests that experience is always aesthetic, it also gives us all the tools to explain why not all experience is aesthetic *to the same degree* – why, that is, our everyday, mundane experience is not constantly imbued with the pleasure, awe and wonder that we get from engaging with great art. Here we might usefully return to what we said about Figure 3 in Chapter 1 and repeated throughout this work: most percepts are just too consistent with our predictions to surprise us much; others are just too surprising to be made meaningful. Only few percepts afford the right level of reducible ambiguity required to prompt a discernible feeling of cognitive gain; even fewer provide a coherent, organic flow of such moments of disruptions and readjustments of our expectations capable of generating the powerful impression of continuous cognitive progress. The PP proposal about aesthetics is that a novel, a painting, a piece of music, when successful, are such percepts for a specific observer. They are "progress niches", portions of our world designed to leverage our permeability to the statistical structure of our environment and promote further learning. This is what makes them special and sets them apart from most of our ordinary encounters with the world. If everything that is not completely expected or completely random will engender learning, artworks are designed to maximise this experience, a goal they achieve by means of the kind of stratagems that we have been examining in the previous three chapters. In doing so, they provide us with better experiences, experiences more vivid, structured and organised, regimented to enhance our sense of progression. Here we might have a principled explanation of the often-made remark that in art ordinary words, sounds, objects, materials are experienced "as if for the first time", with the vividness and sensory plenitude of the first encounter.

Recognising the fundamental continuity between art and the more mundane cases of meaning-making does not deprive aesthetic experiences of their particularity, nor does it threaten the specificity of aesthetics as a domain of inquiry. Aesthetics, one might say, considers experience in its most adventurous conditions. Its purview is composed by those cases where perception and cognition are progressing exploratorily, by means of tentative hypotheses: poetry more than ordinary speech, narrative more than chitchat, dance more than ordinary movement. This without failing to recognise, however, the continuity between the tentative and the mundane, together with the fact that the boundary between the two is always a movable one, and that, given enough time, the first fades into the second. This is the most reasonable way, I think, to approach the vexed question of aesthetic experience, acknowledging what is particular about paradigmatic encounters with great artworks without succumbing to the temptation to consider such encounters as belonging to some special, qualitatively different realm of experience.

[3] It is indeed an often-repeated platitude among philosophical sceptics that psychological and neuroscientific explanations of how artworks work as perceptual stimuli, being equally valid for all perceptual stimuli, cannot contribute to our understanding of how artworks work *qua* artworks (see e.g. Hyman, 2010; Noë, 2015). As we shall see, the story we have articulated seems to avoid these objections.

This continuity between effective art and enhanced learning experiences opens up several interesting related conjectures that will need to be explored in-depth in future work. For example: if our tendency to be fascinated by artworks is in continuity with the drive of children towards stimuli that maximise their learning, is there not in this latter something aesthetic? Are not children, when selecting which stimuli to attend to, which toys to play with or which motor task to accomplish, exercising an aesthetic sensitivity? If humans more generally are inclined towards shaping their perceptual encounters to maximise learning, are they not effectively "artifying" their sensory flow, trying to produce the kind of experience that they encounter in art in a privileged way? Are therefore agents that maximise their learning transforming their sensorium in a work of art? And if this is the case, where does the artistry ultimately lie? In the object or the perceiver, in the artist or the art consumer? These questions will have to be addressed more systematically elsewhere. It suffices for us here to have established with some plausibility the guiding hypothesis that we formulated at the beginning of our inquiry, namely that art and learning are linked by a fundamental relationship – one that makes our engagement with art part and parcel of our drive towards learning and gives situations that maximise learning the traits of paradigmatic aesthetic experiences.

5.2 The Feeling of Having Learnt Reconsidered

We started our inquiry into art and learning in Chapter 1 by trying to characterise, as well as we could back then, the feeling of having learnt that accompanies our engagement with successful artworks. It is this feeling, we suggested, that is primarily responsible for the widespread intuition that we learn from art, and it is this feeling that we elected as our explanatory target. Of this feeling, we tried to isolate a few distinct but related features. We said that the engagement with successful artworks is *self-sustaining* and *self-perpetuating*, that it has the traits of a *transformative experience*, that it has the feel of a *cognitive improvement*, and that it is *pleasurable*. In the course of our inquiry, did we gain some insight into these features of our experience of art? Can our Bayesian/PP story account for these important facts about the phenomenology of our aesthetic encounters?

Let us start with the self-sustaining and self-perpetuating character of our engagement with art. When an artwork is acting effectively, we said, we are compelled to devote to it our serious and sustained attention. We feel that the artwork is continuously disclosing – or might still disclose – something new, that there are still things that it has not told us yet and might tell us if we keep attending to it. As a result, we are compelled to prolong our engagement with it. The contemplation of beautiful objects, says Kant, "strengthens and reproduces itself" (1987 [1790], §12). In our proposal, this feature of our engagement with art receives a straightforward explanation. Our whole discussion in the previous chapters points to the fact that effective artworks are those that yield promise to disclose new learnable causal patterns. Artworks are stimuli that look "like they are going to make sense": they suggest that what now looks disordered and random will turn out, and is gradually turning out, to be orderly and structured. It is by the promise of this achievable-but-yet-to-be-achieved order that successful artworks maintain our sustained attention and keep us asking for more. If the story we have told is on track, in fact, humans have a sense of whether they are progressing in making their environmental stimulations more orderly and predictable – a sense that expresses itself in perceivable patterns of tensions and resolutions, openings and closings, moments of puzzlement and moments of revelation as our guesses about the causal structure of our world become more or less likely. What the Bayesian/PP apparatus has allowed us to do is just to capture in a principled way these implicitly sensed dynamics. It has endowed us with tools to follow with some precision the guessing process by means of which we distil causal patterns in our sensorium at different levels of abstractions. This has allowed us to ascertain that artists take great pains to provide us with consistent and enhanced experiences of successful discovery of causal patterns, presumably exploiting their own implicit knowledge of these dynamics. Looked through Bayesian/PP lenses, literary language, narrative, visual art, music and certain motor activities all stand revealed as ways to capture and retain the attention of beings like us, geared towards optimal information seeking.

The transformative character of our engagement with art also seems to receive, in our picture, a clear characterisation. Art, we were told, challenges our existing view of the world, leading us to revise or expand it in significant ways. As a result, after the experience, we feel that the artwork has left us not quite the same as we were before. If our PP story is on track, this might be exactly true. Remember that, according to this story, we are probabilistic models of our world. Our brain embodies in its structure and dynamics a plastic and everchanging model of how we expect things in the world to be. Now for such a model to change (to be "transformed"), a particular kind of environment must obtain. If environmental stimulations remain close to what the model predicts, then quite evidently the model does not change much; if on the other hand environmental stimulations appear to be non-modellable, then the model cannot change much either. Again, we make progress as models of our world in environments that present new but graspable patterns – environment, that is, in which we can put our modelling capacities to work. Now, if humans are experts at creating such environments ("progress niches", as we called them) for themselves, and if artworks are a preeminent example of such environments (as our entire discussion has aimed to show), then artworks are also preeminent vehicles for self-transformation. By exposing ourselves to these peculiarly effective progress niches, we are turning our minds into perpetually moving targets.

It is important to stress that this process has powerful existential ramifications, some of which might get lost amidst the technical talk of probabilities and Bayesian belief updating. If we are models of our world, in modelling it we are also creating ourselves. In other words, in giving structure to our sensorium, we are also determining what sort of creatures we are. This means that perception and cognition are effectively processes of self-discovery and self-determination, in which we are trying to establish at the same time the shape of our world and our own identity as models of our world. The oscillations in uncertainty we have been examining in the previous chapters, then, acquire an existential connotation: experiencing increasing uncertainty about the causal structure of our sensorium is also experiencing increasing uncertainty about what we are, and reducing the uncertainty about the causal structure in our sensorium means also discovering what we are. Art, therefore, in allowing us to exercise our modelling capacities, is affording us experiences of self-discovery of this kind, where in structuring the stimuli in question we are structuring ourselves, gaining repeated intuitions of our own evermoving nature.[4]

This should have also made clear why a successful engagement with art is not just lived as a transformation, but as a transformation for the better, or why, in other words, the shifts in worldview that great artworks bring about have the taste of an improvement and a success. In making our sensorium more orderly and predictable, we succeeded in establishing ourselves as viable points of view on the world. We have not only given shape and consistency to what is out there: we have given shape and consistency to ourselves too. It makes sense therefore for such a process to be considered a success, and to be felt as such. Whether the particular shape we have assumed as a result of the experience (the particular configuration of our brain structures and dynamics) will turn out to be adaptive and advantageous over the long term, we cannot say. As we noticed, there is nothing in the brain marking our getting closer to the truth. What the brain appears to be exquisitely sensitive to is the predictability of its sensory stimulations. So if we are getting signals that our sensory world is becoming more orderly and predictable, we cannot but interpret this (in ways no doubt complex and that remain to be clarified) as a sign that the point of view that we have established on the world is holding, and becoming more and more likely. As "self-evidencing" creatures (Hohwy, 2016), embodied models of the world constantly looking for confirmation, we are getting confirmations of the fact that we exist.

What we have said so far should also have made clear however that the success we are talking about cannot really be characterised as simply "cognitive". If perception and cognition are always informed by this underlying existential concern, then one should expect them to be deeply imbued with affect. In particular, if our existence as embodied models of the world is at stake whenever we are searching for patterns in our sensorium, then one should expect failures to find them to be accompanied by negative affect, and successes to be accompanied by positive affect.[5] In Chapter 1 (section 1.3.3) we have seen that this is in fact the hypothesis of a growing stream of research that is linking affective valence to the dynamics of the reduction of uncertainty about

[4] In his lectures on aesthetics, Hegel seems to suggest something similar when he observes (1975, p. 31): "man brings himself before himself by practical activity, since he has the impulse, in whatever is directly given to him, in what is present to him externally, to produce himself and therein equally to recognise himself. This aim he achieves by altering external things whereon he impresses the seal of his inner being and in which he now finds again his own characteristics. Man does this in order, as a free subject, to strip the external world of its inflexible foreignness and to enjoy in the shape of things only an external realization of himself. Even a child's first impulse involves this practical alteration of external things; a boy throws stones into the river and now marvels at the circles drawn in the water as an effect in which he gains an intuition of something that is his own doing. This need runs through the most diversiform phenomena up to that mode of self-production in external things which is present in the work of art."

[5] As Van de Cruys (2017, p. 16) notes, in PP "value and information are intertwined by construction — courtesy of our existence as biological organisms. Taking the organism as an (extensible) model of its environment, epistemic coherence is paramount and emotions emerge as the dynamics of attaining this predictive coherence or error reduction.".

the causes of our sensory states. According to this developing view, as we said, the pleasure of finding a structure in our sensorium (of which the paradigmatic “Aha!” experience would be just one of the most noticeable instances) has an existential connotation: it is the pleasure of a creature that is succeeding in maintaining itself as a viable model of its world. Now we are able to see that this same line of reasoning might go a long way in explaining the last feature of a successful engagement with the arts that we are considering: the fact that it is pleasurable. It has been lamented that all too often aesthetic appreciation is conceived as a cold intellectual game, disconnected from our bodily and affective concerns (see e.g. Shusterman, 2012). But if we adopt a PP perspective, we certainly do not run this risk. The rich possibilities of pattern-finding afforded by art are in this view as many possibilities for existential realisation, with the pleasurable affective correlates that this entails. What we call aesthetic pleasure would then be the fulfilment of a creature that has reaffirmed its existence as a viable model of the world.[6] The pleasure we get from the arts would not be disconnected from bodily concerns but rather informed and generated by them.

These latter considerations may prompt the aesthetician to ask a further question: where does our PP story about art and aesthetics leave us with respect to the vexed question of whether our aesthetic encounters are “disinterested”, or “pursued for their own sake”? After Kant (1987 [1790]), it is customary (even if very controversial) to think that judgements of taste and the experience on which they are based are free from any connection to interest. Does PP have anything to say on the matter? On the face of it, the PP story seems to refute the Kantian idea of a disinterested engagement with the arts. According to the PP story, it seems, no experience is disinterested. Perception and cognition are always informed by the agent’s existential concern of maintaining its viability as a model of the world. Aesthetic pleasure is what we get when this concern appears to be satisfied, and aesthetic taste is what directs us towards regions of the input space that can satisfy it. In approaching and enjoying art, then, we are manifesting not just interest, but the strongest interest we could possibly have. On the other hand, however, if this story is on track, to ask why we care about art is tantamount to asking why we care about maintaining ourselves as viable models of our world, or, in other words, why we care about existing. And the continuation of our existence is both something we are deeply interested in pursuing and something we arguably pursue for its own sake, as an end in itself. The empirical observation that we like stimuli and activities that are in that sweet spot of unpredictability (of which art is, as we by now know, a preeminent example) and that we pursue them with intrinsic motivation and with no further aim might therefore find, in light of PP, a more profound explanation.

These conclusions are admittedly very speculative, and will have to be more carefully spelled out before they can coalesce into a full-fledged PP account of aesthetic experience. Once we leave the clear formalisms of Bayesian cognitive science as applied to art and we try to discuss their broader philosophical implications, we enter a much fuzzier territory all to be explored.

[6] See Van de Cruys and Wagemans (2011, p. 1055): “When a successful, sublime aesthetic experience is described as a selfless state of harmony between the viewer and the world, we might take this quite literally: the beholder has advanced in tuning the self to the world.”

What is sufficiently clear however is that the PP story seems well positioned to capture some of the conundrums of an experience that is both cognitive and affective, disinterested but geared towards a fundamental existential imperative. The way PP brings together perception, cognition, affect and existential concerns seems to offer a key to understanding many aspects of our aesthetic encounters, including, ultimately, why we find artistic stimuli so valuable, attractive and engrossing. What is even more clear is that the PP story we articulated has the tools to explain the feeling of having learnt from which we began our inquiry. Under the lens of PP, as we saw, all the features of this pleasurable, self-sustaining transformative experience that has the appearance of a cognitive gain seem to find their proper place.

5.3 Does Art Make Us Better, After All?

If the above account is on track, therefore, we have managed to characterise in some detail the experience we undergo while engaging with effective artworks. We have, that is, explained how and why art changes us, generating the powerful experiences that we elected as our explanatory target at the beginning of our inquiry. A curiosity might however remain at this point: granted that art changes us in the way we described, does it also change us *for the better*, improving our general epistemic position and getting us "closer to the truth"? After all, I have characterised the change brought about by an effective artwork as something with the appearance and feel of a cognitive success. What kind of cognitive success, the philosopher might ask, does not involve getting closer to the truth?

In the view we have articulated, the question of whether art brings us closer to the truth reduces itself to the question of whether the patterns of causal regularities that we discover are *really there* in the world, instead of being just conjured up by our pattern-hungry brains. Or, if you prefer, it is a matter of assessing whether the expectations (beliefs, hypotheses, predictions, best guesses) that we acquire by being exposed to art turn out to be confirmed by experience. Now, as I have noted above in a couple of circumstances, this problem looks to me remarkably similar to Hume's problem of induction, and, as such, I am not sure that anything conclusive can be said about it. Whether in each individual case the world would behave as we expect is something destined to remain uncertain. If our story is on track, this uncertainty is indeed just part and parcel of what it means for a creature to live – that is, to defend its own embodied sense of how the world is. This is why, as we noticed, the question of whether we learn from art in this sense seems to be rather uninteresting, destined as it is to decompose itself into a myriad of smaller empirical questions that will need to be ascertained separately. Furthermore, there is the additional complication that the agent has a say about the kind of world it meets, and therefore about whether its expectations will turn out to be fulfilled. By selecting the kind of environment it encounters, the agent effectively establishes whether or not its own beliefs will remain viable. From a PP perspective, this is again just part and parcel of what it means to be alive: to bring about that world whose statistical properties are compatible with your existence as a model of it.

The question of whether art changes us for the better seems therefore uninteresting at worst, and at best made very complex by loops of reciprocal influence between individual and environment. Fortunately, however, we do not need to entangle ourselves into intricate epistemological issues to see that what the arts provide is not “learning” in any of the truth-oriented senses most commonly invoked in current philosophical discussion. It suffices to consider a couple of hypothetical scenarios.

Say you have little or no knowledge of nineteenth-century Russia and get hold in some way of an entire library of history books and historical fictions about nineteenth-century Russia. Except that every single fact in these books is pure fantasy, pure fabrication. Now further suppose that these fabrications are not contradicting any of your scarce previous beliefs about nineteenth-century Russia, and that all of them are mutually consistent with one another. As you read through the library, book after book, you will discover and come to embody in your brain patterns of statistical regularities. If for example in all the books you have read so far a man called Aladdin has conquered Moscow in 1865 you will come to embody this pattern of information and expect it to be reflected in your further readings about nineteenth-century Russia. That is, you will have learnt in a perfectly legitimate psychological and neuroscientific sense. But (I take) this is certainly not "learning" in any sense dear to the philosopher: not only have you not acquired a single true belief from your readings: you have not even acquired any better beliefs. At most, you have replaced no beliefs with false beliefs, thereby worsening your epistemic position. Furthermore, in this fake library scenario, the more such regularities you discover and come to embody, the more you learn in the psychological and neuroscientific sense. So your learning will be maximal not if you acquire only true beliefs, but if you keep reading books with an optimum of unpredictability (not too close, not too distant from the ones you have already read), irrespective of whether the facts that these books describe are true or false.

Details of this scenario might be changed to show that the same might happen with any kind of stimuli. Suppose for example that you are examining a book about an animal you have heard of but never seen before, and suppose that each page of this book has an illustration of this animal seen from a different angle or depicting a different behaviour it supposedly engages in. But (as it often happened with medieval bestiaries describing creatures from distant lands that neither the reader nor the author had ever encountered) the animal in the illustrations of your book does not quite look like the one in nature. Moreover, the illustrations of your book are arranged so that, little by little, they diverge ever more from how the animal in nature really appears and behaves. In following these illustrations, you will therefore acquire and revise expectations about how the animal ought to look like and behave, getting a progressively more precise idea of it while at the same time, as it happens, moving further and further away from how things in the world are. This is certainly learning in the sense we have defined, but perhaps not quite what the philosopher extolling the cognitive value of art had in mind.

We can devise similar examples by imagining, for example, musical pieces or complex motor sequences that depart, gradually but steadily, from the statistical properties of our auditory and proprioceptive environments respectively. In each case, learning in a psychological and neuroscientific sense consists in coming to embody ever more complex and detailed systems of probabilities, even if this brings us further away from the general statistical properties of the world. It is indeed difficult to see what else learning and adaptation might be if not movements from the general statistical properties of the environment towards the properties of the particular niche the creature is adapting to. Maximising learning means therefore furthering the process by which a creature determines the creature that it is by selecting the kind of environment it is exposed to.

Here I think we are getting closer to the realms of art, a place where we may get mastery of increasingly complex, human-made systems of probabilities perhaps without further aim, just because they afford us occasions for self-realisation. In other words, our engagement with the arts might be just the rarefied and extreme prosecution of that process of self-creation by means of which the agent individuates itself by bringing about the ever more selective and unlikely environment that can further this very process of individuation. The advancement of this process depends not on the agent's ability to move "closer to the truth", but on its ability to carve out for itself a "progress niche" that offers further possibilities for self-determination. In saying this we have come, I believe, as close as we can get here in describing the kind of cognitive successes brought about by art.

5.4 The Role of Art in the Study of the Mind

Before we close our inquiry into art and learning, it is worth pointing out some promising directions for interdisciplinary research that our discussion seems to have disclosed. Throughout this work, I have tried to show that a knowledge of the dynamics of inference and belief updating can tell us a lot about how art wins and maintains our interest, what is the cognitive value that we attribute to it, how it tends to be structured, and how it makes us feel. We cannot really understand how art affects us, I suggested, without getting into the details of our workings as cognitive agents, and therefore engaging with disciplines that have those details as the main object of their research – at the very least, psychology, neuroscience and cognitive science. But if this is the case, the opposite may also be true: in studying how art affects us, we may also make discoveries about how we function as cognitive agents, discoveries which psychology, neuroscience and cognitive science could benefit from. In fact, I have often suggested in the course of our discussion that the way in which artworks are organised may reflect not just mere aesthetic preferences, but fundamental cognitive needs. I have also often characterised artists as experts in cognition, possessing an implicit knowledge of the dynamics of inference and belief updating. In the first section of this chapter, I have also suggested that, thanks to the link between art and learning that we have established, aesthetics could be understood as being continuous with the study of experience. It is time to draw some conclusions from these scattered remarks and see whether, under the right framework, aesthetics could become a key partner in the study of various aspects of the human mind.

What could philosophical aesthetics and the study and practice of art tell the psychologist, the neuroscientist and the cognitive scientist? One thing they may certainly do is help computational and neuroscientific frameworks to capture and connect with phenomenology and person-level experience. A common criticism leveraged against neuroscientific frameworks is indeed that, by focusing on sub-personal processes, they fail to account for our rich person-level, phenomenal experience. In this, PP is no exception. As Clark (2013, p. 197) notes (and as we discussed in Chapter 1, section 1.3.3), PP is still very much only "a story about the brain's way of encoding information about the world. It is not directly a story about how things seem to agents deploying that means". After all, as Clark observes, the world does not look as if it is encoded as an intertwined set of probability distributions. In other words, it is still very doubtful whether PP as a theory of sub-personal neural dynamics has anything to say about our conscious experience.

But we have seen that, in art, the dynamics of inference and belief updating seem to reveal themselves in precise affective and phenomenal correlates. Indeed, if there is a lesson that the arts can offer to the PP proponent it is that the prediction error dynamics that she conceives rather abstractly have very vivid effects on a perceiver, especially when they are shaped and channelled by the artist to serve her purposes. As art consumers, we can say if a piece of prose is fluent or disfluent; we can feel the ups and downs of uncertainty while engrossed in a suspenseful narrative; we can feel the visual tension of an unbalanced picture; we can tell when a verse or a melody closes incisively and when it remains suspended; we can tell, when watching a film, if the editing is continuous or discontinuous, invisible or marked. In other words, art is the place where predictions are built, violated and manipulated deliberately, in order to make a difference to the phenomenal experience of an observer. Such phenomenal experience is indeed what the artist is aiming to shape. In contrast with most of our everyday experience, where the fluctuations of uncertainty are more unsystematic, the perceptual flow offered by artworks is more regimented, often by means of quite articulated systems of norms and conventions (think about rhymes and metrics in poetry or the tonal system in music). Artworks could therefore be seen as powerful tools to investigate the workings of the predictive brain, as they afford a unique perspective on how predictions are formed and deployed in the perception of richly structured sensory streams. There is, in fact, an increasing interest in neuroscientific circles for this kind of research. Music has so far been privileged, as the set of expectations of western music listeners (the tonal system) is relatively well-defined, and it has long been known that harmonic and melodic violations of such expectations engender both characteristic ERP responses in the brain and conscious phenomenal effects.[7] But other arts could prove useful too. Literature, for example, seems to be another case in point – as our discussion in Chapter 2 and Chapter 3 has tried to show. By revealing other systematic relationships between neural dynamics and aspects of our conscious experience, this line of inquiry might well contribute to reducing the distance

[7] See for example Koelsch, Vuust and Friston (2019), according to which "music offers a most illuminating paradigm to understand the fundaments of the predictive brain, largely because every type of music is based on predictable regularities" (p. 63).

between sub-personal processes and our rich person-level experience, and might even inform a theory of conscious experience on the lines of PP. In sum, by paying closer attention to how in art expectations are formed and manipulated to produce deeply felt effects, the psychologist and the neuroscientist could gain insight into conscious cognitive dynamics of which the artist has already an intuitive mastery and the aesthetician at least a good command.

The character of person-level experience is however not the only thing art could help the sciences of mind clarify. After all, if the story we have articulated is on track, in manipulating the dynamics of inference and belief updating, artworks and artists are also manipulating a host of other related psychological variables, including our affective experience, our curiosity, our motivation to engage with the stimulus in question, the way we employ our attentional resources and the details we will remember most after the experience. In exploring art through a PP lens, therefore we can not only elucidate art, but also investigate, through art, several issues of general interest for mainstream psychology and neuroscience, including affect, valence, motivation, attention, memory, creativity and problem-solving. Moreover, if artworks are really devices that optimise learning, and if artists are really architects of optimal experiences as our story seems to suggest, then the study of art and aesthetics may have important implications for areas as diverse as education, wellbeing and psychopathology. While these strands of research could also be pursued separately and outside a PP perspective, the interest in PP (and the reason why I proposed to concentrate on it) lies in its apparent capacity to accommodate all the above-mentioned elements of our cognitive makeup in a coherent fashion. As we have often remarked, under this explanatorily ambitious framework elements as different as perception, cognition, learning, affect, motor control, planning and decision-making seem to hold together in a particularly convincing way, so that exploring one of them might yield insights into all the others at the same time.

Finally, a third area in which, in light of our discussion, the study of art and aesthetics might prove rewarding for the sciences of mind is research on cultural and material evolution and the impact that these have on the evolution of our cognitive capacities. The last few decades of debate in cognitive science and the philosophy of mind have accustomed us to the idea that our cognition might be not skull-bounded but extended in important ways into human-made environments able to supplement, scaffold and augment it in various ways (see Clark and Chalmers, 1998; Clark, 2008). Together with this awareness of the extended character of cognition has come the recognition that the human mind might be as much the product of human-made environments as it is their cause. The environments we build, it is said, alter our minds, leading to new environmental modifications that spur further mental development, in repeated swells of reciprocal influence (see Dennett, 1991; Clark, 1997; Knappett, 2005; Malafouris, 2013). As we noted in Chapter 1, the Bayesian/PP story we have articulated is very much in line with these acquisitions. The human-built environments with their tools and sociocultural practices greatly expand the range of new statistical regularities our plastic brains are exposed to. In structuring and restructuring them, we effectively structure and restructure the probabilistic model of the world that we embody, thus producing new changes in the environment that drive new expansions of our models. If our story is on track, in fact, humans actively try to speed up this self-fuelling process by building and searching for “progress niches” that can maximise their learning. Now, if artworks are a prime example of these humanly engineered environments created to drive new learning (and if indeed all such enhanced learning processes tend to assume an aesthetic character, as we have suggested above), then the study of art and aesthetics may be crucial for our understanding of how our minds and our cultural and material environment coevolve in mutually determining developmental trajectories.[8] The history of aesthetic practices, in turn, becomes the history of how humans have triggered in themselves changes in their patterns of perception, thought and action – that is, the history of the human mind.

All these suggestions are also very speculative and largely still to be gauged by theoretical and empirical research. Together, however, they delineate a rich and fast-growing research programme whose promises we shall now say a few closing words about, and towards whose establishment this whole work may be intended as a modest, preliminary contribution.

[8] See Constant et al. (2021) for some preliminary attempts in this direction made from a PP perspective.

Conclusions

Let us recapitulate here in closing the whole path of our inquiry into art and learning, from the questions and problems that prompted it to the tentative answers and conclusions that we have reached. In doing so, we shall perhaps clarify what questions remain to be answered and find hints of the road ahead.

We started by confronting the old philosophical question of whether art is cognitively valuable, or, in other words, whether we learn from it. The debate around this question, I suggested, is not fuelled primarily by rational arguments or empirical evidence, but rather by our widespread and obstinate intuition of readers, viewers and listeners that artworks affect us in beneficial ways. When engaging with effective artworks, I suggested, we experience a characteristic feeling of having learnt: we live a pleasurable, self-sustaining transformative experience that has the appearance of a cognitive gain. Our question became then how to best describe the process we undergo in such circumstances. In search of an answer, we first turned to the contemporary philosophical debate on art and learning. There we found characterisations of learning that, for their stress on truth, their emphasis on the propositional character of knowledge, and their disconnect from the phenomenology of learning are not very helpful in illuminating the way art affects us. We therefore turned to disciplines that have learning as a paramount concern and approach it with more descriptive intents: psychology, neuroscience and cognitive science.

We found out there that the notion of insightful problem-solving may offer a good preliminary description of the process by which creatures like us come to find and internalise new patterns of perception, thought and behaviour, a description that fits well the phenomenology of discovery and insight that characterises our engagement with the arts. To further formalise and articulate this description, we turned to a strand of contemporary work in cognitive science that conceives the brain as an ever-changing probabilistic model, forever busy updating its guesses about the causal structure of its environment. We saw that this process of belief updating can be conceived in Bayesian terms, and that there might be reasons to think that the brain implements it in a hierarchical fashion, in an effort to maintain itself as a viable model of the world. This provided us with both a precise notion of learning (namely, the updating of a hierarchical probabilistic model of the world) and a precise idea of the kinds of percepts that maximise learning (namely, percepts that diverge from the agent's current model while remaining modellable). We then saw that humans seem to naturally seek out this kind of percepts, carving out for themselves "progress niches" that speed up their developmental trajectories.

Equipped with an idea of what learning is and what are the stimuli most conducive to it, we then set out to establish whether artworks of various kinds are in fact stimuli of this sort. Our examination led us to test the Bayesian/PP apparatus for inference and belief updating on literary utterances, narratives, paintings, musical pieces and skilful actions. In all these areas, we received confirmations of the same story: effective artworks are crafted to offer beings like us a rich supply of graspable causal structures, making them experience a consistent feeling of cognitive gain.

This latter fact turned out to reveal a lot about what we demand from artworks, how they tend to be structured and how they make us feel. The fine-grained discussions of particular art forms led us, finally, to draw some general conclusions. We concluded that there is indeed a fundamental relationship between art and learning, a relationship that, while we were burdened by the theoretical preoccupations and the conceptual apparatus of contemporary epistemology, we were unable to discern. Effective artworks maximise learning in the perfectly good sense of the world current in psychology and neuroscience, and we seem to seek out and value them for that reason. This vindicates both philosophical accounts as old as Aristotle and our intuitions as art consumers about the cognitive value of art. It also opens, as we saw, interesting possibilities for interdisciplinary research. The link we established between art and our cognitive needs and workings offers hints of an interdisciplinary research programme in which art and science can fruitfully collaborate to unveil crucial aspects of the human mind.

It is with a few words on the present state and future prospects of this programme that I want to conclude. The encounter between art and the sciences of mind that we have delineated in this work, although foreshadowed in its main lines by a long philosophical tradition, is substantially novel and largely to be defined. It transcends the question of art and learning and seems to have far-reaching implications for our understanding of both the arts and the human mind, implications that are all to be assessed. This is why the discussion in this work has sometimes been more programmatic than expository, pointing to possible directions of inquiry rather than presenting established acquisitions. The great and fast-growing interdisciplinary interest around those themes is however very encouraging. The PP framework is at the cutting edge of present-day cognitive science, and its philosophical implications are currently being hotly debated. The application of this framework to art and aesthetics is a particularly novel and burgeoning area of research. As we noted, to an increasing number of interdisciplinary researchers, PP seems to offer a promising way to unify all the arts under a common analytical framework, as well as a fresh reconceptualization of long-debated issues in aesthetics such as the nature of aesthetic pleasure, the character of aesthetic normativity, the role of interest and affect in aesthetic experience and (as we saw) the cognitive value of art. At the same time, art and aesthetics are becoming increasingly interesting subjects for psychologists, neuroscientists and cognitive scientists working within the PP framework. Artworks are beginning to be seen as sources of insights into how predictions are formed and deployed in the processing of richly structured sensory streams, and traditional notions of philosophical aesthetics are animating a discussion that could lead to a better understanding of such issues as affect, valence, motivation, learning, attention, curiosity and exploratory behaviours in human agents. While scientists, philosophers and art historians are beginning to assess these theoretical prospects, a growing number of international research groups are setting up new empirical studies to test the hypotheses of the PP approach to art and aesthetics. Talks and conferences on those themes are beginning to be organised. The ferment around these themes is no doubt still somewhat uncoordinated, and much more work will be in order before this stream of research can coalesce into a full-fledged PP approach to art and aesthetics. This work has tried, among other things, to render such an approach more specific and concrete, clarifying its methodology, its theoretical commitments and some of the directions in which I believe it should proceed. Despite ample margin for discussion, a few directives for such an approach are, I think, clear.

The first is that it should preserve and expand the particular articulation of perception, cognition, affect and existential concerns offered by the PP framework. In PP, as we saw, perception and cognition can be seen as two sides of the same hierarchical predictive process, always informed by an underlying existential imperative, and thereby also intrinsically affective. This makes PP particularly apt to describe those layered and potent transformative experiences that art affords. A PP story about aesthetics should leverage these conceptual resources of the PP framework to describe how the arts manage to engage the full spectrum of our inferential abilities, from low-level sensory inference to high-level conceptual judgement, in experiences that are both cognitive and affective and have often powerful existential ramifications.

The second desideratum for a PP approach to aesthetics is that it exploits the possibilities that PP seems to offer to see the arts and aesthetics as being in continuity with the most mundane instances of meaning-making. The arts, we have suggested, just offer more vivid and exploratory instances of the processes by which we give structure and organisation to our sensorium. Aesthetics, we have said, is the study of perception and cognition in their more adventurous conditions. I believe that it is only by recognising this fact that we can develop an approach that truly generalises across the variety of the arts and human activities and does justice to the full scope of the aesthetic domain. The recognition of this fact ought to make aesthetics, as we said, a key element in the study of the human mind.

This brings us to the third desideratum for such an approach: that the arts and aesthetics play in it not the passive role of an explanandum, but the active role of partners in an explanatory enterprise that has the human mind and experience as its (perhaps forever receding) targets. In other words, a PP approach to the arts and aesthetics should find a way to combine the formal rigour and empirical preoccupations of the sciences with the phenomenological insights and the rich body of implicit knowledge about the subtleties of experience that artists, aestheticians and art historians have to offer. As Seth puts it in his paper on PP and art (2019, p. 399), the focus in this sort of interdisciplinary enterprise should be "on art and neuroscience together exploring phenomenology, and not on a neuroscience *of* art or *of* aesthetics".

These are, I believe, some of the guidelines for the work ahead. It will not always be easy to balance the phenomenological insights that come from the arts and humanities and the need for rigour and empirical evidence. The present work might indeed be a testament to the difficulties of such an enterprise more than a proof of its feasibility. But this is the road that I believe will have to be pursued if we do not want to end up with a story either overly reductionistic or overly vague and without empirical grounds. Here again the PP framework seems to be particularly well-suited to keep these two necessities together. If previous encounters between art and neuroscience have often been criticized for being overly reductionistic, PP has so far eschewed these charges, as it seems to provide both sound biological foundations and a new descriptive language to effectively connect brain research and mental, social and historical levels of analysis. This leaves us with some hope that, through PP, artists, philosophers and art historians will really be able to enter into meaningful dialogue with psychologists, neuroscientists and cognitive scientists.

The consequence of this methodologically sound encounter between the arts and the sciences of mind could be far-reaching for both sides, delivering entirely new ways in which our study of human cognition can be brought to bear on our study and practice of the arts, and vice versa. Even in its present form, this encounter has allowed us to approach the problem of the cognitive value of art from a renewed perspective, and to provide the beginning of a fresh take on neighbouring issues such as the nature of aesthetic pleasure, the role of affect and interest in our engagement with the arts, and the scope of aesthetic experience. In all these respects, PP seems to point to a fundamental continuity between aesthetic experience and experience more generally, giving new strengths to the original vocation of aesthetics as the study of perception and cognition. Under the light of PP, arts as diverse as literature, visual art, music and skilful action reveal striking similarities and can be fruitfully compared on the basis of their shared perceptual (i.e., *aesthetic*) character. This expansive understanding of aesthetics and aesthetic experience provided by PP is set to profit not only the aesthetician but also the psychologist, the cognitive scientist and the neuroscientist, who can now start drawing in a systematic way from the rich body of implicit knowledge about the dynamics of experience provided by the arts.

References

Aaronson, D., and Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1), 56-70.

Adams, R. A., Shipp, S., and Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, 218(3), 611-643.

Andersen, M. M., Kiverstein, J., Miller, M., and Roepstorff, A. (forthcoming). Play in predictive minds: A cognitive theory of play. *Psychological Review*.

Anderson, R. C., and Ortony, A. (1975). On putting apples into bottles—A problem of polysemy. *Cognitive Psychology*, 7(2), 167-180.

Aristotle. (1984). *The Complete Works of Aristotle: Revised Oxford Translation* (J. Barnes, Ed.). Princeton, NJ: Princeton University Press.

Arnheim, R. (1974 [1954]) *Art and Visual Perception. A Psychology of the Creative Eye*. Berkeley, CA: University of California Press.

Auble, P. M., Franks, J. J., and Soraci, S. A. (1979). Effort toward comprehension: Elaboration or "aha"?. *Memory & Cognition*, 7(6), 426-434.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrell, N. S., and Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 13(4), 471-481.

Baumgarten, A. G. (1936 [1750]). *Aesthetica*. Torino: Laterza.

Baxter, C. (1997). *Burning Down the House: Essays on Fiction*. St. Paul, MI: Graywolf Press.

Beardsley, M. C. (1958). *Aesthetics. Problems in the Philosophy of Criticism*. New York and Burlingame: Harcourt, Brace & World.

Benjamin, W. (1969), The storyteller. Reflections on the works of Nikolai Leskov. In H. Arendt (Ed.), *Illuminations. Essays and Reflections* (pp. 83-107). New York: Harvard University Press and Harcourt.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception and Psychophysics*, 8 (5), 279–286.

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Biederman, I., and Vessel, E. A. (2006). Perceptual pleasure and the brain: A novel theory explains why the brain craves information and seeks it through the senses. *American Scientist*, 94(3), 247-253.

Black, M. (1962). *Models and Metaphors: Studies in the Philosophy of Language*. Ithaca, NY: Cornell University Press.

Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755), 20170341.

Bonawitz, E. B., van Schijndel, T. J., Friel, D., and Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, 64(4), 215-234.

Brogaard, B. (2019). What can neuroscience tell us about reference? In J. Gundel and B. Abbott (Eds.) *The Oxford Handbook of Reference* (pp. 365-383). Oxford: Oxford University Press.

Brooks, C. (1939). *Modern Poetry and the Tradition*. Chapel Hill, NC: University of North Carolina Press.

Brooks, C. (1970 [1947]). *The Well Wrought Urn: Studies in the Structure of Poetry*. New York: Harcourt, Brace & World.

Brooks, P. (1984). *Reading for the Plot: Design and Intention in Narrative*. Cambridge, MA: Harvard University Press.

Camp, E. (2004). The generality constraint and categorical restrictions. *Philosophical Quarterly*, 54, 210–231.

Camp, E. (2006). Metaphor and that certain 'je ne sais quoi'. *Philosophical Studies*, 129(1), 1-25.

Carroll, N. (2001). On the narrative connection. In *Beyond Aesthetics: Philosophical Essays* (pp. 118–132). New York: Cambridge University Press.

Carroll, N. (2007). Narrative closure. *Philosophical Studies*, 135(1), 1-15.

Carroll, N., Moore, M., and Seeley, W. P. (2012). The philosophy of art and aesthetics, psychology, and neuroscience. In A. P. Shimamura and S. E. Palmer (Eds.) *Aesthetic Science. Connecting Minds, Brains, and Experience* (pp. 31-62). Oxford: Oxford University Press.

Carroll, N. and Seeley, W. P. (2013). Cognitivism, psychology, and neuroscience: Movies as attentional engines. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring Cognition at the Movies* (pp. 53-75). Oxford: Oxford University Press.

Cavell, S. (1976 [1969]). *Must We Mean What We Say? A Book of Essays*. Cambridge: Cambridge University Press.

Cheung, V. K., Harrison, P. M., Meyer, L., Pearce, M. T., Haynes, J. D., and Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology*, 29(23), 4084-4092.

Chmiel, A., and Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature. *Psychology of Music*, 45(6), 886-909.

Chomsky, N. (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, N. (1982). A note on the creative aspect of language use. *The Philosophical Review*, 91(3), 423-434.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204.

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Clark, A. (2018). Beyond the 'Bayesian blur': Predictive processing and the nature of subjective experience. *Journal of Consciousness Studies*, 25, 71–87.

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7-19.

Cochrane, T. (2021). *The Aesthetic Value of the World*. Oxford: Oxford University Press.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science*, 7(2), 137-163.

Constant, A., Tschantz, A. D. D., Millidge, B., Criado-Boado, F., Martinez, L. M., Müeller, J., and Clark, A. (2021). The acquisition of culturally patterned attention styles under active inference. *Frontiers in Neurorobotics*, 15, 729665.

Cook, C., Goodman, N. D., and Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, 120(3), 341-349.

Cools, A. and Verheyen, L. (2019). Literature and the nugget of knowledge. An interview with Derek Attridge and Peter Lamarque. *Aesthetic Investigations*, 3(1), 9-27.

Coons, E. and Kraehenbuehl, D. (1958). Information as Measure of Structure in Music. In *Journal of Music Theory*, 2(2), 127-161.

Costandi, M. (2016). *Neuroplasticity*. Cambridge, MA: MIT Press.

Cross, E. S., Kirsch, L., Ticini, L. F., and Schütz-Bosbach, S. (2011). The impact of aesthetic evaluation and physical ability on dance perception. *Frontiers in Human Neuroscience*, 5, 102.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper & Row.

Csikszentmihalyi, M., and Robinson, R. (1990). *The Art of Seeing. An Interpretation of the Aesthetic Encounter*. Malibu, CA: J. Paul Getty Museum/Getty Center for Education in the Arts.

Currie, G. (2006). Narrative representation of causes. *The Journal of Aesthetics and Art Criticism*, 64(3), 309-316.

Currie, G. (2011). Empathy for objects. In A. Coplan and P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 82-95). Oxford: Oxford University Press.

Currie, G. (2014). Creativity and the insight that literature brings. In E. S. Paul and S. B. Kaufman (Eds.). *The Philosophy of Creativity: New Essays* (pp. 39-61). Oxford: Oxford University Press.

Currie, G. (2016). Imagination and learning. In A. Kind (Ed.), *The Routledge Handbook of Philosophy of Imagination* (pp. 407-419). London: Routledge.

Currie, G. (2020). *Imagining and Knowing: The Shape of Fiction*. Oxford: Oxford University Press.

Currie, G. and Frascaroli, F. (2021). Poetry and the possibility of paraphrase. *The Journal of Aesthetics and Art Criticism*, 79(4), 428-439.

Davidson, D. (1978). What metaphors mean. *Critical inquiry*, 5(1), 31-47.

Delplanque, J., De Loof, E., Janssens, C., and Verguts, T. (2019). The sound of beauty: How complexity determines aesthetic preference. *Acta Psychologica*, 192, 146-152.

Deming, R. E. (Ed.). (2005). *James Joyce. The Critical Heritage*. *Volume 2, 1928-1941*. London and New York: Routledge.

Dennett, D. (1991). *Consciousness Explained.* Boston, MA: Little Brown.

Deterding, S., Andersen, M. M., Kiverstein, J., and Miller, M. (2022). Mastering uncertainty: A predictive processing account of enjoying uncertain success in video game play. *Frontiers in Psychology*, 4214.

Dewey, J. (2005 [1934]). *Art as Experience*. New York: Perigee.

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, 58(5), 100-113.

Eco, U. (1979a). *Lector in Fabula. La Cooperazione Interpretativa nei Testi Narrativi*. Milano: Bompiani.

Eco, U. (1979b). *The Role of the Reader. Explorations in the Semiotics of Texts*. Bloomington: Indiana University Press.

Elgin, C. Z. (2002). Art in the advancement of understanding. *American Philosophical Quarterly*, 39(1), 1-12.

Elgin, C. Z. (2007). Understanding and the facts. *Philosophical Studies*, 132(1), 33-42.

Elgin, C. Z. (2009). Is understanding factive?. In A. Haddock, A. Millar and D. Pritchard (Eds.), *Epistemic Value* (pp. 322-330). New York: Oxford University Press.

Emerson, R. W. (1987 [1850]). *Representative Men: Seven Lectures*. Cambridge, MA: Harvard University Press.

Evans, D. R. (1970). Conceptual complexity, arousal, and epistemic behaviour. *Canadian Journal of Psychology/Revue Canadienne de Psychologie,* 24(4), 249-260.

Feagin, S. L. (2007). On Noël Carroll on narrative closure. *Philosophical Studies*, *135*(1), 17-25.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., and Perry Jr, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4), 400-409.

Fitz, H., and Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, 111, 15-52.

Form, S. (2019). Reaching wuthering heights with brave new words: The influence of originality of words on the success of outstanding best-sellers. *The Journal of Creative Behavior*, 53(4), 508-518.

Forster, E. M. (2022 [1927]). *Aspects of the Novel*. New York: Dover Publications.

Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2013). Word surprisal predicts N400 amplitude during reading. In M. Butt, S. Hussain (Eds.), *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 878-883). Sofia: Association for Computational Linguistics.

Freedberg, D. (1989). *The Power of Images: Studies in the History and Theory of Response*. Chicago: University of Chicago Press.

Freedberg, D., and Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in cognitive sciences*, *11*(5), 197-203.

Frege, G. (1903). *Grundgesetze der Arithmetik: Begriffsschriftlich Abgeleitet* (Vol. 2). Jena: Hermann Pole.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological Sciences*, 360(1456), 815-836.

Friston, K. J. (2010). The free-energy principle: A unified brain theory?. *Nature Reviews Neuroscience*, 11(2), 127-138.

Friston, K. J. (2013a). Life as we know it. *Journal of The Royal Society Interface*, 10(86), 20130475.

Friston, K. J. (2013b). The fantastic organ. *Brain*, 136(4), 1328-1332.

Friston, K. J. (2018). Does predictive coding have a future?. *Nature Neuroscience*, 21, 1019–1021.

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, 102(3), 227-260.

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., and Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural Computation*, 29(10), 2633-2683.

Friston, K. J., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1), 137-160.

Friston, K. J., Rosch, R., Parr, T., Price, C., and Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, 90, 486-501.

Gallese, V. (2017). Visions of the body. Embodied simulation and aesthetic experience. *Aisthesis. Pratiche, Linguaggi e Saperi dell'Estetico*, 10(1), 41-50.

Gallie, R. (1964). *Philosophy and Historical Understanding*. London: Chatto & Windus.

Gallistel, C. R., and King, A. P. (2010). *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. Hoboken, NJ: John Wiley & Sons.

Genette, G. (1972). *Figure III*. Paris: Seuil.

Gerken, L., Balcomb, F. K., and Minton, J. L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, 14(5), 972-979.

Gettier, E. L. (1963). Is justified true belief knowledge?. *Analysis*, 23 (6), 121–123

Gibson, J. (2003). Between truth and triviality. *The British Journal of Aesthetics*, 43(3), 224-237.

Gibson, J. (2008). Cognitivism and the Arts. *Philosophy Compass*, 3(4), 573-589.

Gick, M. L., and Lockhart, R. S. (1995). Cognitive and affective components of insight. In R. J. Sternberg and J. E. Davidson (Eds.), *The Nature of Insight* (pp. 197-228). Cambridge, MA: MIT Press.

Goffin, K., and Friend, S. (2022). Learning implicit biases from fiction. *The Journal of Aesthetics and Art Criticism*, 80(2), 129-139.

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., and Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: a reward for learning?. *Journal of Neuroscience*, 39(47), 9397-9409.

Gombrich, E. H. (1960). *Art and Illusion*. Princeton, NJ: Princeton University Press.

Gombrich, E. H. (1984). *The Sense of Order: A Study in the Psychology of Decorative Art*. New York: Phaidon Press.

Green, M. C. and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701-721.

Gregory, R. L. (1997). *Eye and Brain*. Princeton and Oxford: Princeton University Press.

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 59–100). Cambridge: Cambridge University Press.

Groupe µ. (1977). *Rhétorique de la Poésie*. Paris: Seuil.

Groupe µ. (1981 [1970]). *A General Rhetoric* (P. B. Burrell and Edgar M. Slotkin, Trads.). Baltimore: Johns Hopkins University Press.

Hansen, N. C., and Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. Frontiers *in Psychology*, 5, 1052.

Hegel, G. W. F. (1975). *Aesthetics*. *Lectures on Fine Art*, Vol. 1 (T.M. Knox, Trans.). Oxford: Clarendon Press.

Heidegger, M. (1971 [1960]). The origin of the work of art (A. Hofstadter, Trans.). In M. Heidegger, *Poetry, Language*, *Thought* (pp. 15-86). New York: HarperCollins.

Heilbron, M., and Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex?. *Neuroscience*, 389, 54-73.

Helmholtz, H. (2013 [1867]). *Treatise of Physiological Optics*, Vol. 3 (A. Gullstrand, J. von Kries and C. Ladd-Franklin, Trans.). New York: Dover Publications.

Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., and Ramstead, M. J. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation*, 33(2), 398-446.

Hills, A. (2016). Understanding why. *Nous*, 50 (4), 661-688.

Hirsch, H. V. B. and Spinelli, D. (1970). Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*, 168(3933), 869–71.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, 50(2), 259-285.

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, 35(2), 209-223.

Hurley, S. L. (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Hyman, J. (2010). Art and neuroscience. In R. Frigg and M. C. Hunter (Eds.) *Beyond Mimesis and Convention*. *Representation in Art and Science* (pp. 245-261). New York: Springer.

Ichikawa, J. J. and Steup, M. (2017). The analysis of knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition). URL = <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>.

Ichino, A., and Currie, G. (2017). Truth and trust in fiction. In E. Sullivan-Bissett, H. Bradley and P. Noordhof (Eds.), *Art and Belief* (pp. 63-82). Oxford: Oxford University Press.

Iser, W. (1980 [1976]). *The Act of Reading. A Theory of Aesthetic Response*. Baltimore: Johns Hopkins University Press.

Jackson, F. (2011). On Gettier holdouts. *Mind & Language*, 26(4), 468-481.

Jakobson, R. (1923). *O Cheshskom Stikhe Preimushchestvenno v Sopostavlenii s Russkim*. Berlin: Opoyaz.

Joffily, M., and Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Computational Biology*, 9(6), e1003094.

Johnson, M. (2018). *The Aesthetics of Meaning and Thought*. *The Bodily Roots of Philosophy, Science, Morality, and Art.* Chicago: University of Chicago Press.

Joyce, J. (1961). *James Joyce's Scribbledehobble; The Ur-Workbook for "Finnegans Wake"* (T. Connolly, Ed.). Evanston, IL: Northwestern University Press.

Judge, J., and Nanay, B. (2021). Expectations. In T. McAuley, N. Nielsen, J. Levinson, and A. Phillips-Hutton (Eds.). *The Oxford Handbook of Western Music and Philosophy* (pp. 997-1018). Oxford: Oxford University Press.

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.

Kammann, R. (1966). Verbal complexity and preferences in poetry. *Journal of Verbal Learning and Verbal Behavior* 5(6), 536-540.

Kamp, H., Genabith, J. V., and Reyle, U. (2011). Discourse representation theory. In D. Gabbay, and F. Guenthner (Eds.), *Handbook of Philosophical Logic*, Vol. 15 (pp. 125-394). Dordrecht: Springer.

Kant, I. (1987 [1790]). *Critique of Judgement* (W. S. Pluhar, Trans.). Indianapolis/Cambridge: Hackett Publishing Company.

Keenan, J. M., Baillet, S. D., and Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 115-126.

Kesner, L. (2014). The predictive mind and the experience of visual art work. *Frontiers in Psychology*, 5, 1417.

Kidd, C., and Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, 88(3), 449-460.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One*, 7(5), e36399.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795-1804.

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, 8(3), 159-166.

Kirsch, L. P., Urgesi, C., and Cross, E. S. (2016). Shaping and reshaping the aesthetic brain: Emerging perspectives on the neurobiology of embodied aesthetics. *Neuroscience & Biobehavioral Reviews*, 62, 56-68.

Kiverstein, J., Miller, M., and Rietveld, E. (2019). The feeling of grip: novelty, error dynamics, and the predictive brain. *Synthese*, 196(7), 2847-2869.

Koehler, W. (1927 [1921]). *The Mentality of Apes* (E. Winter, Trans.). London: Kegan Paul, Trench, Trubner & CO.

Koelsch, S., Vuust, P., and Friston, K. J. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, 23(1), pp. 63-77.

Knappett, C. (2005). *Thinking Through Material Culture: An Interdisciplinary Perspective*. Philadelphia, PA: University of Pennsylvania Press.

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neuroscience*, 27, 712–719.

Knill, D. C. and Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, 31(1), 32-59.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., and Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *Neuroimage*, 33(1), 343-361.

Kuperberg, G. R., Paczynski, M., and Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23(5), 1230-1246.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.

Kutas, M. and Schmitt, B. M. (2003). Language in microvolts, in M. T. Bannich and M. Mach (Eds.), *Mind, Brain and Language: Multidisciplinary Perspectives* (pp. 171-209.). Mahwah: Lawrence Eribaum Associates.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.

Kvanvig, J. (2009). The value of understanding. In A. Haddock, A. Millar, and D. Pritchard (Eds.), *Epistemic Value* (95-111)*.* Oxford: Oxford University Press.

La Fontaine, J. de (1895 [1668-1694]). *Fables*. London, Paris and Boston: Hachette.

Lamarque, P., and Olsen, S. H. (1994). *Truth, Fiction, and Literature: A Philosophical Perspective*, Oxford: Clarendon Press.

Leder, H., Bär, S., and Topolinski, S. (2012). Covert painting simulations influence aesthetic appreciation of artworks. *Psychological Science*, 23(12), 1479-1481.

Lepore, E., and Stone, M. (2015). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press.

Livingstone, M. (2002). *Vision and Art: The Biology of Seeing*. New York: Harry N. Abrams.

Lotman, J. M. (1977 [1971]). *The Structure of the Artistic Text* (R. Vroon, Trans.). Ann Arbor: Michigan University Press.

Lotman, J. M. (1990). *Universe of the Mind* (A. Shukman, Trans.). Bloomington and Indianapolis: Indiana University Press.

Lotman, J. M. (2009 [1992]). *Culture and Explosion* (W. Clark, Trans.). Berlin and New York: Mouton de Gruyer.

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, 54, 1-95.

Luchins, A. S. and Luchins, E. H. (1959). *Rigidity of Behavior: A Variational Approach to the Effect of Einstellung*. Eugene: University of Oregon Books.

Mack, M. (2012). *How Literature Changes the Way We Think*. London: Continuum.

Malafouris, L. (2013). *How Things Shape the Mind*. Cambridge, MA: MIT Press.

Matthews, T. E., Witek, M. A., Heggli, O. A., Penhune, V. B., and Vuust, P. (2019). The sensation of groove is affected by the interaction of rhythmic and harmonic complexity. *PLoS One*, 14(1), e0204539.

Matthews, T. E., Witek, M. A., Lund, T., Vuust, P., and Penhune, V. B. (2020). The sensation of groove engages motor and reward networks. *NeuroImage*, 214, 116768.

Mencke, I., Omigie, D., Wald-Fuhrmann, M., and Brattico, E. (2019). Atonal music: Can uncertainty lead to pleasure?. *Frontiers in Neuroscience*, 12, 979.

Menninghaus, W., Wagner, V., Wassiliwizky, E., Schindler, I., Hanich, J., Jacobsen, T., and Koelsch, S. (2019). What are aesthetic emotions?. *Psychological Review*, 126(2), 171.

Merleau-Ponty, M. (1964 [1960]). *Signs* (R. McCleary, Trans.). Chicago: Northwestern University Press.

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism*, 15(4), 412-424.

Meyer, L. B. (2008 [1956]). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Miller, M., Andersen, M., Schoeller, F. and Kiverstein, J. (forthcoming). Getting a kick out of film: Aesthetic pleasure and play in prediction error minimizing agents.

Miller, M., Kiverstein, J., and Rietveld, E. (2022). The predictive dynamics of happiness and well-being. *Emotion Review*, 14(1), 15-30.

Mink, L. O. (1970). History and fiction as modes of comprehension. *New Literary History*, 1(3), 541-558.

Muth, C., and Carbon, C. C. (2016). SeIns: Semantic instability in art. *Art & Perception*, 4(1-2), 145-184.

Nannicelli, T. (2020). *Artistic Creation and Ethical Criticism*. Oxford: Oxford University Press.

Noë, A. (2015). *Strange Tools: Art and Human Nature*. New York: Hill and Wang.

Nussbaum, M. C. (1990). *Love's Knowledge: Essays on Philosophy and Literature*. New York and Oxford: Oxford University Press.

Öllinger, M., Jones, G., and Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology*, 55(4), 269.

Olmos, P. (Ed.). (2017). *Narration as Argument*. Cham: Springer.

Oudeyer, P. Y., and Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 1, 6.

Oudeyer, P. Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, 11(2), 265-286.

Pearce, M. T., and Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4), 625-652.

Pelowski, M., and Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology*, 29, 80–97.

Piaget, J. (1952). *The Origins of Intelligence in Children* (M. Cook, Trans.). New York: W. W. Norton & Co.

Plato. (1997). *Plato: Complete Works*. (J. M. Cooper, and D. S. Hutchinson, Eds.). Indianapolis: Hackett Publishing.

Plumer, G. (2015). On novels as arguments. *Informal Logic*, 35(4), 488-507.

Plutarch. (1969). How the young man should study poetry. In *Plutarch's Moralia* (F. C. Babbitt, Trans.), Vol. 1 (pp.74-197). Cambridge, MA: Harvard University Press.

Pritchard, D. (2014). Knowledge and understanding. In A. Fairweather (Ed.), *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science* (315-327). Cham: Springer.

Propp, V. (1968 [1928]). *Morphology of the Folk Tale* (L. Scott, Trans.). Austin, TX: University of Texas Press.

Quine, W. V. O. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, 60 (1), 20–43.

Ramstead, M., Weise, W., Miller, M., and Friston, K. (forthcoming). Deep neurophenomenology: An active inference account of some features of conscious experience and of their disturbance in major depressive disorder. In T. Cheng, R. Sato and J. Hohwy, *Expected Experiences: The Predictive Mind in an Uncertain World*. London: Routledge.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.

Reid, L. A. (1964). Art, truth and reality. *The British Journal of Aesthetics*, 4 (4), 321–331.

Richards, I. A. (1936). *The Philosophy of Rhetoric*. New York and London: Oxford University Press.

Ricoeur, P. (1975). *La Métaphore Vive*. Paris: Seuil.

Riggs, W. D. (2009). Understanding, knowledge, and the Meno requirement. in A. Haddock, A. Millar and D. Pritchard (Eds.), *Epistemic Value* (pp. 331-338). New York: Oxford University Press.

Ryan, M. L. (2007). Toward a definition of narrative. In D. Herman (Ed.), *The Cambridge Companion to Narrative* (pp. 22-35). Cambridge: Cambridge University Press.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Saka, P. (2007). The argument from ignorance against truth-conditional semantics. *American Philosophical Quarterly*, 44(2), 157-169.

Sarasso, P., Neppi-Modona, M., Sacco, K., and Ronga, I. (2020). 'Stopping for knowledge': The sense of beauty in the perception-action cycle. *Neuroscience & Biobehavioral Reviews*, 118, 723-738.

Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., and Scherer, K. R. (2017). Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PloS One*, 12(6), e0178899.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE* Transactions *on Autonomous Mental Development*, 2(3), 230-247.

Schoeller, F., and Perlovsky, L. (2016). Aesthetic chills: Knowledge-acquisition, meaning-making, and aesthetic emotions. *Frontiers in Psychology*, 7, 1093.

Schulz, L. E., and Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, 43(4), 1045.

Schultz, R. A. (1979). Analogues of argument in fictional narrative. *Poetics*, 8(1-2), 231-244.

Seeley, W. P. (2020). *Attentional Engines. A Perceptual Theory of the Arts*. Oxford: Oxford University Press.

Seth, A. (2019). From unconscious inference to the beholder's share: Predictive perception and human experience. *European Review*, 27(3), pp. 378-410.

Seth, A. (2021). *Being You. A New Science of Consciousness*. London: Faber & Faber.

Shen, W., Yuan, Y., Liu, C., and Luo, J. (2016). In search of the 'Aha!' experience: Elucidating the emotionality of insight problem-solving. *British Journal of Psychology*, 107(2), 281-298.

Shusterman, R. (2012). *Thinking Through the Body: Essays on Somaesthetics*. Cambridge: Cambridge University Press.

Silwa, P. (2015). Understanding and knowing. *Proceedings of the Aristotelian Society*, 115 (1), 57-74.

Smith, L. B., Jayaraman, S., Clerkin, E., and Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4), 325-336.

Sperber, D. and Wilson, D. (1995). *Relevance*: *Communication and Cognition* (2nd edition). Oxford: Blackwell Publishing.

Sporns, O. (2007). What neuro-robotic models can teach us about neural and cognitive development. In D. Mareschal, S. Sirois, G. Westermann and M. H. Johnson (Eds.), *Neuroconstructivism: Perspectives and Prospects*, Vol. 2 (pp. 179–204). Oxford: Oxford University Press.

Stanley, J., and Willlamson, T. (2001). Knowing how. *The Journal of Philosophy*, 98(8), 411-444.

Stein, G. (2019 [1933]). *Autobiography of Alice B. Toklas*. Toronto: Ryerson University Press.

Stolnitz, J. (1992). On the cognitive triviality of art. *British Journal of Aesthetics*, 32(3), 191-200.

Temperley, D. (2007). *Music and Probability*. Cambridge, MA: MIT Press.

Temperley, D. (2014). Information flow and repetition in music. *Journal of Music Theory*, 58(2), 155-178.

Temperley, D. (2019). Uniform information density in music. *Music Theory Online*, 25(2).

Tomashevsky, B. (1925 [1999]). *Teorija literatury. Poetika*. Moskva: Aspekt Press.

Topolinski, S., and Reber, R. (2010). Gaining insight into the "Aha" experience*. Current Directions in Psychological Science*, 19(6), 402-405.

Valéry, P. (1936). *Variété III*. Paris: Gallimard.

Van de Cruys, S. (2017). Affective value in the predictive mind. In T. Metzinger and W. Wiese (Eds.). *Philosophy and Predictive Processing* (pp. 1-24). Frankfurt Am Main: MIND Group.

Van de Cruys, S., Bervoets, J., and Moors, A. (forthcoming). Preferences need inferences: learning, valuation, and curiosity in aesthetic experience. In M. Nadal and M. Skov (Eds.), *The Routledge International Handbook of Neuroaesthetics*. London: Routledge.

Van de Cruys, S., and Wagemans, J. (2011). Putting reward in art: A tentative prediction error account of visual art. *I-Perception*, 2(9), 1035-1062.

Vander Elst, O. F., Vuust, P., and Kringelbach, M. L. (2021). Sweet anticipation and positive emotions in music, groove, and dance. *Current Opinion in Behavioral Sciences*, 39, 79-84.

Van Geert, E., and Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, 14(2), 135-155.

Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18(4), 380-393.

Van Petten, C., and Kutas, M. (1991). Influences of semantic and syntactic context on open-and closed-class words. *Memory & Cognition*, 19(1), 95-112.

Van Petten, C., and Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176-190.

Velleman, J. D. (2003). Narrative explanation. *The Philosophical Review*, 112(1), 1-25.

Vidmar Jovanović, I. (2021). Applied ethical criticism of narrative art. *Etica & Politica*, XXIII (3), 443-459.

Vuust, P., and Witek, M. A. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, 5, 1111.

Vygotsky L. (1978). *Mind in Society. The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Walsh, K. S., McGovern, D. P., Clark, A., and O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, 1464(1), 242-268.

Walton, K. L. (1970). Categories of art. *The Philosophical Review*, 79(3), 334-367.

Wertheimer, M. (1959). *Productive Thinking*. New York: Harper.

Williamson, T. (2002). *Knowledge and its Limits*. Oxford: Oxford University Press.

Wilson, D. and Sperber, D. (2012). *Meaning and Relevance*. Cambridge: Cambridge University Press.

Wilson, D., and Carston, R. (2019). Pragmatics and the challenge of 'non-propositional' effects. *Journal of Pragmatics*, 145, 31-38.

Witek, M. A., Clarke, E. F., Wallentin, M., Kringelbach, M. L., and Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PloS one*, 9(4), e94446.

Wundt, W. (1874). *Grundziige der Physiologischen Psychologie*. Leipzig: Engelmann.

Zagzebski, L. (2001). Recovering understanding. In M. Steup (Ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue* (pp. 236-252). New York: Oxford University Press.