Candidate number: ***11419***
Undergraduate philosophy thesis 2002

Is a naturalistic account of reason compatible with its objectivity?

Abstract 1
1 – Introduction to Nagel's views 2
What do we mean by 'reason'? 2
What is objectivity? 3
Truth and certainty 4
Thoughts we cannot get outside of 5
Reason - ontological and epistemological theses 6
2 – What is naturalism? 7
Naturalism - ontological and epistemological theses 7
3 – Ontological and epistemological clashes between reason + naturalism 8
The ontological thesis of reason 10
The epistemological thesis of reason 16
Empirical evidence of irrationality 17
Evolutionary theory 22
Degrees of rationality 33
Connectionism 36
Finite beings and infinite thoughts 51
4 – Conclusions 53
Bibliography 55

Abstract

In *The Last Word*, Thomas Nagel argues strongly that even "contingent, biological creatures such as ourselves" can have access to "universally valid methods of objective thought"[1], i.e. *reason*. I consider how this rather pure notion of reason is threatened and can be reconciled with the naturalistic approach in general, and in particular with a near-future naturalistic world view, characterised mainly in terms of recent psychological expermimental evidence, neo-Darwinist evolutionary theory and connectionism.

[1] Nagel (1997), ch 1.

1 – Introduction to Nagel's views

In passionate, almost righteous terms, Thomas Nagel's most recent book, *The Last Word*, is an attempt to expunge the subjectivism that is "epidemic in the weaker regions of our culture"[2], by defending our capacity to engage in "universally valid methods of objective thought", through reason. Reason is the means by which we can form beliefs that are not 'constructed', or based contingently in consensus, but that everyone must accept

> *by virtue of their generality and their position in the hierarchy of justification and criticism*[3].

What do we mean by 'reason'?

It may help to start by stating the obvious, by distinguishing between:

1. a '*reason*'

   as a propositional justification, that can play different roles which can be broadly divided into explanatory roles (having a logical connection to, and so justifying, the having of other beliefs) and normative/motivational roles (reasons as causes for action[4])

2. '*reason*', as in the mental capacity for rationality

   which need not be a unitary module in any sense, but rather any application of our mental faculties by which we can apprehend and form objective reasons

[2] ibid. pg 4.

[3] ibid., ch 1.

[4] see Davidson, e.g. 'Actions, reasons and causes' (1963) in Davidson (1980).

3. a line of *reasoning*

   where we chain together sets of reasons to justify further beliefs, desires, intentions or actions

What is objectivity?

Before going any further, we need to unpack the underlying notion of 'objectivity'. In order to give reason normative force to reason as a source of authority, both within oneself and by which we can persuade others, Nagel makes the central claim that there are principles and ways of thinking whose validity does not rest on a point of view, i.e. that are not relativised or need to be qualified as true only *for me* or *for us*. The difficulty and confusion, as Nagel sees it, is that "to be rational we have to take responsibility for our thoughts while denying that they are just expressions of our point of view"[5].

As Nagel summarises it most neatly:

> *The aim is to arrive at principles that are universal and exceptionless – to be able to come up with reasons that apply in all relevantly similar situations, and to have reasons of similar generality that tell us when situations are relevantly similar*[6].

Importantly then, reasons, in the sense that Nagel wants us to use the term, must be applicable for anyone within a given set of circumstances. When the question being considered does not directly involve a perspective, as when considering a mathematical proof, then seeing a reason as applicable in all "relevantly similar situations" may not seem unduly problematic, since we can all but ignore the circumstances. However, if the domain of discussion necessarily involves

[5] Nagel (1997), ch 1.

[6] ibid. ch 1.

a perspective (first-person or otherwise), as in ethics, where one is always trying to answer the question, ‘What should *I* do?’, the extent to which the answer is relativised to circumstances becomes of considerable importance[7].

In this essay, I do not intend to discuss such applications of reason to particular domains, such as ethics, but rather to focus upon the generic notion of reason as an objective capacity considering issues that minimise reference to circumstances, and its compatibility with a contemporary naturalistic account.

Truth and certainty

In his discussions of objectivity in *The Last Word*, Nagel largely skirts the issue of truth and certainty. In *The View From Nowhere*, he states that:

> *We must be resigned to achieving [truth] to a very limited extent, and without certainty*

and Moore comments that:

> *The subjectivity of a belief does not, in itself, impugn its truth. (There are familiar arguments to the effect that even our belief that grass is green is subjective.)*[8]

I mention these statements because together they show that the authority of reason may not extend to certain true, inherently subjective statements, and may not provide certainty. Perhaps the best that we can do is aim for non-relativised principles and ways of thinking that are universal and exceptionless. By talking in these terms, we support but do not commit ourselves to the view that we converge upon truth as we become more and more objective.

[7] See Nagel (1997), ch 5 sections III and IV for further discussion of Hume and the passions, and of Williams and our internal motivational set.

[8] Moore (1999).

Moreover, we can still make literal sense of the attacks on consensus-based pragmatisms such as Rorty's, which claim that there is nothing more to objectivity than solidarity with your speech community, since we can still talk in terms of truths we don't yet know or don't believe, or falsities which will never be revealed. Arguably though, the pragmatist would just view these as more useful beliefs that we have yet to discover. However, Nagel argues that such a 'phenomenological reduction' would be trying to get outside of thoughts that underpin our entire means of thinking and regard them merely as appearances, which cannot be done.

Thoughts we cannot get outside of

Nagel formulates his broadest attack on all forms of skepticism in terms of 'thoughts we can't get outside of'. It is perhaps best illustrated by his brain-scrambling argument, based around Descartes' evil demon: if my brain is being scrambled, then I cannot rely on any of my thoughts, nor can I rule out any thoughts as invalid, however, inconceivable. In such a situation, says the skeptic, I would not be able to accord objective validity to any hierarchy among my thoughts.

Nagel argues in response that there just isn't room for skepticism about our most basic processes of thought, since we are relying on these to formulate the skepticism itself: "If my brains are being scrambled, none of my inferences are valid, including this one"[9]. He thinks that the true philosophical point of the *cogito* is that there are some thoughts which we cannot get *outside* of. They enter "inevitably and directly into any process of considering ourselves from the outside"[10]. There are more than one such thoughts – it's not just 'I exist', but also all of logic and maths, and possibly practical reasoning and moral reasoning.

[9] Nagel (1997), ch 4 section II.

[10] ibid. ch 2.

Next, I want to draw a distinction that can be drawn about reason as depicted in Nagel's account, between the ontological and epistemological theses of reason. This is paralleled by a similar distinction of levels in the naturalistic account. Equipped with these two-tier accounts of reason and naturalism, we will be in a position to consider how naturalistic and rational pictures of the world can be reconciled.

Reason - ontological and epistemological theses

The distinction between what I will term the ontological and epistemological theses of reason is crucial to the discussion.

By the ontological thesis of reason, I mean the universal norms which form the authority to which we appeal with our reason. These norms are what govern "the logical relations among propositions"[11] that our thoughts obey, and he hypothesises, arise out of "some systematic aspect of the natural order that would make the appearance of minds in harmony with the universe something to be expected". Nagel confesses that he finds congenial a number of alarmingly Platonist, antireductionist, realist excerpts that he quotes from Peirce, but which he is concerned rest on a religious or quasi-religious world picture. If we were to reject the ontological thesis of reason, and said that there are no universal norms of reason to which our mental capacities can attach and apply to circumstances, then we would be left perhaps with a world view like Rorty's pragmatism. The task here is to consider whether this ontological thesis of reason can be reconciled at all with an equally strong ontological thesis of naturalism.

The epistemological thesis of reason relates to our limited capacity as finite beings to be rational. If we accept the ontological thesis of reason, then the epistemological thesis concerns the extent to which we are able to correctly apprehend the dictates of these universal norms of reason. If we reject the ontological thesis, and adopt some pragmatist (anti-)metaphysics, then the epistemological thesis concerns how well we are able to form useful beliefs given our situation, and recognise more useful belief-sets when faced with them.

As I will try and show, the extent to which we consider naturalism and reason to be compatible depends on whether we accept one or both of these aspects of Nagel's account of reason.

2 – What is naturalism?

Naturalism - ontological and epistemological theses

Naturalism is both an ontological and an epistemological thesis, often taken together:

> *all of reality is natural, that everything that exists is amenable to scientific study*[12]

In essence, the ontological thesis considers that:

> *the world of nature should form a single sphere without incursions from outside by souls or spirits, divine or human, and without having to accommodate strange entities like non-natural values or substantive abstract universals*[13].

'Natural entities' includes theoretical entities which cannot be directly observed, but whose existence is postulated (whether real or merely as instrumental constructs) to explain various phenomena. In totality, the different methods and levels of explanation should form a continuous chain, amenable ultimately to empirical testing.

[11] ibid. ch 7, section I.

[12] *Bloomsbury Guide to Human Thought*, Bloomsbury 1993.

[13] *The Oxford Companion to Philosophy*, Oxford University Press 1995.

The epistemological thesis states that everything in the world is amenable to the scientific method. Broadly speaking, the scientific method involves formulating a hypothesis to explain a set of preliminary observations, then testing (and so confirming or falsifying) the hypothesis against empirical evidence, and finally codifying the results as laws and theories ready to be refined and subjected again to the same process. Of course, we can take this epistemological thesis more or less strongly. We could say most strongly that nothing in the world is amenable to understanding except by scientific study, or we could say only that the world is most easily or best understood by the method of scientific study. This epistemological claim fits closely with the ontological claim, since it makes sense that if everything in the world is natural, the methods of natural science would be the best for investigating and learning about the world.

However, we can take one claim without the other and maintain some intelligibility. We could say that the world is natural, but that our unanalysed intuitions and senses are the best sources of knowledge we have, which might make more sense perhaps for a highly evolved but less intelligent creature, although it sounds like an implausible approach for us. On the other hand, we could imagine that the world contains non-natural entities (e.g. God), but that natural science remains our best means of apprehending it, as a scientist like Newton more or less believed.

3 – Ontological and epistemological clashes between reason + naturalism

There is an obvious clash here between Nagel's realism (the ontological thesis of reason, of irreducible norms of reason with which we somehow engage), and the naturalist antagonism towards any such entities as "strange … substantive abstract universals". We might hope that the situation is somewhat similar to trying to understand how there could be objective norms of morality in a naturalistic world, which is a more familiar question. It is the ontological thesis of naturalism that is referred to (usually pejoratively) as ‘scientism’, a “special form of idealism”, which:

> *assumes that everything there is must be understandable by the employment of scientific theories like those we have developed to date – physical and evolutionary biology are current paradigms – as if the present age were not just another in the series*[14]

There is a second clash relating to the two epistemological claims, both of which seek to be our primary means of apprehending the world and justifying our understanding. relates to the epistemological claim of naturalism. According to Nagel, naturalistic accounts of reasoning seek:

> *an understanding of the world [which] could close over itself by including us and our methods of thought and understanding within its scope*[15]

but

> *this hope cannot be realised, because the primary position will always be occupied by our* employment *of reason and understanding … even when we make reasoning the object of our investigation.*

The question then becomes whether our primary means of understanding ourselves and the world is through reason, or the methods of science. The fact that the two are inextricably related gives promise of an easier reconciliation than the two ontological theses.

[14] Nagel (1986).

[15] Nagel (1997), ch 7, section III.

I intend to discuss the compatibility between reason's objectivity and naturalism by assuming both ontological and epistemological naturalism, and seeing how this affects the account of reason.

The ontological thesis of reason

I want to consider first whether Nagel's ontological thesis of reason resists attack.

Nagel's anti-skeptical arguments ignore a remaining area of logical space, by characterising skepticism in terms of rationally forming objective beliefs:

> *[skepticism] is always the product of reasoning to the conclusion that various mutually incompatible alternative possibilities are all equally compatible with one's actual epistemic situation, and that it is therefore impossible to decide among them on rational grounds*[16]

He has described the situation as though one has the choice between asserting the objectivity of reason, or trying to deny it by incoherently relying on it in one's skeptical argument.

However, it seems possible that one can simply refrain from believing, one way or the other, in the objectivity of reason, and treating all propositions to which we assent as doubtful, attaining the "state of mental rest owing to which we neither deny nor affirm anything"[17] of Pyrrhonic skepticism. Nagel tries to forestall this possibility by arguing against the intelligibility of simultaneously dismissing the objective validity of *p* while believing that *p* is true. But that is not what is being done here – if a skeptic is prepared to forego reliance on any logical coherence of thought and belief, and maintain perpetual non-belief, then they are able to avoid

[16] ibid. ch 5, section II.

[17] Sextus Empiricus, *Outlines of Pyrrhonism.*

both horns of Nagel's dilemma. Of course, to phrase the skeptic's argument in the way I've done here, either out loud or on paper, is to employ reason, and so to fall foul of Nagel's primary thrust. Unfortunately of course, such an approach would hamstring the naturalistic account as much as the rational objectivist's account.

Along similar lines, it could be said that Nagel is unjustified in shifting the burden of proof to the subjectivist so readily. In fact, I think his own anti-skeptical argument can almost be turned against him. He tries to use the brain-scrambling thought experiment to show that the brain-scrambled thinker will have no position from which to argue against reason, except by adopting the position of a reasoner. But if we start off from a position of skepticism, a limbo in which we have no idea if the norms of reason exist or are accessible to us, the objectivist will equally have no position from which to argue for reason, except by presumptuously adopting the position of a reasoner. For after all, the objectivist is also relying on reason without having proven it when producing his anti-skeptical arguments. I accept that Nagel can respond to the argument in this paragraph by pointing out my unavoidable use of reason. But the skeptic can equally respond that Nagel would be completely unable to convince the Pyrrhonic skeptic mentioned above, who has not adopted reason in any form, any more than Nagel would be able to convince a baby not to be a skeptic. If one refrains from the use of reason, then the very possibility of reason can be called into doubt.

I think Moore makes a similar point. He discusses the contradiction inherent in a skeptic's saying that any thought or belief can only be justified as being historically contingent or culturally local. Is this judgement supposed to apply to itself? Wouldn't such judgements of relativity require a position of greater objectivity from which to be made? If so, Nagel thinks that this would leave us without the possibility of thinking anything at all. He says that such claims are like saying ‘Everything is subjective’ – these break down whether they are considered to be objective or subjective themselves.

In Moore’s opinion, the argument that such a general subjectivist statement is self-refuting is a “standard but to my mind facile objection”[18] since:

> *If the claim is subjective, there can still be reason to accept it, if only subjective reason. It does not rule out any objective claim. At least it does not need to rule out any objective claim if there are no such things. It need only rule out other subjective claims, which it certainly does: it rules out the claim, from the same point of view, that some of our beliefs are objective.*

At first blush, this seems quite persuasive. Subjectivist general statements are only self-defeating if there is an objective general statement for them to run up against; but if there is no such objective claim, there can still be subjective reason to accept the subjectivist statement. Of course, we would have to consider what we mean by subjective reason – perhaps we could characterise it terms of internal coherence, elegance, explanatory power etc. or some such list, or even in terms of some problematic notion like “idealised rational acceptability”[19].

However, I think that Nagel’s argument can burrow deeper than Moore realises here, since this is an example of a thought we cannot get outside of. Nagel can argue that in forming this entire counter-objection, the subjectivist must have rested upon reason in some form. Here, the thought we cannot get outside of is an entire way of thinking which we cannot do without, namely logic (even if not formalised into a symbolic, context-free formula). This seems quite clear to me – even if we allowed Moore’s contention that the subjective claim “does not need

[18] Moore (1999).

[19] Putnam (1981).

to rule out any objective claim if there are no such things", this contention itself is intended to be an objective principle that is not simply true *for Moore* or for *us*, but true objectively.

As I will argue below, this still leaves the option open to the subjectivist to fall back on 'subjective reason' alone, happily accepting that he would be giving his readers no objective reason to be persuaded of his claim. If we take 'subjective reason' to mean something like internal coherence or elegance, then we have effectively relegated reasons to pragmatic beliefs. In doing this, we would be treading the fine line between incoherently making the objective statement that there are no objective statements, and making the case for relativism/pragmatism.

As I will argue further below, I think that Nagel is at his weakest here, in defence of his bedrock claim that:

> *there can be no justification of the fundamental principles of deductive reasoning - the simple laws of logic are the last word*[20]

Simply put, we are asking Nagel to point to how he knows that modus tollens is an objectively valid mode of thought. What is there about the dictates of reason that we can be so sure is engaging with objective principles, beyond the self-evidence of rational statements. Ultimately, if a subjectivist claims that all of Nagel's arguments are persuasive *to him*, but that he still has no reason to believe that they are objectively authoritative, it seems that there is nothing that Nagel can say to convince the subjectivist of the existence of universal norms, to which the subjectivist cannot reply, 'Yes, *I* find that argument convincing, but how do I know that it is really, *objectively* valid if I have no underlying objective reason for believing in underlying

[20] Bermudez (1999).

objective reasons? The fact that I am employing reason in the way that we Western philosophers usually do, does not mean that I know my reasoning is *objectively* valid.'

Nagel's almost Hegelian, "quasi-religious" answer is that:

> *the capacity of the universe to generate organisms with minds capable of understanding the universe is itself somehow a fundamental feature of the universe*[21]

although he stresses that this

> *at no point [implies] the existence of a divine person, or a world soul*

so, he veers sharply away from Hegel's more literal idea of a universal, all-encompassing self-consciousness, or *Geist*. In an attempt bridge a growing gap between his position and the naturalistic one, he describes this as a "mind-friendly cosmology", i.e. natural "laws that explain the *possibility* of intelligent life". Essentially, he is trying to argue that the universe is constituted such that rational beings exist, since none of the candidate explanations (God, subjectivism, or evolution) he considers for the "cosmic authority problem" satisfy him. Indeed, he recognises the parallels with the anthropic principle in cosmology, stated in its strong form as "the Universe (and hence the fundamental parameters on which it depends) must be such as to admit the creation of observers within it at some stage"[22]. Unfortunately, I think he is entirely correct in thinking that any such anthropic prediction will always be subject to the worry that there is a deeper and more satisfying underlying theory.

I think that Nagel has shown skepticism to be an incoherent position for anyone who trusts their thought processes. That is, skeptical *arguments* (based on reason) are necessarily self-refuting. He has further shown that any agent embarking on a Cartesian project of pure enquiry

[21] Nagel (1997), ch 7, section I.

must, of necessity, trust those thought processes in order to progress from Pyrrhonic skepticism at all.

However, I think these objections demonstrate that if we are seeking a foothold in the dark from which we can justify those thought processes, he has failed to provide an all-powerful means of persuading the Pyrrhonic skeptic. I suspect that he would throw his hands up in frustration at the obtuseness of an objector demanding an argument from nothing that can bootstrap oneself into the position where that argument will have normative force. But it seems to me that that is what is needed in order to justify the ontological thesis of reason.

As a result, I will sketch the position of pragmatism, along the lines that Rorty has advocated, as a possible stance from which to continue the discussion of the epistemological thesis of reason.

Rorty's pragmatism rejects the Platonic notion of truth as the complete set of universal, objective, incontrovertible (and possibly inaccessible) beliefs. As he puts it, "the picture which holds traditional philosophy captive is that of the mind as a great mirror, containing various representations – some accurate, some not"[23], with philosophy peering into and polishing this mirror in the hope of seeing a better, clearer image of the universe reflected in their own minds. Instead, he wants us to "treat beliefs not as representations but as habits of action, and words not as representations but as tools". "There is no point in asking whether a belief represents … either mental or physical reality". Rather, we should ask, "For what purposes would it be useful to hold that belief?". "The purpose of inquiry is to achieve agreement among human beings about what to do, to bring about consensus on the ends to be achieved and the

[22] Nagel (1986), ch 5, section on 'Evolutionary epistemology'.

[23] Rorty (1999).

means to be used to achieve those ends." In this way, world-views are like clothing. If things get cold, we wear the warmest clothing we have. If we find a better, warmer fabric, we wear that instead. If the climate changes, we don a new, more suitable garment. There is no single, true clothing (belief/belief-set, world view etc.) best suited to the environment (read universe), no one ideal fabric that we are converging upon, just a series of adapting cognitive apparatuses. To a limited extent, it is even possible to talk of degrees of objectivity, with respect to the variety of individuals, communities or races which hold a given belief – but crucially, this only a consensus-based objectivity. In short, it's just the sort of linguistic, agent-relative subjectivism that Nagel despises.

The pragmatic position that I want to adopt is based on the useful naturalistic belief that the assumption that there is some sort of systematic order in the universe (the ontological thesis of naturalism), and that we can and must trust our thought processes (the epistemological thesis of reason) to operate effectively, if falteringly, within the scientific method (the epistemological thesis of naturalism). We cannot be sure of this, or that our conclusions are universally and objectively valid, but I will argue that our success in manipulating and understand the world, reason and our place in it through science, gives us good reason to continue to employ naturalistic beliefs.

The epistemological thesis of reason

If we had accepted the ontological thesis of reason, we would now be in the difficult situation of considering how as contingent, biological beings, we are able to access objectively valid thoughts with infinite range, such as modus tollens or arithmetic, and whether these objective norms could be explained in naturalistic terms. However, having found available arguments for the ontological thesis insufficiently persuasive to a devil's advocate full-blown skeptic, it was necessary to adopt pragmatism in its stead in the meantime. As a result, we have to view modus tollens and arithmetic as enormously useful beliefs whose objective validity is irrelevant to our efficacy. In this case, the difficulty of our apprehending such beliefs and ways of thinking becomes less intractable. If the epistemological thesis of reason is shown to be similarly unsturdy, we will be have to consider ourselves finite beings whose faculties are ill-suited to forming useful beliefs.

The way I intend to consider our capacity for rationality is by first reviewing some of the salient empirical evidence. I have focused mostly on our failures, since our success at reasoning is well-known: logic, mathematics, science and philosophy are the most obvious examples. I will then discuss a number of noticeable limitations on our reasoning capacities, some of which are apparent a priori, and others through introspection and interaction between reasonable people. Finally, I want to see how well our contemporary naturalistic picture is able to explain our success and failure as reasoners, focusing on evolutionary theory and connectionism as largely exhausting a modern scientific understanding of mind.

Empirical evidence of irrationality

Despite the impression that Nagel sometimes gives with grandiose claims like:

> *I am justified in trusting [my reasoning capacity] simply in itself – that is, believing what it* tells *me, in virtue of the* content *of the arguments it delivers*[24]

human rationality is limited and highly fallible, and any theory of reason needs to accept and explain this. To his credit, he acknowledges that we can make temporary mistakes in our reasoning, and that it is often possible to discredit appeals to the objectivity of reason by showing that their true sources lie elsewhere, in prejudices, social conventions, unexamined assumptions or tricks of language. He also reiterates the important but simple distinction between criticisms of reasoning and challenges to reason itself. A given line of *reasoning* can be mistaken and can be challenged, but the means by which we correct these mistakes and the position from which we pose these challenges assume reason itself as a source of non-subjective authority.

But I think that some of the empirical evidence of human irrationality requires more explanation than glib dismissals of mistakes as failures of concentration or the result of being finite creatures. Our minds clearly develop through our lifetimes, and it seems plausible to look for further evidence that our rational capacity has evolved to fit our ecological niche, and furthermore that rationality admits of grades. However, just as many intelligent adults when questioned, "think that a ball flying out of a spiral tube will continue in a spiral path"[25], but when the same people are shown an animation of a ball flying out of a spiral tube in a spiral trajectory, they burst out laughing. This illustrates that even when our intuitions or initial judgements are wrong, with a little help our errors sometimes become easily apparent, just as in the following probability examples[26]:

- All gambling and playing the lottery (the 'stupidity tax') – since the house must profit, the players, on average, must lose.
- People are more afraid of flying than driving, even though plane travel is statistically far safer. The same misjudged wariness with regard to nuclear power rather than coal, or pesticide residues and food additives (which "pose trivial risks of cancer compared

[24] Nagel (1997), ch 7 section II.

[25] Pinker (1997), ch 5, pg 302.

[26] Tversky and Kahneman, 'Extensions versus intuitve reasoning: The conjunction fallacy in probability judgement', in *Psychological Review* (1983), 90, pp 293-315

to the thousands of natural carcinogens that plants have evolved to deter the bugs that eat them").

- Narratising some sort of memory into independent events, e.g. coin-tossing

The brain appears to adopt rules of thumb in place of theorems, e.g. the more memorable an event, the more likely it is to happen (plane crashes). Of course, such behaviour can often be understood to some degree in other ways. Gambling is a thrill, and plane crashes are disturbingly horrific.

However, people's performance at falsifying hypotheses is similarly flawed. Wason (1966) told subjects that a set of cards had letters on one side and numbers on the other, and asked them to test the rule 'If a card has a *D* on one side, it has a *3* on the other', a simple P-implies-Q statement. The subjects were shown four cards and were asked which ones they would have to turn over to see if the rule was true."

D F 3 7

"Most people choose either the *D* card or the *D* card and the *3* card. The correct answer is *D* and *7*. 'P implies Q' is false only if P is true and Q is false." Only about 5-10% get it right. It's not because people assume it's an 'iff' statement, otherwise they'd turn over all the cards. People seem to be "confirming their prejudices rather than seeking evidence that could falsify them". Cosmides[27] has found that rephrasing the experiment with real-world events helps, e.g. phrasing a logically-identical problem in terms of a bouncer checking for under-age drinking – but only when the rule is a contract, an exchange of benefits. She cites this sensitivity as support for Trivers' prediction that humans, as "the most conspicuous altruists in the animal kingdom, should have evolved a hypertrophied cheater-detector algorithm".

The discussions about probability and falsifying hypotheses serve to make two more points. They show that our thought processes seem to be biased towards having or forming beliefs and habits of thought that are useful, efficient, approximate the truth and are perhaps specialised for the world in which we evolved. Lastly, we sometimes seem able to employ higher-order processes (involving language, abstraction/formalisation and step-by-step breakdowns) to validate and quantify these semi-intuitive conclusions and form the sort of systematic frameworks that exemplify Nagel's defence of our rational capacities. It is only in forming these frameworks that we are able to step back and devise the mathematics of probability or mechanics to see our error. The ability to do this could be seen as supporting some sort of objective norms, but could perhaps also be seen as simply a useful way of evolving to seize upon whatever order and regularity happens to exist in our environment.

Borrowing from Chomsky[28], Cohen[29] distinguishes between performance and competence, which amounts to the distinction between how well you actually do something, and how well you are (potentially) capable of doing that task. In the same way that a superb sportsman may have an off-day because of lack of sleep, or nerves, our *performance* as reasoners (e.g. in various psychological tests) may be significantly inferior to our *competence* on a good day. In contrast to Cosmides, Cohen considers the inferential failings demonstrated in the above experiments to result much more from "either the presentation of the problem, or from subjects' inability to properly encode the logical structure of the task being presented", both of which are failures of *performance* rather than competence. The overall thrust of Cohen's conclusion is that the research on human inferential shortcomings should be construed as

[27] Barkow, Cosmides and Tooby (1992).

[28] Chomsky, N. (1965), *Aspects of a theory of syntax*, MIT Press.

[29] Cohen (1981).

showing how subjects can be vulnerable to "cognitive illusions" when problems are presented in unfamiliar ways that interfere with their inferential performance. There is something to be said for these criticisms of the experimental method, but as I will argue further below, Cosmides' conclusions are far too plausible to be discounted without much stronger evidence against them.

Because of the gap between performance and competence for native speakers of a language (who can reliably identify grammatical sentences, but frequently make slapdash errors in speech), Cohen argues that the only way to define the grammar of a language is through the careful intuitions of what Rorty might term an ideal community of speakers of that language. The job of linguists is to systematically describe the sum of careful, intuitive grammatical judgements given by just such a group of intelligent, fluent (probably native) speakers. Where obvious schisms do appear between groups of speakers, then we can say that we have distinguished dialects within the language. In the same way:

> *the only way to tell that modus ponens and modus tollens are valid inference rules is that competent thinkers judge arguments of this form to be good ones. Note that this does not mean that competent thinkers will never be misled by the presentation of an argument and fail to recognize that modus tollens is an applicable inference rule.*

This is quite in keeping with a consensus- (or in this case, ideal community-) based definition of what is rational. There are no norms on which we are converging, only a set of more or less shared and useful evolved intuitions that have been described and improved upon from Aristotle through to Frege. This non-objective characterisation of rationality also makes long-standing, ineliminable differences of opinion between like-minded, intelligent, reasonable people much more intelligible: there just is no final, objective answer on which the debate must settle.

Evolutionary theory

I want to discuss evolutionary theory now, because the naturalistic account rests so heavily on it as a means of explaining how it is that we can be cognitively so well adapted to the world, without recourse to God or Kant's Copernican revolution that portrays the world as adapted to us (see Nozick's account below).

The fundamental tenet of evolutionary theory (or 'neo-Darwinism') is the principle of natural selection, whereby parental characteristics that vary across organisms play a role in non-random differential reproduction. That is, 'adaptive' variations are those which increase an organism's (or other organisms with similar genes, strictly speaking) propensity to survive and reproduce (its 'fitness'). This process gradually gives rise to diverse forms leading ultimately, through selective adaptation to specific niches and environments, to the emergence of new species. If humans are the end-product of a natural, non-teleological process of evolution that has resulted in the particular, contingent bodies and brains that we have, then our reasoning abilities and limitations should be largely explicable in terms of evolutionary theory too.

Curiously, evolutionary theory has been employed by both objectivists and subjectivists to support their claims. For example, Stich[30] cites Quine, Dennett and Fodor as implying that evolution selects for rationality and that irrationality is empirically impossible or unlikely, e.g.:

> *creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die out before reproducing their kind.*[31]

However, it is also extremely plausible that the opposite is the case, i.e. that evolved creatures (including ourselves) are highly unlikely to be objectively rational.

[30] Stich (1990), ch 3.

Stich[32] considers that there are two main reasons why people think that evolution insures rationality:

1. Evolution produces organisms with good approximations to optimally well-designed characteristics or systems
2. An optimally well-designed cognitive system is a rational cognitive system

He systematically undermines takes apart both of these premises. I will not consider the arguments in as much detail, but I will try and recount, evaluate and supplement them to a limited extent.

The notions of 'fitness' and 'optimality' are central to any evolutionary theory. A system is 'well-designed' if it enhances fitness more than any alternative. Of course, this is problematic because of the difficulties of deciding what counts as an alternative. No doubt a predator which had evolved a high velocity rifle as an extra limb would be at an enormous evolutionary advantage, but as I will discuss below, the march of evolution is restricted to a sequence of gradual changes, each adaptive in their own right – although the end result of a fully-functional rifle would be highly adaptive, all of the intermediary stages (growing a long, perfectly straight, protruding barrel, the combustion mechanism, an organ for ammunition manufacture etc.) would be highly maladaptive right up until they were all brought together.

Stich considers a number of technical arguments against a naïve faith in evolution as an infallible optimiser. He points out that natural selection is not the only process that causes changes of gene frequency in populations (which is how biologists define evolution). Mutation, migration and random drift all affect gene frequencies, to a greater or lesser degree.

[31] Quine (1969).

For instance, a random event or disaster might wipe out a large proportion of a population, including all the carriers of a particular fit gene, allowing a less fit gene to take hold in the population.

Natural selection does not necessarily choose the best genes in the gene pool anyway – Stich discusses meiotic drive, the effect of combined recessive genes, pleiotropy and heterozygote superiority. Each of these phenomena can lead to less optimal members of the available gene pool being selected for. In the case of meiotic drive, for example, "certain genes have the capacity to 'cheat' in meiosis ["the process that produces sperm and eggs"] and end up significantly over-represented in the sperm or eggs", and so "obviously, such a gene will spread quickly through a population, even if the phenotypic effects of the gene are harmful".

Lastly, it is necessary to show that our cognitive system is a product of biological evolution. After all, "even if it were the case that natural selection is a flawless optimiser and that it is the only cause of biological evolution, it would still not follow that our system of inferential strategies is optimally well-designed" unless "evolutionary factors are the only [or major] ones that have shaped our current inferential strategies". In order for natural selection to shape a characteristic, there must be variation in the population that affects reproductive success in a systematic way, and this variation must be under genetic control either directly or indirectly. Stich considers clothing styles and language as examples in which there is great diversity within the human population, that may have some impact on fitness, but this diversity is not genetically based. "Had I been born elsewhere, I would now have the ability to speak Lapp or Korean rather than English." And the processes by which languages spread are almost entirely independent of biological evolution, depending far more on social and historical factors, for

[32] Stich (1990), ch 3.

example. Similarly, it may be that “the [inferential] strategies a person employs, like the language he/she speaks, are determined in large measure by environmental variables”.

Assuming that we could show that evolution produces near-optimal systems, the crucial second step necessary to show that evolution selects for rational systems is to show that an optimal system is a rational one.

The first way of doing this uses the traditional methods of analytic epistemology. Analytic epistemology is the approach of grounding cognitive assessment in the analysis or explication of our ordinary evaluative concepts. It involves showing that ‘optimality’ and ‘rationality’ are conceptually, analytically identical, that is, “if we analyse what we ordinarily mean when we say that one inferential system is more rational than another … we will find we mean that one is more fitness enhancing than the other”, i.e. “the claim that optimally well-designed cognitive systems are rational is a conceptual truth”. It is intuitively clear that this is not what we mean by ‘rationality’, and this will become clearer soon as the second approach which might support this view falls down too.

The second means of showing that an optimal system is a rational one would be to argue that the rationality of an inferential system is a function of its reliability and consistency. (Here, we have to depart from Stich’s approach, since he has not yet adopted a pragmatist stance, and so couches his description of rationalism in terms of truth and reliabilism.) We need to show that inferential strategies that generally yield the best beliefs are fitness enhancing. Of course, there will always be cases when using the most reliable belief-forming strategy will not be fitness-enhancing, e.g. getting the time of your train wrong (by using an reliable strategy), arriving late and missing it, only to find that it crashed en route, in which case you would have died.

However, subjectivists need to show that a *generally* less reliable system could exceed a more reliable system. Following Stich, I will utilise Sober's distinction between internal and external fitness.

The internal fitness of an inferential system relates to how economically it achieves its effects, in terms of the demands made on the organism's memory, processing capacity, energy/resources, time etc. The declining marginal utility of increasingly accurate beliefs has to be considered when evaluating an expensive, reliable system as opposed to a less expensive, less reliable system. A good example of such a trade-off might be the fact that our cognitive representation of space employs Euclidean geometry, which isn't as 'true' as Einsteinian relativistic geometry, in that the latter provides a better quantitative model of the world. It seems reasonable to think that a relativistic representation of space would require considerably more processing capacity, in that it would require the agent to take account of the speed of an object when predicting how heavy it will be to pick up, for instance. The deviations between a relativistic and a Euclidean system are quite negligible at the human scale, and would offer almost no selection advantage whatsoever. Clearly, we would be highly likely to evolve towards any such cognitively cheaper, equally adaptive though less true system.

The external fitness of a system relates to how conducive to survival and reproductive success it is. One might consider the greater adaptive value of an over-sensitive strategy which produces plenty of false positives which have only minor negative implications for fitness but assiduously avoids potentially fatal false negatives. Such a system might well yield false beliefs more and true beliefs less, but still be favoured by natural selection, preferring 'reliability-when-it-counts-most' over 'overall-reliability'.

Nozick's account supplements Stich's arguments nicely. He sees the relationship between evolution and rationality as informing problems aired by Descartes and Hume. Hume's problem of induction addressed the impossibility of finding a rational (deductive) argument for why (inductive) reasoning works. Descartes questioned why self-evident propositions, as discerned by the natural light of reason, must correspond to reality.

After all, if reason and the facts are independent variables, why should they be correlated at all? Kant responded that since we cannot show why our reason would conform to objects, it must be that we perceive the objects to be the way they are because they are constructed by our faculties. In other words, our knowledge is not of things in themselves, but only of an empirical reality shaped by our constitution. Kant termed this upheaval a ‘Copernican revolution’, although confusingly, in contrast to Copernicus’ effect on astronomy, Kant is reaffirming an anthropomorphic perspective on the universe.

Nozick is overturning the Kantian dependence of the facts on reason. His ‘evolutionary hypothesis’ amounts to saying that it is reason that is the dependent variable, shaped by the facts. Our inferential system has evolved to become specialised for common past situations and stabilities in our environment. He suggests that there was selection for recognising as valid certain kinds of connections that are factual, which come to seem to us as more than just factual. Thus, the neural architecture for a given factual connection that appears regularly and stably in our environment may be modified over evolutionary time so that our descendants learn it faster.

Thus, “reason tells us about reality, because reality shapes reason, selecting for what seems ‘evident’”[33]. However, just because a certain factual connection has been consistent in the past and we have evolved to see it as a valid and increasingly self-evident basis for inference, does not guarantee that it will continue to hold in the future. Moreover, the question is not just

[33] Nozick (1993), ch 4.

whether the stable regularities of the past continue to hold in the future, but also whether evolution has picked out the ‘right’ regularities or given us ‘green’ in a ‘grue’ world. We may come to see the given sequence of thought as increasingly self-evidently certain (because we are selected to do so, because in a stable world such semi-automatic inference-making is adaptive), but this does not guarantee that it ever was strictly true.

Importantly then, he is not saying that it is the capacity to recognise independently existing valid rational connections that is selected for. Rationality can be seen as a biological adaptation with a function. It was never the function of rationality to justify certain of our most basic, stable, useful assumptions, because all we needed was to utilise them as trust-worthy, predictive regularities. These basic, sub-rational assumptions include the list of philosophical problems we’ve been least successful with: the problems of induction; of other minds; of the external world; and of justifying rationality.

We may still be able to sharpen our goals and procedures though, at least to some extent. Evolutionary theorising may help us understand what sort of rational system would be adaptive, and consequently why our rational system is the way it is. The fact that rationality wasn’t designed to justify itself or its framework assumptions does not mean that it can’t, or that we can’t turn our rationality upon itself. Rationality is self-conscious, in that it attempts to correct biases in the information it is supplied, and in its processes of reasoning. Nozick claims that “Whatever the initial functions of reasons were, we can use our ability to employ reasons to formulate new properties of reasons and to shape our utilisation of reasons to exhibit these properties. According to Nozick, we can, that is, modify and alter the functions of reasons, and hence of rationality.” After all, although psychological experiments show how often most people fail to reason well, for example about probability, the very fact that we have been able (through centuries of reflection) to formalise and so correct such faulty reasoning lends hope to improving upon these biologically-instilled assumptions, e.g. Euclidean geometry.

However, Nozick stresses that his evolutionary account of why we find certain thought processes rationally self-evident does not provide a reason-independent explanation of reason, since after all, "the evolutionary explanation itself is something we arrive at, in part, by the use of reason to support evolutionary theory in general and also this particular application of it … Hence, the account is not part of first philosophy; it is part of our current ongoing scientific view."[34] By conceding a lower epistemic status for the account, he avoids any circularity in his explanation.

However, although Nozick regards his account as a "proposal of a possible naturalistic explanation of the existence of reason that would, if it were true, make our reliance on reason 'objectively' reasonable", I think that Nagel is right to feel that "the idea that our rational capacity was the product of natural selection would render reasoning far less trustworthy than Nozick suggests". As he says, in order for his objectivist position to hold, "I have to be able to believe that the evolutionary explanation is consistent with the proposition that I follow the rules of logic because they are correct – not *merely* because I am biologically programmed to do so". Importantly though, this biological programming, though contingent in a sense, is extremely likely to align our thought along the 'fault lines' of the universe, and so it seems plausible that as the as complexity and generality of our mental representations increases, so can our trust in its objectivity.

Nozick's account, with its de-emphasis on objectivity, and its focus on a naturalistic explanation of reason, is an ideal example of the sort of reconciliation that can be sought if we

[34] Nozick (1993), pg 112.

suspend belief in the ontological thesis of rationality, and simply try and build up the most objective picture we can through science (which necessarily incorporates reason). This is one interpretation of what Quine meant by 'epistemology naturalised'[35].

We are now in a position to consider a theory that fulfils Nagel's basic requirements for an evolutionary account of our rational capacities:

1. a "general analysis" of rationality "into a limited set of functional elements"
2. considering "the relation between this set of capacities and the simpler habits of mind that might plausibly have carried selective advantage in the period when the human brain evolved"[36]

Pinker's[37] main aim is to show how an evolutionary account can explain modern empirical experiments showing the limitations and successes of people's reasoning, as well as considering the original adaptive role of the faculties we can use for science, maths, chess etc. He argues that we "[build] parochial inference models that exploit eons-old regularities in their own subject matters", e.g. "recognising objects, making tools, learning the local language, finding a mate, predicting an animal's movement, finding [our] way" – the "subject-specific intelligence of our species" that Tooby and Cosmides[38] call "ecological rationality". He argues in terms of both internal and external fitness (though not in those terms) that our brains have been shaped for fitness, not for truth.

Following Jackendoff, he considers these sentences (amongst many others):

[35] Quine (1969).

[36] Nagel (1997), ch 2.

[37] Pinker (1997), ch 5.

[38] Barkow, Tooby and Cosmides (1992).

The messenger *went from* Paris *to* Istanbul

The inheritance finally *went to* Fred

The light *went from* green *to* red

The meeting *went from* 3:00 *to* 4:00

Pinker argues that our entire linguistic faculty is based on concrete inferential machinery that has been co-opted to represent new, more abstract domains. The concepts of space and force "appear to be the vocabulary and syntax of mentalese, the [combinatorial] language of thought". He speculates whether if "ancestral circuits for reasoning about space and force were copied, the copy's connections to the eyes and muscles were severed and references to the physical world were bleached out", "the circuits could serve as a scaffolding whose slots are filled with symbols for more abstract concerns like states, possessions, ideas and desires". As evidence, he considers Premack's experiments on chimps which showed that they could pick out the object which plays a causal role linking before-and-after pictures. Also, "space and force metaphors have been reinvented time and again in dozens of language families across the globe". "Preschool children spontaneously coin their own metaphors in which space and motion symbolise possession, communication, time and causation" (Bowerman), e.g. 'Can I have any reading behind the dinner?'. Thus, our minds aren't really adapted to think about arbitrary abstract entities, so much as having inherited "a pad of forms that capture the key features of encounters among objects and forces, and the features of other consequential themes of the human condition such as fighting, food and health". We can adapt these inherited forms to more abstruse domains.

In a similar way, mathematician Sanders Mac Lane[39] followed Nagel's two-part prescription of breaking rationality up into functional elements and considering what selective advantages these modules might have had. Mac Lane "speculated that basic human activities were the inspiration for every branch of mathematics":

counting	→	arithmetic
measuring	→	real numbers, calculus, analysis
shaping	→	geometry, topology
forming	→	(as in architecture) symmetry, group theory
estimating	→	probability, measure theory, statistics
moving	→	mechanics, calculus, dynamics
calculating	→	algebra, numerical analysis
proving	→	logic
puzzling	→	combinatorics, number theory
grouping	→	set theory, combinatorics

But we may not be biologically designed for (and it would be surprising if we were) large number words, large sets, the base-10 system, fractions, multicolumn addition, carrying, multiplication/division, radicals and exponents. These skills develop slowly and unevenly, perhaps by applying the sense of number to things that at first feel like the wrong kind of subject matter, and by practicing (chunking and automaticity – fitting together over-learned routines).

[39] ibid.

If anything, Pinker's success at expounding a probably non-objectivist account while more or less adhering to Nagel's prescription highlights the difficulty for an objectivist of explaining how objective rationality would/could conceivably have evolved. Indeed, Nagel's position at the end of *The Last Word* seems somewhat analogous to his 'explanatory gap' position in philosophy of mind, where current naturalistic accounts are insufficient to ground rationality in the way that it requires, i.e. "simply in itself – that is, believing what it *tells* me, in virtue of the *content* of the arguments it delivers". Unfortunately though, I find his objectivist instincts about reason considerably less defensible and intuitively evident.

Degrees of rationality

Having shown that our reasoning capacities are easily subject to 'cognitive illusions', and that we appear to be employing 'rules of thumb' as handy but often non-rational short-cuts, we need to reformulate our conception of our rational capacities to incorporate degrees of rationality (which also allows us to make sense of our common sense knowledge that some people are better reasoners than others, and that our reasoning can improve over time).

Cherniak[40] provides one approach to considering degrees of rationality. In particular, he is attacking an idealised, (more or less) all-or-nothing conception of rationality, as is perhaps exemplified by Davidson[41] when he says that we need a "large degree of consistency" but actually for a more or less ideal consistency. However, as Cherniak point out, in reality, implicit inconsistency can be very difficult to unmask if the logical relations are convoluted, or if ideal rationality demands that we automatically notice some relation between different beliefs which may be compartmentalised distantly in our memory.

[40] Cherniak (1986), ch 1.

[41] Davidson, D. (1971), 'Psychology as philosophy' in Davidson (1980).

He is looking to explain why intentional explanations (the attribution of a cognitive systems of beliefs, desires, perceptions etc.) are so successful as a means of predicting and understanding others' behaviour. He wants to show that either too weak or too strong a conception of rationality is insufficient to explain the success of these intentional explanations, as well as being wholly inapplicable to human beings in the real world.

His 'minimal general conditions for rationality' have to lie between the too-weak 'assent theory of belief' and the too-strong 'ideal conditions of rationality'. The assent theory of belief considers that:

> *An agent believes "all and only those statements which he would affirm", i.e. that believing a proposition consists simply in having an accompanying "feeling of assent*[42]

Almost anything goes in such a caricatured theory, since it places no inherent consistency constraints, and no system by which inferences can be drawn from a given set of beliefs. As a result, it is quite unable to explain the predictive success of assuming intentionality in other people, since an agent is free to hold any beliefs he chooses – or at least, there is no systematic way of predicting, deducing or explaining which beliefs such an agent would have.

At the opposite end of the spectrum, Cherniak characterises the ideal general rationality criterion as:

> *An ideally-rational agent with a particular belief-desire set would:*
>
> > *make all of the sound inferences from his belief set*
> >
> > *eliminate all inconsistencies that arise in his belief-set*

[42] Cherniak (1986) ch 1.

> *undertake all actions which would, according to his beliefs, tend to satisfy his desires* (termed 'apparently appropriate actions')

This leaves no room for 'sloppiness'. Sloppiness in Cherniak's sense is almost a technical term, encompassing all of the factors which undermine our deductive ability. These include: laziness or carelessness; the difficulty of the deduction to be made (i.e. whether it is convoluted, indirect, requiring numerous unrelated-seeming premises); cognitive limitations (e.g. short-term memory); time constraints; and so fundamentally, the 'finitary predicament'. We have finite-sized brains, a finite time available to us, and so we are restricted in the number and range of inferences we can consider, let alone draw. The reason that these idealisations are made is that they allow us to simplify to a manageable level human behaviour sufficiently to formalise it in disciplines which deal with an enormous mass of human interactions, like economics, and to make cleaner philosophical distinctions.

Cherniak considers the Goldbach conjecture. We have a set of axioms, a conjectured inference, and yet we are unable to tell whether the inference follows deductively. Appeals to more prosaic cognitive limitations like short-term memory, carelessness or simply failing to take into account relevant premises by accident cannot explain our failure. In one sense, the problem is simply that the space of possible mathematical proofs is far too big for us to be able to search through it. But the space in which we operate on a daily basis when acting rationally is also far huger than we can possible search, as is the space of mathematical propositions that mathematicians somehow navigate through when inventing brilliant new proofs and mathematical domains.

We need a way of placing further constraints on the minimal rationality conditions that Cherniak suggests:

> *A minimally-rational agent with a particular belief-desire set would:*
>
> *make some, but not necessarily all of the sound inferences from his belief set*
>
> *eliminate some (but not necessarily all) inconsistencies that arise in his belief-set*
>
> *attempt some, but not necessarily all, of those actions which would, according to his beliefs, tend to satisfy his desires*
>
> *not attempt most (but not necessarily all) of the actions which are inappropriate given that belief-desire set* (termed the corresponding 'negative rationality' requirement)

This account is too sparse – it does not explain why some inferences are easier, sounder, more relevant and useful to draw than others, why we consider different beliefs to be more closely related (just as Hume noted through introspection with his catalogue of the relations of ideas[43]) etc. Cherniak elaborates a theory of human memory structure, which goes some way towards incorporating these features, but rather than go into this, I want to consider connectionism as a theory about the workings and physical implementation of mind, since I believe it can accommodate these feature of our experience more readily and powerfully than a high-level approach.

Moreover, it will allow us to consider in greater depth how a modern naturalistic account can shed light on the epistemological thesis of reason, i.e. on the successes and failures of our rational capacities, and so give us on idea of where the limits of our reason lie.

<u>Connectionism</u>

At root, connectionism amounts to the thesis that the brain is a dynamical system, like a mathematically modellable complex of levers and pulleys, or in this case, neurons and synapses. The high-level behaviour of the system seems to emerge like magic out of a morass of low-level interactions, just as the seemingly-centralised wheeling and coordination of a flock of birds results from each bird paying attention to purely local rules, e.g. the position and speed of its immediate neighbours. When connectionist systems are modelled on a computer, they are often termed ‘neural networks’.

More specifically, connectionism refers to the family of theories that aim to understand mental abilities in terms of formalised neuron-level models of the brain. These usually employ large numbers of nodes (neurons), with weighted inter-connections (synapses). The firing rate of a neuron is usually some non-linear function (e.g. sigmoid) of its activity, which is calculated as the weighted sum of the firing rates of neurons that synapse onto it. In this way, activity is propagated over time (milliseconds, in practice) in parallel from the input neurons eventually to the output neurons.

Input neurons are defined as those whose activation is (at least partially) determined by the external environment (in the case of the brain, various sensory receptors), and output neurons are those which affect some change in the system’s behaviour in that environment (e.g. motor neurons connected to muscle) – hidden neurons are those whose activity is invisible to the environment.

What makes neural networks interesting is their ability to self-organise, or ‘learn’, by modifying their weights according to a learning algorithm. The simplest are the Hebbian-type learning rules[44], which are based on the principle:

> the synapse between two neurons should be strengthened if the neurons fire simultaneously

[43] Hume, D. (1739), *A treatise of human nature,* (ed. Selby-Bigge), Clarendon Press.

[44] Hebb, D.O. (1949). *The organization of behavior,* Wiley.

This can be implemented in a pattern-associator, an architecture for associating a set of input patterns with a set of pre-specified output patterns. Innumerable improvements and revisions have been employed, and the Hebbian rule really only works well for orthogonal (i.e. uncorrelated) input patterns, but its human-like robustness and ability to generalise are notable. When presented with a novel pattern which is similar but not identical to a learned input pattern, its output will be similar or identical to the learned output pattern. It can be seen to generalise to new data, and form prototypes based on families of resemblance between input patterns, both of which features had to be explicitly, inelegantly and inefficiently built into previous symbolic models.

I want now to mention a second, stronger sense in which the term ‘connectionism’ is used as a thesis about the workings of the mind. The stronger claim, as espoused by Smolensky, can be stated negatively: a symbolic, cognitive-level description cannot fully capture (i.e. specify in law-like terms) our mental activity. That is, if we want to fully understand (i.e. account for or predict) the workings of the mind, we cannot talk at the level of psychology, but must (at least partially) descend towards the neural level. Smolensky maintains that a sub-symbolic level consisting of non-semantically evaluable constituents or micro-features of symbols exists, above the neural level, at which we will be able to fully specify (i.e. capture nomologically) mental activity.

If we reject this stronger thesis of connectionism, we are left with the (more or less incontrovertible) physiological evidence that the brain’s approximately 10^{11} neurons, linked by about 10^{14} synapses, form the substrate of computation, in the same way that the microchip is the physical substrate in a modern PC. In rejecting the stronger thesis of connectionism, we are demoting neurons to playing the generic role of a Universal Turing Machine[45] (a system that can, given enough time, emulate *any* machine whose behaviour is susceptible to being described algorithmically) implementing implementing the symbols and algorithms posited by psychologists and classical AI researchers.

However, I find Smolensky's view highly congenial – it seems implausible to me that the labyrinthine workings of the brain can be cleanly distilled down to a manageable number of discrete boxes (or 'modules', in Fodor's sense), each with an informationally-encapsulated, specific domain/function etc.[46] I will use this stronger sense of 'connectionism' from now on, since I consider it interesting, powerful and plausible, and only really open to a single extra objection, the systematicity objection.

The brain is a more or less analog system. It is a dynamical system operating in real time (as opposed to discrete time-steps), based on continuous variables like membrane voltage potential, synaptic weight strength etc. (although admittedly at the atomic level, the quantity of neurotransmitter at a given synapse is discrete, but this is a moot point). It seems intuitive that since the computations being performed by the system are analog, and the outputs also analog, that a neural system could not give discrete responses – at best, the system might respond with a very high tendency in one direction or another, but the neurons are not binary, and do not give 'true' or 'false' answers, only high or low firing rates. As a result, the sort of binary

[45] Turing, A., 'On Computable Numbers with an application to the Entscheidungsproblem', in Proc. London Math. Soc. (1936), Ser. 2, vol. 42

[46] Fodor, (1983). Fodor defines a 'module' in terms of nine features, of which I have mentioned two of the most important. The others are: mandatory; central systems have limited access to the representations computed by input systems; fast; informationally encapsulated; input systems have "shallow" outputs; associated with fixed neural architecture; exhibit characteristic and specific breakdown patterns; their ontological development exhibits a characteristic pace and sequencing.

formal logic that mathematicians, logicians and rationalists employ seems inappropriate for such a system. More fundamentally, it seems as though such a system could never be definite, in the way Nagel requires. If it were to turn out that our minds are inherently probabilistic, and could only consider a proposition to be 99.9% true, or infer the correct consequences of a belief most of the time, then reason's primary position as an ultimately trustworthy source of authority would be fundamentally, irrecoverably undermined.

Fortunately though, our rationality can not, I believe, be so easily undermined. The objection rests on a confusion between the neural and behavioural levels, that is, between the way that individual neurons operate and the way the overall, dynamical system that they comprise operates. A crucial aspect of a connectionist system's dynamics relates to its non-linearity. The most obvious source of this non-linearity is in the activation function relating neuronal activity (membrane voltage) with firing rate (or strictly, the temporal pattern of action potentials produced). As mentioned above, the activation of a neuron can be expressed more or less as the weighted (according to the strength of the synapse) sum of all its inputs. The firing rate is not, however, linearly proportional to this activity. A low activation may produce the occasional lonely action potential. However, as the activity increases, the firing rate will increase non-linearly, up to an asymptote, determined by the bare minimum 'absolutely refractory period' between action potentials that a neuron requires to "recharge", so to speak. This non-linear function could take many forms, such as binary, a threshold linear model, sigmoid or logarithmic[47]. All that matters is that it is not simply linear.

This non-linearity gives rise to peculiar dynamics at a high-level, i.e. ensembles of neurons collectively forming a distributed representation, which can begin to seem more and more

[47] Rolls (1999).

discrete. We can understand this intuitively if we consider that each neuron will be only slightly activated if its input neurons are not firing vigorously, and so will in its turn hardly fire at all. However, if its input neurons are firing rapidly, its output will be especially high. Consequently, at a high level, after numerous successive computations have been performed, a more-or-less binary output could easily result.

It should of course be noted that the real situation in the brain is considerably more complicated than has been outlined here. The brain makes use of graded and patterned firing rates, rather than simply treating the signal as a mean or 'rate' code over a short period of time, and so incorporating the possible informational content of temporal synchronisation, e.g. as employed in sound localisation. All consideration of inhibitory neurons, neurons with spontaneously high firing rates, the effects of random noise, and competing or inhibitory modules etc. has been stripped from the account to make the essential point that the brain can be considered to work in a discrete way at a high level.

Perhaps the broadest criticism of all such approaches stems from Godel's theorem, most famously advocated with relation to the mind-body problem by Lucas, and more recently, Penrose[48]. Godel's theory states that in a formal system of above a certain complexity, there will always be formally-undecidable, true propositions, i.e. statements that are true, but which cannot be proved within the system. This thwarted attempts like Russell and Whitehead's Principia Mathematica to found the whole of mathematics on a minimal set of principles (axioms). It also poses problems for connectionist systems. Part of the appeal of a connectionist system is that it can be seen as a Universal Turing Machine. Consequently

[48] Penrose (1995).

though, formally non-computable functions cannot be implemented finitely by such a system. Penrose argues that the brain (i.e. people) *can* do this, and so our minds must be more than Turing machines.

Of course, Penrose is himself a physicist and mathematician foremost, and so very much a believer in reason's objectivism. He proposes that there must be more going on in the brain than we're currently aware of at the sub-neural level - he speculates that there may be quantum effects in microtubules in the brain that allow us to perform non-computable functions, that allow us to see outside the system, where even a very fast machine (such as Deep Blue) would flounder and fail to make the meta-inference.

This is a tricky area to discuss, since Penrose's extensive proof would be the initial point of attack, but the mathematics are beyond the scope of this paper. However, it is worth noting that there is very little empirical evidence at all to support Penrose's substantive claims (about the quantum micro-tubules) and that the issue of human fallibility may complicate the picture of the brain as a normal formal system. If Penrose were to prove right, then almost all of the debate currently centring around the capabilities of purely connectionist systems becomes almost irrelevant, because the nature of such a quantum system would probably be unimaginably different and more powerful. If anything, I would be more tempted to ask what limitations such a hypothetical system would have, and whether our brains are actually much more limited-seeming than one would expect of such a system.

Tooby and Cosmides' notions of ecological rationality speculate that we are genetically hard-wired to be cognitively well-suited to certain domains of action. Nozick's stronger notion of chains of reasoning that have become automaticised to seem ‘self-evident’ requires our genes to be able to quite precisely specify neural representations for such ideas and behaviour.

Our current understanding of our genes’ influence is in terms an enormously complicated interaction between the genotype and the environment, which results in the eventual phenotype. That is, the way we are and the way our body becomes is an interaction of our genes and the experiences we have. In terms of neural development, this can be expressed in terms of internal (developmental) and external (learning) processes.

There is some debate about the degree of control our genes could have over low-level synaptic organisation, or whether in fact the neural constraints are very broad, determining only architectural or timing parameters perhaps[49]. This is ultimately an empirical issue, but one that is unlikely to be categorically settled for a considerable time. It seems implausible to me though that our DNA would code for such regularities as the assumption of other minds, or of an external world), but the possibility cannot be dismissed. This certainly makes things easier for Tooby & Cosmides, and for Nozick.

I want now to discuss a deeper concern: to what extent could a connectionist system be as general in its domain of applicability as Nagel’s rationality requires?

Computational models have demonstrated that simple logic gates (like AND or OR) can be easily simulated by neural networks. Indeed, much more complicated functions can be replicated too. However, these might be considered to be misleadingly simple cases, since the number of possible permutations is small enough to be contained inside the training set. The system can learn, like a finite state machine, a set of prescribed absolute responses for the given input patterns. This is clearly not an option for most problems. One of the major strengths of a connectionist system is that it can generalise. It forms prototypes from the data, and is able to gauge the similarity between given patterns. As a result, it is able to respond appropriately to novel patterns, and so degrade 'gracefully'. Connectionist systems, unlike the programs running on most desktop computers today, are robust. By this, I mean that unexpected, erroneous or corrupt data does not bring the system to its knees. If a neural network is fed damaged or incomplete data, it will settle into the closest attractor available, based on the weight organisation that has arisen from its training.

Reason's principles aim to "apply in all relevantly similar situations", and have "reasons of similar generality that tell us when situations are relevantly similar". This seems perhaps to require too much of a connectionist network. The problem can perhaps best be phrased as an empirical question: 'Is the data set to which our brains have been exposed sufficiently broad and representative for us to be able to reason reliably about the areas to which we apply it?' It requires an implausible stretch of the imagination to explain how our senses could provide the data by means of which we could learn to reason mathematically or logically.

This can be seen as another way of asking the same question that led Alfred Wallace (the lesser known co-discover of evolution) astray: "why would early man require a brain capable of playing chess and writing poetry?". Despite conceiving evolution in more or less the same way as Darwin, and at the same time, Wallace remained a creationist about intelligence because he considered modern man's intelligence to be superior to that of early *homo sapiens* (the savage languages "contain no words for abstract conceptions; the utter want of foresight of the savage man beyond his simplest necessities; his inability to combine, or to compare, or to reason on

[49] Elman et al. (1996).

any general subject that does not immediately appeal to his senses"[50]), and indeed to be far beyond what is necessary to sustain such a forager lifestyle. The fact that early and modern man are, at least phylogenetically (that is, as a species), more or less cognitively equals, can be explained in a number of ways. I have tried to cover the most important reasons *why* we might have evolved to be rational in the section on evolution, so in this connectionist section I am interested primarily in *how* it is that our physiology could be understood as implementing this rationality.

The main answer to both similar questions, of the extent to which a connectionist system could be as general in its domain of applicability as Nagel's rationality requires, and of how a connectionist system originally designed for a forager lifestyle could be capable of playing chess and reasoning formally is *plasticity*.

At this point, we have to remember a very obvious point: people's reasoning improves with time. This is partly through the basic genetic and developmental processes that govern our improved hand-eye coordination through youth, or puberty, for example. However, as evidenced by the effects of education, human cognitive capacities can be trained in certain directions, allowing us to build enormous pyramid-like conceptual toolkits. Maths is probably the most obvious example. To take a very basic example, we learn what the 'addition' operator means through continual, repetitive usage, practicing sums as small children. Over time, somehow, this process 'chunks' into a simple, atomic 'concept' or automaticised 'habit of thought' that we can use unthinkingly when trying to master more complicated concepts which build upon it, e.g. multiplication, or addition of complex numbers. Moroever, it may be the fact that when learning, for example, the addition operator, we use it in a growing multitude of

[50] Wallace, cited in Pinker (1997).

situations: perhaps first in simple sums, with larger numbers, in conjunction with other operators, with negative numbers and fractions, algebra etc. By employing it in various situations, we are viewing and growing the novel abstract space from a variety of different perspectives – it may be this that gives our mental representations so much power and abstraction[51].

What we are actually doing is building new, abstract spaces within which we become increasingly adept at operating[52]. We see this process going on every day – when we learn a new word, there is an acclimatisation period where it becomes necessary to reiterate the definition every time we encounter the word, but through usage and repeated encounters, it nestles into our vocabulary web. Variants of this process are going on when we learn new languages, mathematics, formal logic, analytic philosophical reasoning etc. Much more is going on than simply learning new words – we are creating new domains within which certain mental operations are easy or appropriate, just as it can be easier to express one idea in one language than another, or through an image rather than words. These domains piggyback upon and inter-weave with each other, and we shouldn't expect to see obvious delineations at the neural level. As we progress through education, even long beyond the point at which our brains are undergoing developmental (i.e. internally-prescribed) changes, our ability to reason improves. We are continually forming new conceptual spaces, and this improvement is incremental. This is related to the reason that maths, for instance, requires an element of trudging practice that cannot be avoided. An essential part of learning a new theory or technique is practicing it, repeatedly, with different problems. In this way, we are expanding our set of training data to be more representative of a given problem domain, and in the

[51] For a superb and persuasive exposition of this idea, see Minsky, 'Why people think that computers can't'.

[52] Plunkett, personal communication.

process expanding the generalisation ability of our reasoning. This is exactly what philosophers are doing during study and when reading each other's work – expanding their training data.

Nagel wonders how contingent, biological beings can have access to universally valid methods of objective thought. The answer that I am trying to build up is that finite, living beings access truths of infinite range in an incremental way, even down to a bacterial level. As representations become more abstract and complex, coupled with the connectionist learning algorithms that are exhibited by any creature with a nervous system, then the nature of even a contingent being's representations becomes more powerful. I see reason as being an extension of this with larger brains, only discontinuous insofar as we have language as a means of representing and communicating greater abstractions explicitly.

When we reason, or indeed form a sentence, we relate a series of symbols, whether at the level of morphemes, words, clauses or propositions, inter-changeably together according to certain rules, or morphology or syntax. According to Fodor and Pylyshyn[53], certain thoughts are intrinsically connected, that is, were a normal cognitive agent to lack some thoughts that cognitive agent would also lack certain other thoughts. They posit a semantically and syntactically combinatorial language of thought as an explanation. The notion of systematicity is often tied up with discussion of language, since syntax is the paradigmatic example. It makes sense intuitively to say that we would not understand what a noun is if we did not use adjectives and verbs, and that there is something common syntactically about all our usage of

[53] Fodor, J., and McLaughlin, B. (1990) 'Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work.' *Cognition*, **35**, 183-204. Also: Fodor, J., and Pylyshyn, Z. (1988). 'Connectionism and cognitive architecture: A critique.' *Cognition*, **28**, 3-71. Also in Macdonald and Macdonald (1995).

nouns that must be represented cognitively for us to be able to manipulate them in the syntactically systematic fashion that we do.

The fact that a connectionist model can be trained, for example, to recognize 'John loves Mary' without being able to recognize 'Mary loves John' implies that it has not formed the proper individual representations of 'John', 'Mary' and 'loves' that Fodor argues that we require in order for us to understand any sentences containing them.

Although in principle, this seems like quite a telling objection, it need not be. Smolensky proposes one solution. Effectively, it involves a distributed representation composed of pairs of neurons (or more likely, pairs of mini-ensembles). One of the pair specifies the content (e.g. the word), and the other specifies the role being played (i.e. the position in the sentence). A string of such pairs could thus specify both:

(loves, 2) (John, 1) (Mary, 3) = John loves Mary

or

(loves, 2) (John, 3) (Mary, 1) = Mary loves John

Admittedly, this solution is inelegant, probably impractical and inflexible, and biologically implausible, but it does neatly settle the central issue raised by Fodor, of how a formal syntax *could* be implemented in a distributed connectionist system. The fact that such a system can be devised also conveniently settles a second, less powerful objection known as the productivity objection, concerning our ability to produce an infinite number of grammatical sentences (by adding adjectives, subordinate clauses, conjunctions etc.), that I think can easily be met in this or other ways.

Fodor’s attack is most problematic for early ‘localist’ connectionist models, where each neuron represented a separate concept, e.g. a 26-node ensemble where each letter was signalled by its own neuron:

> ‘10000000000000000000000000’ = ‘A’
>
> ‘01000000000000000000000000’ = ‘B’
>
> etc.

or straw man models of the mental lexicon (the way in which words are stored in the brain) which hypothesised separate, modular lists of words categorised by part of speech, for instance. In contrast, the content of distributed models is represented as a vector, i.e. a collective ensemble of neurons, where no neuron on its own represents an identifiable, ‘semantically evaluable’ concept. This is what Smolensky is referring to when he talks of a sub-symbolic level.

The reason I raise this issue is because I see it as fundamental to any disussion of connectionist implementations of rationality. Logic provides the most obvious parallel to syntactic manipulation of language – all instances of modus ponens share a common syntactic structure, and we would not say that someone who accepted the infererence from P and (P → Q) to Q, but not from (P v Q) and ((P v Q) → R) to R, understood modus ponens. (Of course, given our empirical evidence about people’s ability to implement this with more complicated, embellished scenarios, we have to be lenient.)

Indeed, at a much higher level, it seems to me that justifying a belief with reasons rests on a similarly intrinsic connection between thoughts, and a means of combining and manipulating them systematically. As Nagel describes it, the aim of reason is to “be able to come up with reasons that apply in all relevantly similar situations, and to have reasons of similar generality that tell us when situations are relevantly similar". This is where a connectionist system's powers of self-organisation, learning and generalisation from self-generated prototypes seems to lend enormous support for its cause.

The debate continues to rage, centred around the issues raised by Fodor and Smolensky, over whether or not connectionist systems can display the combinatorial manipulations that both language and reason require. As the number of neurons involved increases, their learning algorithms become more sophisticated and their training environments more life-like and suitable for the task, the power of connectionist models is bound to increase substantially. However, Fodor's queries are important because they would still apply, if they held. Fortunately, it seems clear to me that they don't, at least not a priori. However, in order for connectionist models to be applicable as an explanation of how our brains work, they need to be biologically plausible, that is, they need to be set up to employ similar algorithms that the brain uses and to function under similar constraints (of accuracy, time, spatial organisation etc.). Currently, the most powerful connectionist models employ algorithms like back-propagation of error, which require information to be available immediately across an entire ensemble of neurons, and to travel backwards down a synapse, neither of which real neurons can do. This raises questions about whether more biologically-plausible models, employing only local learning algorithms like the Hebbian mentioned above (which only involve the pre- and post-synaptic neurons), can reproduce such success. Ultimately, this is an empirical question, but it seems reasonable to me to hope that continuing developments, aided by neuroscience, and greater computational power and complexity, can bridge the gap between current biologically-plausible models and what is required to understand how a connectionist system could demonstrate human-level rational capacity.

I contend that some variant on these claims will remain the dominant way of thinking about the mind and brain for the foreseeable future, and that this should inform our understanding of rationality in a number of ways. To some degree, adherence to this picture narrows down what we can be capable of as connectionist-implemented rationalists - most notably, it serves as a constant reminder of our finitude (see my discussion of Cherniak's 'minimal rationality').

It also emphasises the differences between ourselves and digital computers, e.g. in the way we search a space. Computers painstakingly plod through point by point, and can only speed their search through algorithms that prioritise certain areas. Connectionist systems 'settle' into an attractor based on their synaptic organisation, the inputs and current activities of their neurons. Because both processing and memory are both a function of representation, the system learns as it processes, and processes better and more efficiently as a result of its learning. This is one of the reasons that a long period of learning is a pre-requisite to expert and creative performance in any domain[54]. This system of self-organisation explains almost in a single sweep many of the relations between ideas that Hume and others have identified through introspection, and gives us a much richer picture of the sort of constraints acting on a minimally rational agent than could be devised pain-stakingly at a high level. Moreover, the system naturally finds the structure in its environment, adding further power to evolution's self-organising mechanism.

Finite beings and infinite thoughts

Even though we have rejected the idea of objective norms of reason by which we can steer, we still have to explain how as finite beings we are able to internalise ways of thinking with infinite range (e.g. counting) as part of our mental toolkit.

In a sense, I see this question as creating a problem where none exists. If we take a fairly simple example like a bird, it is evident that a finite being is responding in a more or less infinite range of behaviour to an infinite range of conditions in its environment, just by making an enormous number of continual, minute motor responses to the unceasing eddies, wind currents and fluctuations in the air around it in flight. I see the fact that there are an infinite number of causes, and an infinite number of effects, and the fact that every form of life can process and respond to this causal plenum of possibility as the root of the answer to Nagel's question.

I admit that it requires a leap of imagination, and ideally considerable empirical evidence, to see something like our use of syntax as a hugely complex extension to this theme. But after all, soon after we can understand a certain, finite, small number of words (perhaps only a few hundred), at the age of about three years old[55], we can already produce an almost infinite number of sentences, just by chaining adjectives together or using conjunctions to concatenate sentences endlessly. And despite Nagel's description of counting as entering into an "infinite mathematical landscape", it really amounts to a very simplified form of syntax. Language, logic and reason are presumably enormously more complex than the motor control of a bird's wings, and it is this difference of degree that gives rise to the apparent difference in kind.

I think that once this basic point is accepted, it will become apparent that reacting to (and so, in some way, forming a mental model that can cope with) the infinite with finite resources, is a quite tractable problem.

[54] Various studies have shown that even geniuses require at least ten years of dedicated study in their chosen field (whether music, sport, mathematics etc.) before they begin to produce world-class work..

[55] Altmann, G. (1997), chapter 4. By between two and three years old, infants have a production vocabulary of over 300 words, and are beginning to use basic syntax, i.e. word order to make semantic distinctions.

If one considers a fractal visually[56], it might seem that containing or reducing such an endlessly ramifying, beautiful, infinite spectacle is as much a mystery as 'how contingent, biological creatures such as ourselves' can explore and manipulate infinite logical space. But we know that a fractal can be wholly reduced to a single, simple equation. This is the power of symbols and abstraction, and actually in a slightly more direct way, the isomorphism between a connectionist representation in our brains and the world outside (via our senses), is analagous. As Nagel more or less says, the secret to reason must lie in its formality, in that it applies ways of thinking to different problems by seeing how they are relevantly similar. This is another way of describing the sort of association, pattern-matching, prototype-generation and robust generalisation that emerges out of connectionist systems in a quite ummysterious and formalisable manner. The question, 'how is it possible for finite beings like us to think infinite thoughts', if we strip away free will, language, the ability to produce arguments in a form that we can understand – can be answered really by considering that the infinite is ordered, contains (infinite) redundancy, and can be processed by a very, very simple finite device – just as a calculator can count too[57].

4 – Conclusions

Reason, seen in this way, is just another process (or better, the interaction between a number of neural ensembles), like all of matter and all of life. That our rational capacities are so well

[56] See http://www.softsource.com/softsource/fractal.html, or search for 'fractal images' at http://www.google.com.

[57] I see no reason to accept that calculators can't count because their intentionality is derived (e.g. Searle, J. (1980) Minds, brains, and programs, in *Behavioral and Brain Sciences*, 3 (3): 417-457) – see Hofstadter, D. (1979) on isomorphism), or because we can't empathetically interpret what it's doing in terms of our own capacities, as Nagel argues.

suited to our world, our mental representations so powerful, adaptable and reflexive, is the result of numerous nested levels of self-organisation: learning in a connectionist network, the interaction between genotype and environment, the chemical and biological interactions out of which our body and DNA result, and even lower.

I have succumbed to "the constant temptation towards reductionism – the explanation of reason in terms of something more fundamental", but I don't feel that I'm guilty of trying to reducing away the irreducible. Nagel's strictures on reduction are that "any reduction to something else must leave us with a more credible world picture than one that keeps them in, unreduced" and that no external view of a practice should collapse how it feels from the inside or make it mysterious. I don't feel that this explanation of how we, as finite beings, are able to function so effectively in our world, even without objective norms of reason out there to guide us, diminishes reason by reducing it.

The ontological thesis of reason and the ontological thesis of naturalism clash irreconcilably. The ontological thesis of reason is subject to a number of objections, which together make it less plausible than its rejection. Unless further arguments can redeem it, it makes sense to adopt a pragmatic stance within which to locate the epistemological thesis of reason. Our use of reason is integral to the scientific method, and I have tried to show that both a priori and in terms of a particular, contemporary naturalistic picture, can in turn be explained *by* science, at the very least as a means of producing useful beliefs that fit within the useful framework of logic. Reason's objectivity then too becomes an empirical question, which can only be answered if and when we encounter other rational minds with whom we can compare ourselves.

Bibliography

Altmann, G. (1997), *The Ascent of Babel*, Oxford University Press

Bermudez, J. L., 'Psychologism And Psychology: Thomas Nagel's *The Last Word'*, in *Inquiry* (1999); 42(3-4), pp 487-504

Barkow, Cosmides and Tooby ed. (1992), *The adapted mind: evolutionary psychology and the generation of culture*, Oxford University Press

Cherniak, C. (1986), *Minimal rationality*, MIT Press

Clark, A., 'Minimal rationalism' in *Mind* (Oct 1993), Vol. 102. 408

Cohen, L. J., 'Can human irrationality be experimentally demonstrated?', in *Behavioral and Brain Sciences* (1981)

Crane, T. and Mellor, D. H., 'There is no question of physicalism' in *Mind* (April 1990), vol. 99, 394

Churchland, P. S., 'Epistemology in the age of neuroscience', *The Journal of Philosophy* (1987)

Damasio, A. (1995), *Descartes' error*, Papermac

Dancy and Sosa, ed. (1992), *A companion to epistemology*, Blackwell reference

Davidson, D. (1980), *Essays on Actions and Events*, Oxford University Press

Davidson, D. (1984), 'On the very idea of a conceptual scheme', in *Inquiries into Truth and Representation* 2nd ed., Oxford University Press

Dennett, D., 'The case for Rorts' in Brandom, R., ed. (2000), *Rorty and his critics*, Blackwell

Elman et al. (1996), *Rethinking Innateness*, MIT Press

Fodor, J. (1983), *Modularity of mind*, MIT Press

Gowans, C. (2000), *Moral Disagreements*, Routledge

Hofstadter, D. (1979), *Gödel, Escher, Bach: an eternal golden braid*, Basic Books

Haack, S. (1993), *Evidence and enquiry*, Blackwell

Jaynes, J. (1993), *The origin of consciousness in the breakdown of the bicameral mind*, Penguin

Kornblith, H. (1994), *Naturalised epistemology*, 2nd edn., MIT Press

Macdonald and Macdonald, ed. (1995), *Connectionism*, Blackwell

Moore, A. W., in *Mind* (April 1999), vol 108, pp 382-394

Minsky, M., 'Why people think that computers can't', first published in *AI Magazine*, vol. 3 no. 4, Fall 1982. Reprinted in *The Computer Culture*, ed. Donnelly (1985), Associated Univ. Presses, Cranbury NJ

Minsky, M., 'Jokes and the Logic of the Cognitive Unconscious' in *Cognitive Constraints on Communication*, Vaina and Hintikka, ed. (1981), Reidel

Nagel, T. (1997), *The Last Word*, Oxford University Press

Nagel, T. (1986), *The View from Nowhere*, Oxford University Press

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

Papineau, D., 'The evolution of knowledge', in Carruthers and Chamberlain, ed. (2000), *Evolution and the human mind*, Cambridge University

Plantinga, A. (1993), *Warrant and Proper Function*, Oxford University Press

Penrose, R. (1995), *Shadows of the mind*, Oxford University Press

Pinker, S. (1997), *How the mind works*, Norton

Putnam, H. (1981), 'Why reason can't be naturalised' in *Realism and Reason: Philosophical papers vol 3* (pg 194)

Putnam, 'Two conceptions of rationality' in *Reason, Truth and History* (1981), Cambridge University Press

Putnam, H. (1990), *Realism with a human face*, Harvard University Press

Quine, W. V. O., (1969), 'Epistemology naturalized', in *Ontological relativity and other essays*, Columbia University Press

Raz, 'Explaining normativity: On rationality and the justification of reason' in *Engaging reason* (1999), Oxford University Press

Rolls, E. T. (1999), *The Brain and Emotion*, Oxford University Press

Rorty, R. (1999), *Philosophy and Social hope*, Penguin

Rorty, R., 'Putnam and the relativist menace' in *The Journal of Philosophy* (Sept 1993), Vol. XC, No. 9

Smolensky, P., 'On the proper treatment of connectionism' in *Behavioral and Brain Sciences* (1988), 11(1):1-74.

Stich, S., (1990), *The Fragmentation of Reason: preface to a pragmatic theory of cognitive evaluation*, MIT Press