Philosophica 57 (1996, 1) pp. 33-51

EVOLUTION AND SELF EVIDENCE

William S. Robinson

ABSTRACT

Robert Nozick (1993) has offered an evolutionary account of self evident beliefs that comes into conflict with a "mild realist" (Dennett, 1991a) view of beliefs. This paper summarizes both views, and explains the conflict. Emergence is examined. Mild realism is found to embrace "emergence" in an acceptable sense, and to eschew it in its problematic sense. Nozick's cases of self evident beliefs are examined and difficulties in his account are explained. An alternative approach is developed that avoids the difficulties in Nozick's account and is compatible with mild realism.

Robert Nozick (1993) has given an interesting evolutionary account of the development of self evident beliefs. In broad strokes, his idea is that long term constancy of environmental features permits evolution of structures that take advantage of those features, and beliefs that reflect those features are examples of such advantageous structures. For truly pervasive and constant features, quickness in arriving at a corresponding belief and stability of the belief in the face of superficial variety could confer increments in fitness. Self evidence of belief would promote quickness and stability. Thus, self evidence of belief could itself be selected for; and this may explain why we find certain matters to be self evident.

In examining this account of self evident beliefs, I shall begin by describing what Dennett (1991a) has called a "mild realist" view of the nature of beliefs in general. We shall find that this view is in conflict with a "strong realist" causal assumption about beliefs upon which Nozick's account of self evidence rests. This conflict by itself might, of course, be taken as a reason to doubt mild realism. However, after clarifying the relation of mild realism and "emergence" in part II, we shall

find, in part III, that there are reasons independent of mild realism to question Nozick's account. Discussion of these difficulties leads to the development, in part IV, of a positive understanding of an approach to evolutionary explanations that fits well with a mild realist stance.

I

The view of belief summarized here has been developed in Robinson (1988; 1986; 1990; 1995). On this view, beliefs are properties of (whole) subjects -- typically people, but also some other animals. This is to be contrasted with a common practice of identifying beliefs with brain states. The difficulties of this practice have been explained in detail in Robinson (1990); here, I will explain the point by reference to Dennett's (1991b) Multiple Drafts Model (MDM). At any one time, according to MDM, I have many representations, and even many complex representations that have a quasi-judgmental form. Not all of the quasi-judgmental representations are compatible and many of them are suppressed before making any contribution to behaviour. For both reasons, not all of our quasi-judgmental representations can be regarded as our *beliefs*. But this conclusion implies that our having a certain belief does not consist just in our having a quasi-judgmental representation. On the MDM, a subject S's having a belief that *p* involves at least S's having a quasi-judgmental representation that survives competition with other such representations for access to behavioral control. This view, however, makes having a belief to be a global fact about the whole organization of S's system of quasi-judgmental representations and behavioral control mechanisms.

Let us add to this the fact that many beliefs -- what we might call our nonperceptual beliefs -- extend over considerable periods of time. This fact is crucial for the predictive utility of belief attributions, of which much is made by, for example, Fodor (1987). The more frequently we postulate a change of nonperceptual beliefs, the less attribution of such beliefs can serve as a guide to expectations of future behaviour. Or, to put the point positively, predictive utility of belief attributions depends on nonperceptual beliefs' remaining stable over the time between prediction and action. Since attributions of nonperceptual beliefs are indeed useful, those beliefs do, in general, possess the required stability. Thus, in general, for nonperceptual beliefs, S's believing that *p* is a fact not only

about S's global organization at one time, but about the organization of S's representing and behavioral control mechanisms over a considerable period of time.

Attributing a belief with a particular content requires us to use a sentence to specify that content. This fact does not imply that subjects to which beliefs are attributed have language. What it does imply is that their actions exhibit a pattern that resembles a pattern of behaviour (in appropriate circumstances) that is indicated by the sentence that language users use in ascribing the belief. Thus, a cat can believe (roughly) that food is behind a cupboard door; it shows its belief by meowing by that cupboard or scratching at the door when it is hungry. We do have to be careful here, because any sentence we use will over-describe the cat's belief in some way -- the cat, after all, presumably lacks representations that correspond fully to our concepts of *food*, *cupboard*, *door*, etc. Nonetheless, the cat can exhibit a pattern of behaviour that is similar to the pattern we would exhibit if we believed that food lies behind the cupboard door and were hungry -- always allowing for the fact that the similarity must be adjusted to the behavioral repertoire of a cat.

When we attribute beliefs, we are attributing expected patterns of behaviour -- that is, patterns of behaviour that are expected under some circumstances.[1] Thus, there must *be* patterns in our behaviour. Some of these patterns will require language -- e.g., one cannot believe that mind supervenes on the physical unless one is capable of coherently discussing the concepts involved in this claim. But neither belief attributions nor language itself could develop unless there were *some* patterns in our behaviour that did not require language use. Thus, if we are to think of ourselves as subjects that are organized in such a way as to be able to have beliefs, we must first think of ourselves as subjects that are organized in such a way as to be able to have language-independent patterns of behaviour.

A further step now suggests itself. Language independent patterns of behaviour can hardly depend on linguistic organization. Thus, if we agree with the foregoing, we must hold that there is a level of brain organization that is non-linguistic in character, but capable of sustaining patterns of behaviour -- patterns of behaviour that may be attributed to subjects by the use of sentences, but that are explained by a pre-linguistic type of brain organization. When I say "non-" or "pre-linguistic" here, I do not mean nonrepresentational. There is abundant evidence that animals repre-

sent objects that are not present, and if a cat goes to the kitchen when it is hungry, or a dog to the door when it needs relief, the explanation of these patterns will undoubtedly involve representation of kitchens or doors (or some roughly corresponding feline or canine representations). What I do mean to emphasize is that the organization of whatever representations are necessary must be thought of on some principles *other* than those of linguistically organized concepts in sentence-like groupings. The latter correspond to the form of the *report* of what is believed, but cannot occur in the (causal) *explanation* of the pattern that is reported.

Some of our actions, of course, do depend on language. For example, in signing a contract, I agree to be bound by its words. Such high-level cases depend on a complex background. (a) As we have just seen, there must be patterns of behaviour that arise from pre-linguistically organized representations. (b) Language must be learned. This requires connecting words to (at least) things, actions, properties, changes, and facts. On pain of regress, some of this connecting must be produced by processes that, while involving representations, are not yet linguistic. Parsimony suggests that we think of these processes as variations (by small changes or small additions) on the same pre-linguistic organizing mechanisms that are necessary for the prelinguistic patterns of behaviour in (a). (c) Only after these first two stages are present can we suppose that linguistic organization can enter into the *explanation* of (as opposed to the reporting of) patterns in our behaviour. (d) Only after the language has become well developed do we have the basis for metalinguistic beliefs, or for the reporting of another's beliefs. These are closely related matters. In order to have a full concept of belief, one has to distinguish between the truth of *p* and the fact that S believes that *p*; and this means that one has to be able to understand and be able to work with what "*p*" says without asserting (for oneself) that *p*.[2]

We are now in a position to clarify the conflict, alluded to above, between mild realism and Nozick's account of self evidence. We often say that people did things *because* they believed *p* and desired that *q*. I have no quarrel with such statements, and in so far as they lead to correct expectations about people's further behaviour they are quite in order. I believe, however, that they are very superficial statements, and I have argued (Robinson, 1995), that we err if we mistake them as having any serious *causal* import. In short, beliefs are not causes of behaviour, they are patterns of behaviour. Causation is to be sought in the underlying

brain organization, which is an organization of representations, but not of beliefs. But if beliefs are not causes of behaviour, then they are not what natural selection is selecting for, nor could their self evidence be selected for. We can, of course, say that the principles of brain organization that are selected for are those that lead to quickly acquired and stable patterns of behaviour that are useful given longstanding and pervasive environmental facts. What remains is the question what this has to do with self evidence of belief.

II

Before turning to this question about self evidence, I want briefly to consider whether the account of belief just described makes beliefs "emergent". In one innocuous sense, the answer is affirmative. This sense is just that, on the foregoing account, beliefs are attributable only to whole persons. Whole persons are composites of body parts (notably including neurons and clusters of neurons) standing in a very complex set of relations. Beliefs are attributable only to composites thus organized and related to other things, and not to any of the parts of such composites.

Most accounts of "emergence", however, build into this concept considerably more than what we have just described.[3] Parts of trees are not generally trees, nor parts of walls walls; but trees and walls hardly qualify as emergents. What makes for emergence is *irreducibility*, that is, inability to be explained by the properties of whatever emergents are said to emerge *from*. To take a traditional case, if conscious experiences, e.g., pains, are said to *emerge* from neural firings, that would imply that the properties and relations of neurons and their activations cannot explain why it is that, given those properties and relations, pains should occur.

What is irreducible need not, as a matter of logic, have any causal role. But those who have made claims for emergence have generally thought that they had evidence for their emergents and, moreover, evidence in the form of events that would not have occurred but for the proposed emergents. That is, they have generally attributed causal roles to emergents. This attribution of causal roles, combined with irreducibility, gives us a sense of "emergence" that we may call "robust emergence". What we now need to see is how robust emergence leads to a difficult problem.

Both things having emergent properties and things on which these have effects are composed of parts that lack the emergent properties. Consider the nonemergent properties that are supposed to occur as effects of an emergent property. If the emergent property of the cause both makes a causal contribution and is not reducible to a complex of nonemergent properties of the cause, then the nonemergent properties of the effect cannot follow from the laws of nature applied to the nonemergent properties of the cause. This situation guarantees that the nonemergent effects will appear to be violations of the laws of nature. To illustrate, suppose pains are regarded as (irreducible) emergents, and held to have effects upon neurons that eventually lead to pain behaviour. Their irreducibility implies that we cannot explain why they occur, given their neural bases; their possession of causal properties implies that they are necessary to cause the neural events leading to pain behaviour. These two points imply (see Robinson, 1982) that the neural predecessors of pains will not be sufficient to explain why the neural events that lead to pain behaviour occur; these latter events will thus appear to violate the electrochemical laws that ought to apply to the chemicals and membranes of which neurons are composed and the solutions that surround them.

Meehl and Sellars (1956) showed that this situation need not involve any contradiction. Contradiction can be seen to be avoided by imagining a partitioning of the explanatory space. Within the domain where the relations required for emergents do not obtain, one set of laws governs items at the basis level (in the case of our example, these would be chemicals, solution strengths, membrane constructions, and so on). Within the domain where the relations that give rise to emergents do obtain, a different set of laws governing the basis level elements holds sway. So long as the domains are distinct, there is no contradiction.

The division of the laws of chemistry into those holding for the domain of conscious brains and those holding in nonconscious (or nonliving) structures is a radical solution that few have been willing to accept. Alternative views that embrace robust emergence often merely obscure the difficulty, or propose solutions that are scarcely less problematic. It is thus fortunate that the account of belief given in part I does *not* imply that beliefs are emergent in the robust sense. In particular, it does not imply that beliefs cannot be explained by neural architecture and activations. There is, of course, no *practical* possibility of doing this: patterns of behaviour are readily observable, their neural causes are not. But the

account I have given does assume that our patterns of circumstance-dependent behaviour do have causes that lie in our inputs and the construction of our neural systems, and it looks forward to learning more and more about the principles that enable the neural apparatus to maintain useful patterns of behaviour over long periods of time while conforming to the ordinary laws of organic chemistry.

III

Let us now return to the matter of explaining the self evidence of certain of our beliefs. We shall proceed by considering Nozick's examples. Let us begin with the longstanding (to say the least!) and pervasive fact that unsystematically acquired samples are apt to resemble the sampled population. A species whose internal organization allows this fact to influence its members' behaviour will have, *ceteris paribus*, an advantage over its rivals that are not similarly endowed. Thus, if a species has survived for some considerable time, it is extremely likely that successful action of an individual member of that species in a saliently distinctive set of circumstances will raise the probability of the same action by that individual in similar circumstances.[4] Repeated success will further increase the probability of repeating the action in similar circumstances and, perhaps, broaden the set of circumstances that will elicit that action.

It may very well occur to us to describe the individuals just imagined as "acting inductively". We must, however, be careful. What we have actually described is a capacity for operant conditioning; and this is evidently insufficient for attributing to an organism (which might be a cat or a chimp) a belief in the principle of induction. To hold such a belief one must be able to represent sets of circumstances in general, and actions in general, and form a representation of the very general fact that actions that are successful in a set of circumstances will likely be successful if repeated in the same or similar circumstances. Now, I do not wish to argue that it is impossible that such a general representation could occur outside of a system that would be properly called "linguistic". Perhaps a pattern of neural activation could come to be caused by any case in which several properties and a successful action have gone together more than once or twice, and perhaps this could contribute some beneficial effect, even in a nonlinguistic organism. But something *like* a

linguistic level of representation seems necessary. And even if we stretch our imaginations to allow nonlinguistic, general representations, we face the following problem: It looks as if the belief -- the general representation -- has nothing to do with behaviour and so, cannot be selected for. What is evolutionarily important is that we *act* inductively: our believing the principle of induction seems, so far as fitness goes, an afterthought. Not only is its self evidence not explained, its occurring as a belief at all is not explained by evolution. Let me spell this out.

We have seen that in order for there to be beliefs, there must be coherent, lasting patterns of behaviour that are to be explained by some (as yet not well understood) pre-linguistic mechanisms. However language is learned, there must be some mechanisms (as yet even less well understood) that adjust language to what it is about. It is, presumably, the operation of these mechanisms that accounts for the concordance between our belief in the principle of induction and our behaving inductively (i.e., repeating what has worked in similar circumstances). And, since mechanisms that support the development of language *have* developed, and thus presumably have had some evolutionary advantage, we can surmise that the mechanisms that adjust language to what it is about have been selected for. But, as far as I can see, the fact that we believe the principle of induction is only a consequence of these mechanisms. If so, we have a (very preliminary) sketch of an explanation for our believing the principle of induction; but this kind of explanation holds out no hope of explaining the self evidence of the principle of induction through selection for such a self evident belief.

I am, of course, not denying that there must *somehow* be an explanation of self evidence. I am arguing only that the appeal to natural selection operating on beliefs does not seem capable of providing it. Evolution gives us the mechanisms of language; what we need to see, but cannot yet see, is why these mechanisms yield such self evidence as they do, for certain claims.

It may be that induction is a bad example. Considered abstractly, its self evidence is perhaps not as strong as the self evidence of some other cases. We can imagine what it would be like for it to be false in general, and we can imagine particular cases, like Thanksgiving turkeys, where its application would lead to disaster. Nozick has other examples: Euclidean geometry and principles of deductive logic (p. 110), and belief in other minds and the external world (p. 121). Perhaps we will have

better success with some of these other cases.

Consider Euclidean geometry. We can certainly agree with Nozick that it might have been neurally costly and practically useless for earth-bound organisms to have learned to behave as if space were not exactly Euclidean. The following passage, however, goes far beyond this point of agreement.

'Given Euclidean geometry's close approximation to the truth, and given the attendant advantages of its seeming self-evidently true to us -- advantages including quickness of inference, believing serviceable (approximate) truths, and avoiding other more divergent falsehoods -- we can imagine Euclidean geometry's seeming self-evident as having been selected for; we can imagine selection for that geometry as our form of sensibility (110)'

The problem in this passage is that there is a conflation of perception and thought. The error involved in treating light as travelling in straight lines (even near massive bodies) will never make the difference between successful and unsuccessful fighting, fleeing, feeding or mating, and this does support the idea that evolution should not have made our perceptual apparatus represent anything other than an inert space describable by straight lines and Euclidean planes. But these considerations can be presumed to apply to the perceptual apparatus of animals who lack any beliefs, not to mention self evident ones, about the truth of geometrical propositions.[5] Thus, they do not explain self evidence of *belief* in any geometrical propositions. For an account of geometrical beliefs, I think we will need to add a theory that explains how our beliefs about geometry are constrained to match our pre-linguistically developed perceptual space. While we do not now have such a theory, I believe we have a significant clue, namely that such conformity is necessary if language is to be developed at all. Of course, in that case one can certainly say that mechanisms that produce the conformity are selected for: or better, that any language-like mechanisms that failed to produce the conformity would have been useless and selected against. But this concession still leaves us with a beginning of an explanation that is quite different from what Nozick seems to have in mind. Specifically, the sketch that I have suggested does not suppose that quickness of (conscious) inference or belief in approximate truths plays any causal role in the development of self evidence of geometrical truths. (If one tries to rescue Nozick here by supposing that he means to refer to unconscious inferences, one threatens

to trivialize the issue. For example, nonlinguistic animals will have to be counted as inferring conclusions from their geometrical beliefs.)

It may be objected that the line I am taking here is inconsistent with my description of beliefs as patterns of behaviour. This objection can be answered by paying careful attention to the contents of the beliefs that we have reason to attribute to nonlinguistic animals, and by distinguishing these from other belief contents. While there are many complications, I think it is acceptable to think of some nonlinguistic animals as believing that there is prey over there (even, in some cases, where "over there" is out of sight). Circumventing an obstacle to get to the prey may reveal that an animal has a Euclidean perceptual space, and even that it can make a kind of spatial inference. For example, Tolman and Honzik (1930) showed that rats that have learned three routes to food, each longer than the previous one, will proceed immediately to the third route when they encounter a blockage of the first route that occurs after the place where the second route joins it, whereas they will try the second route if the blockage in the first route occurs before the place where the routes join. In this case, we may well attribute beliefs about whether the second route is open or closed. But neither having beliefs of this kind nor being able to generate them is the same thing as believing that space is "flat", or that diagonals of a rectangle are equal, or even that, in general, blockages in a point common to two routes are always blockages of both routes.

Similar considerations apply to simple arithmetical beliefs and principles of deductive logic. A nonlinguistic animal, or a human infant, may see two attractive objects go into a box, and one come out. Exploration of the box, or staring at it, may convince us that the animal or infant believes that there is still an attractive object in the box. It seems correct to say that this cognitive capacity depends on representing objects, representing something numerical about them, and even that it depends on subtractive processing. For all that, it seems evident that there is not sufficient basis in what has been said for attributing a belief that $2 - 1 = 1$. Likewise, if every A in an organism's experience has had a B behind it, and it tries to get behind a new A when we have reason to think it wants a B, we *may* say that it has modus ponens "built into it". This is quite different, however, from attributing to it either a belief that if A and if A then B, then B, or the self evidence of such a belief.

Let us turn to a belief in other minds. This seems to me to raise two

distinct issues, (a) belief in others' conscious experiences and (b) belief in others' propositional attitudes, most notably, their beliefs and desires. Let us take these in turn.

The immediate difficulty about others' conscious experiences, from an evolutionary point of view, is that they are additional to the behaviour upon which evolutionary considerations might be thought to depend. To explain this, let us begin by noting that quick recognition of others' incipient behaviour can enhance fitness. People can better avoid attacks from others if they recognize when such attacks are likely. They are more likely to succeed in the hunt, or in battle, if they can anticipate what other members of their group are about to do. However, none of these things requires the belief that others have conscious experiences. Even if one adds a premise that conscious experiences are required in order to cause behaviour, others cannot perceive those causes, and so cannot use them in predicting others' behaviour. Thus, quite apart from the kinds of considerations we have already advanced in this paper, we ought to be suspicious of an evolutionary account of belief in others' conscious experiences.

Belief in others' beliefs requires a different commentary. Beliefs, I have held, are (certain) patterns in behaviour. If we can perceive such patterns, and use them to anticipate others' behaviour, we may increase our own fitness, and possession of this ability across a group may enhance group fitness. Thus, there is some reason to suppose that ability to recognize beliefs in others could have been selected for. We should note, however, that allowing this does not commit us to any causal role for beliefs. I have already suggested that attributing beliefs depends on several other levels of cognition. Attributing causal roles directly to beliefs tends to mask the complexity of the achievement of belief attribution. Thus, it would be better to say that what is selected for are the abilities that, once in place, permit the recognition of complex patterns in the behaviour of others.

I have pointed out that belief contents can be of various kinds and that having beliefs with some contents requires language while having beliefs with other contents does not. The case may be similar with attributions of beliefs to others. For example, some evidence suggests that (non-language using) chimpanzees can learn to follow the directions of an experimenter whom they have seen watching the placement of food (without themselves being able to see where the food was placed) and to

ignore the directions of an experimenter whose view of the food placing they have seen to have been obstructed (Povinelli, Nelson, and Boysen, 1990).[6] It may be that a supportable summary of this evidence is that (non-language using) chimpanzees can learn to form beliefs about the knowledge of experimenters (and, therefore, about some belief of an experimenter's). If so, language possession is not necessary to have a belief about a belief. Still, it may be that in order to have a belief about *some* beliefs, language possession is required. Consider, for example, beliefs that are about knowledge in general (as opposed to beliefs that are about some particular subject's knowledge of some particular fact). In this case, it would seem that the behavioral pattern that constitutes the (general) belief includes some linguistic performance (e.g., a disposition to say that what is known must be justifiable, or true, or believed). Because of this inclusion, the ability to have a belief that is about this (general) belief would also seem to require the ability to recognize linguistic patterns.

Let us return to the question of self evidence of beliefs about other minds. Why am I so sure, to the point of not being able to take the denial seriously, that you have conscious experiences? The reason for the belief has often been represented as involving an inference from similarity of behaviour in my own and others' cases. As is well known, difficulties have been found in this procedure. Elsewhere (Robinson, forthcoming) I have argued for a different approach: we believe others have conscious experiences because we believe they are made very much like ourselves. From this point of view, similarity of behaviour is only confirmatory evidence: the primary evidence is that we look similar, feel similar (i.e., another's body feels much the same to my touch as does my own body), eat the same things, bleed when cut, and so on. Our belief in others' conscious experiences is, on this account, a conclusion from the principle that like causes have like effects. The propensity to treat any individual other person (or mammal) we may encounter as sentient can be regarded as a consequence of a built in -- i.e., selected for -- propensity to treat things that are much alike in many respects as alike in some further respect.

The self evidence of our belief in other people's having beliefs could be taken as merely the result of a very well confirmed induction. That is, by the time the philosophical question is encountered, most people have years of observation of other people behind them, and so are in a good

position to have noticed patterns in others' behaviour and to have come to be able to form generally satisfiable expectations on the basis of partial observation of such patterns. The inductive character of belief in others' beliefs is further supported by the general point just outlined, that is, the point that we have an immense amount of evidence for the likeness of others to ourselves. Once we have mastered the concept of belief as applied to ourselves, we immediately have a further reason for attribution of beliefs to others, i.e., they are extremely similar to us.

We have been sketching explanations for our beliefs about other minds. The conclusion I draw from our discussion is that the explanation for beliefs about beliefs does not require an evolutionary account that invokes a special mechanism for such beliefs, and the explanation for conscious experiences positively resists that kind of account. Of course, our belief in other minds and its self evidence are consequences of *some* brain mechanisms, and these developed under selectional pressures; but this fact does not suggest either causal roles for beliefs or the kind of account of self evidence that Nozick offers.

Let us turn, finally, to the self evidence of our belief in an external world. What we should say here depends on just what we think this belief is. We might suppose that it is the belief that the world is mind independent. It seems very unpromising to try an evolutionary account of such a belief, as it depends on possessing the concept of mind, and thus must have arisen quite late in human development. However, when Nozick discusses this belief, he gives a version that any worthy Berkelian could easily accommodate:

[T]hose cousins of our ancestors who could not manage to learn that there was an independently existing "external world," one whose objects continued on trajectories or in place even when unobserved, did not fare as well as those who quickly recognized obdurate realities. (121)

There is something right about this; but what is right surely applies equally well to nonlinguistic animals. They too need to track prey, predators, or potential mates when they disappear behind rocks or into the tall grass. Shall we then say that they believe in an external world? I believe we *could* say that without too much risk of being misleading; but without further commentary it seems too general, or too sophisticated, to attribute to most nonlinguistic animals. If one doesn't think one's cat has *considered* the question whether there is an external world (external to what, exactly?) it is decidedly odd to think that it believes, or finds self evident,

an answer to it. (On the other hand, it seems quite in order to say that the cat that chases a mouse up to a hole, and then positions itself motionlessly facing the hole, believes there is (or may be?) a mouse in there.)

The conclusions that this discussion of cases leads me to are these. We need to distinguish between the embodiment of certain detection skills and habits of expectation, on the one hand, and belief in general propositions that sum up many cases of such habits of expectation, on the other. Natural selection may have a pretty direct connection to the skills and the habits, but it has no such direct connection to general beliefs. General beliefs seem to depend on language. Of course, the mechanisms of language had to be developed over evolutionary time, and can be presumed to have been selected for. It is a further problem to account for the meshing of our general beliefs and our habits of expectation; but it seems likely that this account will be a special case under the general need for language to mesh with nonlinguistic reality in order to prove useful. The picture I have sketched does not intend to place development of the beliefs we have considered wholly outside the realm of matters to which evolution is relevant, nor does it do so. But it does make the relation of evolution to belief in fundamental principles considerably less direct than Nozick's descriptions suggest. It does not require beliefs to have the kind of causal role that selectionist mechanism demands. It promotes an interest in genuine explanations by pointing us toward (a) mechanisms that we share with animals in dealing with the world and (b) mechanisms that adjust all language to realities. Attention to the levels at which such mechanisms operate will, I believe, produce, in the end, an account that is much more explanatorily satisfying than the kind of account that Nozick suggests.

IV

The approach taken here favours explanation by a plurality of mechanisms that are individually not very rich in their consequences but can be deployed in many contexts, in contrast with explanations that invoke relatively specific propensities toward highly complex behaviour. One reason for this preference concerns explanatory depth or elegance: whenever we *can* give an explanation in terms of a plurality of less rich but more widely deployable mechanisms, the more powerful but more specif-

ic propensity becomes explanatorily redundant. That is, it becomes explainable by reference to mechanisms that we need to account for other things, and thus must in any case retain in our whole theory. A further consideration is that satisfaction with a richer, more specific explanation may cause us to be less assiduous than we should be in seeking out more fundamental explanations.

A case that can be used to illustrate these points is provided by Nozick's treatment of desire for accumulation of wealth. I should say at the outset that he offers this treatment only as a *possible* account. I too think it is a possible account, so strictly speaking we are in agreement about this case. Nonetheless, it serves well as an illustration of the difference of approaches to which I want to call attention.

Briefly, the possibility that Nozick explores is that we may have a strong desire for accumulation of wealth because of the following three facts. (a) Except in industrialized societies in the last 150 years, the wealthy have tended to leave more offspring than their less wealthy contemporaries (presumably because of better care, access to food and, often, more wives). (b) People with a strong desire for wealth tend to be wealthier than those who lack it. (c) A genetic disposition toward a strong desire for wealth is heritable. On these assumptions, strong desire for wealth could have been selected for.

As I have said, such an account seems possible. Nozick himself suggests an alternative, namely, "a social explanation in terms of the institutions that shape people's psychological concerns and motivations and the way particular motivations aid the functioning and propagation of those institutions" (126). However, the approach I have taken in this paper suggests that a third kind of alternative should be explored and found unworkable before accepting either of Nozick's proposals. This kind of alternative seeks to derive the strong interest in wealth from more basic interests in acquiring necessities of life, and the benefits of planning and foresight. It seems that we will in any case need to hold that these desires and abilities were selected for; and it seems likely that they might explain the desire for wealth, because foresight would connect acquisition now with ability to secure necessities of life in the future. If this explanation is successful, we would not need to suppose that an additional desire for wealth must be postulated as a heritable disposition.

It is possible to lose track of the issue here by supposing that the desire for wealth can be decomposed into an ensemble of abilities such

as the ones just indicated. It may then seem a merely verbal point whether we say that a disposition is what is selected for, or that an ensemble of abilities is selected for. We can reinstate a sense of distinction here by observing that there is a clear difference between the following two scenarios. (1) Each member of an ensemble of desires and abilities is independently selected for, so that the persistence of the ensemble follows from the evolutionary pressures to maintain each of its components. (2) The persistence of the components of the ensemble is explainable only through the fact that they confer advantage only when they occur together.[7] This second situation is indeed possible, and would support a claim that the ensemble (or its corresponding disposition) is what is being selected for. The point I want to make here, however, is that we are entitled to (2) only when we have tried to supply an account along the lines of (1) and have found such an account to be unavailable.

It is worth emphasizing that in arguing for a preference for accounts like those in (1) I am not supposing that they must always be available. Nor should it be thought that am I supposing that the mechanisms in accounts like (1) are "general purpose" mechanisms. Suitability for solving problems of whatever sort is not the same thing as suitability for solving a certain specific problem that recurs in many different contexts. A hypothetical illustration may be helpful, and can be gleaned from Cosmides and Tooby (1994, p. 331). They suggest a possible preference for participation in coalitions. If we think of such a preference as depending on a single mechanism, or an ensemble of mechanisms that individually have no other useful function, we must imagine a relatively large felicitous mutation. If we keep (1) in mind, we will be more likely to look for components that might exist outside a political context. E.g., perhaps one thing necessary for acceptability of coalitions is a degree of trust that, under certain conditions, somewhat exceeds what is warranted by one's actual evidence. A mechanism that produces such trust is perhaps useful in hunting parties, food storage systems, and many other cooperative activities. Such a mechanism would be specialized, and not a general problem solver, but could be a component of the evolutionary solution to many problems. Development of such a limited mechanism would seem to require less presumption of a large, felicitous mutation than development of a preference for participation in coalitions. This would make explanation by such a limited mechanism preferable, provided that it can be found.

There are two remarks that should be made before concluding this section. The first is a recognition that the approach I have just been recommending poses a problem for which I have no solution. If we have a number of specialized mechanisms for relatively small tasks, and these contribute to the accomplishment of many larger tasks, we must suppose that these mechanisms will be accessed on appropriate occasions and that, over the course of evolutionary time or individual lifetimes, their outputs can become connected to new behavioral outputs. Here, there is an element of generality; that is, theories of how appropriate accessing and development is brought about will not be explanatory unless they apply to many mechanisms and tasks. Unfortunately, no such theory is available. However, this problem of appropriate connection of mechanisms is one that arises for a wide class of views. For example, even if our models contain mechanisms that have rich consequences but work only on special problems, we must face the fact that many situations will present several special problems. Thus, we will have to have something in our theory to handle integration of the outputs of several mechanisms, or selective temporary suppression of some of them. It is to be hoped that clear understanding of this problem of appropriate connection, and recognition of its pervasiveness, will stimulate efforts to solve it.

The second remark concerns the resolution of the conflict described in part I, i.e., the conflict between mild realism and Nozick's account of self evidence. Mild realism emphasizes the global nature of the patterns that constitute belief and, by implication, deprives beliefs of a causal role that is often attributed to them. The explanation of the organization of behaviour that constitutes believing is to be sought, according to mild realism, not in the effects of beliefs, but in relatively local facts about brain events and principles of their interconnection.[8] In the last two sections of this paper there have been several occasions on which we have discussed beliefs without attributing causal roles to them. We have also found some difficulties in Nozick's account. Finally, we have seen reasons in support a view that once again directs our explanatory search in the direction of relatively simple mechanisms. Taken together, these latter developments seem to me to support a resolution of the conflict described in part I that favours mild realism.

Iowa State University

NOTES

1. A whole literature on the difficulties of intentionality arises from the simple observation that some of the circumstances are desires and other beliefs. I have discussed these matters in Robinson (1988, 1986, and 1990).
2. I offer these distinctions only as necessary, not as exhaustive of our levels of organization.
3. See Kim (1993) for conceptual analysis of and historical references on emergence. The account given in the text follows this source, albeit in a very summary fashion.
4. This remark hides a deep problem about what counts as relevant similarity. I must forego discussion of this problem here.
5. It is interesting to speculate on possible differences of perceptual space for echolocating bats, or for certain birds of prey (on this last see Akins, 1993).
6. Three of the four chimpanzees used in this experiment had been previously exposed to various experiments related to learning and cognition, but not to experiments involving linguistic abilities. The fourth subject had been a subject in studies involving linguistic abilities. Interestingly, this subject, possibly due to relatively advanced age, was the *least* competent of the four subjects in the task described in the text. Thus, even in the light of the possible linguistic involvement of this fourth subject, the view expressed here, that this experiment may show an ability available to *non-language using* animals, is well justified.
7. Such components might still be empirically separable by double dissociation under experimental conditions.
8. Perhaps I should say that this implication holds on my own view of mild realism, and not on Dennett's -- although I think there are difficulties in his position. See Robinson (1995) for discussion.

REFERENCES

Akins K. A. (1993), "A Bat without Qualities?", in Davies, M. and Humphreys, G. W., eds., *Consciousness*. Oxford: Blackwell.

Cosmides L. and Tooby J. (1994), "Better than Rational: Evolutionary Psychology and the Invisible Hand", *American Economic Association Papers and Proceedings*, pp. 327-332.

Dennett D. (1991a), "Real Patterns", *The Journal of Philosophy* **88**, pp. 27-51.

Dennett D. (1991b), *Consciousness Explained* Boston: Little, Brown & Co.

Fodor J. (1987), *Psychosemantics* Cambridge, MA: MIT Press.

Kim J. (1993), *Supervenience and Mind* Cambridge: Cambridge University Press.

Meehl P. and Sellars W. (1956), "The Concept of Emergence", in Feigl, H. and Scriven, M., eds., *Minnesota Studies in the Philosophy of Science, vol. I* Minneapolis: University of Minnesota Press.

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

Povinelli D. J., Nelson K. E., and Boysen, S. T. (1990), "Inferences about Guessing and Knowing by Chimpanzees (*Pan Troglodytes*), *Journal of Comparative Psychology* **104**, pp. 203-210.

Robinson W. (1982), "Causation, Sensations, and Knowledge", *Mind* **91**, pp. 524-540.

Robinson W. (1986), "Ascription, Intentionality and Understanding", *The Monist* **69**, pp. 584-597.

Robinson W. (1988), *Brains and People* Philadelphia: Temple University Press.

Robinson W. (1990), "States and Beliefs", *Mind* **99**, pp. 33-51.

Robinson W. (1995), "Mild Realism, Causation, and Folk Psychology", *Philosophical Psychology* **8**, pp. 167-187.

Robinson W. (forthcoming) "Some Nonhuman Animals Can Have Pain in a Morally Relevant Sense", *Biology and Philosophy*.

Tolman E. C. and Honzik C. H. (1930), "Introduction and Removal of Reward and Maze Performance in Rats", *University of California Publications in Psychology* **4**, pp. 257-75.