

## **Epistemic Feelings**

Ronald de Sousa

University of Toronto

*[Penultimate draft of an chapter forthcoming in in Epistemology and Emotions, ed. Georg Brun, Ulvi Doguoglu and Dominique Kuenzle. Ashgate Pub.]*

### **ABSTRACT**

Somewhere along the course of evolution, and at some time in any one of us on the way from zygote to adult, some forms of detection became beliefs, and some tropisms turned into deliberate desires. Two transitions are involved: from functional responses to intentional ones, and from non-conscious processes to conscious ones that presuppose language and are powered by neocortical resources. Unconscious and functional mental processes remain, and constitute an “intuitive” system that collaborates uneasily with the conscious intentionality of the “analytic” system. Emotions bridge these divides: in particular, specific feelings affect inference, cognition and metacognition. In what follows, after a brief reminder of the crucial role of emotions to rational thought and action in general, I first look at how fear affects belief. I then narrow my focus to some intrinsically epistemic feelings. These include specialized variants of fear and greed; and feelings of doubt, certainty, knowing and familiarity. I shall also describe some surprising recent finding about the influence of oxytocin on trust, and about the direct influence of social conformity on perception and belief.

## Epistemic Feelings

Ronald de Sousa

Somewhere along the course of evolution, and at some time in any one of us on the way from zygote to adult, some forms of detection became beliefs, and some tropisms turned into deliberate desires. Two transitions are involved: from functional responses to intentional ones, and from non-conscious processes to conscious ones that presuppose language and are powered by neocortical resources. Unconscious and functional mental processes remain, and constitute an “intuitive” system that collaborates uneasily with the conscious intentionality of the “analytic” system. Emotions bridge these divides: in particular, specific feelings affect inference, cognition and metacognition. In what follows, after a brief reminder of the crucial role of emotions to rational thought and action in general, I first look at how fear affects belief. I then narrow my focus to some examples of what I shall refer to as *epistemic feelings*. These include specialized variants of fear and greed; and feelings of doubt, certainty, knowing and familiarity. I shall also describe some surprising recent finding about the influence of oxytocin on trust, and about the direct influence of social conformity on perception and belief.

### 1. Emotions and Rational Deliberation

It is no longer controversial that emotions play a crucial role in the process of rational deliberation. Many of the ways that emotions condition our reasoning both about what to think and about what to do follow, as I have argued elsewhere, from the hypothesis that “emotions are species of determinate patterns of salience among objects of attention, lines of inquiry, and inferential strategies” (de Sousa 1987, 196). Catherine Elgin’s paper in the present volume presents a number of vivid illustrations of that thesis in its more positive aspects. (Elgin

forthcoming 2008). Less desirable consequences also follow, for the reasons lucidly elaborated, also in this volume, by Peter Goldie (forthcoming 2008). A further consequence of the emotions' grip on our patterns of attention is that they help to tame the combinatorial explosion theoretically confronting the perfectly rational deliberating agent. As far as pure reason is concerned, every decision problem affords a virtually infinite set of plausible aims and an unlimited number of possible means. How is one to encompass that virtually infinite space and select among possible ends and means? This problem is related to the so-called "updating" or "frame problem" of artificial intelligence (Pylyshyn 1987). What has become recognized as specifically philosophical version of the problem is that a purely rational being would have to deliberate about when to stop deliberating, and decide what to ignore in the process of making a decision. That would threaten an intractable regress of deliberation at every turn. Curiously, however, we are not generally plagued by that particular source of indecision in practice. Emotional constraints on salience and attention come close to simulating, for the duration of any emotional episode, the informational encapsulation of our sense organs; the result is that we generally do not spend any time making the second-order decision as to what to include in our decision procedures.<sup>1</sup>

---

<sup>1</sup> This thesis has been criticized by Dylan Evans, who has labeled it the "search hypothesis of emotions", on the grounds that neither the concept of emotions nor the statement of the search hypothesis can be delineated with sufficient precision for the hypothesis to be anything but vacuous. (Evans 2004). In the context of the hypothesis about salience just recalled in the text, however, it seems plain that emotions, by narrowing the focus of attention, drastically reduce the search space for any given problem.

I shall say no more here about the general topic of the influence of emotions on our practical and intellectual agendas, as it derives from the control of salience. But ordinary emotions sometimes affect cognition in different and more direct ways. In the next section, I consider certain peculiarities of the way our beliefs are affected by the emotion of fear. It is the first item in a cabinet of curiosities of which the present paper is intended to offer a brief guided tour. All exhibits featured in this cabinet involve “epistemic feelings,” of which central examples are feelings of knowing, of doubt, of certainty, and of familiarity. Some more familiar emotions such as fear, greed, or trust, or varieties thereof, function like epistemic feelings in affecting conviction, inference, or cognitive strategies more or less directly. Where they function in this way, these emotions too will count as epistemic feelings. Feelings, in the sense intended here, differ from full-fledged emotions in two ways. Unlike emotions, they can be attributed at a subpersonal level, whereas emotions are typically attributed only to persons. A closely related point is that full-fledged emotions are more complex than the sort of feelings I shall be concerned with. Nevertheless, the latter share four points of resemblance with more robust emotions. First, they resemble standard cases of calm emotions in involving *evaluative appraisals*. Those appraisals bear on restricted domains of epistemic values, specifically on the quality of one’s knowledge, on the extent of what one has learned, and how much confidence can be placed in what one believes. But other emotions also bear on specific domains: fear evaluates risk, indignation assesses social or moral conduct, and so forth. Second, the feelings in question also resemble full-fledged emotions in respect of the “Janus-faced” character of the information they provide. They tell us something both about the subject and about something to which the subject is responding. Third, many characterizations of emotion insist on their link to agency and “action tendencies” (Frijda 2004); epistemic feelings, as we shall see, also play a role in the

guidance of (intellectual) activity. The feeling of knowing, for example, acts to stimulate or discourage us from further racking one's brains in attempted recovery. And fourth, both feelings and emotions have a characteristic phenomenology, but are sometimes legitimately attributed despite the lack of any identifiable conscious awareness. .

## **2. Risk, Fear, and the Philebus Principle**

Besides steam rolling attention and motivating self-deception, fear can play a rather different role. This is what is illustrated in the first item in my cabinet of curiosities. In some cases fear acts not merely to change the subjective probabilities of certain prospects, as seems required by its biological function of signalling danger. It acts in other ways as well, with some perverse results.

It seems plausible to speculate that the function of fear is not only to alert one to danger as a red light might do, but to track the extent of the risk envisaged by means either of its intensity or of its influence on the agent's preference rankings. Fear could then be described as an instinctive *measure of risk*. As we all know but seldom remember, however, it is not a realistic one. Terrorist attacks in the US since January 2001 have taken the lives of some 3500 persons, mostly on September 11 2001. In the same six-year period, road accidents have killed about seventy-five times as many, as have guns; and iatrogenic diseases or medical errors have killed more than the last two together.<sup>2</sup> Conservatively, then, Americans are about three hundred times more likely to die of gun wounds, traffic accident, or medical error as from a terrorist attack. In theory, many people know this; yet no politician could get elected by advocating billions of

dollars of deficit spending for a “war” against those far greater dangers. Terrorists ably bank on those discrepancies, achieving huge effects at remarkably little cost. This seems difficult to explain on the basis of the usual economic assumptions about the rationality of normal agents.

One clue to understanding it is that while fear may be a response to perceived risk, it has its own intrinsic experiential quality, including, for most people not currently sitting in a movie theatre, a strong negative valence. So to the classical Bayesian calculation of the expected desirability of a given prospect, which takes account of the probabilities and desirabilities of negative and positive outcomes, we should add the intrinsic desirability of the emotion that results from the contemplation of the prospective event in question. Since that emotion is itself a response to the expected desirability of that same event, the calculation has a self-referential character somewhat reminiscent of Hofstadter’s Law: *It always takes longer than you expect, even when you take into account Hofstadter’s law* (Hofstadter 1980, 152). Though fear is a measure of the risk contemplated, it is also a disvalue that must be added to the prospect of that risk. And unlike the negative outcome envisaged – say, a terrorist attack – it represents an actual disvalue, certain to be endured.

Consider another familiar example. In strictly Bayesian terms, one might represent the expected desirability of flying as opposed to driving to some destination as involving cost, time, and the undesirability and level of risk of an accident leading to delay or death. On the basis of

-----

<sup>2</sup>A pro-gun physician wishing to place gun ownership on a par with medical succour estimates the number of deaths due to medical error at 100,000 a year (Feckler 2005).

those factors and reasonable estimates of their probabilities, flying is definitely the better option.<sup>3</sup> Adding in the factor of fear, however, may significantly lower the expected desirability of flying in relation to driving, providing fear of flying is sufficiently higher than fear of driving.

This may provide a partial explanation of the fact that decisions taken in relation to certain prospects, such as flying or travelling to places threatened with acts of terrorism, are disproportionate to the dangers involved. But it doesn't follow that such responses are rational. Unless one is in the grip of a postulate holding that all general tendencies to thought, feeling and action bequeathed to us by evolution are infallible, one will recognize these cases as paradigms of emotion-induced irrationality.

We can be more specific about the type of irrationality involved. It consists, I suggest, in a violation of what I have proposed to call the *Philebus Principle* (de Sousa 2000). The name of this principle derives from the controversial suggestion made by Plato in his dialogue of that name that pleasures can be true or false (Plato 1997b). Plato argues that this holds not derivatively, in virtue of being based on false propositions, but literally insofar as pleasures can themselves *represent* other pleasures correctly or incorrectly. In the context of the present argument, the relevant example concerns pleasures of anticipation. These, Plato claims, should be proportionate in their valence and intensity to the pleasures that the anticipated events will yield. Here's how the principle might be formulated:

***PP:** A pleasure taken in anticipation of X should be proportional to the pleasure that will be actually yielded when experiencing X.*

---

<sup>3</sup>About ten times safer per mile driven or flown overall, and five hundred times safer than driving if you fly only in large commercial airliners. See <http://philip.greenspun.com/flying/safety>.

Of course, it is always possible that one might be disappointed by the reality of a seemingly bright prospect. But Plato's principle does seem to capture a legitimate principle of rationality, insofar as pleasures of anticipation act also to motivate our responses. A merely random connection between – putting it now more generally – emotions of anticipation and the anticipated emotions would be seriously disruptive to all rational planning. Still, one might ask, if the Philebus principle is a principle of rationality, what *sort* of rationality is it? The rationale just given for it is practical, yet from the point of view of natural selection, one can see why it would sometimes pay to violate it. It might be simply inefficient always to calibrate levels of fear exactly with apprehended risk. When a risk is very low, it is better not to clutter one's emotional landscape with any response at all. We don't bother to consider the risk of a Tsunami on every ocean beach, though it is presumably not zero. Conversely, some people just don't gamble: gambling and not gambling, if the bet is fair, are by definition equivalent in their expected desirability. It's therefore plausible to attribute the decision to bet or not to bet to an emotional attitude that belongs to a different "channel" than the bare calculation of expected utility. An example might be a common decision procedure that takes account only of thresholds, and ignores gradations between them. Emotional attitudes may determine the thresholds, but only in some regions will these attitudes be felt as distinct degrees of emotion. Thus when a danger rises above a certain threshold, we sometimes decline to "take the risk" at all rather than make any attempt to compute it accurately. Only in the middle zone do we experience fear, but even there, the fear isn't finely calibrated. We might, however, make a judgment on where to locate the threshold: is the risk is worth assessing? Is fear an appropriate emotion in the circumstances? Such judgments will be based not on fear but rather on a "feeling of rightness" of which more in a moment.

An additional anomaly in the way that fear bears on the evaluation of future prospects is brought out by two economists in a study of responses to terrorism. Gary Becker and Yona Rubinstein point out that:

an exogenous shock to the underlying probabilities affects agents' choices via two different channels: (i) the *risk* channel: a change in the underlying probabilities keeping (marginal) utility in each state constant; (ii) the *fear* channel: a change in the underlying probabilities *also* determines agents' optimal choice by affecting the expected utility from consumption in each state. (Becker and Rubinstein 2004, 4).

In their scheme, then, emotion has a distorting effect by virtue of its double life, directly affecting both parameters – desirability and subjective probability – in the Bayesian formula. As a measure of risk, fear should affect only probability: its causal influence on value or desirability seems to constitute a kind of leakage that can't be rationally defended. We shall encounter below (in §5) a converse case in which a feeling which seems relevant only to desirability appears to have a direct causal influence on subjective probability.

### **3. Epistemic Feelings**

#### **3.1 Historical Antecedents**

I now focus more narrowly on some feelings that enter into the epistemic processes of inquiry, knowledge and metacognition. Some of these feelings have attracted attention only quite recently. Nevertheless they can be introduced by recalling three particularly historic moments in the history of philosophy.

The first two moments both occur in Plato's *Meno*, which contain two significant allusions to specifically epistemic emotions. First, Meno experiences *despair* or *dejection* brought on by

the apparent impossibility of learning (Plato 1997a). We can't learn anything new, it seems, because unless we already know the object of our quest we will not recognize it when we see it. Socrates urges him to view his dejection as a good thing. Awareness of our ignorance is the beginning of wisdom, providing it is used to spur inquiry. Second, when Meno's slave boy recognizes the correctness of Socrates's demonstration of the way to double the square, Plato diagnoses the slave boy's response as arising from his recollection. The slave boy presumably experiences a *feeling of familiarity*, characteristic of recollection as opposed to novel encounters or his own clueless previous guesses.

The third historic historical moment is that of Descartes's *cogito*. When Descartes noticed that "I exist" must be true at least whenever I'm thinking about it, he wondered by what trick he had achieved the peculiar *certainty* of that conclusion (Descartes 1986). His answer was that the thought of his existence was clear and distinct, and he inferred that the truth of anything else was as good as guaranteed if it too was clear and distinct. As far as I'm aware, this is the first explicit use of something like a criterial *feeling of rightness* to justify a knowledge claim.

Unfortunately, Descartes was mistaken in thinking that this feeling of rightness, which he calls "clarity and distinctness", is a sufficient criterion of truth. Or perhaps he merely failed to notice that the sufficiency of that feeling as a criterion cannot itself claim clarity and distinctness. Rather, as Hookway remarks, the mysterious "immediacy" of the conviction that I exist is merely a reflection of "the paucity of our epistemic knowledge." (Hookway 2008 forthcoming, ms. 8).

### **3.2. The Two-Track Mind**

In truth, the deficiency of epistemic knowledge is not exceptional among forms of self-knowledge. Cognitive science has explored innumerable ways in which we remain "strangers to

ourselves” (Wilson 2002). We know amazingly little about the motives for which we do what we do or believe what we believe, and much of what we think we know we have confabulated (Hirstein 2005). From an evolutionary point of view, there are perfectly good “design” reasons for this: for the “workspace” of consciousness (Baars 1997) is very narrow, and would cause a distressing bottleneck if the processing of our thoughts and intentions all had to pass through it. This creates something of a dilemma, for if the epistemic machinations of the unconscious go on unmonitored, they may well result, from the point of view of conscious planning, in time wasted in irrelevancies. On the other hand, requiring that every proposition that figures as a premise in an argument itself be grounded in a sound argument is a recipe for vicious regress.

If this no longer strikes us as a paradox, it is because it is now something of a commonplace that mental life involves two relatively distinct levels of mental processing. Though details vary among some two dozen different versions of the view (Stanovich 2004, 34–36), most agree on a number of features of the two systems: the first or “primary” system is characterized by “automaticity, modularity, and heuristic processing”. I’ll refer to it as the “*intuitive*” system, ignoring the multiple opportunities for misreadings afforded by that term. The second, “*analytic*” system, tends to be “rule-based, often language based, computationally expensive” (Stanovich 2004,36). Typically, though not invariably, the processes and sometimes the behavioural manifestations of the intuitive system occur without being monitored in consciousness, whereas (again, typically rather than invariably) the analytic system seems to operate largely in the full light of consciousness.

The place of the emotions in this classification is ambiguous. On the one hand, emotions are traditionally held to be “primitive”, hasty, relatively automatic and often impervious to reason, all of which fit the label “intuitive” rather comfortably. Recent brain science concurs,

insofar as the regions of the brain associated with what are often referred to as basic emotions<sup>4</sup> – the limbic system, the amygdala – are not part of the neo-cortex, where most articulate cognitive activity appears to be generated. Against this, an accumulating number of considerations support the view that emotions lie at the heart of the cognitive itself. Most easily available to common sense is the dependence of many individual emotions on specific beliefs. That suggests that emotions, like beliefs, are typically answerable to the analytic system. Second, emotions are (again, typically though not invariably) conscious, and consciousness is generally associated with the analytic system. Furthermore, some emotions, such as *interest*, or the more elusive “feeling of knowing”, appear to be involved in knowledge and inference.

Both Christopher Hookway (1998) and Asher Koriat (2000) have suggested that there are a number of specifically epistemic feelings that have not until recently attracted sufficient attention from epistemologists. These epistemic feelings bridge the mind’s two tracks, as we find certain deliverances of consciousness popping up ready-made, as it were, to take up positions as unargued yet unquestioned premises in explicit arguments. Hookway rightly comments that “the fact that something serves as a first premise for conscious reasoning is compatible with its resulting from a process of unconscious reasoning (or other processes).” (Hookway, 2008 [this volume?, ms. p.3])

### 3.3 Some Varieties of Epistemic Feelings

---

<sup>4</sup> A minimal consensus on what counts as a “basic” emotion is that it should be like Paul Ekman’s “affect programs”, grounded in specific brain circuits and processes; and cross-culturally identifiable. For discussion, see (Ekman and Davidson 1994), and particularly the contributions in that volume by Ekman, LeDoux, and Panksepp.

Before proceeding further, it will be useful to sketch some varieties of epistemic feelings. They can be characterized according to whether or not they bear on a specific propositional object, and by the phase at which they impinge on the pursuit of knowledge. Rather than an exhaustive classification, this is rather intended as a suggestion about what to attending to when looking at the way an epistemic feeling affects cognition.

(1). Feelings such as *wonder* or *curiosity*, motivates inquiry, but they need not presuppose specific suppositions or existing beliefs. Though I can wonder about whether a certain specific proposition is true, I can also be curious about a range of questions or topics without having formulated any clear Yes-or-No questions.

(2). Feelings such as *doubt* also motivate inquiry but they bear on hypotheses already entertained, propositions that have already made some claim on our assent.

(3). In direct antagonism to feelings of doubt, there are feelings of *certainty* or *rightness*, which can also bear on specific beliefs or on the validity of *inferences*. The feeling of certainty freezes inquiry, to the extent that we may now feel we have the answer I was looking for. But it frees us for action based on the factual or normative propositions to which we now assent. Certainty about one proposition obviously doesn't preclude further inquiry into other questions. Close to the feeling of certainty which focuses on propositions, we shall encounter a a distinct feeling of "trust" targeting *persons* directly and only indirectly bearing on propositions. Strictly speaking, trust is not an epistemic feeling, but because of the importance of testimony in the formation of most of our beliefs, it seems worth including in my purview, and I will shortly illustrate how trust, like many full-fledged emotions, can be curiously susceptible to direct chemical influence.

(4). The *feeling of knowing* seems to bear on specific propositions, but without being able to specify what these are. It is a metacognitive feeling that induces us to believe that we know something before we are able to retrieve what we know. Familiar variants are the “Tip-of-the-tongue” phenomenon or the conviction that one has learned something.

In what follows, I say a little more about some items in each of these four classes, particularly the last two. Along the way, I shall exhibit more items in my cabinet of curiosities. These are offered to illustrate the role of feelings as bridging the intuitive and analytic systems, and also to convey the sense that in some cases we seem to see into the very point where physiological process and subjective feeling coincide.

Aristotle opened his book on Primary Philosophy by remarking that “all men by nature desire to know”. And Descartes (1984) listed the emotion of *wonder* at the top of his list of the passions. Its tamer relative, *interest* (Silvia, 2006), also belongs to the first class. The feelings in this category get us started: they motivate inquiry, operating at a stage when we are in want of a hypothesis or explanation. But these emotions are not the only ones, among standard emotions, that can affect some of our deliberate or non-deliberate cognitive strategies.

### **3.4. Cognitive Foraging**

While my next exhibit introduces epistemic feelings proper, it also extends the range of what one might have thought of under that heading. The reason is that it focuses on what we can think of as specifically epistemic species of two standard emotions, namely *greed* and *fear*, which we don’t immediately associate with inference. It also illustrates the surprisingly direct influence of emotional chemistry on inference.

Angela Yu and Peter Dayan looked at the application to science of the trade-off between exploration and exploitation (Yu and Dayan 2005; Cohen and Aston-Jones 2005). Exploitation of known resources is safe but likely to yield diminishing returns: one can see it as expressing a form of fear of risk or preference for a sure thing. On the other hand, giving up well-trodden paths for the sake of exploration may yield a jackpot of discovery, but it is inherently risky: it can be seen as a variant of greed, a willingness to take risks for the sake of large gain.

That trade-off is well known of students of ant foraging. (Bonabeau, Dorigo and Théraulaz 1999). An ant faced with an established path may either follow it, in the expectation of finding food where many others have already found it, or else strike out in an unexplored direction. The latter option is risky but will pay off, if not for the individual at least for the colony, when the original sources of food are exhausted. This is a good example, then, of a mechanism first instantiated at the most basic level of animal foraging. What is surprising is that it seems to apply equally well to sophisticated scientific cognitive strategies, where it appears still to be controlled by a simple chemical trigger. Yu and Dayan found that the balance between the tendency to explore and the tendency to exploit in the cognitive domain is apparently regulated in part by specific neuromodulators. Scientific research commonly gives rise to emotional states: the fear of risk; the lure of the unknown; the disappointment generated by pleasures that don't pan out. But Yu and Dayan make no reference to the subjects' subjective assessment of the situation. Instead, they found that functional equivalents of these emotions can affect cognitive strategies without necessarily emerging as such into conscious deliberation.

The set up of their experiment was briefly as follows. Subjects were presented with several series of cards containing colour-coded arrows that might or might not be reliable indicators of the location of a given target. Over a number of trials, one of the colours served as a positive but

unreliable indicator. During some phases of the experiment, subjects allowed themselves to be guided by those cues that seemed relevant, taking account of the uncertainties involved. But the experimenters then manipulated uncertainty at the metalevel: from uncertainty about a cue's validity, subjects were suddenly faced with uncertainty about the cue's identity. When faced with an unpredictable change in the identity of the relevant cue, subjects had to adapt to situations of "unexpected uncertainty." Yu and Dayan looked at the activity of neurotransmitters that corresponded to the two different conditions of uncertainty:

the modulators acetylcholine and norepinephrine play a major role in the brain's implementation of these uncertainty computations. Acetylcholine signals *expected* uncertainty, coming from known unreliability of predictive cues within a context. Norepinephrine signals *unexpected* uncertainty, as when unsignaled context switches produce strongly unexpected observations. These uncertainty signals interact to enable optimal inference and learning in noisy and changeable environments. (Yu and Dayan 2005, 681).

Given other evidence that epinephrine and norepinephrine are implicated in "a range of cognitive processes" (ibid.), their experiment encourages the speculation that the same chemical mechanisms underlie, in part, both the phenomenology of emotion in different modes of uncertainty and the different processes of decision-making that function in those modes. The reference to greed and fear is more than a metaphor, though less perhaps than a literal identity, encouraging the surmise that we are dealing with epistemic species of the generic emotions of fear and greed.

### 3.5. Doubt and Certainty

Recall that the second category of epistemic feelings bear on hypotheses or beliefs already formulated, but can also contribute usefully to motivating further inquiry, if only in a negative way. Hookway (2008 forthcoming) lists *doubt* and *anxiety*. These are obviously close to the feelings of dejection and confusion already mentioned in connection with Plato's *Meno*. But doubt and anxiety are natural antagonists to the feeling of the third kind, comprising the *feeling of rightness* and the *feeling of certainty* bearing on specific propositions entertained at the time, as well as on the validity of inference. That, as we saw, was something like the feeling aroused in Descartes by the contemplation of the *Cogito*.

Doubt and certainty are both equally indispensable to rational inquiry. Once seized of any random belief, I would have no motive to undertake any further inquiry if I were incapable of experiencing doubt. Freedom from doubt is lauded as Faith, one of the three Theological Virtues, by those who have an interest in promoting mental docility in their followers; but it is notoriously unfriendly to scientific progress. On the other hand, if I didn't have the feeling of certainty about a valid inference, I wouldn't rely on it; and if I lack the feeling of conviction about a conclusion, I would remain, like Hamlet, unable to take any action based on it. (Hookway 1998). Insofar as any valid argument is intended to persuade, it relies for its impact on the listener's feeling that its conclusion is not more implausible than its premises. If that is not the case, the listener is justified in turning the argument on its head to constitute a reductio of that among its premises that least arouses a feeling of certainty.

It took a couple of millennia before anyone had a convincing refutation of Zeno's argument from the possibility of infinite dichotomy against the reality of space and motion. Nevertheless, though it might not have been clear just which of its premises was false and why, it

was plain enough to most people that the conjunction of its premises was no more certain than the denial of the argument's conclusion. So it wasn't because everyone was irrational that scarcely anyone doubted that space and motion were possible. Unless we spontaneously recognized the validity of some basic pattern of inference, such as Modus Ponens or Modus Tollens,<sup>5</sup> no instruction manual could save our inferences from sinking into logical quicksand. That is the lesson Lewis Carroll's story of Achilles and the Tortoise: if we required every applicable rule of inference to be written down as a premise, the simplest inference would require us to endorse an infinite number of finite steps, as we would need another rule to tell us that the present inference was an instance of the rule last specified, and so on forever. (Carroll 1895). Whatever our level of logical sophistication, there are inferences of which we feel the validity, others which we immediately feel are invalid. At the level of conscious policy, therefore, the dialectic of doubt and certainty is a fruitful one. But that dialectic is manifested below the level of conscious deliberation: epistemic feelings seem to serve precisely the function of providing premises elaborated at the sub-personal or intuitive level for use in explicit inferences.

---

<sup>5</sup>These two patterns are not necessarily on a par as a matter of psychological fact. The point made here, like the point made in the *Meno*, holds a priori: it is that unless some patterns of transition are built into the architecture of the brain, no process of reasoning can get going. What those patterns are is an empirical question: in theory – and sometimes in practice – they might include a disposition to *believe whatever you are told*.

### 3.6. The Chemistry of Trust

A near relative of feelings of rightness and certainty is the *feeling of trust*. It differs from rightness in that its object is neither a specific proposition nor the validity of inference, but a *person*. In a recent article in *Nature* which got almost as widely reported in the press as the research that indicated a genetic cause for women's differential capacity to experience orgasm, researchers at the University of Zürich have shown that "intranasal administration of oxytocin, a neuropeptide that plays a key role in social attachment and affiliation in non-human mammals, causes a substantial increase in trust among humans." (Kosfeld et al. 2005, 673). Their results also support the conclusion that "the effect of oxytocin on trust is not due to a general increase in the readiness to bear risks. On the contrary, oxytocin specifically affects an individual's willingness to accept social risks arising through interpersonal interactions" (Ibid.). The experimental set-up in each run of the experiment involved two subjects, an "investor" and a "trustee". Both received twelve monetary units, and each unit invested by the investor was tripled by the experimenter. Thus if the investor handed over all of his 12 units, the trustee now had 48, comprising his original 12 plus the tripled investment. He could then return any amount to the investor.

Is trust an emotion? Like love, trust is often manifested dispositionally rather than by occurrent qualities of awareness. When it is a full-fledged emotion, however, the subject is expected to be able to report that she feels trusting. In this case, the ascription of trust is based on functional criteria rather than on any first person report of experience, of which there is no mention in the authors' account of the case. They speak of "modulating trusting behaviour"

rather than generating a feeling of trust. It seems safe to say, therefore, that we are dealing with a feeling operating at the sub-personal level.

However that may be, trust is a nice bridge between the strictly epistemic and the strategic; for although it inclines the subjects to believe identifiable propositions – in this case, the proposition that investment in this particular trustee would prove profitable for himself – it does so but only in the context of a transaction envisaged with a person. For the effects of oxytocin on the investor was not matched by any effect on the trustee. That showed that the causal factor responsible for the effect wasn't a general increase in benevolence or good feeling. For a vague benevolence would have presumably led to larger returns from the trustees as well as a larger amount risked by the investor. Another significant control contrasted the original set-up with a situation in which the investor was told that a computer, not a person, would determine the return to the investor. Again, in the alternative situation the oxytocin had no significant effect, despite the fact that the sheer probability of getting or not getting a return wasn't affected. This seems to show that oxytocin didn't simply shift the estimate of risk down a notch; rather, it worked specifically on the feeling component of trust.

The authors note that there is substantial independent evidence "that oxytocin promotes prosocial approach behaviour by inhibiting defensive behaviours" (Kosfeld et al. 2005, 675). In the light of this known effect of oxytocin on social approach in other mammals, they tend to minimize the specific effect on belief: "the investors given oxytocin show more trusting behaviour but do not hold significantly different beliefs about the trustworthiness of others." (Ibid.) That is paradoxical, if we assume that in either case the behaviour of the investor follows a roughly Bayesian strategy. It can be partly though not wholly explained, according to the authors, by appealing to an evaluative rather than a strictly cognitive appraisal: what the

chemical has done is help the investors "overcome their betrayal aversion in social interactions." (Ibid.). Still, the *consequence* of the diminished "betrayal aversion" is equivalent to a change in the probability measure of the expectation of return. So we have here a kind of primitive, purely causal case of direct biological influence over a process that is functionally equivalent to making an inference, even though no explicit inference is made.

In any case, as might be expected, the effect of the oxytocin is not determining. It may contribute to a Svengali effect, but cannot guarantee its success and could hardly be credited with one all by itself. (The median amount entrusted by investors who had absorbed oxytocin was 25% higher than those sprayed with a placebo.) The precise of the chemical influence on feelings, and thereby on cognitive processes, invites further research, both in the sort of case just mentioned and in cases of bona fide valid inference.

### **3.7. The Feeling of Knowing**

Feelings of the fourth class have only recently come under scrutiny. These are metacognitive in their import, bearing not on beliefs already acquired or on hypotheses currently being entertained, but rather on the existence of hidden knowledge: on whether *we already know* something, *without knowing what it is we know*. *Feelings of knowing* include the familiar tip-of-the-tongue feeling, as well as the judgment that one has learned something.

Asher Koriat has elaborated the hypothesis that feelings of knowing "serve to interface between implicit-unconscious-automatic processes on the one hand, and explicit-conscious-controlled processes on the other." (Koriat 2000, 152). Under favourable conditions, such feelings appear to afford a moderately reliable indication that an item of information is

retrievable from memory. This provides guidance on whether or not further attempts at retrieval are likely to be worth while.<sup>6</sup>

Apart from the way in which it emerges into consciousness from unavailable sources, what is most remarkable about the epistemic feelings is the subtlety and complexity of their mechanisms which remain, precisely because they normally escape awareness, exceedingly hard to analyse. There is some evidence that in the course of skilled physical activity humans' animal sense of what to expect follows Bayesian lines more accurately than we could ever consciously compute. In this respect, humans behave no differently than other animals whose inferential resources don't include explicit reasoning in language, Thus when a tennis player returns a ball, anticipation of the ball's position at the moment of impact depends on a Bayesian computation guided on the one hand by prior expectations and on the other by current sensory input. Körding and Wolpert (2004) report that the respective contributions of these two parameters are finely tuned to take account of the amount of uncertainty about the accuracy of the sensory output that may arise from poor light conditions, implementing the Bayesian calculus with remarkable accuracy. Aided by computers, we may come dispassionately to do as well in aiming rockets at Mars; but so long as our conscious deliberations are left to their own unaided and explicit devices, we can only rely on a kind of arbitrage among epistemic feelings, which itself can only

---

<sup>6</sup>Just as other emotions can be "misleading" (Goldie 2008 forthcoming), however, it is relatively easy to trip up the feeling of knowing. Koriat notes that under certain experimental conditions "impressive results documenting systematic discrepancies between metacognitive feelings and actual performance have been obtained not only with regard to [feeling of knowing] judgments [...] but also with regard to judgments of learning [...] and confidence judgments". (Koriat 2000, 161).

proceed on the basis of more of those same feelings. Those who urge us to “trust our intuitions” have this much on their side: that it will take a good deal of hard work to understand exactly how our minds do what appears to take no work at all. The findings of Körding and his colleagues serve to show that while the complete analytic understanding of these processes may be hard to achieve, it is not doomed to remain forever out of reach. For the moment, though, while we may be able to provide mechanical criteria for the assessment of validity in deductive arguments, we are very far from knowing how to do this for inductive arguments.

#### **4. The Normativity of Correct Inference**

The correctness of inference remains a normative matter. How does this fact relate to our propensity to experience feelings of knowing or feelings of certainty? At the most basic level, the equipment at our disposal for assessing the goodness of a cognitive claim or information consists in learned and innate dispositions to epistemic feelings, as well as social responsiveness. We innately possess, and refine through learning, certain dispositions to respond with feelings of rightness to the deliveries of our senses and of the emergence in consciousness of certain propositions. In addition, our feelings of rightness are subject to psycho-social facts about the power of social sanctions over individuals.

In a moment, I shall adduce a striking example of the depth of that power, not merely over individual emotions, but even over the content of perception. But it must be acknowledged at the outset that social pressure is sometimes – indeed, perhaps *often* – deplorably wrong-headed. The key to doing better lies in acknowledging something deeper than social pressure: an original evolutionary basis for some of the “intuitive” judgments that are codified in social consensus. The point is clear in its application to our capacity to apprehend logical validity: while that

ability is far from infallible, we have seen that it cannot dispense with the feeling of rightness. But this conclusion is not confined to the domain of logical inference. Let me illustrate this in terms of one more exhibit in my gallery.

#### **4.1. Obsessive Compulsive Disorder (OCD)**

In OCD, or obsessive compulsive disorder, patients experience a recurring need to check on whether they have closed the door, washed their hands, or accomplished some other – sometimes necessary but often utterly trivial – task. I lay no claim to a diagnosis or explanation of this syndrome; but at least some aspects of it invite a description in terms of a failure of the feeling of certainty or rightness to be triggered by the memory of the required action. I closed the door. I *remember* closing the door. I may even be sure that I closed the door. But the memory fails to trigger the right sort of feeling. There is some sort of disconnection of the normal feeling of certainty from the recent memory of having taken necessary precautions. The patient suffering from OCD does not lack the memory of having done what was needed, but lacks those normal feelings of rightness or certainty that would normally be triggered by that memory.

OCD used to be thought of as a neurotic syndrome calling for psychoanalytic diagnosis and therapy. Remarkably, however, some of these cases are capable of clearing up under the influence of a drug such as Prozac (Kramer 1993). A crucial epistemic feeling can be triggered or at least facilitated by a simple chemical agent, which apparently determines the presence or absence of conviction in a particular proposition. This suggests that this apparent complexity is sometimes an illusion. The feeling of doubt – or the failure of expected feelings of certainty – manifests itself on the cusp of consciousness in such a way as to play a crucial epistemic role. In

its susceptibility to a drug targeting specific receptors for certain neurotransmitters, it also leaves traces of its underlying physiological or chemical nature.

A close neighbour of the feeling of certainty in the space of epistemic feelings is the *feeling of familiarity*. Certainty and familiarity are in part related as knowledge by description is to knowledge by acquaintance, since certainty relates to items of putative knowledge, while familiarity bears on particulars, as objects of perception or as apprehended scenes and situations. Most of us have experienced “*déjà vu*” feelings, the misplaced feeling of familiarity that gives us the erroneous impression of having been here before. In pathological cases, following brain damage in certain fronto-lateral circuits of the brain, these can become insistent and deeply disruptive (Moulin et al. 2005). We can surmise that there might be an explanation for the existence of a mechanism for generating familiarity. There may have been, in the long run of evolution, a correlation between familiarity and beneficence. But from the epistemological point of view the tendency of the former to cause inference to the latter can hardly count as a justification. Yet the feeling of familiarity plays an important part in our lives even in cases where it is not conscious as such. In some cases where it breaks down, we get a pathological condition known as the Capgras Syndrome. This is the next exhibit on my tour.

#### **4.2. The Capgras Syndrome**

Patients afflicted with the Capgras syndrome persist in believing that a person close to them – wife, or father – is an impostor. They never accuse persons of being impostors if they are merely acquainted with them, but not emotionally significant. (Ramachandran and Blakeslee 1998; Pacherie 2005; Mangan 2001).

The best explanation for this strange disorder is that a direct link normally exists between the facial recognition mechanism and the areas controlling the appropriate emotional responses (particularly the amygdala). The sight of a close relative – a parent, in the case of Ramachandran’s patient Arthur – normally triggers an affective response, which is itself subject to a “familiarity” evaluation. In Arthur’s case, the direct link to the area in charge of generating the affective response is missing. As a result, the affective response to his father is not produced. This sets up a incongruity between the strictly cognitive familiarity check that applies to the face and the missing familiarity check applied to the expected affective response. The Capgras delusion is then no more than a perfectly reasonable inference (though of course one that is neither conscious nor explicit): *I get a characteristic thrill when my father appears; I’m not getting that father-thrill now; therefore the person before me is not my father. Yet he looks exactly like my father. Therefore he is an impostor, a stranger who looks just like my father.*

The present hypothesis explains why the “impostor” delusion occurs only with persons to whom the person is emotionally bound. It doesn’t occur with mere acquaintances, because in most cases of recognition a more or less indifferent emotional reaction is normal, not aberrant. That suggests that the emotional aspect of recognition is subject to an independent familiarity marker. Where the person recognized is both familiar and affectively significant, both markers, the heart and the head, as it were, must pass the required ID check.

This interpretation of the Capgras syndrome raises suggestive questions: could an animal be subject to the Capgras syndrome? If we think not, is it because their brains are organized differently, or because they only need the emotive marker – “the heart”? If so, is this evidence that what constitutes the “head” is causally tied to the presence of language? Or could one make

a quasi-analytic, *theoretical* case for the impossibility of Capgras syndrome affecting a being without language?

Leaving the reader to pursue these idle speculations if she pleases, I will confine myself to an observation about the object of the feeling of familiarity. The passage from what I rendered as “I didn’t get the father-thrill” to what I rendered as “he’s an impostor” looks like an inference. But it is not experienced subjectively as an inference, but as intuitive conviction about the supposed father. Second, while the feeling of familiarity acts as a marker, it doesn’t present itself as a marker of correct *inference* as such. In the OCD case, by contrast, it seems to be doing just that. For the absence of the feeling of rightness in OCD signals unreliability in the inference that goes from “I remember locking it” to “I locked it.” What is conveyed here by the feeling is a form of metacognition, though it is not recognized as such; for the presence or absence of the marker normally indicates that I am recognizing someone who has a significant role in my life. Here again, one might raise the question of whether that sort of duality requires a two-track mind; whether animals could have OCD, and what role is played in by the explicit thought represented in the sentence “I locked it”.

## **5. Social Influence: The Price of Non-Conformism**

The last item in my cabinet of curiosities is perhaps the oddest. It stems from a new take on some famous experiments on the power of social conformism done in the 1950’s. Solomon Asch had found that when asked to make a judgment of a visual quantity, some 40% of subjects went along with the false judgment of their peers (Asch 1962). In the new variant, a group of researchers explored “potency of social pressure in inducing conformity and how information that originates from humans, versus inanimate sources, alters either perception or decision

making and the neural basis for such changes” (Berns et al. 2005, 250). The surprising aspect of their findings is that they seem to show that social influence is exerted *directly* on perception: “the effects of social conformity are exerted on the very same brain regions that perform the task.” (Berns et al. 2005, 251) Using fMRI data, they found “highly suggestive” the “lack of concomitant activity changes in more frontal areas,” (Ibid.) where one might have expected activity if the subject’s judgment had resulted from a decision to override one’s own judgment in favour of that of the majority. Berns et al. concluded that no special cognitive activity was detected in the cortex of those who conformed to others’ false opinion. Rather, it was those who saw and stood up for an independent truth who suffered emotional disturbance. In other words, the distorting effect of conformity did not require any calculation of costs and benefits. Resisting conformity to stand by the truth, by contrast, did have an emotional cost. What is intriguing about this discouraging finding is that the emotion involved (though the authors make no attempt to pin-point it in our repertoire of normal emotions) do not seem to be among those listed above – in terms of simple common-sense – as epistemic emotions.

## **6. Concluding remarks**

In the various empirical observations I have cited, the feelings in question were sometimes objects of awareness, but at other times they were inferred from their effects on subsequent belief or behaviour, or from their neurophysiological underpinnings. This may seem like cheating: to be sure, the mere activation of neuromodulating chemicals, as instantiated in the cases I have just sketched, can’t be assimilated to the presence of a feeling, let alone an emotion. Emotions, as I have already observed, are typically ascribed at the *personal* level; the activation of neuromodulators is a *subpersonal* phenomenon. But some of what we are learning about the

involvement of specific parts or functions of the brain in reasoning, illustrated by the exhibits in my gallery, implicates just such subpersonal factors, and it is not clear to what extent we are justified in assuming that if the same factors are involved in emotion, that shows that emotions, as commonly conceived, are involved in reasoning. I have used the word “feeling” precisely in order not to be pinned down on where, in the personal/subpersonal dichotomy, they rightly belong. So long as we remain willing to countenance a range of feelings, akin to yet distinct from full-fledge emotions on standard lists, it seems we can hazard four concluding observations.

*1. On the connection between biology and rationality.*

Social pressure, however powerful, won't suffice to guarantee the normative correctness of our inclinations to draw inference. A sort of core of basic procedures – perhaps Modus Ponens and Modus Tollens – had to be installed by natural selection.<sup>7</sup> We have a native disposition to recognize some inferences as valid. This disposition is separable from our capacity to provide an explicit and conclusive argument for their validity. When codified by rules of language, logic or mathematics, however, inference patterns become less reliable rather than more; the reason is that the fit between the “native” disposition and its implementation in explicit language is not itself part of that system of mechanisms on which natural selection has put its certificate of warranty.

---

<sup>7</sup> This should not be read as taking sides in favour of the existence of an abstract topic-neutral logical capability. That some procedures must be taken to trigger the feeling of certainty is quite compatible with the modularist view that different modules trigger this separately in different domains.

## *2. On the normativity of inference.*

There is a prima facie presumption of functionality to any heritable disposition the complexity of which makes it unlikely to be accidental. But what has been put in place by natural selection, however useful to our ancestors in the environment of evolutionary adaptation, may be not be worthy now of any evaluative endorsement. Justification can't be infallibly grounded. It has to run in circles, the capacity of which to inspire respect depends directly on their size. Ultimately we trust our epistemic feelings to tell us when the circle is big enough for comfort. The degree of comfort we feel in that big circle is the currency in which the normative warrant of inferences is ultimately cashed out.

## *3. Some emotional mechanisms serve to guide cognitive strategies and inferences at the proximate level.*

The surprising lesson of the sort of recent psychological and brain research of which I have exhibited a sampling is that in some cases a relatively abstract inference is triggered by what appears to be a fairly simple chemical agent. It would obviously be greatly exaggerated to conclude from the Zürich experiments that trust was simply triggered by oxytocin, or from those of Yu and Dayan that strategic research decisions were determined by noradrenaline. But these experiments are part of an accumulating body of evidence that suggests that emotional factors, more obviously linked to non-cortical brain and hormonal activities, are important to our judgments of what inferences are or are not acceptable.

## *4. The link to the two-track mind.*

I have suggested that we must take seriously the hypothesis that there are two communicating but largely independent levels of processing: the "two track mind", and serve crucial functions in

managing the traffic between them. Emotions are almost omnipresent in both, but much remains to be done in science and philosophy before the detail of that involvement is clear. Epistemic feelings cast a special light on this, because while they appear to hug the ground of neurophysiology more tightly than emotions in general, they are also manifestly important to different aspects of cognition which we generally assume are most prominently governed by “reason”.

In the final analysis, the increasing precision of our understanding of the brain mechanisms underlying the actual and the normatively correct practices studied by psychology and epistemology may blur the very image I have sought to sketch. The idea that one should be able to distinguish specific contributions of feelings to cognitive strategies, inference, and knowledge, and even that there is always a clear answer to the question of what inferences have been drawn, presupposes that we can draw clear lines between the mind’s two tracks, between emotional and merely evaluative determinants of decision-making, and between the effects that brain chemicals have on reasoning and those they have on feelings. But the fine-grained picture, when it emerges, may overwrite the lines drawn in the sand by the presuppositions implicit in our concepts of inference, of normativity, of feeling, and of emotion.<sup>8</sup>

---

<sup>8</sup> This paper has benefited as a result of many judicious comments by the Editors, for which I am most grateful.

## References

- Asch, Solomon. 1962. *Social Psychology*. Englewood Cliffs, N.J.: Prentice-Hall.
- Baars, B.J. 1997. "In the Theatre of Consciousness: Global Workspace Theory, a Rigorous Scientific Theory of Consciousness." *Journal of Consciousness Studies* 4(4): 292-309.
- Becker, Gary S., and Yona Rubinstein. 2004. "Fear and the Response to Terrorism: An Economic Analysis." "Preliminary and Incomplete" Version Online at <http://www.ilr.cornell.edu/international/events/upload/BeckerrubinsteinPaper.pdf>.
- Berns, Gregory S., Jonathan Chappelow, Caroline F. Zink, Giuseppe Pagnoni, Megan E. Martin-Skurski, and Jim Richards. 2005. "Neurobiological Correlates of Social Conformity and Independence During Mental Rotation." *Biological Psychiatry* 58: 245-253.
- Bonabeau, Eric, Marco Dorigo, and Guy Théraulaz. 1999. "Ant Foraging Behavior, Combinatorial Optimization, and Routing in Communications Networks." In *Swarm Intelligence: From Natural to Artificial Systems*. Oxford: Oxford University Press.
- Carroll, Lewis. 1895. "What the Tortoise Said to Achilles." *Mind* 4: 278-80. Available online at <http://www.ditext.com/carroll/tortoise.html>.
- Cohen, Jonathan D., and Gary Aston-Jones. 2005. "Decision Amid Uncertainty." *Nature* 436 (28 July): 471-72.
- de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, MA: MIT Press, a Bradford Book.
- de Sousa, Ronald. 2003. "Paradoxical emotions." In *Weakness of Will and Practical Irrationality*, ed. S. Stroud and C. Tappolet, 274-97. Oxford; New York: Oxford University Press.

- de Sousa, Ronald. 2000. "Deux Maximes de Rationalité Émotive." In *Emotion und Vernunft; Émotion et Rationalité*. *Sudia Philosophica* vol 59, ed. Emil Angehrn and Bernard Baertschi. Bern; Stuttgart; Wien: Paul Haupt.
- Descartes, René. [1641] 1986. *Meditations on First Philosophy with Selections from the Objections and Replies*. Trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press.
- Descartes, René. [1649] 1984. *The Passions of the Soul*. Vol. 1 of *The Philosophical Writings of Descartes*. Trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press.
- Ekman, Paul, and Richard.J. Davidson, eds. 1994. *The Nature of Emotions: Fundamental Questions*. New York, Oxford: Oxford University Press.
- Elgin, Catherine Z. **Forthcoming** 2008. "Emotion and Understanding." In *Epistemology and Emotions*, Georg Brun, Ulvi Doguoglu, and Dominique Kuenzle, eds. Aldershot: Ashgate.
- Evans, Dylan. 2004. "The Search Hypothesis of Emotion." In *Evolution, Rationality, and Emotion*, ed. Dylan Evans. Oxford: Oxford University Press.
- Feckler, Martin L. 2005. "Firearms in America: The Facts." *Newsmax.Com*, 10 August.
- Frijda, Nico. 2004. "Emotions and Action". In *Feelings and Emotions: the Amsterdam Symposium*, ed. Nico Frijda, Anthony Manstead and Agneta Fischer
- Gigerenzer, G., P. Todd, and ABC Research Group. 1999. *Simple Heuristics that Make us Smart*. New York: Oxford University Press.
- Goldie, Peter. **Forthcoming** 2008. "Misleading Emotions." In *Epistemology and Emotions*, Georg Brun, Ulvi Doguoglu, and Dominique Kuenzle, eds. Aldershot: Ashgate.

- Hirstein, William. 2005. *Brain Fiction: Self-Deception and the Riddle of Confabulation*. Cambridge: MIT Press.
- Hofstadter, Douglas R. 1980. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Random House.
- Hookway, Christopher. 1998. "Doubt: Affective States and the Regulation of Inquiry." *Canadian Journal of Philosophy, Supplementary Volume 24*: 203-26. *Pragmatism*. Vol. 24, ed. Cheryl Misak.
- Hookway, Christopher. **Forthcoming** 2008. "Epistemic Immediacy, Doubt and Anxiety: on the Role of Emotions in Epistemic Evaluation". In *Epistemology and Emotions*, Georg Brun, Ulvi Doguoglu, and Dominique Kuenzle, eds. Aldershot: Ashgate.
- Koriat, Asher. 2000. "The Feeling of Knowing: Some Metatheoretical Implications for Consciousness and Control." *Consciousness and Cognition 9*: 149-71.
- Körding, Konrad, and Daniel M. Wolpert. 2004. "Bayesian Integration in Sensorimotor Learning." *Nature 427*(6971, 1915 January): 244-47.
- Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr. 2005. "Oxytocin Increases Trust in Humans." *Nature 435*(2 June): 673-76.
- Kramer, Peter D. 1993. *Listening to Prozac: A Psychiatrist Explores Antidepressant Drugs and the Remaking of the Self*. London, New York: Penguin Books.
- Mangan, Bruce. 2001. "Sensation's Ghost: The Non-Sensory 'fringe' of Consciousness." *Psyche 7*. <http://psyche.cs.monash.edu.au/v7/psyche-7-18-mangan.html>.
- Moulin, Christopher J.A., Martin A. Conway, Rebecca G. Thompson, Niamh James, and Roy W. Jones. 2005. "Disordered Memory Awareness: Recollective Confabulation in Two Cases of Persistent Déjà Vecu." *Neuropsychologia 43*: 1362-78.

- Pacherie, Elizabeth. 2005. "Perception, Emotions and Delusions: The Case of the Capgras Delusion." Paris: Unpublished.
- Plato. 1997a. "Meno" trans. G.M. Grube. In *Complete Works*, ed. John M. Cooper. Indianapolis: Hackett.
- Plato. 1997b. "Philebus" trans. Dorothea Frede. In *Complete Works*, ed. John M. Cooper. Indianapolis: Hackett.
- Pylyshyn, Zenon, ed. 1987. *The Robot's Dilemma: The Frame Problem and Other Problems of Holism in Artificial Intelligence*. Norwood, NJ: Ablex.
- Silvia, Paul. 2006. *The Psychology of Interest*. New York: Oxford University Press.
- Stanovich, Keith E. 2004. *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. Chicago: University of Chicago Press.
- Wilson, Timothy D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA; London: Harvard University Press, Belnap.
- Yu, Angela P., and Peter Dayan. 2005. "Uncertainty, Neuromodulation, and Attention." *Neuron* 46: 681-92.