MINDFUL BELIEF:
Accountability, Expertise and Cognitive Kinds

Josefa Toribio

Department of Philosophy

Sycamore 026

Indiana University

Bloomington, IN 47405.

USA

e-mail: jtoribio@indiana.edu

MINDFUL BELIEF: Accountability, Expertise and Cognitive Kinds.

ABSTRACT:

It is sometimes said that humans are unlike other animals in at least one crucial respect. We do not simply form beliefs, desires and other mental states, but are capable of caring about our mental states in a distinctive way. We can care about the justification of our beliefs, and about the desirability of our desires. This kind of observation is usually made in discussions of free will and moral responsibility. But it has profound consequences, or so I shall argue, for our conception of the very nature of beliefs and other mental states. Suitably developed, it allows us to draw a line between two distinct ways in which a creature may possess a belief, represent a scene, and fall into error. The first way (which I shall call the 'mindless' way) involves little more than an encoding of information in some way designed to guide appropriate response. This is the common heritage of humans, and many other animals. The second way (which I shall call the 'mindful' way) requires that the creature be capable in addition of a special kind of second-order reflection, and (importantly) be expert at detecting the kinds of situation in which such reflection is called for. The differences between these two ways of 'believing that P' are sufficiently deep and significant to warrant (or so I claim) our treating them as two distinct classes of mental states. For it is only courtesy of the second layer of complexity, I shall argue, that it becomes proper to hold someone accountable for their beliefs or other mental states, and it is this fact of (something like) accountability that in turn raises the most significant challenge for philosophical attempts to give naturalized accounts of meaning, belief, and mentality.

INTRODUCTION

Mental states (perceptions, beliefs, and emotions) all come, I shall argue, in two distinct varieties. One variety —which I shall call 'mindless'— involves little more than an encoding of information in a way poised to guide action or response. The other variety —which I shall dub 'mindful'— is marked by the availability of the mental state for critical reflection and revision. A being who can be brought to see that one of her beliefs, perceptions (or even emotions) is inappropriate, given the available evidence or the prevailing circumstances is displaying mental states of the second (mindful) variety.

The distinction between beliefs that are susceptible to arguments and evidence-based rational revision and those that are not is, of course, highly prominent in the literature on epistemic justification. A belief is properly justified, it is sometimes argued, only if it has been formed in some reason-and-evidence sensitive manner, or (on more complex accounts that bring us more closely still to the kind of thesis I shall be defending) if the agent is able to reflect critically on the belief itself (Kim, 1994). I shall argue, however, that the ability to discharge certain doxastic duties (especially those involving critical reflection) can be equally illuminating with respect to a taxonomy of mental states. More specifically, I show that the openness of a mental state to various kinds of reason-based correction is crucial to the determination of the normative status of the mental state itself and thus to our regarding it as belonging to a specific cognitive kind. The failure to factor this into our conception of the nature of mental states (rather than simply their justification) is especially damaging, I shall finally suggest, in discussions of the prospects for naturalized accounts of belief, meaning, and content. For the challenges posed by the naturalization projects are radically different in the 'mindful' and 'mindless' types of case.

1. EPISTEMIC RESPONSIBILITY

The target of the present discussion is, as remarked above, a broadly responsibilist notion of a distinctive sub-class of beliefs and mental states. It will be useful to begin, however, with a brief review of epistemic responsibilism: the view that being *justified* in believing that *p* crucially involves the fulfillment of certain duties. The major competitor to this view is the 'reliabilist' approach linking epistemic justification to the *de facto* truth conduciveness of the procedures, methods, and mechanisms of belief formation[1]. A responsibilist (or deontological[2]) approach by contrast is one that seeks to understand epistemological terms such as justification, evidence, warrant, etc. in some very strongly normative fashion, i.e., as relating to notions of requirement, duty, blame, obligation, and the like. This normative dimension can be seen in the following passage (Chisholm, 1982, p. 7):

> The simplest way of setting forth the vocabulary of the theory of evidence, or epistemology, is to take as undefined the locution, "__ is more reasonable than __ for S at t" (or alternatively, "__ is epistemically preferable to __ for S at t"). Epistemic reasonability could be understood in terms of the general requirement to try to have the largest possible set of logically independent beliefs that is such that the true beliefs outnumber the false beliefs. The principles of epistemic preferability are the principles one should follow if one is to fulfill this requirement.

Being justified in believing that *p* is thus —for Chisholm— a matter of degree. It is also, and most importantly, a matter of conducting our enquiries

---

[1] There are, of course, as many reliabilist and responsibilist views of epistemic justification as advocates of either. The reliabilist group includes, among others, D. M. Armstrong (1973) and Alvin Goldman (1986, 1987, 1988) For responsibilism see e.g. L. BonJour (1985), R. Chisholm (1977, 1982) and S. Cohen (1984). For a very detailed map of current positions, especially of what it is known in the literature as 'virtue epistemology' see G. Axtell (1997).

[2] I shall use the terms 'responsibilism' / 'responsibilist' and 'deontologism' / 'deontologist' as having, roughly, the same sense unless otherwise noted. For a precise characterization of responsibilism as a subclass of deontologism see Lorraine Code (1987).

in such a way as to ensure that we tend, in general, only to accept those beliefs that are most likely to be true. To reach what Chisholm calls 'positive epistemic status' is thus for an agent to fulfill a certain kind of epistemic responsibility, that of trying to succeed in achieving a certain state of intellectual excellence by bringing about a situation in which one's beliefs are likely to be mostly true. This connection between epistemic justification and epistemic responsibility is sometimes cashed out in terms of (un)blameworthiness. 'A subject (S) is justified in believing that p' is presented as having the same truth-conditions as 'S deserves not to be blamed for believing that p' (see e.g. Bedford Naylor, 1988).

At this point, the major problem afflicting epistemic responsibilism becomes visible. For the question arises how to unpack this notion of being, or failing to be, blame-worthy. And it has seemed to many that blame-worthiness must in turn depend on the extraordinarily problematic notion of voluntary control: that a subject deserves to be blamed for believing that p only if she has voluntary control over the entertainment of her belief. Deontological accounts thus seem to require the truth of doxastic voluntarism, i.e., the truth of the idea that believing or not believing that p is within the agent's power[3].

Consider, however, a promising alternative approach. Kim (1994) defends a version of epistemic responsibilism that highlights not the initial formation of a belief (an occurrence which, rather plausibly, is seldom if ever within an agent's direct control), but the agent's later duties to critically reflect on her beliefs. In this latter kind of case, it seems plausible to suppose that the act of critical reflection is one that an agent can properly speaking will: we can decide that such-and-such a belief requires critical reflection, and take appropriate action.

---

3 For a thorough discussion, see Alston (1988).

Granted, there is always room for meta-level worrying. Even if I can be properly said to will my act of critical reflection in a way in which I cannot properly be said to will my first-order beliefs, it is not obviously the case that I can decide *when* to will such acts. At this point, the fan of these versions of epistemic responsibilism might invoke a Frankfurt-style consideration (e.g. Frankfurt, 1971) distinguishing the will the agent has from the will the agent wills herself to have. An agent who has the will she wants, and who decides to critically reflect on a certain belief, may well be exhibiting all the voluntary choice that responsibility requires.

It is not my purpose, however, to delve any deeper into the attractions and pitfalls of the 'critical reflection' versions of the deontological approach. Instead, I want to highlight one further, and quite deep problem. It is the problem of knowing which beliefs require critical reflection. For as Tidman (1996) correctly points out, we cannot be called upon to critically examine *all* of our beliefs. Yet, it is surely not the case that the unexamined beliefs are thereby unjustified. The spirit of the proposal, clearly, is that we are duty-bound to critically examine those beliefs in *need* of critical examination. But how are we to know which these are?[4] This is a problem to which we shall later return.

So far, then, I have sketched one promising version of epistemic responsibilism: a version which locates our epistemic duties in a somewhat non-standard place. The justification of belief turns, on this account, not on the fulfillment of epistemic duties relating to the initial formation of beliefs, but on subsequent duties to critically reflect on the beliefs we come, by whatever means, to hold. In the next section, I develop a closely related account not of epistemic justification but of the nature of beliefs (and other mental states) themselves.

---

[4]   For an illuminating recent discussion, see Katzoff (2000).

2.   MINDFUL BELIEVING

In the natural world, there seems to be a reasonably sharp boundary between those systems capable of critically examining their own mental states, and those which, while no doubt possessing mental states of various kinds, are always disposed to accept them at face value. A rat or a snake may learn to be cautious, and (after some negative reinforcement) cease to respond to surface cues in the same way as before. But it seems unlikely (though not, I suppose, impossible) that these animals engage in anything that might properly be seen as critical reflection on the relation between their beliefs and the evidence they have for them.

It is my contention that this marks a difference not merely[5] in the justifiedness of beliefs, but in the nature of the belief-states themselves. Beliefs that are prone to self-administered critical evaluation are conceptually special, and constitute (if you will) a *cognitive kind*. The failure to distinguish this special class of mental states results, I shall suggest, in a variety of unnecessary misunderstandings, especially regarding the prospects of naturalized accounts of beliefs (and other mental states).

The observation, then, is that not all states of believing (to focus on the basic case) involve contents to which the agent bears a certain crucial relationship, viz. one in which the content is itself a potential object of critical reflection. The cognitive or conceptual role of mental states which exhibit this kind of self-critical openness is so importantly different from the cognitive or conceptual role of mental states not thus critically open that it becomes fruitful to regard the mental states themselves as exemplars of different cognitive kinds.

---

[5]   Perhaps not even —I am not committed to epistemic responsibilism, but merely deploy it as a structural analogy.

Haugeland (1998) marks the crucial distinction using his parable of the permuting family. We are to imagine a human family, each member characterized by a set of features such as visual appearance, smell, voice, gait, etc. Then, one day, the permuting begins (Haugeland, 1998, p. 261):

> Sister sounds like Father, moves like Grandma, and smells like Kid Brother. Even the parts could mix up: Mother's head (but Father's hair) on Uncle's torso with Baby's limbs —or just two heads with no limb or torso at all (sounding like a truck and smelling like a watermelon). And moments later they switch again, with new divisions and new participants. What would you say?

I know what *I* would say. I'd say "have I gone mad?" And that, in essence, is Haugeland's point. The human agent, he suggests, would quite likely tend to reject what they seemed to perceive. The perceptual beliefs would be subject to, and fail to meet the standards of, critical scrutiny. But what, he goes on to ask, about the (presumably un-permuted) family dog. Here is Haugeland's guess (Haugeland, 1998, p. 262):

> I think the dog would *bark* ... would be disoriented and distressed, maybe even frightened. But I can't imagine any part of a dog's reaction amounting to a rejection of the scene, a discounting of its reality, on the grounds that it's impossible.

It is Haugeland's claim that the lack of such critical reactions shows that the dog "holds no objects to constitutive standards and therefore understands nothing" (*op. cit.*, p. 262). The dog does not see things *as objects*. Likewise if, in our perception of the world, we fail to satisfy the standards for recognizing things as objects, we will possess what Haugeland calls 'mundane skills", but lack "constitutive skills". Mundane skills are basic kinds of know-how, like running, hitting, catching, and recognizing a strike in baseball. Constitutive skills focus on the acceptability or legality of actions, such as proper batting

order, etc. (*op. cit.*, pp. 323-324). They are, in essence, skills at recognizing the legal and the rational, and are thereby poised to support various kinds of *cognitive policing*, both of the self and of others in the community.

Without such cognitive policing, Haugeland believes, it is strictly improper to speak of a system's understanding, thinking, or (if language-using) meaning anything by its words. Surely, however, this is a little too strong. Better to allow that the cognitively unpoliced system may nevertheless count as a locus of various world-and-self directed mental states, but that these mental states are simply *different* from (many of) our own.

Consider, for example, the case of Capgras Syndrome. Capgras sufferers come to believe that a spouse or loved one has been replaced by a perceptually indistinguishable copy (often a robot)[6]. The sufferer apparently does not (indeed, cannot[7]) adopt (towards this belief) the critical stance

---

[6] One explanation of the syndrome is that the Capgras patient suddenly lacks (due to some neural disturbance) her standard affective response to the other person, and that she rationalizes the mismatch between surface cues and affective response by positing an impostor. See e.g. Ellis & Young, 1990. Stone & Young (1997) argue that the patient is thus exhibiting a rational response to an anomalous perceptual experience. What remains clear, however, is that (for whatever reason) these patients are incapable of the kind of potentially belief-revising critical activity which Haugeland posits in the case of the 'permuting family' (but see note below). Yet, despite this, Stone & Young argue that the belief *formation* process is sufficiently well-constrained and conducted so as to yield a genuine state of believing: the Capgras "patients do really believe the things they say they believe" (Stone and Young, 1997, p. 334). The question of whether the process of belief fixation that results in the delusional belief is in some sense a rational process is thus complex. But the contrast with Haugeland's cameo is sharp nonetheless.

[7] At least in most cases. Ramachandran and Blakeslee (1998) report a case of Capgras Syndrome in which the patient appears, at some point, to take a mildly critical stance toward his belief that his parents are imposters. Thus, when asked why he thinks his father is an imposter, he replied: "That is what is so surprising, doctor. Why should anyone want to pretend to be my father? Maybe my real father employed him to take care of me, paid him some money so that he could pay my bills" (Ramachandran and Blakeslee, 1998, p. 159). But notice he does not question his senses —the critical attitude is "outward looking" and falls far short of the *self*-critical stance Haugeland requires. Even the patient's momentary attempt to rationalize the situation —momentary because the original delusion very quickly returns— does not invalidate my main point regarding Capgras Syndrome patients' beliefs. For two reasons: a) the patient's (brief) belief-revising activity highlights the fact that —pace Haugeland— some sort of understanding is present, but b) the fact that the patient is still unable to reject the whole situation as impossible, i.e., the fact that, in his attempt to rationalize it, he still cannot help but presupposing the existence of two people: his *real* father and the imposter, illustrates that the patient's belief does not meet the (constitutive) standards that would turn it into the kind of semantic state for which he can be held responsible.

favored by Haugeland. Yet, it seems highly implausible to suppose that no thoughts or understanding are genuinely present.

Doubtless there are many moves and counter-moves to pursue at this point. But rather than getting submerged in this local dialectic, I propose instead a division in the space of genuine mental states. Let us say that S's belief that *p* is *mindful* iff the following two conditions are met: i) S is a well functioning cognitive system and ii) S could reasonably be expected to fulfill her critical duties relative to *p*. To fulfill one's critical duties is, in this context, to undertake the various commitments involved in holding or withholding a belief and to be ready to engage in evidence and reason weighing activity concerning one's right to that belief. When these conditions for mindful believing are met, then praise or blame become appropriate with respect to actions predicated upon the belief.

Thus consider first a compelling case of 'cognitive deficiency'. Take, for example, the person suffering from Capgras Syndrome who falsely believes her spouse to be a robot. It seems likely that she cannot be blamed for the *wrongness* of this belief because at least one of the two conditions for mindful believing has not been met: the subject or, more accurately, the subject's cognitive sub-system involved in the production of this particular belief is not functioning properly[8]. In a case like this, we can talk of a *pathological* belief, i.e., a belief which is the outcome of some physical malfunction in the subject's cognitive sub-system which produced the belief in question. A belief for which the subject is therefore not responsible.

Of course, the concept of cognitive deficiency is not clear-cut. Alston (1988) describes the puzzling case of a college student who can't correctly understand Locke's views. The case is supposed to illustrate the ultimate

---

[8]   In so restricting physical malfunction to the relevant sub-cognitive system that brings about the belief in question, we avoid the obvious objection that *all* beliefs sustained by e.g. a Capgras Syndrome patient are pathological. Obviously, the same subject can have other kinds of beliefs, if e.g. these are brought about by a well-functioning sub-cognitive mechanism (see below).

insufficiency of epistemic responsibility for epistemic justification[9]. As Alston depicts the situation, the suggestion is that the student can't help but believe something false. In particular, the student believes:

(L)   According to Locke, everything is a matter of opinion,

instead of believing

(K)   According to Locke, one's knowledge is restricted to one's ideas.

Alston insists that there is nothing the student can do to improve his understanding of Locke. The student " ... doesn't have what it takes to follow abstract philosophical exposition or reasoning" (*op. cit.*, p. 287). This seems to suggest an interpretation of the case as indeed an example of (mild) cognitive incapacity. But now consider a different reading of the student's cognitive deficiency: one involving what might be better dubbed cognitive *lassitude* or cognitive *thoughtlessness*. Here we are assuming that the student has the potential to identify the mistake in his own belief, but (for whatever reasons) he simply fails to use this potential. Here, we can claim that in some concrete sense —to be pursued in Section 3— the student *should have known better*. It now makes sense to place the student within the space of responsibility, expecting him to be capable of *learning by way of reasoned discourse*. The student's belief, on this scenario, belongs to the category of *mindful* beliefs.

So far I have portrayed cases of pathological beliefs, i.e., cases in which cognitive sub-mechanisms involved in the production of the belief are malfunctioning and cases of mindful beliefs, i.e., situations involving subjects who *can* be held responsible because the two conditions of mindful believing are clearly met. I next introduce a further category, which I'll call *mindless* beliefs. Mindless beliefs are those held by subjects who are well functioning

---

[9]   See Steup, 1988, pp. 78-81 for a critical discussion of this example within an epistemological framework.

cognitive systems but who fail to meet the second condition of mindfulness[10]. Certain perceptual beliefs belong to this category. Think, for example, of forming the perceptual belief that there are three steps in front of you as you walk out of your front door. Your body movements adjust accordingly, but you fall on the fourth step. Some kind of perceptual illusion has taken place. The environment has played a 'trick' on you. This is the kind of situation in which, despite normal functioning, you need accrue no responsibility for your slip. You are not responsible because you could not reasonably have been expected to critically examine the belief that led to the fall, nor (even had you examined it) to recognize that the perceptual belief was probably in error.

It is important to highlight here that the relevant failure in this case — i.e., the failure that makes a belief like this one belong to the mindless and not to the pathological category— lies in the environmental conditions, not in any *internal* malfunction. The distinction is parallel to e.g. the legal distinction between an insanity and an (excusable) ignorance plea[11]. As in the (excusable) ignorance scenario, the way in which the perceptual illusion arose has nothing to do with any *faulty* cognitive sub-mechanism.

We can, of course, go on to learn about the illusion, even becoming expert at spotting the unusual conditions which promote it and taking extra care. To do so —to reach the point at which the perceptual array directly presents itself[12] to you as one requiring extra caution— is to cross the boundary, to convert the one-time mindless belief into a mindful one. *Now*, if

---

10   Here, again, one may want to introduce the restriction stated above regarding the cognitive sub-mechanisms actually involved in bringing about the belief in question. I omit this further specification in order to keep the flow of the text clearer.

11   It is also worth mentioning that this account presupposes that the cognitive system stops short of the surrounding environment. Although that view has been criticized lately (see e.g. Varela, Thompson, and Rosch, 1991; Clark, 1997), a discussion of how these recent views on embodied and embedded cognition might affect our treatment of belief lies beyond the scope of this paper.

12   By 'directly presents itself' I mean that the *type* of perceptual array can be spotted as illusion-fostering without any external prompting, e.g., being told by a trustworthy person that it is illusion-fostering.

you fall again, due to the illusion, we may indeed hold you somewhat more responsible. For now, unlike before, you really should have known better.

But isn't the mindless / mindful characterization then best described as applying to the agent, rather than the belief? After all, it is the agent who is open to correction when she holds a mindful belief[13]. This kind of objection would only be threatening if the general notion of belief discussed here were *just* that of a mental state with a particular content. However, a belief, as it is understood in this paper —and as it is usually understood elsewhere— comprises both the representational content of the mental state itself and the subject's attitude toward that content. The availability for critical reflection that comes with the notion of semantic responsibility aims at capturing the precise nature of that special attitude toward a representational content that constitutes what I have called *mindful belief*. Since the notion of mindful belief involves both the content and the agent's attitude toward it, a difference in attitude can indeed imply a difference in the class of belief at hand. The individuation of the *content* of such a belief may be achieved in a variety of ways —situation semantics, information semantics, causal theory, etc; nothing in this paper favors or undermines any of those accounts. But what determines the relevant normative status attached to a particular belief is the different conceptual role played by such a representational content within the subject's overall cognitive economy.

In the same vein, to talk about the kinds of beliefs that the notion of semantic responsibility generates as pertaining to the attitude component of a belief seems to suggest that such a responsibility is an internal matter. And yet the idea of a subject's accountability being based on the subject's expertise seems to locate this notion in some kind of *external* realm. I believe the assumed dichotomy here is not without problems. But for present purposes, all that matters is that the tension is only apparent. The notion of intellectual

---

[13]   I thank an anonymous referee for pointing out this to me.

obligations involved in the idea of possession of this kind of expertise has to be understood against the background of a paradigmatically non-propositional ability; the ability to realize that a particular situation requires the agent to step back and critically appraise the situation. As such, the mechanisms by which an agent realizes that some critical appraisal is needed are thoroughly sub-personal. It would thus be a mistake to characterize this notion of expertise in externalist terms since the expertise consists in a particular way of the subject's becoming aware; it involves, in other words, a change in the subject's stream of consciousness. This, hopefully, will become clearer below, as I develop the notion of expertise in connection with the idea of critical 'pop-out'.

It is in cases where the normative status of a belief belongs to the mindful category that naturalization projects face the greatest difficulties. As will become clear in the last section of the paper, the difficulty lies, not in giving an account of the representational content of a belief but in properly explaining the relation of accountability between agent and content captured by the notion of semantic responsibility.

This possibility —for a particular belief to fall on different sides of the mindless / mindful boundary depending on whether or not conditions external to the content of the belief itself are met— also illustrates a related and important point. In order to ascertain the kind of cognitive role that a belief plays in a particular situation, and hence to discover under which of our tripartite taxonomy it falls, we don't necessarily have to look at the total state of the system / agent holding the belief. It might be that, for instance, a depressed person or a person with brain damage impairing some aspect of her cognitive apparatus is still capable of meeting all three conditions of semantic responsibility regarding some particular belief or set of beliefs[14]. That would be the case if we were to consider e.g. a Capgras Syndrome

---

14   I thank an anonymous referee for pointing out this to me.

patient's belief about any matter not related to the alleged replacement of her beloved one by an impostor. In a case like this, what counts is a) whether or not the cognitive sub-mechanisms involved in the production of that subset of beliefs are working properly and b) whether the sort of forward-looking tuneability by way of reasons characteristic of mindful beliefs is in place. That kind of tuneability may be present regarding a particular belief even for subjects who are quite incapable —due to either temporary or partial damage to their cognitive systems— of meeting those conditions for other beliefs.

In the same way, we need not expect all a subject's beliefs within a given domain to change from one type (e.g. mindless) to another (e.g. mindful) *en masse*, even if the agent becomes an *expert*[15] in that particular domain. Again, what matters is a relation of accountability that involves the agent's attitude toward the content of a *particular* belief. To return to our original example: the subject who learns to spot the unusual conditions which promote a perceptual illusion regarding the number of steps at her door could still hold a mindless belief —rooted in a similar optical effect— regarding the number of windows in her neighbor's house.

Obviously, a great deal of work is here being done by a notion of what an agent could reasonably have been expected to produce by way of on-the-spot critical reflective response. It could be the case, for instance, that a subject actually exceeds reasonable expectations. That may happen when e.g. a Downs Syndrome person displays a surprisingly critical attitude toward some limited subject, or even a single belief. It would be a mistake, however, to deny the normative status of mindful to this subset of beliefs based on an *overall* consideration of the subject's cognitive capabilities since what matters is whether the conditions that conform the notion of semantic responsibility are met for each of the beliefs considered. If the person suffering from Downs Syndrome can, nevertheless, display a particular kind of expertise regarding

---

15   This notion of *expertise* will be unpacked in more detail in the next section.

a given domain, one which is not impaired by those cognitive deficits that actually make her a Downs Syndrome patient, then it makes sense to claim that the subject could 'reasonably be expected' to fulfill her critical duties relative to beliefs in that domain. This notion, and its attached pitfalls, are further examined in Section 3. Notice finally that the taxonomy we have constructed (pathological, mindless and mindful) is applicable not just to beliefs but to emotions, hopes, fears, and (as far as I can tell) the entire panoply of what might intuitively be classified as mental states. A certain feeling, for example, might be first classified by an agent as a feeling of depression, and then subjected to a kind of critical appraisal (perhaps as part of a treatment method). The sudden upsurge of anger you feel when another driver cuts you off on the highway might likewise be appraised and rejected as inappropriate.

The class of mindful mental states thus comprises all those kinds of mental contents capable of participating in what is sometimes glossed as the 'space of reasons' (Sellars, 1963). Such states are subject to a certain 'constitutive ideal of rationality' (cf. Davidson, 1970, p. 98), and answer to what Brandom (1994, p. 5) calls a "mastery of proprieties of theoretical and practical *inference*". In entertaining this kind of belief we reveal ourselves as participants in a variety of strongly normative practices. By that I mean not only that mindful beliefs have content, are either true or false, express propositions, etc. —the same can be said of the mindless category of beliefs—, but that the bearer of mindful beliefs relate to them in the relationship of accountability captured by the notion of semantic responsibility. It is this extra layer of normative commitment that makes mindful beliefs, as I'll argue in the final section, the very hardest cases for resolutely naturalistic treatment.

The distinction between mindful and mindless beliefs is probably closest to Daniel Dennett's useful (and underappreciated) distinction between

beliefs and opinions (see Dennett, 1978, in which he credits R. de Sousa, 1971). Dennett follows de Sousa in depicting belief as the basic "less intellectual" phenomenon and assent to a proposition as the "fancier" one involved in, for example, both making up and changing one's mind. Assenting to a proposition, on Dennett's account, involves taking an attitude to (roughly speaking) a mentally rehearsed sentence, while being in the more basic state of simply believing such-and-such is determined by the pattern of non-verbal actions to which a creature is disposed. Phenomena such as weakness of will depend, Dennett argues, on the possibility of conflict between the beliefs implied by a pattern of actions and the beliefs explicitly endorsed in sentential formulations —a possibility which is simply not present in the simpler case (cf. Dennett, 1978, p. 307).

Dennett labels the two kinds of cognitive state beliefs and opinions, where opinions are "the verbally infected ... states of language-users" and beliefs are "the deeper states" shared with frogs, dogs, and so on (cf. Dennett, 1987, p. 233). So deeply different are these two kinds of cognitive state that we should, Dennett, suggests "simply view [some of our intuitions] as about a different category of mental state, the state of assent, i.e. opinion, not belief" (Dennett, 1978, p. 304).

Dennett's proposal is thus quite close, both in motivation and execution, to our own. But there are two notable differences. The first, and least important, is that unlike Dennett I want to paint (what Dennett calls) 'opinion' as, indeed, a kind of belief. To do otherwise is, I think, to depart too far from entrenched usage (this may, indeed, help explain why Dennett's own distinction has never fully taken root in the literature). But second, and much more importantly, the mindful / mindless distinction I champion is only contingently related to capacities of language-use and sentential rehearsal. It may well be that the capacity to treat one's own beliefs and mental states as objects of critical appraisal is greatly enhanced by the capacity to speak and

share a public language, and by various cultural and communal practices of verbal rehearsal, exchange and critique. But there are, I want to claim, no conceptual entailments in this vicinity: it is perfectly possible for a non language-user to know when and how to treat some of its own mental states as hostage to error, and to step back and reappraise those states accordingly, It is, of course, notoriously and inevitably hard to convey, in words, exactly how such a process might go. But one can imagine, perhaps, a kind of "architectural intelligence" that thinks 'in images' about possible building designs, and is able, at crucial moments, to critically re-examine aspects of an emerging plan to check whether certain constraints on good design (the arrangement of load-bearing walls and columns, etc.) are actually being followed. The example is fanciful. But the general intuition is not: it is that Dennett is absolutely right to distinguish two deep categories of belief-like mental state, but wrong to tie the difference directly to facts about "linguistic infection" and the mental rehearsal of sentences. The deep divide is, rather, determined by a profound difference in the patterns of normative commitment and critical sensitivity in relation to which the state is poised to participate. Language-use may expand and amplify the scope of these more "norm-hungry" mental states. But it in no way *constitutes* their presence.

3.   KNOWING WHEN TO WORRY

Mindful beliefs (and other mindful mental states) are subject to revision *via* distinctive processes of critical appraisal. And it is this openness to rational (reason and evidence based) appraisal that allows them to play a special kind of cognitive role. They are suited to participate in chains of reasoned thought and argument, and —when expressed in language— in extended and iterated episodes of interpersonal argument and truth-seeking. Human science is,

perhaps, the crystalline version of this special kind of cognitive economy. This fundamental division in the space of mental states —quite possibly the most fundamental division for many philosophical purposes— thus also cross-cuts the much-discussed divides between the conceptual and the non-conceptual (see Toribio, forthcoming) and even the conscious / non-conscious.

Two major worries, however, rapidly loom. The first one concerns the status of those mental states sustained by systems, e.g. animals, which do not seem capable of critically examining their own mental states and which, according to the taxonomy defended here, naturally belong to the mindless kind. It could be argued that, when a trainer teaches an animal to respond to certain cues, the animal engages in chains of inference. Suppose, for instance, that a trainer teaches a dog to navigate a maze —at the end of which some desirable food is found— by teaching the dog to turn left and right at blue and red signs, respectively[16]. Is it not correct to claim that the success of this training consists in the dog having learnt to *infer* which way to turn from identifying colors?

Such a scenario, would only threaten to blur the mindless / mindful distinction given a very weak interpretation of the notion of inference, the kind of interpretation that e.g. is usually applied when discussing the representational products of sub-personal cognitive functioning. Think, for instance, of peripheral sensory neurons like the rods in the retina. The amount of neurotransmitter these cells release varies as a result of changes in light intensity. There is a weak sense in which these neurons may be said to *represent* the property of there being a certain temporal and spatial density of photons at a certain retinal location and a correspondingly weak sense in which the system *infers* certain conclusions concerning the intensity of

---

[16] This example was provided by an anonymous referee whom I thank for pointing out this concern to me.

ambient light from these representational states. Notice, however, that the contents of those representations cannot play the role that we have assigned to the notion of mindful belief. Since the system cannot alter these voltage discharges by way of *reason*, nor is it able to spot the kind of situation in which such corrections would be advisable, the system is not semantically responsible for its representations, and that is, in the end, the mark of mindfulness regarding a particular mental state.

To explain the success of training by saying that the dog has learnt to *infer* which way to turn from identifying colors makes sense only if the notion of inference that we are using is characterized as a case of *mere* forward-looking tuneability. It would, however, be a mistake to claim that those mental representations are mindful based on that simple kind of tuneability. We would only be justified in making that stronger claim if the dog were capable of engaging in a rather more complex inferential process, that of forward-looking tuneability *by way of understanding reasons*.

The second worry concerns the notion of *cognitive culpability* that looks to be implied by the stress on (roughly speaking) our reasonable expectations of cognitive policing. For the distinguishing mark of a mindful mental state, as I have described it, is that it is a state for which the agent can, in a certain sense, be held to account: a state such that if it is in error, we can hold the agent responsible for the error (or, more broadly, for the inappropriateness of the state).

But, as Tidman (1996) nicely pointed out (see Section 1 above) in his analysis of the notion of epistemic justification, we cannot reasonably expect an agent to be continuously critically assessing all her beliefs and mental states. We cannot even expect an agent to be continuously critically assessing all her *mindful* beliefs and mental states, i.e., all those sufficiently open to such appraisal in the first place. So just what *is* being required of the *mindfully believing* agent? I shall argue that what is actually required is the possession of

simply one more (in Haugeland's terms) mundane skill —viz., the mundane (but meta-level) skill of knowing when to worry. Mindful beliefs, mindful language-use, and mindful mental states in general are all constituted, in this account, by the presence of a well-tuned capacity to *know when to worry*: a capacity which, I shall argue (possibly counter to Haugeland's analysis) involves nothing more than the development of one more kind of expertise — a kind of expertise especially well nurtured, however, by the special linguistic and social environments characteristic of the human species.

To begin pinning this down, let's once again consider some cases. A subject is viewing a straight rod through a prism that causes it to appear gently curved. It is an interesting fact that, under such circumstances, the rod, if touched, will also *feel* curved[17]. Now imagine that three agents are confronted with the experimental scenario in which they are shown the rod and proceed to view and touch it, with viewing mediated by the prism. The first agent is a young child. The second is a non-specialist adult. The third is a light and optics expert. Now imagine, finally, that the agents are placed in a forced choice situation in which they must rapidly judge whether the rod is straight or curved. If they judge correctly, a large charitable donation will be made by a rich company. If they judge incorrectly, the donation is forfeit.

Suppose, as seems likely, that the young child (overly influenced by the information from touch) gets it wrong. It seems unreasonable to expect anything else, so the action is based on a perceptual belief that is (in our terms) mindless. Not so, however, with the optics expert. If she makes the wrong call (overcome, perhaps, by the stress and anxiety of the situation), we feel she really should have known better. In particular, she should have known that this was a situation in which to worry: one in which to critically appraise her own perceptually-based inclinations. The non-specialist adult presents an intuitively fuzzy case. It is not immediately clear whether she

---

[17]   For this and other similar examples, see R. Welch and D. Warren (1986).

should have known to worry. In seeking to resolve such a case, we might appeal to general details of her education and communal surround, or introduce a variety of mitigating factors. It is quite possible, indeed, that certain cases are intrinsically fuzzy, and that in those cases there is no clear-cut answer to the question of whether the agent "should have known better". This uncertain middle, given the relative clarity of the two ends of our continuum, is, I believe, to be embraced. For it allows our taxonomy of mental states to be fuzzy at exactly the points where we are intuitively uncertain how best to understand what it is that the agent does or does not know, and hence what exactly her cognitive duties actually are.

The underlying explanation of the uncertain cases is, I think, instructive. What *makes* a mental state mindful, on this account, is in part the relations it bears to bodies of expertise that enable the agent to know when to worry. But such expertise comes in degrees, and so, in consequence, do the notions of mindfulness and mindlessness as applied to mental states.

We can now begin to resolve some of the problem presented earlier. An agent is able to know when to worry when she is, in a certain sense, an expert in a domain. Part of what it takes to be an expert cook, chess-player, or air traffic controller is learning, by trial and error and extended practice, how to spot the kinds of situation in which you need to step back and worry rather than simply trusting your automatic responses. The process of *spotting* these situations must, however (on pain of regress) itself be automatic. The "care-needed" situations must simply "pop-out" to the expert. This crucial cognitive skill, which I shall call "critical pop-out" is —it seems to me— somewhat unjustly neglected in most philosophical and cognitive scientific discussion[18]. Haugeland (1998), despite his insistence on both skillful know-how and the

---

[18] Exceptions include DesAutels (1996) account of learning to perceive moral situations and Clark's (1996) criticism of simple connectionist models of moral cognition. There are signs, in addition, that new cognitive neuroscientific studies —e.g., neuroimaging studies tracking the interactions between automatic and more reflective responses, such as Drevets and Raichle (1998) and Raichle (1998)— may be starting to target these kinds of phenomena.

importance of a certain kind of "vigilance" (watching out for the illegal moves in chess, etc.), never quite connects the two in the way just described, namely, as requiring a further skill, similar in kind to any other expert skill, at detecting situations in which vigilance is called for. Churchland (1996) develops an account of moral cognition which assimilates it to a characteristic kind of 'connectionist' pattern-matching skill, tuned by long real-world experience, but likewise neglects the further (and arguably) morally crucial skill of knowing when not to trust the outputs of the basic pattern-matching process itself (see Clark, 1996).

If "critical pop-out" (the skill of knowing when to critically appraise your own and other's beliefs, assertions, and mental states) is indeed just another kind of expertise, acquired by training and experience (and realized in, let's assume, something like connectionist-style networks in the brain), a question arises as to why human agents, and not, apparently, dogs and cats, and so on, are able to develop and deploy this form of expertise. This is an empirical question which I am not competent to pursue in any detail, but it seems likely that the distinctive training regimes made available by public language and human culture play a large and empirically important role. This may require, in turn, some kind of "bootstrapping" story about the development of human language, since grasp of meaning is itself (on my account) a typical mindful achievement: we become sensitive to the conditions governing the correct use of words, and are poised to critically appraise our own and other's uses accordingly. Nonetheless, a story might well be told (see e.g. Dennett, 1996) in which initially mindless uses of grunts and tokens become progressively mindful as practices of communal usage, correction and training take root. Once such uses are in place, language looks increasingly capable of playing a role envisioned by Jackendoff (1996) —the role of enabling us to freeze our thoughts and ideas in a format which makes them cheaply and widely available as objects for further inspection, analysis

and argument. It is this kind of rich social and linguistic environment, I suspect, that provides the essential training ground in which to nurture skills of self-reflection and (ultimately) critical pop-out. Whether there could, in practice, be non-language involving versions of such a training ground, I do not know. Though I am convinced (as argued earlier) that the role of public language is not constitutive with respect to the phenomenon of mindful believing, but is at most developmentally central and helpful in amplifying the scope of naturalness of critical intelligence.

Notice, finally, that the notion of "critical pop-out" is not to be assimilated to the much simpler notion of (if you like) "novelty pop-out". Novel situations do, to be sure, tend to engage our critical and reflective faculties. But the kind of expertise I am envisioning is also naturally displayed in situations with which we are quite familiar. For example, the experienced air traffic controller who immediately (and just in time) recognizes that her command to drop 300 feet and turn 30° west would cause a head-on collision, or the logician who spots the elementary error in her own proof. Of course, we do not always *succeed* in identifying such errors. But we are culpable when we fail to do so, precisely *because* these are mindfully held beliefs, properly subject to ongoing, pop-out based checking and appraisal. This, on the account I have been developing, is where the buck of responsibility finally comes to rest. For here we possess the requisite knowledge and skills, and we are, generally speaking, capable of deploying them so as to correct the error. Failure to do so (causally explicable though it must always be) is, in the only substantial sense available to us, a kind of moral failure. It is failure to live up to a standard to which we may properly be held to account: a standard which is, in a deep and abiding sense, our own.

4.   NATURALIZING THE MENTAL

I would like to end by commenting on one important implication of the story I have been sketching. It concerns the shape and prospects of naturalization projects in semantics and philosophy of mind. Such projects, it now seems are really targeting two quite distinct classes of phenomena: what we have been calling 'mindless' and 'mindful' mental states. But naturalizing the former is relatively unproblematic, while naturalizing the latter is (if it is possible at all) enormously more complex. By keeping the two classes of mental phenomena distinct, we are able to see what is wrong with certain anti-naturalist arguments while keeping sight of what is most problematic in the large and various realms of the mental.

Thus consider Haugeland's (1998) rejection of teleologically based attempts to naturalize the notion of misrepresentation. Such attempts depict mental representations as 'teleofunctional items', i.e., as items which are produced by biological mechanisms that have been designed or selected during evolutionary history to perform some 'proper function'. Given this characterization of representation in terms of proper function, Haugeland argues that such views cannot account for the crucial distinction between malfunction and misrepresentation. Here is how Haugeland (p. 310) presents his case

> 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.

We can agree with Haugeland on the need to maintain a distinction between getting things wrong as the result of a malfunction in the system and getting things wrong when the system works normally. But we should disagree about the final verdict for teleological accounts.

The problem is that Haugeland's ontology is too austere to accommodate the genuine (but limited) virtues of this kind of naturalization project. For Haugeland, "when there is no malfunction, it's as 'right' as it can be. In other words, there can be no biological basis for understanding a system as functioning properly, but nevertheless misinforming" (*ibid.*). Haugeland thus rejects the idea that the bird in any sense misrepresented the orange butterfly as edible. But this is, I want to suggest, needlessly extreme. For the example is now nicely captured as one of *mindless* misrepresentation. The environment has played a trick on the bird somewhat analogous to the one involved in the perceptual illusion (Section 2) regarding the number of steps at your front door. Remember that in that situation you were —by hypothesis— functioning normally and you certainly *misrepresented* the world in a specific way, viz. as containing three steps instead of four. Yet you were not *responsible* for the belief that led to your fall. Haugeland's example, likewise, displays (I want to say) a case of misrepresentation without malfunction, and without culpable error: a case of what might thus be called *mindless* misrepresentation. And here lies the problem with Haugeland's analysis: it does not leave space for this important category. Yet without this distinction we tend to be blind to the real (but limited) value of many naturalization projects in semantics.

The problem is not that a teleological view of content cannot account for the distinction between malfunction and misrepresentation. Teleological accounts can and do support that distinction. Where they fall short is in relation to a notion, not just of misrepresentation (misinforming), but of *culpable* misrepresentation —error that the system (agent, creature) could

reasonably have been expected to avoid. In other words, teleological accounts can and do support the distinction between malfunction and *mindless* misrepresentation, but they fail to account for the more *mindful* ways of going wrong (and right).

Since mindful beliefs are the most distinctively human, the spirit of Haugeland's critique remains intact, but the fine print is important because it helps us to appreciate the following general lesson: it's not that naturalized stories cannot account for the normative dimensions of mind and content *tout court*. Rather, they cannot simply redeploy the resources that nicely accommodate the kinds of normativity displayed in the simple 'mindless' cases so as to directly account for the more complex and reason-centered forms of normativity I have been describing.

It might be possible to give a neural and/or computational account of these reason-centered forms of believing. Indeed, nothing I have said so far precludes that possibility. I only intend to show that the structure and nature of standard teleosemantic explanations make them unable to capture the essential mark of mindfulness. A way to show the difficulty with such accounts is to focus on how the notions of success and correctness relate to each other when considering beliefs of the mindless and mindful kind.

To bring this into focus, let's look at the bird example again. 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[19]. The mechanism which detects yellow and triggers an avoidance behavior is properly in place.

The correctness of a belief in the mindful cases, however, is not defined simply in terms of its outcome. An example. Joe, a trained logician, asserts

---

[19]   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).

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 extra-linguistic 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. 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 further 'outcome' or any further behavior of his regarding the size of St. Louis. We 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 beliefs I have called mindful.

To get things right, outside the mindful dimension, all that is required is *success* in a task or class of tasks. To get things wrong, outside this dimension, is to fail in a task or class of tasks. By contrast, agents believe mindfully in cases in which rightness is "irreducible to success in some extralinguistic task" (Taylor, 1995b, p. 103).

We thus gain a clearer picture of the problem with teleological accounts of semantic properties. Such accounts 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 extra-linguistic task. To see that this is so, recall that representation *consumption* is, for e.g. Millikan (1984), more important than representation *production*. The content of a belief 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 were produced when that property is not instantiated.

Since the notions of right and wrong present in teleological views are defined in terms of success or failure in some extra-linguistic task, they can 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 present in the cases of mindful believing.

It is a virtue of the proposed taxonomy, I believe, that it brings such issues to the fore, and that it clearly distinguishes the crucial-but-problematic sub-class of mindful mental states. For it is this sub-class of mental phenomena which must occupy center-stage in philosophical and scientific attempts to understand the distinctive features of the human mind. This is not, of course, to suggest that we should ignore the large class of mindless mental states which are the common heritage of the human, the rat, and the chimpanzee. Indeed, a variety of important conceptual and empirical questions need to be raised concerning the relations and interactions between the two classes of mental phenomena, but this is a task I must leave for another day.

To sum up, I have tried to draw attention to a major divide in the class of mental states. This divide (between what I have dubbed 'mindful' and 'mindless' mental states) cross-cuts familiar classifications such as conscious and non-conscious, or linguistic and non-linguistic mental states, and even (though I have not argued this here) between conceptual and non-conceptual mental states (see Toribio, forthcoming). But it highlights, I claim, precisely

that which is most naturalistically problematic and most distinctively human about our daily mental lives: our ability not simply to know, but to know responsibly; to know in a way such that we may be held accountable for the truth or appropriateness of what is known. I have also suggested, albeit sketchily, that the vexing problem of knowing *when* to be critical of our own cognitive performances itself isolates a new and important kind of expertise, which we develop as part and parcel of the process of learning how to think, reason, and respond in various cognitive domains.

As a closing cameo, imagine a perhaps familiar scene. The expert bridge-player, after watching her opponents bid and make a crucial contract, asks herself (and her partner) some tough questions. She may ask herself whether the contract could have been defeated, and see that the answer is yes, had she but led a low spade on the 8th round, the contract would have been set. But she may (and should) go on to ask herself whether, and if so how, she could have known, at the time, to lead the low spade. It is this second round of questioning that reveals, I suggest, the most potent and distinctive dimension of human thought itself. We are not simply engines of efficient adaptive response. We are not even (simply) engines of counterfactual reason. We are engines that reason about reason itself, and that hold our own reasoning to account. We are able to ask (both of ourselves and of others) "should we have known better"? The cognitive and conceptual role of beliefs and other mental states open to such reason-based critical appraisal is sufficiently distinctive, I have claimed, to warrant treating such mental states as a special class of cognitive elements. These elements bear the mark of the mindful. They resist direct attempts at naturalization, and introduce important layers of complexity into our accounts of misrepresentation, belief, and understanding. We conflate them with their mindless neighbors at our philosophical and scientific peril.

BIBLIOGRAPHY

Alston, W. P. 1988: The Deontological Conception of Epistemic Justification. In J. E. Tomberlin (ed.), *Philosophical Perspectives*, Volume 2, pp. 257-299. Atascadero, CA: Ridgeview Press.

Armstrong, D. M. 1973: *Belief, Truth and Knowledge*, London: Cambridge University Press.

Axtell, Guy 1997: Recent Work on Virtue Epistemology. *American Philosophical Quarterly*, 34 (1), 1-26.

Bedford Naylor, M. 1988: Epistemic Justification. *American Philosophical Quarterly*, 25 (1), 49-58.

BonJour, L. 1985: *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Brandom, R. 1994: *Making It Explicit*. Cambridge, MA: Harvard University Press. .

Chisholm, R. 1977: *Theory of Knowledge*, 2nd. ed. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Chisholm, R. 1982: *The Foundations of Knowing*. Minneapolis: University of Minnesota Press.

Churchland, P. 1996: The Neural Representation of the Social World. In L. May, M. Friedman, and A. Clark (eds.), *Minds and Morals. Essays on Cognitive Science and Ethics*, pp. 91-108. Cambridge, MA: MIT Press.

Clark, A. 1996: Connectionism, Moral Cognition, and Collaborative Problem Solving. In L. May, M. Friedman, and A. Clark (eds.), *Minds and Morals. Essays on Cognitive Science and Ethics*, pp. 109-127. Cambridge, MA: MIT Press.

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

Code, L. 1987: *Epistemic Responsibility*. Hanover and London: University Press of New England.

Cohen, S. 1984: Justification and Truth. *Philosophical Studies*, 46, 279-295.

Davidson, D. 1970: Mental Events. In L. Foster and J. W. Swanson (eds.), *Experience and Theory*, pp. 79-101. London: Duckworth.

Dennett, D. 1978: How To Change Your Mind. In D. Dennett, *Brainstorms*, pp. 300-309. Hove, Sussex: Harvester Press.

Dennett, D. 1987: Reflections: The Language of Thought Reconsidered. In D. Dennett, *The Intentional Stance*, pp. 227-235. Cambridge, MA: MIT Press.

Dennett, D. 1996: *Kinds of Minds. Toward an Understanding of Consciousness.* New York, NY: BasicBooks.

DesAutels, P. 1996: Gestalf Shifts in Moral Perception. In L. May, M. Friedman, and A. Clark (eds.), *Minds and Morals. Essays on Cognitive Science and Ethics*, pp. 129-143. Cambridge, MA: MIT Press.

de Sousa, R. 1971: How To give a Piece of Your Mind: or, the Logic of Belief and Assent. *Review of Metaphysics*, 25, 52-79

Drevets, W. C. and Raichle, M. E. 1998: Reciprocal Suppression of Regional Cerebral Blood Flow during Emotional versus Higher Cognitive Processes: Implications for Interactions between Emotion and Cognition. *Cognition and Emotion*, 12 (3), 353-385.

Ellis, H. D. and Young, A. W. 1990: Accounting for Delusional Misidentifications. *British Journal of Psychiatry*, 157, 239-248.

Frankfurt, H. 1971: Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, 68, 5-20.

Goldman, A. 1986: *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. 1987: What is Justified Belief? In H. Kornblith (ed.). *Naturalizing Epistemology*. Cambridge, MA: MIT Press.

Goldman, A. 1988: Strong and Weak Justification. In J. E. Tomberlin (ed.). *Philosophical Perspectives*, 2: Epistemology, pp.51-69. Atascadero, CA: Ridgeview Press.

Haugeland, J. 1998: Truth and Rule-Following. In J. Haugeland, *Having Thought. Essays in the Metaphysics of Mind*, pp. 305-361. Cambridge, MA: Harvard University Press.

Jackendoff, 1996: *The Architecture of the Language Facility*. Cambridge, MA: MIT Press.

Katzoff, Ch. 2000: Counter-evidence and the duty to critically reflect. *Analysis*, 60.1, 89-96.

Kim, J. 1994: The Deontological Conception of Epistemic Justification and Doxastic Voluntarism. *Analysis*, 54, 282-284.

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

Raichle, M. E. 1998: The Neural Correlates of Consciousnes: An Analysis of Cognitive Skill Learning. *Phil. Trans. R. Soc. Lond. B*, 353, 1889-1901

Sellars, W. 1963: *Science, Perception and Reality*. London: Routledge and Kegan Paul.

Steup, M. 1988: The Deontic Conception of Epistemic Justification. *Philosophical Studies,* 53, 65-84.

Stone, T. and Young A. W. 1997: Delusions and Brain Injury: The Philosophy and Psychology of Belief. *Mind and Language*, 12 3/4, 327-364.

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

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

Tidman, P. 1996: Critical Reflection: an Alleged Epistemic Duty. *Analysis*, 56, 268-276.

Toribio, J. forthcoming: Perceptual Experience and Its Contents. *The Journal of Mind and Behavior.*

Varela, F., Thompson, E., and Rosch, E. 1991: *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.

Welch, R. B. & Warren, D. H. 1986: Intersensory Reactions. In K. Boff, L. Kaufman, J. Thomas (eds.) *Handbook of Perception and Human Performance*, vol. 1-25. New York: Wiley.