Is meaning intrinsically normative?

Pierre Jacob

Pierre Jacob

Is Meaning Intrinsically Normative?[1]

Prima facie, meaning seems to be a normative property in the simple following sense: if something exemplifies a given meaning property, then some normative consequences follow. So, for example, if the French word 'cheval' means HORSE – if it is correctly used to refer to horses –, then it is a *mistake* to use it to refer to things which are not horses. Equivalently, one *ought* to use it to refer to horses and only to horses (provided some idealizations). Similarly, my concept HORSE – the thought constituent expressible by the French word 'cheval' – correctly applies to horses and only to horses. So my concept *ought* to be tokened to refer to horses and only to horses. Or else, I misapply it. The fact that it would be a *mistake* to use the word 'cheval' to refer to non-horses signals the fact that meaning is a normative property or that it has normative implications. So does the fact that one *ought* to use a word in certain circumstances and not in others. And furthermore this norm arises in virtue of the meaning of the word.

The question I want to ask is: Is meaning *intrinsically* normative? As I understand it, to claim [N] that meaning is intrinsically normative is to claim both [N1] that meaning is normative and [N2] that the normativity of meaning is *sui generis*, i.e., that it is irreducibly semantic. Hume is widely taken to have shown that it is a fallacy to derive a moral obligation from premises about matters of fact. Nor can one refute the ethical claim that one ought to keep one's promiss by exhibiting unfulfilled promisses. Almost a hundred years ago, Moore famously criticized what he called the "naturalistic fallacy" which he thought would undermine any attempt at providing a definition of the meaning of the word 'good' in purely descriptive non-normative naturalistic terms. Now, if [N] is correct – if meaning is intrinsically normative –, then presumably any naturalistic attempt at understanding

[1] Thanks to Fred Dretske, Wolfram Hinzen, Wolfgang Künne, Albert Newen, Armin Tatzel for comments at the GAP Conference and afterwards. Thanks to Ned Block, Tim Crane, Pascal Engel, Kati Farkas, Paul Horwich, Barry Loewer, Nenad Miscevic, Peter Railton, François Recanati, John Skorupski and Dan Sperber for conversations on this topic.

meaning properties in terms of non-semantic properties is bound to commit a version of the naturalistic fallacy.

1. Kripke's sceptical paradox and the thesis that meaning is intrinsically normative

The charge of a naturalistic fallacy which concerns me has been brilliantly made by Kripke in his famous (1982) book, *Wittgenstein on Rules and Private Language*. And it is based on thesis [N]. Let me briefly summarize Kripke's challenge as I understand it.

In his book, Kripke argues in favor of a certain thesis which he calls "the sceptical paradox". According to this thesis, no attribution of meaning to an expression of a natural language can be made true by any fact about the mental life and the history of a single speaker of the language. I know – or I think I know – the meaning of the French word 'chien'. I know that the word 'chien' in French applies to dogs. I have learnt its meaning on the basis of a finite amount of instances of application. What is it about me which makes it true that were I to apply the word 'chien' to something which would not be a dog I would misapply it? On account of what mental fact would my usage of the word constitute a mistake? If I am the speaker and if I use the French word 'chien', then no fact about me (about my mind or about my individual history) can make it true that by a token of the word 'chien', I mean to refer to dogs and only to dogs. (Incidentally, Kripke's thesis is stated in terms of the arithmetical predicate 'plus'. But the arithmetical meaning of the predicate is irrelevant.)

First of all, in the process of arguing for this sceptical and paradoxical thesis, Kripke appeals to considerations which have been made familiar by other philosophers such as Putnam (1975) and Burge (1979). However, Kripke's sceptical and paradoxical conclusion interestingly diverges from theirs. Their conclusion is an *externalist* conclusion. As Putnam famously put it, on his and Burge's view, "meaning ain't in the head". In other words, meaning or content is not an intrinsic property of an individual's brain: it supervenes on relations between an individual's brain and properties instantiated in his or her environment. It may even supervene on what other members of the individual's community may think.

Kripke's conclusion is that there are *no facts* about what words mean – no facts about meaning. Even though on the externalist view, meaning turns out to be an extrinsic property of an individual's brain, the externalist conclusion does not challenge the factual character of meaning attributions.

Meaning facts (or properties) may not be intrinsic to an individual's brain. Nonetheless they are or they may be respectable naturalistic facts. But Kripke's sceptical paradox does challenge the factuality of meaning because it claims that meaning attributions do not state or describe any fact; rather they express norms. Meaning attributions do not have truth-conditions. They are not aimed at stating facts, they do not have a fact-stating role. Meaning-ascriptions have assertibility conditions because meaning is intrinsically normative.

Secondly, Kripke offers a solution to the sceptical paradox which, he claims, was Wittgenstein's. He calls his solution a 'sceptical' solution by contrast with what he calls 'straight' solutions. Unlike 'straight' solutions, the 'sceptical' solution is not intended as a rebuttal of the sceptical paradoxical thesis: the solution does not consist in rejecting one of the steps leading to the paradoxical conclusion. Rather, in a Wittgensteinian spirit, the solution is to learn to live with the conclusion. Roughly, semantic norms governing an individual's use of an expression arise from the practices of his or her linguistic community in the following sense. What an individual means by his or her use of a word is a matter of agreeing with the uses of other members of his or her community. Meaning consists in a pattern of agreement between members of a community. This is what it takes to be part of a linguistic community: an individual belongs to a community if his or her uses of words coincide with the uses of others. So, to say of a person that he or she means addition by '+' is, as Hale (1997) puts it, to "acknowledge him or her as a fully-paid member of the community of adders, to convey that he or she can be relied upon not to come up with some bizarre answers".

In the sequel, I will disregard Kripke's 'sceptical solution' to the sceptical paradox. As I see it, the sceptical paradox is clearly a version of thesis [N], the thesis that meaning is intrinsically normative. If meaning is intrinsically normative in the sense that there are no facts about meaning, then entertaining the very idea that meaning properties might arise out of non-semantic properties or that the instantiation of non-semantic properties might explain (or account for) the instantiation of semantic properties is committing a naturalistic fallacy. Since the project of a naturalistic understanding of meaning is close to my heart, I do intend to reject thesis [N]. Given the way I construe it, if one wants to reject [N], one can either deny [N1], the thesis that semantic properties are normative, or one can can deny [N2], the thesis that the normativity of meaning is irreducible. I'll call the first strategy the *deflationary* strategy and the second the *reductionist* strategy. In this paper, I want to provide some rationale in favor of the reductionist strategy by exposing some of the difficulties which, according to me, lie ahead of the deflationary strategy.

In a nutshell, the deflationary strategy as implemented by Horwich in his 1998 book, *Meaning*, involves three main ingredients: (i) the distinction between normative properties and normative consequences of non-normative properties; (ii) a pragmatic explanation of the normative consequences of meaning; and (iii) the lack of a distinction between semantic norms and other norms (e.g., ethical norms). However much I would have liked to embrace deflationism about the normativity of meaning, the goal of my paper is to explain why I cannot accept it.

2. The deflationary distinction between normative properties and normative consequences

Consider the following step in the reasoning leading to Kripke's sceptical paradox:

[C1] If the French word 'chien' refers to dogs or means DOG, then one *ought* to apply it to dogs and only to dogs.

Is conditional [C1] uncontroversial? Granted, it requires a little amount of idealization since French speakers do apply the word 'chien' not just to dogs but to *pictures* of dogs and *sculptures* of dogs as well. As a French speaker, it seems to me that whereas it *would* be incorrect to apply the word 'chien' to a picture of something which is not a dog, it is *not* incorrect to apply the word 'chien' to a picture of a dog which is a picture and therefore not a dog. Still it is, I think, uncontroversial that while it is correct to apply the word 'chien' to a picture of a dog, it is incorrect to apply it to a picture of something which is not a dog. So there is no doubt that some revised version of [C1] is correct. What would, I think, be controversial would be to conclude from [C1] that the meaning of the word 'chien' is *intrinsically* normative. The conclusion that the meaning of the word 'chien' is intrinsically normative could only follow from [C1] via the extra assumption [C2]:

[C2] Unless there are no semantic facts, meaning attributions cannot have normative consequences.

At this point, the deflationist rightly points out that [C2] is not plausible. The reason why [C2] ought to be rejected is that a number of prima facie non-normative properties clearly may have normative consequences. The instantiation of a non-normative property may be subject to normative assessment. With Dretske (2000), consider, for example, the relation expressed by the English verb 'kill'. On many occasions, if *x* and *y* are

human and if *x* kills *y*, then some normative consequences follow. If *x* killed *y* intentionally and with no mitigating circumstances, then *x* ought to be sanctioned. Notice the contrast between the relations expressed respectively by the verb 'kill' and by the verb 'murder'. Whatever the circumstances, if *x* murdered *y*, then *x* ought to be sanctioned. Legal and moral norms are built into the very existence of murders. Legal and moral norms are constitutive of murders. The latter presuppose the former: in the absence of legal and moral norms, there would simply be no murder. Lack of legal and moral norms, however, would not be enough to prevent the killing relation from being instantiated. Unlike the murder relation, the killing relation can be characterized in natural descriptive terms. Similarly, meteorological properties and relations are presumably natural physical properties and relations. Although the instantiation of a storm, a flood or a hurricane is a physical process, nonetheless it can be subject to normative assessment. We can e. g., deplore it if it inteferes with our plans, our goals or our intentions.

According to the deflationist standpoint then, semantic properties are like the killing relation or like meteorological properties. According to deflationism, it is a fact that the French word 'cheval' refers to horses or means HORSE. Semantic properties are non-normative properties whose instantiation has normative consequences or implications. Notice, however, that prima facie there seems to be a difference between the way normative consequences arise from the instantiation of semantic properties and the way normative consequences arise from the instantiation of the killing relation or from the instantiation of meteorological properties. Whereas the instantiation of meterological properties may occasionally be subject to normative assessment, the instantiation of a semantic property *always* has normative implications. It does not seem as if the normative implications of meaning can sometimes be relaxed or loosened. In any case, given the assumption that meaning is not a normative property, it is incumbent upon the deflationist to account for the normative consequences of meaning.

3. A pragmatic explanation of the normative implications of semantic properties

Let us examine the deflationist explanation of the normative implications of meaning. What the deflationist has in mind, I think, is conditional [C3] which is a revised version of [C1]:

[C3] If the French word 'chien' refers to dogs or means DOG, then one *ought* to apply it to dogs and only to dogs *provided* one wants (or intends) to communicate with French speakers.

In other words, on the deflationary account, the normative consequences of meaning are contingent or conditional upon an individual's *intentions*, *desires* or *goals*.

At this stage, I want to make explicit a couple of assumptions on which my subsequent discussion will be predicated. On my view, any creature capable of engaging in intentional actions must be capable of entertaining two kinds of mental representations. It must be able to form beliefs and it must be able to form desires. As Anscombe and Searle have pointed out, beliefs and desires have opposite directions of fit. Beliefs have a mind-to-world direction of fit. Their job is to provide an accurate representation (or a correct picture) of how the world is. Now, unless they were motivationally inert, beliefs could not do their job: a creature whose beliefs would systematically be based on wishful thinking could not survive the test of natural selection. Goals, intentions and desires represent possible or impossible non-actual states of affairs. They have a world-to-mind direction of fit. They represent the world as it ought to be, not as it is. An intentional action is expected by the agent to bring the world in accordance with his or her desire. Unlike beliefs, goals, intentions and desires are motivational states: they are motivationally efficacious.

Now, I take it that truth is to beliefs or utterances what reference is to words or concepts. If meaning is normative, so are reference and truth. Deflationists about meaning deny that meaning is a normative property. Similarly, they deny that truth and reference are normative properties. Hence, they deny that, in Bernard Williams' (1970) words, "beliefs aim at truth". On their view, what is constitutive of beliefs is that they obey the principle of bivalence: beliefs ought not to be true simpliciter, they ought to be true *or* false. This is, according e.g., to Dretske (2000), enough to distinguish beliefs from other propositional attitudes, e.g., desires.

Just as it was incumbent upon the deflationist to account for the normative implications of meaning, it is incumbent upon him to account for the normative implications of truth. Why do we assess differently true beliefs and false ones? Again, the deflationary account of the normative implications of truth is pragmatic and it makes the normative consequences of truth conditional upon a creature's motivational states. Although beliefs do not aim at truth, we nonetheless generally prefer to hold true beliefs rather than false ones because true beliefs are more likely to promote practical success than false ones. Of two creatures, the one whose belief forming mechanisms

are more reliable is the more likely to survive.[2] The pragmatic account of the normative consequences of truth could thus be captured by conditional [C4]:

[C4] If *p* is true (and sufficiently relevant), then one ought to believe that *p* provided that one intends to satisfy one's desires.

Presumably, the deflationist does not want to be saddled with the absurd view that if a proposition is true, then one ought to believe it provided that one intends to satisfy one's desires. Presumably, a person ought to believe only *relevant* truths provided he or she intends to satisfy his or her desires.[3] Conditional [C4] is expressed in terms of an individual's higher-order goal, intention or desire to satisfy his or her lower-order goals or desires. It makes the normative consequences of the truth of a relevant belief conditional upon the agent's higher-order desire to achieve practical success.

4. Criticism of the pragmatic explanation of the normative consequences of semantic properties

First, I want to consider the pragmatic explanation of the normative consequences of meaning. The deflationary view that the normative consequences of meaning are conditional upon an individual's goals or desires is plausible in the case of the meanings of words of public languages because ordinarily, one uses words of his or her language in order to communicate his or her thoughts to his or her conspecifics. So making a speech act by uttering words of a public language is an intentional action. And intentional actions depend on the agent's beliefs and desires.

But what about thoughts? Thoughts, just like words, have content. However, unlike uttering a word of a natural language, thinking or entertaining a thought is not – or at least, not always – an intentional action. It may but it need not be an intentional action. On my view, unlike my utterance of the word 'chien', my tokening of the concept DOG need not be an intentional action. Suppose I token my concept (or mental symbol) DOG upon hearing some dog bark in the neighborhood. My thinking about a dog involves my tokening of the concept DOG. And the tokening of my concept DOG may be automatically prompted or triggered by my processing an acoustic perceptual input. My tokening of my concept DOG or my thinking about dogs is not – not always – under the control of my goal, intentions and desires. If I cannot

[2] The motivational state may be a personal or a sub-personal state.

[3] I take it that it would be strange to require of a person that she believes any truth. That is why I add 'relevant'.

help thinking 'There is a dog in the street', then my entertaining that thought or belief is not an intentional action. Nonetheless, my concept DOG correctly applies to dogs and only dogs in virtue of its content.

I grant, however, that many beliefs may arise as a result of an intentional action. So when a mathematician believes (or knows) that a proposition is true because he just proved it or because he knows how to prove it, then his belief in the truth of the theorem is the outcome of an intentional action. He believes it because he proved it or because he grasped a proof. And he proved it because he decided, he wanted or he intended to prove it. Or he grasped a proof because he decided to learn or study a proof. Similarly, when a scientist comes to believe that a hypothesis is correct either because he has accumulated enough evidence in its favor or he has disproved a rival hypothesis, then his belief arises out of an intentional process. Following Dennett (1978) and Sperber (1997), I shall call such beliefs "reflective" beliefs because they typically involve linguistic communication which involves the grasp of communicative intentions which in turn relies on the metarepresentational ability to think about other people's thoughts.[4] Unlike "reflective" beliefs, however, what I call "intuitive" beliefs typically arise from basic cognitive processes such as perception, memory and rudimentary inferential mechanisms. I assume that, unlike "reflective" beliefs, a creature with no ability to form higher-order thoughts about thoughts can form "intuitive" beliefs.

Now, "intuitive" beliefs involve the intuitive application of concepts (e.g., the concept DOG). Conversely, the mastery of concepts involves knowledge of its conditions of application. Whether one wants to think of knowledge of the application of a concept as explicit or implicit, there are norms for the application of a concept: my concept DOG applies correctly in some circumstances, not in others. Possessing the concept DOG involves knowing in what circumstances it is correct to apply it. Deflationists deny that concepts have normative conditions of application. On their view, the conditions of application of a concept are not normative: it is merely a fact that the concept DOG applies to dogs and only to dogs. However, the conditions of application of a concept may have normative consequences which derive from an individual's desires. My point is that the normative consequences of the application of a concept cannot – not always – be conditional upon an individual's goals or desires. It cannot when the application of a concept is not an intentional action, i.e., something explainable by the agent's motivations. If a concept is a constituent of an "intuitive" belief, then the normative

[4] Dennett (1978) calls 'opinions' what I call reflective beliefs.

distinction between its correct application and its incorrect application cannot be conditional upon the individual's goals or desires.

Secondly, I want to consider a potential rebuttal from the deflationist. Suppose we accept the deflationst suggestion that beliefs do not aim at truth. Suppose rather that one ought to *prefer* true beliefs over false ones because, unlike false beliefs, true beliefs contribute to practical (or reproductive) success. Suppose also that correctly applying a concept contributes to forming true beliefs. Should one not conclude then that the norms for correct applications of a concept derive from one's preference for true beliefs over false ones? If so, then the fact that the application of a concept has normative consequences would indeed be conditional upon an agent's goals or desires.

But from the above two premisses, what does follow is that one ought to prefer applying a concept correctly rather than incorrectly. It does not follow that the norms for correct applications of a concept derive from one's preference for true beliefs. It is one thing to hold a preference for applying concepts correctly. Norms for correct application are something else. If I prefer to form true beliefs over false ones, and given that my correct application of my concept DOG contributes to forming true beliefs about dogs, then it follows that I ought to prefer to apply correctly rather than incorrectly my concept DOG. But the fact that my concept DOG correctly applies to dogs and to nothing else does not depend on my preference for true beliefs. Even if I had a preference for false beliefs, my concept DOG would nonetheless apply correctly to dogs and incorrectly to anything else. The normative distinction between correct and incorrect application of a concept does not derive upon one's preference for true beliefs.

Thirdly, I want to consider the deflationary view that truth is not a normative property of beliefs (or utterances). According to deflationism, beliefs do not aim at truth. According to the pragmatic account of the normative consequences of truth, beliefs must be true or false and we prefer true beliefs to false ones because true beliefs allow us to achieve practical success better than false ones. The reliability of belief-forming mechanisms enhances survival.[5] I have already offered reasons for rejecting the pragmatic account of the normative consequences of meaning and truth. Now, I want to question the deflationary view itself: I want to resist the claim that it is constitutive of beliefs to be true *or* false (or to obey the principle of bivalence), not to be true simpliciter. I claim that to believe that *p* is to believe that *p* is true, not to believe that *p* is true or false. Let me say why: if one discovers that one of one's beliefs is false, then one will either reject it or take it that

[5] Stich (1990) has argued that survival cannot depend only on the reliability of the belief forming mechanisms: there must be a trade-off between reliability and costs.

one ought to reject it. Now a false belief is a belief which is true or false. So if it were constitutive of beliefs to be true or false, then upon discovering that one of one's beliefs is false, then one would have no reason to reject it. One ought not to feel impelled to reject it. But I claim that one would – or one ought to – feel impelled to reject it. If I am right, then I think that, contrary to the deflationary view, in Williams' words, "beliefs *do* aim at truth".

I suppose that the deflationist might want to object to the above little argument in favor of the view that beliefs ought to be true – or aim at truth – by pointing out that if deflationism was correct, then indeed one ought to reject a false belief provided that one prefers to hold true beliefs rather than false ones. On the deflationist view, beliefs ought to be true or false. So, given that true beliefs, not false beliefs, are conducive to practical succcess, it would indeed follow that one ought to reject a false belief, on the assumption that one prefers to have beliefs that are conducive to practical success. But what if one prefers to have false beliefs? Presumably, if one did not prefer true beliefs, then one would not have any reason to reject a false belief. Hence: beliefs ought to be true or false, not just true. And one's preference for true beliefs must be independent from any intrinsic or constitutive property of beliefs.

The question raised by the deflationist rebuttal then is the following: What does it mean – what could it mean – to prefer *false* beliefs over true ones? Could one truly prefer false beliefs? I would like to distinguish two cases: one may have a *local* preference for false beliefs over true ones, and one may have a *global* preference for false beliefs. By local preference, I mean that one may have reasons to have, or it may be adaptive for one to have, *some* false beliefs among one's overall set of beliefs. Equivalently, one may no doubt prompt some one else to have *some* false beliefs. One may, for example, genuinely prefer to hold a subset of false beliefs if e.g., some true belief is too painful or if some subset of beliefs is too costly to form. But this local preference, I think, does not really threaten the claim that beliefs ought to be true – no more so than does the fact that many beliefs *are* in fact false. It is only, it seems to me, against the background of true beliefs that one may prefer to hold a subset of false beliefs, if e.g., either some true belief is too painful or too costly to entertain.

The real problem then has to do with one's *global* preference for false beliefs. If one had such a global preference for false beliefs, then one would (or ought to) reject any true belief and one would (or ought to) accept only false beliefs. The job of the belief-forming mechanism would be to get things wrong systematically. It is very hard, it seems to me, to make sense of such a global preference from either an evolutionary point of view or a (personal) psychological point of view. From an evolutionary point of view, I fail to see how a creature with a global preference for false beliefs could survive. A

creature with a belief forming mechanism that would get things systematically wrong would be dead. From a psychological point of view, the problem seems to me to be that a global preference for false beliefs is threatened by a paradox which is reminiscent of the *liar*'s paradox. Presumably, a global preference for false beliefs would apply to higher-order beliefs about oneself (about one's lower-order beliefs, desires and other attitudes). So suppose that a person who claims to have a global preference for false beliefs were asked: "Is it *true* that you prefer to hold false beliefs? Do you truly believe that you prefer to have false beliefs?" What is this person going to answer? She can only falsely believe of herself that she prefers false beliefs. She cannot truly believe it. If she believes it, then she does not prefer to hold false beliefs. Conversely, if she does not believe that she prefers to hold false beliefs, then she believes it. She can only believe it if she does not believe it. So it cannot be right that she prefers to have false beliefs. This, I think, suggests that no creature can have a global preference for false beliefs. Only local preference for false beliefs makes sense. But, as I said, local preference for false beliefs does not really threaten the thesis that beliefs ought to be true any more than mistakes do.

5. Semantic and non-semantic norms

When an agent performs an intentional action – an action done for a reason –, the possibility arises of there being a *gap* between what the agent does and what he ought to do. If an agent does something for a reason, then his reason motivates his action. His reason is a motivating reason. Not all reasons for an action, however, are motivating reasons. In addition to motivating reasons, there are normative reasons for action. An agent's motivating reasons may, but they need not, coincide with his normative reasons. In other words, normative reasons and motivating reasons may part company. An agent may recognize a deontic or an ethical norm and fail to act in accordance with the norm. Recognition of an ethical norm is therefore consistent with having no desire to comply with it. I may recognize that I ought to do such-and-such and fail to do what the norm prescribes. Conversely, my motivating reason may conflict with some normative reason of which I am perfectly aware. Recognition of a norm without having a desire to comply with the norm simply lacks motivational force. Recognition of an ethical norm is therefore consistent with violation of the norm.

Let us call "valuing" the recognition of an ethical norm or of a normative reason. I think that the very possibility of a gap between recognition of an ethical norm and motivation suggests that valuing – recognition of an ethical

norm – cannot consist in having a first-order ordinary desire. The question then arises whether recognition of an ethical norm – valuing – consists in having a belief or having a higher-order desire (a desire to form ordinary first-order desires). I will not try to adjudicate this dilemma here. In any case, as I already said, I assume that only ordinary first-order desires are genuine motivational states. So neither beliefs nor higher-order desires are, on my view, genuinely motivational states.

Now, consider the relation between truth and what I earlier called "intuitive" beliefs. There is simply no gap between recognition of truth and believing. It is neither open to me to refrain from believing what I take (or know) to be true, nor to believe what I take (or know) to be false. This is presumably what Moore's paradox taught us: I cannot consistently entertain the thought that it is raining but I do not believe it. Nor can I entertain the thought that it is not raining but I do believe it. Of course, I can accept a proposition for the sake of argument. But acceptance of a proposition for the sake of argument – as I already claimed – is not forming an "intuitive" belief. Arguably, the reason why there is no gap between truth and belief is that in the process whereby I recognize the truth of a proposition and I believe it, there is no room for motivation or for desire. And the reason why there is no room for motivation or for desire is that forming an "intuitive" belief is not an intentional action. I do not form an intuitive belief because I intend to or want to. Unlike ethical and deontic norms then, semantic norms are not norms of intentional action. Contrary to deflationism, I think that beliefs aim at truth. Beliefs ought to – their job or their function is to – be true. Contrary to deflationists, I think that truth is a normative property of beliefs.

6. Two objections

Before closing, let me consider two objections. The first objection consists in what seems to be a counterexample to the claim that semantic norms differ from other (e.g., ethical) norms in that they cannot always be willingly violated. It is an objection to the claim that in the case of semantic norms, there may be no gap between recognition of a norm and motivation. The second objection simply denies that what cannot be knowingly violated can be a norm at all.

First of all, there are cases where the "intuitive" belief-forming capacity seems to be at a loss. A stick seen in the water looks "broken". In spite of the fact that I know that unbroken sticks look broken when seen in the water, a stick seen in the water will still appear broken to me. In the famous Müller-

Lyer illusion, two equal segments with two different endings look unequal. Although I may know and therefore believe that two Müller-Lyer segments are equal, nonetheless the two equal segments will still look unequal to me. In such cases, the subject recognizes that visual appearances are deceptive. Although he recognizes that he ought to hold true a proposition, does he not lack the motivation to believe what he ought to?

I think not at all. Although the stick seen in the water seems broken to me, if I know that unbroken sticks seen in the water look broken, then I will discard the visual appearances and accept the belief that the stick which looks broken when seen in the water is in fact not broken. I will accept this "reflective" belief in order to minimize chaos in my overall picture of the world. Now, my choice is not infallible. I may be wrong: *broken* sticks too look broken when seen in the water. If I know that two Müller-Lyer segments are equal (either because I was told or because I measured them), then I will discard the visual appearances and judge them to be equal. So it is simply not true that I recognize a norm but I am not motivated to believe what I take to be true.

The last question I want to consider is the question whether it is constitutive of norms that they can be knowingly violated. I grant that the meaning of a word can be knowingly violated. Although I know that the French word for horses is 'cheval' and the French word for sheep is 'mouton', I can decide to apply the word 'mouton' to horses and the word 'cheval' to sheep. However, I deny that the norms for the correct application of a *concept* can be knowingly violated. I claim that although I can intentionally misapply the word 'mouton' to a horse, I cannot, upon hearing the sound produced by a galloping horse, intentionally misapply my concept SHEEP to categorize the auditory stimulus. I can make a mistake in perception. But I cannot intentionally misapply a concept. As I said, using a word of a public language for the purpose of communication is an intentional action and as such it depends upon one's goals, desires and intentions. But applying a concept as a result of processing a perceptual stimulus is not an intentional action and it does not depend upon one's goals, desires and intentions. Similarly, because forming an "intuitive" belief is not an intentional action, one cannot intentionally form a false "intuitive" belief. As far as "intuitive" beliefs are concerned, one cannot intentionally violate the norm of truth. Since all beliefs aim at truth and "intuitive" beliefs dot not arise out of intentional actions, one cannot intend to form a false "intuitive" belief. Now, the objection goes: what follows is not that semantic norms differ from other (e.g., ethical) norms. What follows is that what I call "semantic norms" are not norms after all.

According to the objection, nothing will be a norm unless it can both strike a person as a reason for action and at the same time lack motivational

force. The issue raised by the objection is, I think, whether all norms do or not apply at the personal level – whether norms must be open to conscious personal inspection – or whether some norms may apply at the subpersonal level. In the remaining time, my goal will be modest: I will merely argue that the notion of a subpersonal norm is not incoherent. I will adduce two sorts of evidence in favor of this modest claim. I will appeal to the computational paradigm in cognitive science and to functional explanations in biology.

First of all, I want to reflect on the fact that the strong claim that all norms must apply at the personal level and be available for conscious inspection should be reminiscent of Searle's (1992) Connection Principle according to which genuine mental states – as opposed to merely physiological processes – must be accessible to consciousness. And it deserves, I think, the same response. Explanations in cognitive science are framed as computational explanations of the behavior of information-processing systems. I assume that not all information-processing systems are conscious systems. But all information-processing systems are physical systems. So they fall under physical laws and physical explanations. Computational explanations presuppose that not everything an information-processing system does can be usefully described by showing how it instantiates physical laws. Some of the behavior of an information-processing system is best explained as a set of computations. Since a little computing device is a physical device, its behavior falls under the laws of physics. If, however, it is capable of performing additions, then only a computational explanation may fully capture its arithmetical behavior. Now, my point is that a computational explanation presupposes norms of computation. Given a pair of natural numbers as input, there is a number which is their sum and which is such that the device *ought* to compute it as output. Norms for computing additions are built into the device. They are not accessible to the device for conscious personal inspection for the simple reason that the device is nonconscious. Nor could the normative implications of arithmetical computations be derivable from the device's motivational states since the device lacks motivational states. Cognitive science requires, it seems to me, that norms can apply at the subpersonal level and hence that they be unavailable for conscious inspection, let alone that they be knowingly violated.

Arguably, a computing device is an artefact. So norms of computations may have been built into the computing device by a human being with beliefs and desires. So the normative implications of arithmetical computations may after all depend upon a human being's motivational states. However, I do not think that any mistake I ever made in computing an addition could be explained as a failure to be motivated by an arithmetical norm which

I recognized. Nor do I think that arithmetical norms depend on motivations.

On the kind of teleosemantic reductionist view which I hold, semantic and cognitive norms do not depend on motivations.[6] They derive from, and are built into, the semantic and cognitive functions which are, on my view, a species of biological functions. Like Millikan (1993) and Neander (1995), I think that, unlike physics and chemistry, biology is loaded with norms. If a planet of the Solar system were to stop gravitating towards the Sun, it would stop being a planet of the Solar system. If a chemical bond between two atoms within a molecule breaks down, then the whole molecule disintegrates or gives rise to another molecule. If, however, a biological organ stops performing its biological function, it does not ipso facto loses its identity. Why? Because a biological organ has a biological function. I accept Wright's (1973) etiological theory of functions according to which functions are selected effects, i. e., they are things which a device can do and which have been promoted by some selection process. It is the function of a mammal's heart to pump blood. A mammal's heart may fail to pump blood. If it does not pump blood, then it does not what it ought to do. If, however, it does not do it job, it is not thereby disqualified as a heart. A diseased heart may misfunction. But if it does, it is still a heart. The crucial point here is that the relevant biological notion of a function is a teleological notion, not a dispositional notion: if an organ (or a biological device) has a function in this teleological sense, then there are things which it is supposed to do or which it ought to do. In other words, there are biological norms and they arise from biological functions. Biological functions in turn arise from the physical process of natural selection at work in evolution.

In sum, my point is that not all norms depend on motivations. Norms of computations (or grammatical norms for that matter) and biological norms do not apply at the personal level. Nor are they be available for conscious inspection. Nonetheless, I want to claim, they are norms. Furthermore, unlike ethical norms, they cannot, I claim, be knowingly violated because whether they are enforced does not depend upon any motivational state.

References

Burge, T.: Individualism and the Mental. In: *Midwest Studies in Philosophy, vol. IV*, ed. by P. French, T. Uehling, Jr. & H.K. Wettstein, University of Minnesota Press, Minneapolis 1979.

[6] See Jacob (1997).

Dennett, D.: How to Change Your Mind. In: D. Dennett, *Brainstorms*, Bradford Books Publishers, Inc., Mongtomery, Vt 1978.
Dretske, F.: Norms, History, and the Constitution of the Mental. In: F. Dretske, *Perception, Knowledge and Belief* , Cambridge University Press, Cambridge 2000.
Hale, B.: Rule-Following, Objectivity and Meaning. In: *A Companion to the Philosophy of Language*, ed. by B. Hale & C. Wright, Blackwell, Oxford 1997.
Horwich, P.: *Meaning*, Oxford University Press, Oxford 1998.
Jacob, P.: *What minds can do*, Cambridge University Press, Cambridge 1997.
Kripke, S.: *Wittgenstein on Following a Rule*, Blackwell, Oxford 1982.
Millikan, R.G.: *White Queen Psychology and Other Essays for Alice*, MIT Press, Cambridge, Mass. 1993.
Neander, K.: Misrepresenting and Malfunctioning, *Philosophical Studies* 79 (1995), 109–41.
Putnam, H.: The Meaning of 'Meaning'. In: H. Putnam, *Philosophical Papers, vol. 2*, Cambridge University Press, Cambridge 1975.
Searle, J.: *The Rediscovery of the Mind*, MIT Press, Cambridge, Mass. 1992.
Sperber, D.: Intuitive and Reflective Beliefs, *Mind and Language* 12 (1997), 67–83.
Stich, S.: *The Fragmentation of Reason*, MIT Press, Cambridge, Mass.1990.
Williams, B.: Deciding to Believe (1971). In: B. Williams, *Problems of the Self* , Cambridge University Press, Cambridge 1973.
Wright, L.: Functions. In: *Conceptual Issues in Evolutionary Biology*, ed. by E. Sober, MIT Press, Cambridge, Mass. 1973.