THEORIA ET HISTORIA SCIENTIARUM, VOL. VIII, N° 1
Ed. Nicolaus Copernicus University 2008

Josefa Toribio

Normativity: the Hard Problem

Introduction

In a recent article, John Haugeland (1998) presents a detailed specification of different kinds of rules and rule-following conceptions. The ultimate goal is to ascertain the nature of the rules involved in semantic behavior, in particular in the semantic phenomenon of truth-telling. In order to achieve his goal, Haugeland develops a complex taxonomy not only of rules but also of the different normative statuses attached to them. His aim is to analyze two familiar ways of understanding basic governing rules (biological and social) and show that both fall short of providing a foundation for the phenomenon of truth-telling. The biological — reductionist— account of semantic properties fails because it leaves no room for the distinction between malfunction and misrepresentation, a distinction that is introduced as constitutive of the kind of normativity involved in truth-telling. The social —anti-reductionist— account of semantic properties fails because there is no room for another important distinction: the distinction between social propriety and objective correctness (truth). His strategy is to shed light on *which* characterization of governing rules would be adequate for truthtelling by showing how these two fail.

My goal in the present paper is to use Haugeland's treatment as a springboard to explore a key problem for biological accounts of semantic properties, namely, how to capture the distinction between malfunction and misrepresentation. While Haugeland generates a complex taxonomy of species and subspecies of rules to support his arguments against the risks of identifying social propriety and objective correctness, he is rather more austere in his analysis of the naturalistic problem. I shall sketch a somewhat richer landscape, in which the space of error- by-misrepresentation (as opposed to malfunction) is itself split into two. On the

one hand, there is the notion of (what I shall call) *'mindless* misrepresentation'. On the other, a more full-blooded species of misrepresentation that I shall call 'cognitive error'. *Mindless* misrepresentation occurs when the system, although not malfunctioning, cannot be held responsible for its mistake or wrong-doing. Such responsibility characterizes instead the kind of mistake or misrepresentation I dub 'cognitive error'. I thus construct a tripartite space comprising (mere) malfunction, *mindless* misrepresentation, and full-blooded misrepresentation (cognitive error). I use this three-way classification of wrongness to revise and modify Haugeland's criticisms of biological teleology and to argue for what I call the 'hard problem of normativity". When applied to linguistic behavior, the *hard problem* is to give an account of cognitive error that respects naturalistic constraints without conflating that notion with either malfunction or *mindless* misrepresentation.

This is how I proceed. In Section 1 I introduce and further subdivide the distinction between malfunction and misrepresentation in the context of Haugeland's critique of biological accounts of semantic properties. In Section 2 I compare my three-way distinction (malfunction / *mindless* misrepresentation / full-blooded misrepresentation) to a related distinction found in the practice and philosophy of law. Thus we need to consider in the legal case, not just the standard (though itself problematic) distinction between culpable crime and acts of insanity, but also a distinction, drawn among charged criminal acts committed by the fully sane, between those actions for which one is held responsible, i.e. guilty, and those (usually involving excusable error or ignorance) for which one is not. This latter kind of case, I shall argue, corresponds nicely to the case (introduced in the previous section) of *mindless* misrepresentation: cases where there need be no malfunction (legal parallel: no insanity), where there is some kind of contravention of a norm, yet no ultimate culpability. In pursuing this parallel I do not, of course, aim to provide a detailed discussion of these tricky and important issues in the practice and the philosophy of law. Rather, the parallel is drawn as a way of clarifying and further motivating the tripartite distinction argued for in Section 1.

Section 3 reconsiders the (alleged) failure of biological (and more particularly, teleological) accounts of semantic properties in the light of the new, finer-grained classification. My contention is that teleological accounts can (pace Haugeland) respect the distinction between malfunction and misrepresentation, but only so long as 'misrepresentation' is understood in the *mindless*, noncognitive fashion. They typically err, however, in conflating the normativity attaching to this non-

[1] I borrow the expression and the spirit behind it from David Chalmers, who distinguishes two kinds of questions —easy and hard— about conscious phenomena. See David Chalmers (1996).

cognitive notion of misrepresentation with the kind exhibited by its full-blooded, cognitive counterpart. The naturalistic explanation of this latter kind of misrepresentation constitutes the truly *hard* problem of normativity.

I. Malfunction vs. Misrepresentation

Following a Fregean tradition, we can say that a thought (and the sentence that expresses it) is defined as that of which we can predicate truth or falsity (see Frege, 1918). Without acknowledging that essential feature, the possibility of communication or description is not intelligible. The very idea of language as a representational system is built upon this normative distinction. Linguistic behavior —in particular linguistic behavior consisting in the assertion of declarative sentences— is thus a norm-governed phenomenon. When immersed in declarative linguistic behavior, we are following rules that establish not only and not mainly) what is appropriate or polite or advisable to say in such-and- such circumstances, but also, and most importantly, how the world would have to be for the sentences to be true. What is the nature of those rules? How do we actually comply with them? These are the kind of questions which Haugeland (1998) attempts to answer, if only in an indirect way. In order to do so, he develops a rather complex taxonomy of rule-following conceptions and their normative statuses. The first distinction that Haugeland draws is between factual and governing rules[2]. Factual rules (also called 'exhibited' or 'descriptive' by Haugeland) rehearse what happens; they describe a certain order as presented to us by the world. Governing rules (also called 'prescriptive' or 'normative' rules) *determine* what happens. They are about how the world *ought* to be, not about how it is.

We can use Anscombe's example to illustrate this distinction. Imagine a detective who follows a list-in-hand shopper writing down everything the shopper buys. The lists of both the shopper and the detective will be identical in the end, but their nature is very different. The detective's list is meant to match the world, i.e. to *reflect* the shopper's actions, whereas the shopper's list *guides* his own actions. He is forcing the world to match the words on his list by putting into his basket only those items already listed (see Anscombe, 1957, pp. 56-57). When considered as a rule, the detective's list captures a pattern of behavior, something that happens to be the case, but the rule (the list) does not have the same normative status as the shopper's list since it does not tell us what the

[2] See also G. E. M. Anscombe (1957), §32, and John Searle, „A Taxonomy of Illocutionary Acts", in J. Searle (1979), pp. 3-4, for a similar distinction.

shopper's behavior *ought to* be. Factual rules have, in Haugeland's interpretation, a *world-to-rule* direction of fit. If the mismatch of rule and world is the only criterion for what we might call making a mistake, then, in the case of the detective, any mistake is in the list (i.e. it could easily be fixed by e.g. deleting one word and writing the 'right' one). The shopper's list, on the contrary, has the (different) normative ingredient of guiding —and not merely reflecting— what the shopper's behavior *ought to* be. This is the essential feature of governing rules: rules which have, according to Haugeland, a *rule-to-world* direction of fit[3]. In this latter case, if there was a mistake, the mistake would not be in the rule (i.e., the list), but (typically) in the shopper's behavior[4]. Changing the names of the items on the list would not normally "make it right" (see Anscombe, 1957, p. 56). Governing rules are the focus of Haugeland's interest.

As Haugeland acknowledges, and as most famously suggested by Wittgenstein (1962), following a rule by way of interpreting an expression of it leaves us facing a potential infinite (and vicious) explanatory regress. Since complying with a rule by interpreting the expressions in which the rule is expressed involves following some other (also governing) rule, how are we to understand compliance with this latter sort of rule? Wittgenstein's solution to the problem was to interpret compliance with some rules in a way that did not require any further compliance with rules *of that sort* (see Wittgenstein, 1962, §84-87; §198-201). As Haugeland puts it: "for each sort of governing rule, at least some rule compliance must be intelligible as *basic*, in the sense that the rule is not complied with by (that is, by means of) complying with some other rule(s) of that sort" (Haugeland, 1998, p. 307). The search for an account of semantic properties in terms of *basic* governing rules is thus the search for those non-semantic rules by means of which semantic behavior can be made intelligible in a non circular way. Since the rules that make linguistic behavior intelligible

---

[3] Searle's terminology, the detective's list has a word-to-world direction of fit whereas the shopper's list has a world-to-word direction of fit (see J. Searle, *op. cit.*, p. 4). Haugeland, however, identifies factual or exhibited rules as having a world-to-rule direction of fit and governing rules as having a rule-to-world direction of fit (see Haugeland, *op. cit.*, pp. 305-306). I don't think there is any deep inconsistency or misunderstanding here. Searle takes the direction of fit to be a consequence of the illocutionary point (purpose) of a type of illocution. Statements, descriptions, assertions, and in general the types of illocutions of which it makes sense to say that they are true or false all have a word(rule)-to-world direction of fit. Requests, commands, and promises, for instance, have a world- to-word(rule) direction of fit. Given that Haugeland's target is the normative apparatus of truthtelling, his interpretation of the dichotomy makes sense even if the reading is confusing at first blush.

[4] There is actually another way in which a governing rule can go wrong. This is the case in which the rule (the list, in Anscombe's example) is somehow misguided. For instance, it would be misguided if the names on the list referred to items impossible (practical impossibility) to find.

are not basic, we have to find a different *sort* of rule, compliance with which would be intelligible without having to invoke any other *semantic* rules. Reductionist accounts of semantic properties are precisely an attempt to provide such *naturalized*, i.e., non-semantic, kinds of rules, and teleological accounts seem especially suitable for this task.

Now, for any kind of system, there is an intuitive and clear difference between the system working properly but getting things wrong vs. the system getting things wrong as a result of malfunctioning in some way or another. If due, let's say, to some optical illusion, I fail to discriminate the twelve steps out of my house (perhaps seeing only ten), and I fall, my *mistake* need not be the result of anything going wrong at the level of my physical or physiological functions. My visual system and neural mechanisms may all be working just as they are supposed to, yet (under these specific conditions) they fail to deliver a veridical representation of the number of steps, much as a perfectly functioning heart may under extreme pressure) fail to pump sufficient blood. In such a case it is surely still correct to say that I *misrepresented* the number of steps. If my fall had instead had its origin in a bad case of arthritis, such talk about misrepresentation would not be appropriate. What we would there have is a case of physical *malfunction* even though the number of the door steps were all correctly perceived. Of course, one can also imagine a case in which the visual system *malfunctions* in a way that causes it to misrepresent. For example, if a sudden neural accident caused double vision resulting in a fall.

Even in these oversimplified examples, it is not difficult to see how very close these several notions run. Under appropriate or normal circumstances, i.e. without optical illusion or neural accident, the visual system performs its proper function and misrepresentation of the door steps typically does not occur. The notion of proper function is thus a favorite among biologically oriented naturalists who use it as a non-semantic way of cashing out the notion of representation. And once veridical representation is cashed out in terms of proper function, misrepresentation quickly looks like a case of malfunctioning (see Millikan, 1984). This assimilation is central to the development of teleological accounts of semantic properties.

Such approaches are also called upon to provide a solution to the so-called disjunction problem'. This problem arises because purely causal representational accounts seem unable to distinguish between someone *mistakenly* using a word and someone *(correctly)* meaning something completely different when using that word, a failure that suggests a potent objection to naturalistic proposals of the causal-informational kind (see Fodor, 1990; Dretske, 1986, 1988). This is especially obvious within causal-informational theories because most situations in which an instantiated property (e.g. COW) counts as the cause of a given belief token are also situations in which other properties, co-instantiated with

the first (e.g. COW OR HORSE-IN-THE-DARK. See Fodor, 1990), are also possible causes of that belief token. The disjunction problem thus arises whenever a theory can't distinguish between a true tokening of a symbol that means something disjunctive, and a false tokening of a symbol that means something non-disjunctive.

The key move in the teleological attempt to solve this problem is to consider evolved structures whose adaptive role is that of gathering the information necessary for a creature's survival. We thus first define a relation that beliefs bear to properties that are sometimes instantiated in the system's environment as the relation 'has as its content that'. The instantiation of a given property then explains why the production of a belief token in certain conditions helps the system perform its proper function. Instantiations of a property causally affect the system and are evolutionarily relevant with respect to the proliferation of the system. Once all this is in place, false beliefs are explained in terms of the direct malfunction of the evolved structures or some alteration of the usual conditions in the environment (cf. Millikan, 1984).

In thus treating misrepresentation as a special case of malfunction, teleological theories become liable to the kind of criticism Haugeland has in mind: biologically defined norms do not support a distinction between malfunctioning and making a mistake without any malfunction. Here is how Haugeland illustrates the point:

> Imagine an insectivorous species of bird that evolved in an environment where most of the yellow butterflies are poisonous, and most others not; and suppose it has developed a mechanism for detecting and avoiding yellow butterflies. Then the point can be put this way: if a bird in good working order (with plenty of light, and so on) detects and rejects a (rare) nonpoisonous yellow butterfly, there can be no grounds for suggesting that it mistook that butterfly for a poisonous one; and similarly, if it detects and accepts a (rare) poisonous orange butterfly ... in such cases, ... there is nothing that the response can „mean" other than whatever actually elicits it in normal birds in normal conditions. (Haugeland, 1998, p. 310).

Haugeland's idea —which is surely correct— is that we need to maintain the distinction between a system getting things wrong as a result of some fault and a system working normally that nonetheless actually makes mistakes. Despite this initial agreement, I believe Haugeland's analysis of the case is compromised by a failure to spot a further (and important) distinction within the space of misrepresentation itself. For Haugeland „there can be no biological basis for understanding a system as functioning properly, but nevertheless misinforming" *(op. cit.*, p. 310). But this seems to me to be wrong. Biological accounts can and do support such a possibility. Where they fall short is in relation to an even stronger notion, viz., a notion not just of misrepresentation (misinforming), but of *culpable* error—error that the system (agent, creature) should have been able to avoid.

To bring this into focus, let's look at the bird example again. I believe there is a strong sense in which it genuinely makes sense —from a teleological view point— to say that the bird *did* make a mistake, even though the bird is not malfunctioning. The bird-species has learnt to use the perception of yellow as a sign of poison and has learnt to apply it appropriately (to avoid it) in a given task (foraging). The 'rightness of the signal' is constituted by its success in this task[5]. The mechanism which detects yellow and triggers an avoidance behavior is properly in place. The bird behaves as if the environment was a certain way (the way the environment *ought* to be given how things have been in the past). Yet, the environment now plays a trick. In this sense, since the task has not been facilitated (perhaps the bird dies of hunger in the first case and from poison in the second), there is a clear sense in which something has gone wrong, something that has nothing to do with bird malfunction.

The case is parallel, 1 suggest, to that of my falling as a result of some optical illusion. Here successful behavior is defined by successful completion of some extra-linguistic task, let's say, that of getting safely from my home to the street. I am —by hypothesis— functioning normally and 1 am properly connected with the environment; my behavior is linked directly to my perceptual experience, and involves no conscious inferential or linguistic reasoning. I am thus not responsible for my fall, but I certainly *misrepresented* the world in a specific way, viz. as containing ten steps instead of twelve. It is a case of misrepresentation without malfunction, and without culpable error: a case of what I am calling *mindless misrepresentation.*

So when *can* we talk of full-blooded misrepresentation? When, indeed, there is something like *culpable error.* When the mistake is in a deep sense a *cognitive* mistake. By that I mean not only that it belongs, or could belong, to the stream of consciousness but also, and more importantly, that the failure of some purely causal process is not criterial for its wrongness. In other words, the correctness of the behavior, in the full-blooded cases, is not defined simply in terms of its outcome. An example. Joe, a trained logician, asserts that St. Louis is larger than London. He knows London is larger than Paris, and just learnt, from a reliable source, that Paris is larger than St. Louis. His assertion is false, and misrepresents the size of St. Louis. His assertion is false not only and not essentially because he may fail in some extralinguistic task involving the size of St. Louis (e.g. allowing himself more time than he needs to go from one neighborhood to another in the city). His assertion is false even if he never had and never would have any perceptual or causal interactions of any kind with St. Louis, London, or Paris.

---

[5] I am here following Charles Taylor in his account of the explanatory direction(s) for the *rightness* of an action. See Taylor (1995a and b). This will be clearer in Section 3 of the paper.

Moreover, he is also answerable for his mistake since he could properly be asked to provide reasons for his belief: reasons that would surely display an inferential network whose failings are both visible to Joe himself and completely independent of any 'outcome', any behavior of his regarding the size of St. Louis. We can thus say of Joe —but not of the insectivorous bird— that he *should have done better.* It is this crucial, but difficult feature —the feature of *culpable* cognitive failure— that characterizes the cases I will call 'full-blooded misrepresentation'.

Once we acknowledge the presence of these two grades of misrepresentation, Haugeland's classification reveals itself as too simplistic to constitute a direct rebuttal of teleological theories of content. The problem with these theories is not that they don't leave room for the distinction between malfunction and misrepresentation. They do. The problem is that they don't leave room for the distinction between full-blooded and *mindless* misrepresentation.

II. Evil vs. Ill

I now turn, briefly, to a different debate where a distinction, closely parallel to that between *mindless* and full-blooded misrepresentation, can be seen to do significant work. Under Anglo-American criminal law, a person is not guilty of a crime unless two conditions are met: it has to be proved that the person has committed a criminal act *(.actus reus),* and also that the person had an evil intent in doing so *(mens rea).* The requirement of having a particular intentional state has important consequences, one of the most important being the idea that certain *abnormal* states of mind —states which show that there was no *evil* intent, such as insanity— can exculpate the person. Under certain circumstances then, some people who have been charged with a crime are not considered guilty; they are not considered legally responsible for their wrongful deeds. 'Insanity' is taken here to be a *medical* concept based on the scientific confirmation of some mental disorder, a disorder that then has to be assessed as exculpatory by a court of law. Since the Anglo-American[6] system is based on past procedures and decisions, the circumstances under which the condition of *mens rea* is not met are legion. Even an approximation to this issue is far beyond the limits of this paper. I would like nevertheless to review —if only superficially— some of these legal resolutions so as to motivate and clarify the distinctions argued for in the previous Section. Let me start with what can be taken to be clear cases of simple malfunction.

[6] From now on, whenever I refer to 'the' legal system, I will be referring to the Anglo-American legal system. This 'regionalism' does not affect my analysis.

For an insanity plea to be successful, the person charged with a crime has to be considered a malfunctioning cognitive system[7]. Insane agents are not deemed legally responsible because their wrong doing is the result of *a. failure* in their cognitive system. They are not punished, although they are, sometimes, held in medical institutions. They are, from the legal point of view, no more responsible for the results of their actions than is a malfunctioning gas heater that accidentally kills people by producing too much carbon monoxide. Their wrongful actions are the consequence of malfunction; the 'wrong of the ill' (Reznek 1997).

Of course, the very notion of insanity is one that provokes much debate in the practice and philosophy of law. The M'Naghten Rules[8] were established more than 150 years ago in an attempt to legislate the criminal responsibility of the mentally ill. Two of the most remarkable features of these rules are that they differentiate medical from legal insanity —the former considered a matter of fact; the latter a matter of evaluation for the court— and that they allow only *cognitive* tests of insanity. The tests are cognitive because they check exclusively whether a person's mental illness prevented her from *knowing* what she was doing[9]. Thus take Mrs. Griffiths, an 85 year old lady suffering from senile dementia, who suffocates her grand-daughter while the baby is asleep. She does not recognize her grand-daughter as a baby (but rather as a wild animal) and is completely unaware of the 'content' of her action. Her mind is not working properly and she is not held responsible for her action in a court of law. The wrongness of her deed is just the 'wrongness' of a malfunctioning system[10].

A criminal, by contrast, is said to be possessed of a properly working cognitive system but nevertheless *does the wrong thing.* These agents are considered fully legally responsible and are punished by the law accordingly.

---

[7] It has also to be proved that the criminal act charged is the product of such malfunction.

[8] The name refers to the case of Daniel M'Naghten's, a Scott who mistook Edward Drummond, private secretary to the English Prime Minister, Sir Robert Peel, for Sir Robert Peel himself, and shot him provoking his death in 1843. M'Naghten's defender used a doctor's testimony to argue that his client suffered from paranoia and lacked self-control. The jury's verdict was not guilty for reasons of insanity. Since this was actually a very liberal use of the excuse of insanity, the House of Lords called upon fifteen judges of the Queen's Bench to articulate in more detail the legal responsibility of those with some mental disorder. The result of such deliberations are the M'Naghten Rules, which have been in use both in England and North America ever since. See Walker (1968).

[9] As opposed to volitional tests which check whether a person's mental illness prevented her from being able to *control* what she was doing.

[10] This is a case in which a neural malfunction causes a state of misrepresentation. The agent is excused because the misrepresentation involves no culpable cognitive error (unlike Joe, the trained logician). What this shows is that *not all* cases of mindless misrepresentation need involve no malfunction. Certain malfunctions may be sufficient for mindless misrepresentation to occur. But malfunction is not necessary, and is frequently not involved.

They are *culpable* law-breakers, whose contraventions of social, legal and moral norms are often thought of as 'evil'. These cases correspond to those of fullblooded misrepresentation: cases where the agent is held responsible for her errors and norm-violations. There is a sense, of course, in which the comparison across these two categories may seem stretched, since the wrong of the evil does not seem to consist in any mismatch between the agents' representations and the world. However, the deeper parallelism emerges once we recall the rule-oriented normative picture with which we began. No matter how 'right' and 'wrong' are morally justified or characterized, legal wrongness involves *breaking a rule,* breaking the commitments —the *'oughts to'*— that constitute the legal basis for the well-being and well-working of a community, often in full knowledge of the legal rule. Our cases of full-blooded misrepresentation involve just that kind of broken commitment. The governing rules that regulate semantic behavior, and especially the semantic phenomenon of truth-telling, are *oughtladen rules*[11]. They are constituted by what the agent ought to say —ought to think— given the meaning of the expressions and sentences she uses, and also given how the world is. The agent is cognitively responsible both for the circumstances under which those expressions are correctly used, and the appropriate consequences of their application, namely their inferential roles in the space of semantic activity. These underlying inferential processes are essentially involved in the understanding-based capacities that constitute the genuinely cognitive sphere of an agent's representational life. The idea of cognitive error or fullblooded misrepresentation thus involves a mismatch between what the agent *ought to say,* given both the meaning of her words and how the world is, and what she actually *says.* In the case of punishable criminal behavior there is a parallel mismatch between what the agent *ought to do* given both how the legal terms have been characterized and how the legal system has been institutionalized, and what the agent actually *does.*

Now, still within this legal scenario, there are also cases which belong to an intermediate position parallel to our notion of *mindless* misrepresentation. Recall that this is the kind of misrepresentation that occurs when well-functioning agents make mistakes which do not involve any cognitive error and for which the agent cannot be held responsible. The parallel legal cases concern well-functioning cognitive agents whose proper perception of a situation leads them to be charged with a criminal act for which they are not held, in the end, legally responsible due to some condition of (excusable) ignorance. Thus consider the case of Ngok Keir, who killed, with his fish spear, a woman cutting durra heads on a field. Keir heard noises outside his house in Sudan and believed a marauding monkey was going to attack him. His perception of the field was accurate, and his inference

I borrow the expression 'ought-laden' from John McDowell (1998).

from the type of noise to the likely presence of a marauding monkey appropriate. The court thus declared Ngok Keir non guilty:

> The evidence shows that monkeys do frequent the durra cultivation in that locality, and that the spearing of such animals is not illegal, and that, when the accused threw his spear at the deceased, he assumed she was a marauding monkey, and did not know that she was a human being. (Katz 1987, p. 165)

Cases like this are interesting in the present context because their features closely resemble those present in *mindless* misrepresentation: something has gone wrong; a real mistake has been made, but, despite the absence of malfunction, it is a mistake for which the subject cannot be held responsible.

III. The Varieties of Normativity

We can now return to Haugeland's criticisms of teleological naturalism. Recall that Haugeland's main complaint was that these projects leave no room for the distinction between simple malfunction and misrepresentation. I have argued that this distinction is too coarse-grained, and that the notion of misrepresentation encompasses at least two distinct classes. How, then, does this finer-grained classification affect Haugeland's project? We can grant Haugeland's broad claim that naturalized approaches to semantic rules do not leave room for an important distinction, but the distinction is not that between malfunction and misrepresentation *(simpliciter)*, but rather that between mindless misrepresentation and cognitive error. Haugeland is right[12] in complaining about a certain amount of confusion in the way teleological views make contact with the normative features of semantic behavior. But such criticisms, by not explicitly acknowledging the naturalizability of what I am calling *mindless* misrepresentation, can easily seem misguided, blind to the true potential of naturalized approaches. The problem is not that teleological views make no contact with normative features. It is rather that they make contact only with a *kind* of normativity (the *normativity* of *mindless* misrepresentation) that is not related to personal level cognitive notions of right and wrong.

Consider Charles Taylor's useful suggestion that a system displays a genuinely semantic, genuinely normative dimension —what he calls 'the semantic dimension'— when the *rightness* of the phenomena *explains* the causal effects they bring about. It is only because T am sorry' means what it means that in

[12] As are other critics of the naturalization projects in semantics, and of teleological theories in particular. See e.g. Brandom (1994); Godfrey-Smith (1994); Peacocke (1990); Pietroski (1992); Toribio (1998)

uttering those words someone can e.g. achieve reconciliation with another person. Outside the semantic dimension, however, the explanatory direction is the opposite; the rightness of a phenomenon *is explained* by the effects it brings about:

> This [my saying „I'm sorry"] was „the right thing to say" because it restored contact. But at the same time, we can say that these words are efficacious in restoring contact because of what they mean. Irreducible rightness enters into the account here, because what the words mean can't be defined by what they bring about. Again, we might imagine that I could also set off a loud explosion in the neighborhood, which would so alarm you that you'd forget about our tiff and welcome my presence. This would then be, from a rather cold-blooded, strategic point of view, the „right move". But the explosion itself „means" nothing. (Taylor 1995b, pp. 104-105)

As suggested in the quote, there is a sense in which, even within the semantic dimension, the explanatory flow goes in both directions, i.e., it is also appropriate to say that the utterance of 'I'm sorry' is right because of what it brings about — in Taylor's example, the restoring of intimacy between two people. However, even though this bidirectionality obtains, it is not criterial for the utterance's *rightness* because, again, „what the words mean cannot be defined by what they bring about". The explanatory direction outside the semantic dimension is 'unidirectional' (see Taylor 1995a, p. 84). *Rightness,* outside the semantic dimension, is explained by *success* in a task or class of tasks. *Wrongness,* outside the semantic dimension, is to fail in a task or class of tasks. This is the only appropriate explanatory *direction.* If I am successful in restoring our intimacy via the explosion, my success is not explained by the explosion's having *the right meaning*. Systems operate outside the semantic dimension when the notion of rightness (or wrongness) appropriate to their behavior is reducible to success in a task. By contrast, systems operate in the semantic dimension when they operate in a sphere of activities in which rightness is irreducible, a sphere of activities in which rightness is „irreducible to success in some extralinguistic task" (Taylor 1995b, p. 103). Only in the case of full-blooded (mis)representation do we encounter notions of rightness and wrongness that sever the link with success and failure in extralinguistic tasks.

By combining Taylor's observations with the tripartite taxonomy suggested earlier, we arrive at a clear diagnosis of the problem with teleological accounts of semantic properties. Such proposals can accommodate only *mindless* misrepresentation, and one indicator of this is that rightness and wrongness remain defined solely in terms of success or failure in some extralinguistic task. To see that this is so, recall the general form of a teleological view[13]: Mental

[13] A detailed rehearsal of teleological views is beyond the scope of this paper but see e.g. Millikan (1984); Neander (1991); Papineau (1990).

representations are understood as 'teleofunctional items', i.e., as items which are produced by biological mechanisms that have been selected during evolutionary history and that are designed to perform some 'proper function'. Some item *A* has a function *F* as its proper function only if either:

> (1) A originated as a „reproduction" ... of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed *F* in the past, and A exists because (causally historically because) of this or these performances. (2) *A* originated as the product of some prior device that, given its circumstances, had performance of *F* as a proper function and that, under those circumstances, normally causes *F* to be performed by means of producing an item like *A*. (Millikan 1989, p. 288)

Representation *consumption* is, for Millikan, more important than representation *production.* The idea is that for something to be a representation at all, it must be a representation for the system itself. Only if certain conditions are met, can we say that a systemic mechanism will perform its proper function in consuming e.g. a belief token. Those conditions are (i) that a certain property P be instantiated when the system's mechanism consumes belief-tokens and (ii) that the instantiation of that property actually explains why the production of belief tokens enables the system to perform its proper function. The content of a belief thus turns not on the causes of the belief so much as on *the advantageous results that the belief brings about.* Tokens of a certain belief type have as their content that e. g. there is a poisonous butterfly in front of the bird if it is required (for evolutionary purposes) that there *be* a poisonous butterfly in front of the bird in order for the token to perform its proper function. Assuming that certain inner or outer tokens corresponding to a belief are produced if and only if a certain property is instantiated (e.g. there is poison around), such belief tokens would be wrong, would be mistaken, if they are produced when that property is not instantiated.

Even from this brief rehearsal, it is clear that the notions of right and wrong present in teleological views are defined in terms of success or failure in some extra-linguistic task. But, if we accept Taylor's characterization, the fact that such representational content is unidirectionally explained by the effects it produces is proof enough that the phenomena thus explained belong to the 'non- semantic dimension'. Teleological views, although introduced and developed as an attempt to account for genuinely cognitive, genuinely semantic (mis)representation, can thus account only for *mindless* (mis)representation. They can indeed 'explain away' the kind of normativity appropriate to that realm, but this is not the kind of normativity immanent in the full-blooded semantic dimension. This last remark brings me (finally) to the title of my paper. Normativity, as Haugeland clearly shows, comes in several flavors: but only one

of them presents a 'hard problem' for the project of naturalization[14]. Thus the normative status attaching to the phenomenon of malfunction is uncontroversially that characteristic of a factual rule. Malfunction is simply accounted for using the explanatory framework found in the natural sciences, i.e., a framework of causal relations and functional roles. This kind of (factual) normativity is taken to be a property that is ushered into being by some specific causal organization of the matter. To account for factual normativity is thus to understand that organization using the same causal / mechanistic explanatory style that we use for other natural phenomena. Missing from standard accounts, however, is an acknowledgment of the somewhat different normative status of what I have been calling *mindless* misrepresentations. These are not simple cases of malfunction, and indeed may often[15] involve no physical malfunction at all. They do, however, involve the deliverance of a genuinely *mistaken* (non-veridical) representation of how things are. Such mistakes count as (merely) *mindless* misrepresentation just in case they cause errors for which we should not hold the agent or creature *accountable.* They are, roughly, cases of *nonculpable misrepresentation.*

Despite their differences, however, simple malfunction and *mindless* misrepresentation share certain explanatory features. In both cases, as I have argued, the agent fails in her attempt to follow *& factual rule,* a failure marked, ultimately, by some lack of practical, extra-linguistic success. In such cases, it is not hard (as we have seen) for standard modes of naturalistic explanation to get a grip on the phenomena. The virtue of teleological projects is thus that they solve what I would like to call the easy problems of normativity, namely, the problems of transforming what is essentially a causal and mechanistic explanatory framework so as to allow for *historical* causality, to allow, in other words, for final causes in a way that then accommodates a kind of normativity within the natural dimension. However, these teleological views do not —and cannot— account for the fullblooded, genuinely semantic notion of (mis)representation, a notion which we have further unpacked using the idea of cognitive (culpable) error.

[14] In this vein, David Chalmers (op. *cit.)* distinguishes two kinds of question about conscious phenomena. One (the 'easy question') concerns functional capacities: how can a physical device access such-and-such a stored memory, achieve such-and-such a feat of perceptual recognition, and so on. We know, broadly speaking, what an explanation of such capacities might look like. It would look pretty much like the kinds of causal, mechanistic, explanations that belong to the natural order. Not so, Chalmers suggests, when we confront the „hard problem" of explaining not (mere) behavior but the feelings and experiences that sometimes accompany it. This problem, Chalmers suggests, cannot, in principle, be resolved by telling familiar kinds of causal or functional or, in general, reductionist stories. The normativity of genuinely semantic phenomena —a: phenomenal consciousness— looks equally irreducible and basic.

[15] Though not necessarily. See footnote 10.

A system that makes a cognitive error breaks a *governing* rule, not a factual one. In general, systems whose behavior is genuinely semantic follow governing rules, and the appropriate explanatory framework for governing rules is one in which the rightness of the rule is genuinely irreducible[16]. The hard problem of normativity is to account for this irreducible, constitutive notion of rightness within some explanatory setting that can provide for the same degree of objectivity as we find in the natural sciences. This is the kind of problem which teleological theories seem unable to solve. The present analysis suggests, moreover, that the sense in which *rightness*, in such cases, is ultimately irreducible is closely tied to the sense in which there are certain errors for which we can be held personally responsible. It is this dimension of, as if were, *culpable* normbreaking, that seems most clearly to distinguish the hard problem of normativity from the rest. It may be, then, that the prospects for a fully naturalized account of normativity go hand-in-hand with the prospects for a naturalized account of responsibility. Whether this is good news or bad, only the future will tell.

Bibliography

Anscombe, Elisabeth (1957) *Intention.* Oxford: Basil Blackwell.

3 random, Robert (1994) *Making it Explicit. Reasoning, Representing, and Discursive Commitment.* Cambridge, Mass.: Harvard University Press.

Chalmers, David (1996) *The Conscious Mind.* Oxford: Oxford University Press.

Dretske, Fred (1986) „Misrepresentation", in R. J. Bogdan, (ed.), Belief: Form, Content and Function. Oxford: Oxford University Press. Reprinted in S. P. Stich and T. A. Warfield (eds.), *Mental Representation. A Reader,* Oxford: Blackwell, 1994, pp. 157-173.

— 1988) *Explaining Behavior. Reasons in a World of Causes.* Cambridge, Mass.: MIT Press.

Fodor. Jerry (1990) *A Theory of Content and Other Essays.* Cambridge, Mass.: MIT Press.

" The situation in philosophy of language and philosophy of mind regarding such genuinely semantic properties is not all that different to the situation in other branches of philosophy such as ethics, political philosophy or the philosophy of law. My brief excursion into legal territory suggests a case in point. Although it is not a trivial task, it seems reasonably feasible to establish appropriate criteria for exculpating conditions such as those involved in cases of insanity and excusable ignorance. What is much more difficult, and what constitutes the 'hard problem' for the construction of a morally sensitive legal system is to find objective criteria which could characterize notions such as right, wrong, good or justice. Those ideas are the basis of the system in the same irreducible way as the notion of right (true) and wrong (false) within semantic discourse: and they are the hardest to account for.

Frege, Gottlob (1918) „Der Gedanke", *Beitrdge zur Philosophic des deutschen Idealismus;* translated by A. M. and Marcelle Quinton as „The Thought: a Logical Inquiry", Mind, LXV (1956), pp. 289-311.

Godfrey-Smith, Peter (1994) „A Continuum of Semantic Optimism", in S. Stich & T. Warfield (eds.) *Mental Representation. A Reader.* Oxford: Basil Blackwell, pp. 259-277.

Haugeland, John (1998) „Truth and Rule-Following", in J. Haugeland, *Having Thought. Essays in the Metaphysics of Mind.* Cambridge, Mass.: Harvard University Press, pp. 305-361.

Katz, Leo (1987), *Bad Acts and Guilty Minds.* Chicago: University of Chicago Press.

McDowell, John (1998) „Having the World in View: Sellars, Kant, and Intentionality", *Journal of Philosophy,* XCV, 6 (June 1998), pp. 431-491.

Millikan, Ruth (1984) *Language. Thought and Other Biological Categories. New Foundations for Realism.* Cambridge: MIT Press.

— (1989) „In Defense of Proper Functions", *Philosophy of Science LVI,* pp. 288. Neander,

Karen (1991) „The Teleological Notion of 'Function'", *Australasian Journal of Philosophy LXIX,* 4, pp. 454-468.

Papineau, David (1990) „Truth and Teleology", in D. Knowles (ed.) *Explanation and its Limits.* Cambridge: Cambridge University Press, pp. 21-43.

Peacocke, Christopher (1990) „Content and Norms in a Natural World", in E. Villanueva (ed.) *Information, Semantics, and Epistemology.* Oxford: Blackwell, pp. 57-76.

Pietroski, Paul (1992) „Intentionality and Teleological Error", *Pacific Philosophical Quarterly* 73 (1992), pp. 267-282.

Reznek, Lawrie (1997) *Evil or III? Justifying the Insanity Defence.* London and New York: Routledge.

Searle, John (1979) *Expression and Meaning.* Cambridge: Cambridge University Press.

Taylor, Charles (1995a) 'The Importance of Herder', in Ch. Taylor, *Philosophical Arguments.* Cambridge, Mass.: Harvard University Press, pp. 79-99.

Taylor, Charles (1995b) 'Heidegger, Language, and Ecology', in Ch. Taylor, *Philosophical Arguments.* Cambridge, Mass.: Harvard University Press, pp. 100-126.

Toribio, Josefa (1998) „Meaning and Other Non-Biological Categories", *Philosophical Papers* (1998): 27 (2), pp. 129-150.

Walker, Nigel (1962) *Crime and Insanity in England.* Edinburgh: Edinburgh University Press.

Wittgenstein, Ludwig (1962) *Philosophical Investigations.* New York: Macmillan.