RATIONAL ANIMALS:

What the boldest lion won't risk.

[Final draft of a paper now published in

Croatian Journal of Philosophy IV-12 (2004) pp. 365-386]

Ronald de Sousa

University of Toronto

sousa@chass.utoronto.ca

Abstract.

I begin with a rather unpromising dispute that Nozick once had with Ian Hacking in the pages of the London Review of Books, in which both vied with one another in their enthusiasm to repudiate the thesis that some human people or peoples are closer than others to animality. I shall attempt to show that one can build, on the basis of Nozick's discussion of rationality, a defense of the view that the capacity for language places human rationality out of reach of a comparison with animals. The difference rests, paradoxically, on the human capacity for *irrationality*. Irrationality depends on the capacity for language, which allows the detachment of explicit thoughts from their underlying dynamic implementation; these, in turn, condition the essential disputability of principles of rationality. That is what places every human potentially—if not actually—on the other side of an unbridgeable gulf that separates us from other animals.

Many are the problems
  for which we find a cure
and many are the snags
  that arise
rationalization
  is rational for sure
but is it also sure
  that it's wise?
Piet Hein

1. A tiff in the *London Review*

In an uncharacteristically ill-tempered review of *The Nature of Rationality*, Ian Hacking accused Robert Nozick of implying that some people (more precisely some *peoples*) were "closer to animality" than others.[1] Here is the core of Hacking's attack:

> *Nozick writes that "our rationality, both individual and co-ordinate, defines and symbolises the distance we have come from mere animality."* [quoting from *NR* p. 181] *Sounds terrific! Do the publishers realise that it appears to follow... that some peoples are closer to animality than others?* (Hacking 1994a, 18).

In reply Nozick (1994) formulated and repudiated the "Inference to Animality Thesis" (IAT) which he took Hacking to be imputing to him:

---

[1] (Hacking 1994a), reviewing (Nozick 1993) (henceforth *NR*). Nozick replied (Nozick 1994), and Hacking issued a somewhat mollified rejoinder (Hacking 1994b). In what follows I'll also refer to (Nozick 1981) as *PE* and to (Nozick 2001) as *INV*.

> *(IAT) If any trait distinguishes people from animals, and if people differ in degree along this trait, then ... some people are closer to animality than others.*

Nozick had indeed, suggested that rationality is "a graded notion, one that speaks of degrees of rationality" (*NR* p. 98). He protested, however, that this does not entail (IAT):

> *I don't think it follows, and have shown that it doesn't. If Hacking also doesn't think it follows then his attempt to pin on me a conclusion that I don't state via an inference that I don't make and that he himself knows to be invalid is even more reprehensible*....(Nozick, 1994, p. 4).

Though Hacking's last contribution to the exchange was irenic, he still gripped fast the notion that what matters is to keep human culture from the contamination of biology:

> *The differences between us matter—they are differences between a philosopher who feels closer to evolutionary biology and cognitive science, and another who feels closer to cultural anthropology. Because of the time-honoured Western connection between rationality and humanity, these differences have political and social meaning.* (Hacking, 1994 p. 4).

I found this exchange puzzling. Why did they both so emphatically protest? On the face of it, (IAT) isn't obviously invalid, neither is it obviously valid. Abstracted, it seems to go like this:

> *(1) If P distinguishes A's from B's, and if A's differ in degree of P, then some A's are closer to B's than others.*

In a number of instantiations (1) is quite plausible. Consider:

> *(1'). If a confidence in the free market distinguishes conservatives from liberals, and if conservatives differ in degree of confidence in the free market, then some conservatives are closer to liberals than others.*

If, as I do, you find (1') reasonable, then if you find (IAT) repugnant it may be because you already believe that conservatives and liberals are on a continuum, while humans and other

animals are not. Compare another instantiation, obtained by replacing A and B with 'women' and 'men', and 'P' with 'femininity':

> *(1") If femininity distinguishes women from men, and if women differ in degree of femininity, then some women are closer to men than others.*

Here, your reaction to the argument will depend, I surmise, on how much of an essentialist you are about gender. If you think that the "gulf" between men and women is an essential and impassable one, you will be unimpressed. No woman is a man, and no man a woman, and there's an end on it. But if you think gender is itself a matter of degree, then you'll be less likely to balk.

Although both Nozick and Hacking are keen to reject (IAT) they do so from what appear to be very different perspectives. Nozick's reason is that the differences among humans are negligible in comparison with the differences between humans and others:

> *"The important gulf between humans and animals is this: all humans are able to learn and use a human language; no animals are. The large differences in linguistic ability among humans are relatively minor by comparison. All people have passed the significant threshold, and the variations do not put some individuals closer to those on the other side of the threshold.*" (Nozick 1994, p. 4).

Maybe it's like losing the lottery: if your number is out by one, that doesn't meaningfully bring you any closer to winning than if all the numbers are wrong: "a miss is as good as a mile." But if different measures lie on a single scale of "degrees of rationality," it's hard to see why it would be like that. Hacking's reason reflects a more specific prejudice against applying biology to the understanding of human beings. Insisting (tendentiously) that "trait" must mean "a genetically determined characteristic or condition," Hacking wants to say that rationality, like anything else that is in essence cultural, is not a "trait" at all, "although I'm sure," he adds, rather confusingly, "only our species could develop that idea of rationality. It is a social product...." (Hacking 1994b, p. 4). But if a social product is such that only our

species can develop it, why is that very capacity not enough to make it a "genetically determined trait"? What is genetically determined, to be sure, is not the specific form that it takes, but the general capacity to develop one form of it or another. Humans can all learn a language, and only humans can. That is a biological fact; but individual languages do not differ in consequence of any biological differences.

Both Nozick and Hacking agree, in the end, that the crucial difference depends on language. So the reasons for making so much of a disagreement that turns out to be merely apparent may just come down to a distinction, which Nozick is at pains to make elsewhere, between asking *what it is rational to believe*, and asking *whether believing something is the rational thing to do* (*NR* 70, 113). As far as the latter is concerned, someone might rationally "avoid investigating certain subjects in a given society because of what he predicts will be the harmful social consequences of the results." Maybe the (IAT) is of that sort: Hacking doesn't want it even raised. For my part, I note this possible conflict between epistemic ends and broader concerns, but I propose to ignore the latter issues entirely. I shall ask only whether Nozick is right in thinking that (IAT) can be shown to be fallacious. His own "proof" was sketchy. Why is it exactly that the faculty of language makes a difference so huge that other differences become imperceptible? That is the question I want to explore in this paper.

2. Measuring Culture.

Here is one way one might attempt to reject (IAT). The form of argument (1) is valid, it might be claimed, only when we are not actually dealing with a biological fact. A biological difference between two species admits of no degrees. Thus (1") works if and only if we think of gender as non-biological. Yet the only difference there *could* be between us and the animals is a biological one. So what's wrong with (IAT) is that it perversely insinuates that the difference between some animals and some people might be of the same kind as some (non-biological) differences between different people. Or else—equally perverse and even more politically incorrect—it is committed to treating differences between *peoples* as biological.

But this begs the question. When a species splits to form two clades, the only sure biological criterion for the species separation is their mutual reproductive isolation. This may follow from any of a great many different sorts of differences: geographic, ecological, perhaps behavioral. But other characteristics will remain more or less continuously (though less and less uniformly), distributed among the members of the both clades. It is gratuitous to suppose that each species resulting from a split will have an essence, some genetic signature possessed only by its members and absolutely absent from all the members of the other.

Culture is a case in point. We're used to thinking that "culture-bound animal," like "rational animal," defines human nature. (Hence Hacking's declaration about "feeling close to cultural anthropology.") But in fact it is now quite generally recognized that other animals do have cultures (Reader and Laland 2003). That doesn't mean the difference between humans and other animals can't be cultural. For present purposes, however,, it would have to be shown that the cultural differences is so large as to constitute an impassable threshold, so that animals couldn't get close by dint of more culture any more than you can get closer to infinity by adding a million.

But that is to look at it the wrong way around. The objectionable claim (AIT) is not about how close animals are to being human, but how close humans can be to animality. An argument of the form of (1) but *going the other way* doesn't seem objectionable:

> (1''') *If* ***having culture*** *distinguishes* ***animals*** *from* ***humans****, and if animals differ in degree of culture, then* ***some animals are closer to humans than others****.*

Most people would accept the conclusion anyway, which may induce endorsement of the argument. As Nozick remarks, we are apt to pursue an "optional stop rule. I do not stop the philosophical reasoning until it leads me where I want to go; then I stop." *(PE* p.2)[2]

---

[2] Elsewhere Nozick also remarks on the happy coincidence that people's belief about relativism correlates very highly with their preference as to whether it is true: "Now it *could* be that we adjust our desires to the realities of the situation, ... However, it seems more likely—doesn't it?—that in this case our ... belief... follows in the footsteps of our desire. This gives one pause" (*INV*, p. 21).

It could be true that while some animals are closer to humanity than others, no humans are closer to animality than others. (Especially for anyone who manages to avoid the evening news.) I will argue, in effect, that this is just how it is. But it must first be acknowledged that the question of whether and how it is possible to measure *quantity of culture* raises some intriguing problems. To answer it, one must first distinguish between two different meanings of 'culture'. In one sense, the word refers to products of the fine arts, classical music and highbrow literature. In this sense, culture is that in terms of which certain classes define themselves as better than others. And whatever we may think of that, the idea that some people are "more cultured" than others certainly implies that culture admits of degrees. But if one were using the term in that sense, one could reasonably be accused of snobbery—or worse—if one used it to justify the claim that some people are further away from animality than others. Anthropologists, eager to distance themselves from ethnocentric prejudice, use the word 'culture' in a rather different sense, in which it designates the totality of practices, attitudes and beliefs that characterize a group or population in contradistinction to others. This sense of culture allows for huge differences between the *features* of different human cultures, but implies no essential differences among humans in *quantity or quality* of culture. The desired gulf between humans and animals might then be secured by showing that while different species of animals have more or less culture, all human cultures have it in *incomparably* greater degree.

This last claim presupposes the possibility of arriving at some sort of quantitative measure of culture. Adapting the structuralist methods of Lévi-Strauss (1976) to the conceptual apparatus of Shannon information theory, we might try to identify the actual number of binary contrasts embodied in any given culture, so as to arrive at a quantitative estimate of a given culture's actual informational content. (How many bits does it take to list all distinguished degrees of kinship? How many bits to spell out differences in social rank, or in food types?....) Thus formulated, it is an empirical question whether the count thus obtained will turn out to be roughly the same or wildly different in different cultures. Until the count is

in, the claim that all cultures are quantitatively equal remains purely dogmatic.[3] One might, however, make good on the thesis that even quantitative differences between human cultures cannot ever bring any one culture closer to animality, if one could show that human culture is *virtually infinite* in a way that no animal culture can be.

Cultural atoms and their replication

If cultural variation is continuous, there isn't going to be anything to count. Quantifying culture could be done only if the basic cultural atoms are discrete. (I'll call them 'memes', because 'meme' is a handy meme, without implying any special commitments to the theory of memetics.) Generally speaking, the existence of discrete units in any system of representation is also what makes it possible for items in that system to be faithfully reproduced. That is the main advantage of digital technology, an idea best viewed as the industrial application of Plato's theory of Forms. If any two things that resemble one another do so in virtue not of any two-term relation between them, but in virtue of a triadic relation holding between the two of them and a paradigm, a Form of which they are both *instances*, then a copy of a copy of a copy, iterated any number of times, is only ever two steps away from the original. In copying one 'a', for example, I look not to its idiosyncratic shape in this font, but to the abstract paradigm A of which it is an instance.

So to the extent that cultural items are faithfully reproducible an indefinite number of times, they must be discrete. This is true of linguistic items such as phonemes, words, and grammatical forms. Language is an essentially digital medium of representation.[4] Whether a

---

[3] Witness, for example, the assertion that "there are no essential differences in the fundamental nature of thought processes among the various living races of men." (Geertz 1973, p. 62).

[4] This probably needs to be qualified. First, it's not obvious that all the semantically significant characteristics of linguistic items are digitally organized. Take intonation, for example. In certain dialects, the high-rising terminal intonation contour normally characteristic of interrogatives has lately come to be used in assertions, as if the ellipsis of a final request for agreement had left its intonation contour behind. ("And that's a fact, isn't ít?" —> "And that's a fáct.") This carries with it an annoying perlocutionary implication of a constant quest for agreement or approval. To those pedants who, like me, find this irritating, it leads to an unnecessary confounding of illocutionary types. But in the

meme is digital or not depends on the existence of a notation that steadies the norm. Nelson Goodman has shown how this works in his elaboration of the distinction between "allographic arts," such as literature, for which a notation supports a criterion of identity that allows *the same work* to exist in many copies, and "autographic arts," of which painting is the paradigm, of which arbitrarily accurate copies can be made but of which there can be only one original (Goodman 1976). There are intermediate cases: music has a notation, but that notation is only suitable to a certain sort of classical European music, and much of the detail about just how a performance ought to sound is conveyed only by "tradition," which does not resort to digital notation.

It seems, then, that memes needn't be strictly digital. Indeed, that very distinction, between what is and what isn't a matter of degree, appears to be a matter of degree. Insofar as memes aren't digital, one can expect them to undergo fairly rapid degradation or transformation in the process of multiple reproduction. Yet they can't count as cultural features of any kind unless they have a sufficiently distinct identity to be assessed as being correct or incorrect. Among the sources of stability are various mechanisms of homeostasis, which include psychological dispositions that favour certain patterns as "attractors" (van Gelder 1998) and may (though they need not) themselves have been conditioned by natural selection. A good example is our perception of color: "Wavelengths are continuous, yet color is psychologically experienced as organized into bands." (*INV* p. 118). A more complicated example is the wide yet roughly recognizable class of memes that fall under the general heading of religion. Like a virus, constantly mutating but seldom losing virulence, the spread and maintenance of religious memes must be facilitated by the evolved capacities and

absence of notation intonation is more difficult to think of as governed by standard norms than phonemics or syntax. Furthermore, though the elements of language are digital, language evolves, like organisms, in a dynamic way. Each child learns, in effect, a new idiolect, under the influence of the unique set of utterances that it hears. There have been interesting attempts to model this in terms of physical models that eschew differentia (Provost and Jennings 2000).

propensities of the human mind. This class of memes bears witness to the fact that our inferential powers can't be assumed to be exclusively conducive to cognitive rationality. Indeed, perhaps it owes part of its success to its defiance of cognitive rationality.[5] An obvious alternative to rational inference is social pressure, joined to an inclination to imitation, which can produce a broad range of not necessarily beneficial conformity to norms (Boyd and Richerson 1992; Fehr and Fischbacher 2003).

3. Culture and rationality.

So how rational is all this likely to have left us? Among the "natural" propensities most useful to science, it seems we should count our predisposition to invent hypotheses about unseen explanantia to explain perceptible phenomena. These, like our propensity to infer the future from the past, is doubtless indispensable to the capacity for scientific discovery no less than to religious belief. We also seem to have predispositions to think certain connections *necessary*, and to follow efficient rules of inference. Where would we be if, like Lewis Carroll's Tortoise, we felt no particular inclination to follow *modus ponens*? These are presumably among the cognitive tools which more disciplined scientific thought builds on and refines.

Such speculations raises three questions:

(a). Is there indeed good reason to think that we have an innate disposition to care about the truth, and distinct innate dispositions to make certain inferences?

(b). If so, is there any reason to think the sort of inference we tend to make are likely to lead us to true belief?

---

[5] Thus Pascal Boyer (2001) suggests that among the features that make religious concepts particularly catchy is their use of familiar and basic categories such as "living thing", conjoined with paradoxical features such as the idea of resurrection or miraculous powers. See also Daniel Sperber (1985) for an "epidemiological" view of cultural transmission.

(c). Does this make us different from other animals in a crucial way, or is there only a difference of degree in the extent to which these predispositions enable us and other animals, respectively, to acquire objective knowledge? Let's look at these questions in turn.

(a). The Impurity of Cognitive Goals

At the biological level, true belief is just a survival strategy. Evolution has programmed us, Nozick suggests (*NR* p. 68), to care for truth, but not necessarily to worry about why we should care for truth.[6] This is just sound strategy on the part of our genes.[7] If you explain what you are after, your pesky underlings are likely to question your motives. Better just get them to think it's something they intrinsically want, so they won't raise the question why. So from the gene's perspective, our concern for truth as an *intrinsic* goal is *instrumentally* useful. When it isn't, that will be because of a conflict with *other* "intrinsic desires," such as lust, about which that same point can be made: lust is the more instrumentally useful to my genes for its seeming intrinsically valuable to me.

Once that remark is made, however—once we become self-conscious about truth—we can question the authority of our desire for truth, just as we question the authority of our instinctive impulses.[8] Why *should* I pursue truth? While any project needs a modicum of

---

[6] More generally: "Having instilled desires that serve to maximize inclusive fitness does not mean that evolution has instilled the desire to be maximally inclusively fit." (*NR* 30).

[7] I will persist in this unrepentant talk of genes, despite a growing literature arguing that this sort of talk of 'genes' is misguided. See for a sampling (Oyama 2000); (Griffiths 2001); (Moss 2003). As far as I can see, while the debate about the relative role of DNA, developmental systems, or cellular morphology in the mechanism of heredity is of great scientific interest, it has little bearing on the philosophical point I am making here. Talk of genes in this context can be taken to be a metonymy for whatever turn out to be the *enduring beneficiaries* of natural selection. Whatever these are, they can't be biological individuals such as ourselves, because sexually "reproducing" individuals are never, in fact, reproduced in nature.

[8] In a fascinating interpretation of the *Genesis* story about the discovery of shame that followed Adam and Eve's eating of the fruit of the tree of knowledge of good and evil, David Velleman (2001) has suggested that what Adam and Eve really discovered was the *possibility of rejecting their natural impulses*. I'm suggesting that language had just that effect in the cognitive domain. Once you are equipped with language, you can decide to accept or reject the goal of truth as an unquestioned intrinsic value, and you can decide to accept or reject the apparently compelling form of a valid logical inference.

relevant information for its successful completion, many don't really require truth for success. Indeed, many practical goals actually *compete* with the pursuit of truth. The worthy goal of serenity, for example, may be fostered by religious delusion, or for that matter lesser forms of trust in epistemically dubious but emotionally comforting propositions. The point I want to stress here is that we have a *choice* in how we weigh the goal of truth against other goals. But this choice arises only insofar as we are able to reflect on the competing claims of various goals. One can posit, to be sure, an innate animal curiosity, such as the one that proverbially killed the cat, on the reasonable assumption that in the cat's phylogenetic past curiosity more often saved it. But if the cat confronts a choice between satisfying curiosity and some other pressing goal, the conflict will resolve itself dynamically, not by anything we could call *reasoning*. No cat, and no ass either, worries about Buridan's ass.

What of our powers of inference? Here the answer must depend on what we are prepared to count as inference. In any animal to whom we are willing to attribute beliefs, we must posit some mechanism of belief change. At the simplest level, however, it would be odd to attribute an *inference* to a worm that responds to a tropism in orienting towards food. ("Food is good for worms. That smells like food over there. Therefore, there is food over there. So I'll move in that direction."). The extent of more elaborate powers of inference has been subjected to a good deal of investigation in our closer cousins among the primates. (See, for example Tomasello and Call 1997). But what is lacking is the *metarepresentation* that is required for beliefs to be examined, criticized, and made the object of self-conscious inferences.

There is now much impressive research suggesting that the roots of our capacity for such metarepresentation is laid very early in social interaction. By the age of two, "children demonstrate their ability to establish self-other equivalence, to take different perspectives on things, and to reflect on and provide normative judgments of their own cognitive activities." (Tomasello and Rakoczy 2003, p. 123). Children's acquisition of a sense of the normativity of language use enables them to them to play with the breaking of linguistic convention. (p.

129). Particularly important to this process are interactions involving "disagreements and misunderstandings" (p. 135)—in other words, opportunities for *corrections*, in that they are "an extremely rich source of information about how one's own understanding of a linguistically expressed perspective on a situation may differ from that of others." (p.137) Such interactions, by creating the possibility of disagreement formulated in verbal terms open to criticism, extend the range of possible beliefs, inferences, preferences and actions.

That the acquisition of *norms* governing belief should come with the acquisition of language itself is hardly seems surprising, but it is not quite analytic, and so it is encouraging to find that empirical research confirms it. But does this give us any reason to suppose that the ways of reasoning we learn in the process of social interaction, are particularly likely to preserve truth?

(b) Do our inferences preserve or generate truth?

Nozick rightly notes that the best we can hope for from natural selection is that it will promote fitness-enhancing strategies. Whether the success of epistemic goals always coincides with fitness is highly dubious: "It may be wondered whether success in action is even an indication of the belief 's truth.... Navigation in accordance with a geocentric theory leads to successful goal achievement, yet that theory is not true, even approximately. Rather, its *consequences* are close to the truth concerning the particular goals of navigation." (*INV* p. 47). And if indeed our standards of rationality are, as Hacking (1994b p. 4) insists, "a social product" knowing how capriciously related to truth and goodness "social products" are likely to be, there is little reason to think our practices will necessarily be rational.

Several philosophers have advocated a principle of charity: "Interpret or translate what the person says and does so as to make the person as *rational* as possible" (*NR* 153). But there are, as Nozick notes, several problems with this principle. First, given the weight of evidence for the view that we are sometimes irrational, it might simply be wrong to ascribe rationality to everyone. Secondly, different groups impose different standards. Appeals to Confucius, or

to Aquinas, or to Mohammed, are each regarded as a conclusive clinching argument in certain non-overlapping circles; prima facie, their incompatibility counts against them. I might add that where someone appears to assert a blatant contradiction, 'charity' may be a misnomer. If someone asserts an explicit contradiction, it will seem natural for most hearers to look for a figurative interpretation, since otherwise there seems no way to make sense of it.[9] So there's no choice but "charity," which rather takes the virtue out of it.

In fact, there is much evidence that our cognitive strategies are not always optimal (Kahneman, Slovic and Tversky 1982; Nisbett and Ross 1980). Critics such as Gerd Gigerenzer (2001) see in such alleged errors just efficient shortcuts, "fast and frugal" rules of inference calculated to give more cognitive bang for the evolutionary buck in the statistical long run. But a cognitive strategy honed over the statistical long run will yield perverse inferences in specific cases. Superstition, for example, is just an application of induction which commits what statisticians call "errors of type II," which is to take chance fluctuations to be statistically significant indications of causal influence. There is no correction for this which doesn't fall into the risk of committing an error of type I, which is to regard the effect of what is in fact a genuine causal factor as a mere outcome of chance.

But why assume that some long-term strategy explains each particular failure? Without an axe to grind on behalf of some adaptationist presupposition about human rationality, the aim of defending each case of systematic irrationality as merely apparent or accidental seems a quixotic one. A more plausible hypothesis is that rationality, like any construction of natural selection or of social collaboration, is likely to have been cobbled together by dint of numerous tinkerings and compromises. Worthy goals, even purely epistemic goals, frequently

---

[9] Some disagree even here. For one defense of the view that some contradictions are true, see (Priest 1998), who makes the important point that whether or not one chooses to countenance the view that some contradictions are true depends on a *theory.* Any theory is by nature debatable, and would remain so even if by some luck accident everyone were to agree on its truth. (I'm only slightly less confident that a theory remains inherently debatable even when everyone agrees on its falsity. But human eccentricity can doubtless be counted on to keep both questions purely counterfactual.)

conflict.The classic ends of *believing truths* and of *avoiding falsehoods* are only one such pair: each one would be easy to satisfy *by itself*, but that wouldn't count as success. The same goes for the goals of avoiding errors of type I and of avoiding errors of type II.[10]

c) Making out the human difference.

Can we still make out, on the basis of the very imperfect and variable endowment of rationality I have just sketched, a difference so categorical between animal and human rationality as to place the two literally beyond measure? Here, in somewhat provocative terms, is the answer I mean to elaborate in the rest of this essay: The crucial threshold that animals do not cross consists in *our capacity to be irrational*. And in turn, the capacity to be irrational rests on our capacity *to speak*.

To start with, I want to frame this suggestion in a broader biological context. Maynard Smith and Szathmáry (1999) have sketched eight crucial transitions in the history of life. Two of those transitions involved a saltation in the an order of magnitude of open possibilities, resulting from the introduction, at each of these two stages, of a digital system of representation. The first was the invention of DNA, without which life itself as we know it could never have acquired the astonishing mix of variety and stability that an ultimately particulate system of heredity made possible. The second was, of course, the invention of language. Thanks to the power of the combinatorial explosion, the number of possible explicit thoughts (like the number of possible proteins made possible by the genetic code) is *superastronomical*.[11]

---

[10] Note that "believing truly and withholding belief in falsehoods" isn't the only way of setting out the duality (Levi 1967) suggest "relief from agnosticism" instead of "believing truly". This may make a subtle difference to the resulting calculus, but it is still the case that the whole-hearted pursuit of one goal is bound to place the other at risk.

[11] Let's say that an *astronomical* number is one that is somewhere between the order of magnitude of the number of particles in the galaxy, estimated at about 10^{65}, and the order of magnitude of particles in the universe, which is generally estimated as around 10^{80}. A *super-astronomical number* is one that is bigger than that. It is startling to remember, then, that thanks to our simple decimal system of representation, the number of possible different integers we could write on a blackboard—though not

This explosion of possible articulate thoughts, I shall argue, is the first of three consequences of the gift of language, constituting three factors crucial to establishing a categorial human difference that reduces differences between humans to negligible proportions. The second factor is the *detachment* of explicit thoughts from their underlying dynamic representation. The third is the essential *disputability* of principles of rationality that is made possible by the first two.

4. What the bravest lion won't risk

"If a lion could speak," Wittgenstein claimed, "we could not understand him" (Wittgenstein 1953, Part II, p. 223). Why not? Wittgenstein's own answer has to do with "forms of life." But this is both vague and obscure. The real reason is this, which is more specific and which I hope soon to make less obscure. It is that *even the bravest lion will not risk an assertion*.

When people speak, as Wittgenstein was keen to point out, they do many things besides making *assertions*. So we might understand the lion's greetings, and the lion's expression of hunger, perhaps even its intention to eat us there and then. These are illocutionary acts our language can perform, but they are not assertions. An assertion standardly expresses a "full belief," by which I mean the kind that admits of no degrees even when it is made with less than complete confidence. A full belief is correct if true and wrong if false. It aims at *truth*, and to other cognitive values that Nozick occasionally lists: "Science *progresses* when it continues to further achieve its goals, by discovering more truths, accepting fewer falsehoods, uncovering deeper explanations, extending the domain of things that can be explained scientifically, unifying the explanations if offers, making more precise and more accurate

all at once!—is superastronomical. Of course, the number of *points on a line* we can draw on the blackboard is much bigger than that, but the difference is that we can't tell most of those apart, whereas in the case of any two integers between 1 and 10^{100} anyone could tell any two apart.

predictions and so forth." (*INV* p. 116).[12] While animals share our interest in getting certain things right which are essential to their well-being, they do not, I suggest, share these epistemic goals.

But surely, it may be objected, the lion needs to get things right too. Yes, but since he makes no assertions, he doesn't need to have unqualified true beliefs. He only needs correct assignments and revision functions for Bayesian probabilities.

But come to think of it, life being never quite certain, why shouldn't we get rid of unqualified beliefs in favor of Bayesian probabilities? This is "radical Bayesianism," which was advocated by Richard Jeffrey (1992, cited *NR* p. 89). Nozick's objection to radical Bayesianism is that "it is unclear that this position ... can be formulated coherently.... The very setup of theories of personal probably and utility, or the background commentary they require, involves the existence or attribution of *beliefs* to the person whose choices are taken to indicate preference or probabilistic judgments." (*NR* p. 95). In other words, there needs to be a meta-level discourse to make sense of *our talk about* Bayesianism. I hope it will emerge how the momentous consequences I am arguing for arise from the insufficiency of radical Bayesianism *once we start talking.* Meanwhile, however, that doesn't mean that radical Bayesianism isn't true of our lion, providing he persists in not talking about it.

One key role of assertions is to figure as premises as practical syllogisms. Oddly, however, Aristotle's practical syllogism appears in *de Motu animalium*.[13] That suggests that it is not intended primarily as a theory of *rational action;* rather, it appears to be meant as a

---

[12] The goal of understanding can perhaps be thought of as independent of the goal of believing truths and avoiding falsehood. Brevity, "chunking," connection, and the hierarchical organization of knowledge are all clear benefits to limited minds, but they may in the end be ancillary goals. It is not obvious that an omniscient God, with instantaneous access to all atomic facts, would *in addition* need or yearn for explanation. And why seek "unity" or "simplicity" when all is present to consciousness at all times?

[13] "Now we see that the movers of the animal are reasoning and phantasia and choice and wish and appetite. And all of these can be reduced to thought and desire." (*De Motu* 700b17: Nussbaum 1978, 38).

theory about animal behaviour in general, on the assumption that human behaviour conforms on this point to that of other animals. But as a theory of animal behaviour it is most unlikely to be right. And it is equally unlikely to be right about human bodily skills, such as returning a tennis ball, which are not mediated by language (Körding and Wolpert 2004). In all these cases, animal or human, where behaviour is not brought about by explicit deliberation, the determining causes of behaviour have nothing to do with any formulations in language or with any logical inferences.

Actually, this may apply to some decisions that appear to be deliberate as well, but where the deliberation is merely rationalization. To those cases, that most unfashionable figure, Sigmund Freud, is still a good guide. Freud deserves a place in this discussion, because his central preoccupation—and his most enduring claim to our attention—is his concern with the scope and nature of rationality. Freud realized that the veneer of rationality often conceals powerful but irrational tectonic layers. Conversely, he also realized that the most irrational acts, relative to the right set of considerations, have an intelligible rationality of their own. But Freud also distinguished between *primary* and *secondary processes* (Freud 1900, ch. 7). This contrast is related to, though it's not identical with, the contrast between what we do deliberately and what we do "unthinkingly." Freud thought it could be characterized in terms of a number of crucial differences, among which was the fact that the primary processes were *not linguistic* and that they did *not submit to logical rules*.[14]

---

[14] In this respect, it might be interesting to explore in some detail, in the light of the "epidemiological" or "memetics" approach to culture already mentioned, the respects in which culture is swayed by a sort of collective primary process. Ideas spread not by logical or even inductive inference; their dissemination is not subject to any limitations of logic, and for that matter of any other constraints or taste or morality, but solely on the basis of *association*, *displacement*, *reaction formation*, and the *pleasure principle* (Freud 1920). In other words, *culture is essentially infantile.* Sometimes, to be sure, an idea spreads because it is a good idea, in evidence or logic. But the milieu in which this is true is actually a highly restricted one. It scarcely happens outside the environment of science, where the *point* of the activity is focused on the epistemic values, even if the energy that drives the activity is driven, like every other activity, as much by greed, egotism, posturing, and the thirst for fame and sex.

Now the distinction I have in mind is not exactly Freud's. But we can at least suppose that the lion does not go in for the sort of deliberation that typically involves full beliefs. Only full beliefs enter into the practical syllogism: there is no place there for degrees of belief, and indeed even degrees of desire are hard to fit in: a desire of greater intensity can only be compared with another of lesser intensity by using an explicitly comparative premise such as "Dry food is good for man; but moist food is *better*."[15]

If the lion doesn't formulate practical syllogisms, how else can he process information and come to make decisions? There are at least two alternative models for non linguistic information processing. One is Bayesian reasoning; the other is the connectionist alternative to the model of top-down programming in artificial intelligence. Both can be construed, as Freud construed primary processes, as *dynamic*: when belief changes in a Bayesian system, or when connection strengths are adjusted in accordance with some self-correcting algorithm in a connectionist system, we can think of this process as involving a number of vectors causally interacting to produce a behaviour determined by the resulting vector. Bayesian reasoning calculates expected desirabilities on the basis of numerical values assigned to outcomes in the dimensions of both desire and belief. Contrast Aristotle's own example of a practical syllogism: "For example, whenever a someone thinks that every man should take walks, and that he is a man, at once he takes a walk." (Nussbaum 1978, 40). This is still assumed by many to be the paradigm of rational deliberation; certainly it is often useful to think this way in practice. But something like the following would surely be more realistic: *Taking a walk*

---

[15] After writing his famous paper on weakness of will (Davidson 1980), Donald Davidson refrained from publishing it for a number of years. When asked why, he once replied that he hoped first to recast the thesis in terms of the more realistic Bayesian theory of decision. He eventually published it in the original form. No doubt he realized that Bayesian theory simply can't make sense of weakness of will. Framed in terms of practical syllogism, weakness of will can be construed as "detaching" the conclusion of a practical argument of which the premises are less comprehensive than another competing argument. (I will say more on detachment below.) By contrast, since Bayesian theory conceives of the role of reasons as essentially dynamic, it can make no room for a sense in which theory stronger decision is defeated by the weaker.

would be nice and may be good for me; but it's rainy and cold; besides, I have a lot of things to do, I can go to-morrow instead; anyway I have life insurance and no history of cardiovascular problems; besides I just don't really feel like it.... Bayesian theory could well represent such a train of though. It works well with probabilities and degrees of desire; it can be thought of as representing the dynamic interaction of various vectors that together determine a resultant that emerges into action. By contrast, the practical syllogism is hopeless for representing this sort of thinking. It can take no account of degrees of belief; it's not much better at degrees of desire; and it has no method for bringing different considerations *together*. Hence there is a gap between our decisions and our verbal accounts of them. These last sometimes come to look like outright confabulations.

5. Forms of Detachment

This lack of fit between our assertions and the underlying states they rest upon brings me to the second crucial element in the picture I am trying to construct. As Nozick points out, the notion of rationality inherits the problem of *detachment* which, as Hempel stressed, plagues statistical statements (*NR* p. 67). Since assessments of probability are relative to their grounds, the detachment of their conclusion requires additional rules. When detached, they are always liable to be confronted with equally well supported negations (Hempel 1965):

> "With an inference based upon a universally true leading principle, we need not worry about the particular occasion of our information. But with one based upon statistical principle, we need to worry about whether *this* occasion of inference is a representative one." (*NR* p. 66 fn.)

In fact, in the case of a practical argument based on a non-deductive inference, there are no fewer than three steps of detachment that have to be justified. We need, first, a principle of "comprehensive evidence" that allows us to infer that the probability assignment dictated by the argument is the best available. This takes us, say, from (1):

(1) The probability of rain in the light of *E* is *p*

to (2):

(2) The probability of rain is *p*.

But then we need a principle of rational belief which takes one from probabilities to full belief. This one will take us from (2) to (3):

(3) It will rain.

I may still have to make a decision, however, about what is the reasonable thing to do on the basis of this belief. Should I assert it? Should I make plans in the light of it? In making plans, I might actually be better off reverting to (2) and plugging the probability into a computation to arrive at the decision with the highest "decision value" (*NR* pp. 45-63). But we often proceed in terms of a practical syllogism built on unqualified assertions, and into those, only (3) will fit. So we need a third rule, a third kind of detachment, if you will, which will take you from a belief to an action, from (3) to (4):

(4) ‘It will rain’ is a reason for action A in the current decision problem.

On the other hand, perhaps not. Maybe a full belief like (3) can always be plugged into any practical syllogism whatever. If so, we need no further transition to (4).

But that conclusion would be rash, even taking into account the context-relativity that Nozick recommends in his Rule 3: "Believe an (admissible) *h* only if its credibility value is high enough, *given the kind of statement it is*." [my emphasis] (*NR* p. 88). For the relevant context here isn't just a question of classifying the utterance, of "whether it is a report of observation, a statement of theoretical science, a belief about past historical events, and so on." (ibid). One should also take account of the purpose for which we believe it. As Nozick points out a page or two later, the statement that a given ticket will lose in a lottery might make it into the set of rational beliefs for the purposes of assertion, or for the purposes of deciding whether to buy it, but consistency requires it to be suspended in the context that conjoins it with the same statement applied to all the other tickets. That consideration has to figure somewhere, though not necessary in a principle that governs how to get from (3) to (4).

It could, instead, figure as an additional consideration, a downward principle, as it were, that has to figure into the passage from (2) to (3) *in the light* of (4).

Actually, it's not clear that Nozick is right that "we need not worry" in the most favorable cases. Even in the case of a valid deductive argument based on a "universally true leading principle," we might worry about whether it is rational to detach the conclusion. For it is rational only insofar as the leading principle is indeed "universally true," and it may be *true* without being sufficiently *obvious* to soothe our worry. The same holds for the validity of the form of argument involved. Faced with any argument, it would seem reasonable, in the light of reasonable assumptions about my own cognitive powers, to apply the following rule (Rx):

> (Rx) Believe the conclusion of an argument only if its falsity is not more obvious than the truth of the conjunction of the premises with the argument's principle of inference.

Thus it certainly *seemed*, for over twenty centuries, that Zeno's arguments for the nonexistence of space, time and motion were unassailable. But most people sensibly went right on believing in space, time and motion. If pressed, they might or might not have identified the premise, or the form of argument, most likely to be flawed. Failing this, however, most people faced with Zeno's puzzles would have agreed that it would seem more reasonable *blindly* to reject the conjunction of their premises and form of inference than to accept the conclusion.

6. Essential Disputability

Aristotelian practical syllogisms quite often work well enough. Yet just as often they function as *rationalizations*. As verbal declarations, they are, quite literally, stories that we tell ourselves about the identification, motivation, explanation, and justification of an action. But as Freud knew,[16] they are far from constituting a full record of any of those aspects of our

---

[16] For a somewhat pessimistic updating of the Freudian view without the theoretical commitments, see also (Moldoveanu and Nohria 2002, p. 23), who cite (Bargh and Chartrand 1999) in support of the thesis that "people's rational plans by and large *follow* their behaviour rather than lead it."

behaviour. Still, they have a part in what we do, a part that is made possible by language, and owes its defining characteristics to precisely the contestability of its meta-level pronouncements.

This is the third element of the solution I am proposing. Nozick speculates that rationality "may have the evolutionary function of enabling organisms to better cope with new and changing current situations of future ones." (*NR* 120) But that conflates a reference to what the lion does—learn to make reliable predictions on the basis of sensory information processed in the light of experience—with the sort of rationality that emerges from "the processes by which societies mold their members....." (*NR* p. 124-5). This latter process, in which the rules of rationality are socially elaborated, must certainly admit of greater or lesser success: "People are not born rational," Nozick goes on in the passage just quoted. The elaboration of rationality depends essentially, as I have urged, on the capacity for metacognition and mutual criticism. "Once we become self-conscious about it, we can improve the accuracy of our given procedures." (*NR* p. 113). We become self-conscious by facing criticisms that can undermine our own confidence in any given belief or mode of inference. The lion, I have said, will not risk assertions; a fortiori, he will not dispute any assertions either. The scientist, by contrast, is ever ready to do so.[17] On this showing, rationality does indeed admit of degrees. And the most intricate forms of it (not to put too fine an evaluative point on it) are actually the creation of a small *subculture* in which disputation lies at the core of civility. Graham's Priest's rejection of the law of non-contradiction; Fox-Keller's doubts about logical arguments in the context of science; and the whole elaborate debate that culminated in Nozick's own suggestions about the best way to deal with the paradoxes generated by Newcomb's problem, or by the lottery paradox: those are examples of

---

[17] Thus Evelyn Fox Keller writes in a recent book: "I had been trained to see arguments based on mathematics and logic as determining, and experimental evidence as fallible. But others, I soon came to realize, regarded logical arguments as suspect. To them, experimental evidence, fallible as it might be, provided a far surer avenue to truth than did mathematical reasoning. " (Keller 2002, p. ix).

the metacognitive level of discourse that is made possible by the invention of a system of digital representation that runs in parallel with, without ever superseding, the dynamic mechanisms that determine animal behaviour. The possibilities opened up by the essentially contestable character of principles of rationality are virtually infinite, in the sense that verbally articulable thoughts are virtually infinite. That is enough to vindicate Nozick's rejection of (IAT) on the basis of our ability to speak.

7. A potted genealogy of rationality

Here, to summarize my argument, is a potted genealogy of rationality.

In a *pre-rational* phase of our evolution (and also, we shouldn't forget it, at the ontogenetic level in each individual infant), behaviour is governed by tropisms: these can be understood functionally, and they are therefore in some sense—a sense that seems fully captured by the etiological view of objective functions—*teleological*.[18] But they involve neither mentality nor rationality. Tropisms have been installed over the course of evolution by natural selection; they require only some form of detecting function and some sort of differential response. But they display no hint of rationality, despite the temptation, understandable before Darwin, to think of organisms equipped with tropisms as products of rational design.

At the first stage where the idea of rationality begins to seem applicable, the process of natural selection has been internalized, to take place over the life of a single individual. This is *learning*, or at least learning of the Skinnerian kind, and can be regarded as a second-order: function, with first order functions of the same etiological sort now getting built up over the lifetime of the individual. These functions might be implemented in various ways—maybe, but not necessarily, by something like a connectionist network, or by some other type of dynamical system. The implementation mechanism doesn't matter. What matters is that such

---

[18] For the sense in which we can speak of objective biological teleology, see (Millikan 1984; 1989).

organisms have a better claim to look rational, in that it seems more plausible, in their case, to identify success and failure in the pursuit of individual goals. They seem to be intelligible in terms of the "intentional stance" (Dennett 1971). But that is in fact an illusion, since there isn't anything about such organisms that can justify disentangling the individual's goal from the entirely general and "vestigial" goal of gene propagation. The internal dynamics that determine the behaviour of organisms without language can't be subjected to the sort of *criticism* that essentially marks access to rationality. If claim your pet is irrational, to be sure, you won't be contradicted; but nor will you be able to show, in the absence of any possible explicit avowal, that you haven't just misinterpreted your pet's system of desires and beliefs.

At the next crucial stage, the acquisition of language enables the formulation of assertions, and the making of inferences in accordance with rules geared to the form of assertions expressing unqualified (non-Bayesian) full beliefs. Only at that stage is it possible for a conflict to arise in which one of the disputants sets itself against the first order Bayesian vectorial sum of dynamic forces. For now the inferences have at least a partial independence from the desire/belief vectors that they are originally meant to represent. Explicit rules may contradict and correct intuition. By the same token, the explicit rules may get it wrong. They also raise a whole new range of problems. Typical of these is the problem of *akrasia*, which can't be formulated in terms of a dynamic conception of the determination of behaviour by a vectorial sum believing and desiring states, however implemented.[19] At this level, language becomes available to criticize inferences and reconstruct alternative routes.

Nozick might seem to disagree when he writes:

> It was never the function of rationality to *justify* these assumptions that embodied stabilities of our evolutionary past but to utilize these assumptions in order to cope with changing conditions and problems *within* the stable framework they established. It

---

[19] See note 15 above.

> should not be surprising that our rational instruments could not provide conclusive reasons or "justification" for these assumptions. (*NR* p. 121).

But while our "rational instruments" can't in themselves provide *conclusive* reasons for our cognitive strategies, we open ourselves, as soon as we start to talk about them, to an explosion of possibilities for endorsement or for disputes about the principles we use. That is what brings us to the third stage, exemplified by the investigation into rationality represented by the entire tradition of critical investigation of proposed principles of rationality of which Nozick's own work is such a noteworthy exponent. At this stage the issue *does* become that of justifying or rejecting general principles. Take, for example, the contrast between Nozick's (1986) and his (1993) takes on Newcomb's problem. The original paradox depended essentially on existing inclinations to accept principles already tried and tested. Both the principle of expected utility maximization and the principle of dominance had already undergone formulation, refinement, and criticism, and been endorsed on the basis of their apparent correctness. Newcomb's problem brought them into conflict; but that was no more of an indictment of human rationality as such than Russell's discovery of the paradoxes of set theory. Before Russell, it seemed that the unrestricted axioms of set theory were obviously obvious. The paradox shows that not to be so. And so the conversation continues, framed in linguistic and mathematical terms unavailable to other animals. Nozick's new proposal for Newcomb's problem, involving a distinction between evidential and causal conditional probability (*NR* p. 45) may or may not be satisfactory; but what is crucial is that it belongs to a new level of metacognition. At that level, in working out "rules" which will look very different from the ones philosophers were hoping for—for we hoped for "rules and procedures that we ourselves could apply to better our own beliefs, *surveyable* rules and procedures" (*NR* p. 76)—we may well discover, and perhaps repeat, at least vicariously through our computers, some of the same selectionist processes that led to the intuitive rules we had applied in practice all along.

We are equally likely, however, to discover new rules or flaws in the old ones. Look again at the case of consistency and closure. "Perhaps the most effective procedures for arriving at a high ratio of truths (and relatively few falsehoods) will yield a set of beliefs that is inconsistent. Hence, if that high ratio of truths is to be maintained, the set of beliefs had better not be deductively closed." (*NR* p. 77) Such is the moral of the "preface paradox." It is reasonable and not excessively modest of me to think it practically certain that at least one in a thousand of my sincere assertions is false. Suppose I publish a book that makes a thousand claims, adding one more statement to the effect that *Something in this book is false*. This is, or entails, a disjunctive proposition inconsistent with the conjunction of all assertions already in the book. It is, moreover, guaranteed to be true, if only in virtue of its own falsehood if all the other statements in the book are true. So this statement is *certain* to increase the ratio of truths in the whole book, at the cost of consistency. On the closely related "lottery paradox," Nozick suggests that we could deal with without giving up the notion of full belief, if we made belief relative to a context. (*NR* pp. 89-93) Relativity to context here must mean, as far as I can see, relativity to context of *discourse*, not just of decision. Once again, the lion won't care . He's still at a safe, unbridgeable distance from the troubles rationality gets us into.

Both Hacking and Nozick appear to be committed to the view that rationality is a "social product." It might then look as though they must also be committed to the view—the mere dogma, as I have claimed—that all cultures are quantitatively and evaluatively equal. Otherwise, the conclusion that different social groups attain to different levels of normative rationality would be hard to resist. For it is hard to see how any "social product" could be limited only to the two states of perfection or nonexistence. Yet Hacking and Nozick can both avoid the dreaded (IAT), providing they are satisfied with a categorial difference between humans and animals that pertains to *potential* rather than *actual* rationality. All that is needed for that to yield a *potentially infinite* difference, I have argued, is the biological predisposition to acquire, in a social setting, the capacity for linguistic expression of beliefs. The very imperfection of the fit between the underlying degree of Bayesian confidence and the

categorical on/off character of unqualified assertion, opens the way for irrationality. Irrationality, in turn, is what we can be taxed with either by others or by critical reflection on ourselves. Reflection leads to the formulation of increasingly sophisticated explicit principles of inference. That process can become indefinitely technical. That is the current or imminent stage at which, as Nozick suggests, philosophers "will be replaced by cognitive and computer scientists, workers in artificial intelligence, and others." (*NR* p. 76) That's because it now becomes a technical question what are the best rules of inference and how they might best be implemented, just as it is a technical question for engineers, how to improve on the birds' methods of flight with fixed-wing machines. The philosopher is henceforth limited to such observations as I've been making in this paper, in which the reader may have noticed a dearth of normative principles of rationality. Similarly, rules such as Nozick's R1-R6, discussed on pp. 85-93 of *NR*, are not technical rules of inference, but rules that specify ranges of cases in which rules of inference should be applied, practical consequences that might be taken into considerations in addition to epistemic ones, conditions on detaching conclusions from their premises, and so forth.

The bad news is that it may no longer really be possible to understand the rational mind. We aspire, Nozick remarks, to "a structurally revealing and relatively brief description of how the content and structure of our current rational beliefs is related to the content and structure of where they came from. What is the pattern of *this* connection? *There may be no such thing to understand.*" (*NR* p. 80 [my emphasis]). That suggests that certain ways of understanding understanding may no longer be valid. Lord Kelvin once wrote: "If I can make a mechanical model [of a thing] I can understand it. As long as I cannot make a mechanical model all the way through I cannot understand." (Quoted in Johnson-Laird 1988, p. 24). But now perhaps the tables are turned: in relation to cognitive science, it may be that *even when we can build a brain* we will not be able to understand it.[20]

---

[20] Compare Christopher Cherniak (1986, p. 128). "For a representation to qualify as being understood

But the good news is that it doesn't matter all that much, because we are able to correct our intuitions: to "find reasons against the unrestricted truth of the framework assumptions, even if they were evolutionary instilled. Recall the example of Euclidian geometry which, even if selected for as 'self-evident', can be discovered to be not strictly accurate." (*NR* p. 124). So at this stage the additional self-consciousness involved will lead us repudiate some of the principles we have taken for granted in the past—in the past, that is, not only of the species, but of the far shorter history, dating roughly since Aristotle, of our efforts to extract, test, and improve on the highly variable and intrinsically questionable principles of rationality so far elaborated. That's what makes for the unbridgeable distance between practitioners of the Queen of the Sciences and the King of the Beasts.

---

by an epistemic agent, the agent must be able to perceive an adequate proportion (of course, not necessarily all) of the interrelations among elements of the set. But as the "mind's dictionary/encyclopedia" grows, it becomes much more difficult just to search, even with cataloguing and cross-referencing of its propositions or theories.... The universe may be not merely inhumanly complex, but "transcendentally" unmanageable for any physically realizable entity..."

References

Bargh, J. A., and T. L. Chartrand. 1999. The unbearable automaticity of being. *American Psychologist* 54(7):462-79.

Boyd, R., and P. J. Richerson. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology* 13:171-95.

Boyer, P. 2001. *Religion explained: the evolutionary origins of religious thought*. London; New York: Random House; Basic Books.

Cherniak, C. 1986. *Minimal Rationality*. Cambridge, Massachusetts: MIT Press.

Davidson, D. 1980. How is weakness of the will possible? In *Essays on actions and events*, 21-43. Oxford: Oxford University Press, Clarendon.

Dennett, D. 1971. Intentional Systems. *Journal of Philosophy* 68:87-106. Reprinted in Brainstorms (1978). Vermont: Bradford Books.

Fehr, E., and U. Fischbacher. 2003. The nature of human altruism. *Nature* 425 (23 October):785-92.

Freud, S. 1900. *The interpretation of dreams*. In *Standard edition of the complete psychological works*, trans. J. Strachey. London: Hogarth Press, Vol. 5.

Freud, S. 1920. *Beyond the pleasure principle.* In *Standard Edition of the psychological works*, trans J. Strachey, i. London: Hogarth Press, Vol. 18.

Geertz, C. 1973. *The interpretation of cultures; selected essays*. New York: Basic.

Gigerenzer, G., and R. Selten. 2001. Rethinking rationality. In *Bounded Rationality*, ed. G. Gigerenzer and R. Selten. Cambridge, MA: MIT Press. Cognet.

Goodman, N. 1976. *Languages of art*. Indianapolis: Hackett.

Griffiths, P. 2001. Genetic Information: a metaphor in search of a theory. *Philosophy of Science* 68(3):394-412.

Hacking, I. 1994a. What's Best. Review of Robert Nozick, Rationality. *London Review of Books*, 27 Jan., 17-18.

Hacking, I. 1994b. What's Best: reply to Nozick. Review of Robert Nozick, Rationality. *London Review of Books*, 24 March, 4.

Hempel, C. G. 1965. Inductive inconsistencies. In *Aspects of scientific explanation and other essays in the philosophy of science*, 55-79. New York; London: The Free Press.

Jeffrey, R. C. 1992. *Probability and the art of judgment*. Cambridge; New York: Cambridge University Press.

Johnson-Laird, P. N. 1988. *The computer and the mind: an introduction to cognitive science*. Cambridge, Massachusets: Harvard University Press.

Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment under uncertainty: heuristics and biases*. Cambridge and New York: Cambridge University Press.

Keller, E. F. 2002. *Making sense of life: explaining development with models, metaphors, and machines*. Cambridge, MA; London: Harvard University Press.

Körding, K., and D. M. Wolpert. 2004. Bayesian integration in sensorimotor learning. *Nature* 427(6971), 1915 January:244-47.

Levi, I. 1967. *Gambling with Truth*. New York: Alfred A. Knopf.

Lévi-Strauss, C. 1976. *Structural anthropology*. M. Layton. New York: Basic Books.

Maynard Smith, J., and E. Szathmáry. 1999. *The origins of life: from the birth of life to the origins of language*. Oxford; New York: Oxford University Press.

Millikan, R. 1984. *Language, thought, and other biological categories*. Cambridge, MA: MIT Press, A Bradford Book.

Millikan, R. G. 1989. In defense of proper functions. *Philosophy of science* 56:288-302.

Moldoveanu, M. C., and N. Nohria. 2002. *Master passions: emotion, narrative, and the development of culture*. Cambridge, MA: MIT Press.

Moss, L. 2003. *What genes can't do*. Cambridge, MA: MIT Press.

Nisbett, R. E., and L. Ross. 1980. *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ.: Prentice-Hall.

Nozick, R. 1981. *Philosophical explanations*. Cambridge, Massachusetts: Harvard University Press, Belknap.

Nozick, R. 1986. Newcomb's problem and two principles of choice. In *Essays in honor of Carl G. Hempel*, N. Rescher. Dordrecht: Reidel.

Nozick, R. 1993. *The Nature of Rationality*. Princeton: Princeton University Press.

Nozick, R. 1994. What's Best. Letter to the editor. *London Review of Books*, 10 March, 4-5.

Nozick, R. 2001. *Invariances: the structure of the objective world*. Cambridge, Massachusetts: Harvard University Press, Belknap.

Nussbaum, M. C. 1978. *Aristotle's* De Motu Animalium*: text with translation, commentary, and interpretive essays*, 430 p. Princeton: Princeton University Press.

Nussbaum, M. C. *Aristotle's De Motu Animalium: text with translation and notes and interpretative essays.* Princeton: Princeton University Press.

Oyama, S. 2000. *The ontogeny of information: developmental systems and evolution*. Second Edition revised and expanded. Durham, NC: Duke University Press.

Prévost, N., and R. Jennings. 2000. Language and intelligent understanding without semantic theory. *Canadian Artificial Intelligence Magazine* 47 (Autumn):9-16.

Priest, G. 1998. What's so bad about contradictions? *Journal of Philosophy* 95(8):410-26.

Reader, S. M., and K. N. Laland. 2003. *Animal innovation: an introduction*, 300 pp. New York; Oxford: Oxford University Press.

Sperber, D. 1985. Anthropology and psychology: towards an epidemiology or representations. Malinowski Memorial Lecture. *Man* N.S. 20:73-89.

Tomasello, M., and J. Call. 1997. *Primate cognition*. Oxord: Oxford University Press.

Tomasello, M., and H. Rakoczy. 2003. What makes human cognition unique? from individual to shared to collective intentionality. *Mind and Language* 18(2), April:121-47.

van Gelder, T. 1998. The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences* 21:615-65.

Velleman, D. 2001. The genesis of shame. *Philosophy and Public Affairs* 30(1):27-52.

Wittgenstein, L. 1953. *Philosophical investigations*. Trans. G. E. M. Anscombe. New York: Macmillan.