The Explanatory Role of the Notion of Representation

Michael Strevens

Draft of November 2004

Abstract

This paper investigates the explanatory role, in scientific theories, of the notion of representation and other kindred notions, such as the notions of information and coding. I develop a simple account of the use of the notion of a representation, or more specifically the notion of an *inner map*, in a class of explanations of animals' navigational capacities. I then show that the account, though simple, is more widely applicable than it seems, and I pose the question whether the same simple account could make sense of the use of representational notions in other kinds of scientific explanation, such as the explanations of neuroscience, genetics, developmental biology, cognitive psychology, and so on. I extend the account of the explanation of capacities to the explanation of behavior, and conclude by discussing representations of goals.

Contents

1. Representations in Explanations

More than ever before, *representation* and kindred notions, such as *information* and *coding*, play a key explanatory role in scientific theories—in cognitive psychology, neuroscience, genetics, developmental biology, cognitive ethology, and many other areas. Philosophers of science have had little to say, however, about the nature of this explanatory role. What does the notion of representation contribute to an explanation, if anything? Does it always play the same explanatory role? How is this role, or are these roles, related to the major theories of scientific explanation, such as the causal and unification accounts? To answer these questions is a major project, worthy of a book or more; my aim in this paper is to make a beginning, at least, to the endeavor.

What I propose to do is to elucidate the explanatory role of representation in one particular kind of explanation, the explanation of certain capacities of animals which might loosely be called *navigational capacities*. I will call my account of the function of representations in these explanations the *R-C model of explanation*, and the explanations themselves *R-C explanations*. The R-C model is developed in section 2; the rest of the paper then examines the model's scope. First, in section 3, I show that the R-C model is more flexible than might be supposed within the realm in which it was developed, the explanation of navigational capacities. Second, in section 4, I propose that the R-C model will likely have considerable relevance to explanations that invoke representation outside the realm of animal navigation. In support of this proposition, I discuss the use of the R-C model to explain behavior, as opposed to capacities, in section 5, and I give an account of the explanatory role of representations of goals, as opposed to maps, in section 6.

Although the explanatory role of representation has not been a major topic within general philosophy of science, it has, of course, been much discussed in the philosophy of mind. Three research projects, in particular, seem relevant:

1. The first, and best known, project concerns not explanation so much as causation, and in particular, the question whether the semantic prop-

erties of representations are causally efficacious. This issue is not one that I will investigate here.

2. A more immediately relevant project is Robert Cummins's account of the explanatory import of representations (Cummins 1983, 1989). I discuss Cummins's work in section 4, where I conclude that his project is complementary to, rather than competing with, my own work.
3. Finally, certain work on naturalistic theories of mental representation, in particular teleological theories, bears on the role of representation in evolutionary and psychological explanation (Millikan 1984; Dretske 1988). I will mention this work where appropriate.

The problem of the explanatory role of representation also appears in other literatures connected to the philosophy of science, such as the recent literature on the question of whether it is apt to say that genes *encode a program* for human development. I will not, however, prolong this paper by discussing these literatures here.

Some preliminaries: the representations discussed in sections 2 and 3 are intuitively representations that some state of affairs—such as a pool of water's being in a particular place—obtains. I call such representations, as have some other writers, *indicative representations*. In an R-C explanation, what is important about an indicative representation is its corresponding state of affairs. I do not assume that the state of affairs exhausts the content of the representation, but it is a consequence of my view that any additional content is, at least in an R-C explanation, explanatorily irrelevant.

Another class of representations might be called *imperative representations*. These are representations of goals and such like. Intuitively, an imperative representation is not supposed to indicate that a state of affairs obtains, but that it in some sense ought to obtain. I will discuss the explanatory role of imperative representations in section 6. As in the indicative case, what is explanatorily important about an imperative representation, on the R-C model, is its corresponding state of affairs.

Finally, some terminology: call notions such as *representation*, *information*, *coding*, and so on *informatic notions*, and any explanation that appeals to such notions an *informatic explanation*.

2. The R-C Model of Explanation

The informatic explanations that I consider in this section are all explanations of animals' navigational capacities that appeal to the animals' *inner maps* of the terrain that they must negotiate in their search for food and other goods. Some of these inner maps are very simple, representing only the direction and distance of the animal's home; others are closer to the paper maps that we ourselves use to get around.

Explanations of this sort are not limited to navigation in the literal sense. A map—a representation of obstacles, pathways, and interesting destinations—may be just as useful for getting around the ecological or social terrain as for getting around the geographical terrain. But navigation in the physical sense will be my focus in what follows.

2.1 Explaining Navigational Capacities

My principal source for informatic explanations of animals' navigational capacities is Gallistel's recent survey of the subject (Gallistel 1990). Gallistel begins his discussion by describing the homing abilities of the foraging desert ant *Cataglyphis bicolor*, which, after tracing a complex tangle of looping paths on its search for food, is able to turn and head straight back to its nest. Experiments show that the ant does not rely on any landmark to find its way home. Rather, the experimenters concluded, the ant's navigational capacity is explained by the fact that the ant has a representation of the direction and distance of its nest that it updates constantly as it moves around the desert. When the time comes, the ant simply turns in the represented direction and moves for the represented distance.

In the rest of his book, Gallistel surveys a large number of cases, involving the navigational capacities of birds and bees, locusts and toads, rats, bats, and fish. The explanations of these capacities appeal both to representations of directions and distances relative to organisms themselves, and to maps that are less "egocentric", representing distances between objects other than the creature itself. Some migratory birds represent the configurations of the constellations in the night sky, using these celestial maps to find their way on their long flights (see section 3.2), while bees appear to use internal maps of local landmarks to orient themselves. Gallistel concludes his discussion of navigational capacities by writing that

> The findings reviewed . . . suggest that the more advanced mobile animals represent the shape of the environment in which they move around and their moment-to-moment location within that shape (p. 173).

What is the general form of these informatic explanations? They posit two kinds of psychological mechanism connecting representations to the world. The first kind of mechanism, which in these simple cases might be called the *detection mechanism*, is responsible for updating an animal's map of its surroundings. In the desert ant, for example, the detection mechanism updates the ant's record of the current direction and distance of its nest as the ant moves around the desert.

The second kind of mechanism, which might be called the *prosecution mechanism*, ensures that, at the right time, the representation gives rise to appropriate behavior, appropriate, that is, for realizing the capacity to be explained. In the ant, the prosecution mechanism is activated by whatever event—for example, finding food—triggers the ant's return home. The mechanism orients the ant in the represented direction and sets it in on its way for the represented distance. Provided that the detection mechanism has done its job, so that the representations are accurate, this behavior will realize the ant's homing capacity.

Two comments: first, obviously the detection and prosecution mechanisms will work in concert with other more general-purpose mechanisms in the ant, such as the walking mechanism. Second, detection and prosecution mechanisms may—and in the case of the ant, Gallistel argues, almost certainly do—make use of intermediate representations (see also section 6).

All the explanations surveyed by Gallistel appeal to detection and prosecution mechanisms. In some sense, it could not be otherwise, for without either mechanism, the representations involved could not contribute to an animal's navigational capacities. Action without accurate representation is blind; accurate representation without appropriate action is useless.

In the simplest explanations of animal navigation, then, there are three components at work: a detection mechanism, a representing mechanism, such as a memory circuit, whose state is set by the detection mechanism, and a prosecution mechanism. My next task is to spell out the properties of each of these mechanisms, in virtue of which they play their explanatory roles.

The representing mechanism does not need to do much in order to play its role. It has simply to register particular states, a task that can in principle be performed by any physical object that has the requisite number of physical states. For example, if distance to a nest is to be represented, then all that is required is that the representing mechanism have a physical state for each distinct value of the distance that it is required to represent. The representing mechanism could be a rock whose temperature corresponds to the distance, a piece of paper on which the distance is inscribed, or, as in the case of the ant, a piece of neural circuitry with the right number of neural states.

Note that it is sometimes useful to distinguish, as I have above, between the representing mechanism, which has one physical state for every state of affairs that may be represented, and a representation, which is one of those physical states, and thus corresponds to a particular state of affairs. A dial is a representing mechanism; the settings of the dial are representations.

Sections 2.2 and 2.3 will be concerned with the explanatory roles of the detection and prosecution mechanisms in Gallistel's explanations. It is the description of these roles that provides the greater part of the R-C model of explanation presented in section 2.4.

Very briefly, the detection and prosecution mechanisms have the following explanatory roles. The role of the detection mechanism, according to an R-C explanation, is to ensure that the representation is accurate. If a certain representation *R* represents a certain state of affairs *S*—say, the state of affairs of an ant's nest being at an angle of 120 degrees to the sun—then the role of the detection mechanism is to make sure that the ant's representing mechanism is in state R if and only if the nest is indeed at a 120 degree angle to the sun. At this initial stage, accurate representation can be identified with covariation: R accurately represents S to the extent that R is tokened (the representing mechanism is in state R) just in case S obtains. For further discussion, see section 3.1.

The role of the prosecution mechanism, according to an R-C explanation, is to exploit the covariation of representation and world so as to realize the capacity that is to be explained. Just what this means is the subject of the following two sections, in which I consider both covariation and exploitation more carefully. Roughly, however, a good prosecution mechanism should execute, when a representing mechanism is in state R, an action that will help to realize the capacity to be explained when the world is in state S. The prosecution mechanism takes advantage of the fact that the animal "knows" that S obtains to prompt the animal to do something that is especially appropriate when S obtains, that is, which is especially helpful in realizing the capacity to be explained when S obtains.

2.2 The Role of the Detection Mechanism: Establishing Covariation

A mental representation that covaries with the state of affairs it represents is a reliable representation. It "tells the truth". I will first define a particularly strong kind of covariation, *absolute covariation*, then a weaker, probabilistic form.

A representation R *absolutely covaries* with a state of affairs S, just in case R is tokened in all and only those cases where S obtains.[1] A representation that absolutely covaries with the state of affairs it represents is completely accurate. One tends to find that the most important representations covary absolutely. In normal circumstances, for example, the desert ant never makes

1. Recall that R is a particular state of a representing mechanism. Suppose that such a mechanism M has representing states $R_1, \ldots, R_n$. Then these states covary absolutely with mutually exclusive states of affairs $S_1, \ldots, S_n$ just in case M is in state R_i if and only if S_i obtains.

a mistake about its position.

This sort of perfection, however desirable, is often not necessary for capacity realization. A mental representation of S will be useful even if it is not always tokened when S obtains, or if it is sometimes tokened when S does not obtain. The more reliable the representation, the better, of course, but what is minimally required, if a representation is to be of any use at all, is that the representation be more reliable than an informed guess. To state this condition, I will introduce some probabilistic apparatus. I will say that a representation R probabilistically covaries with a state of affairs S just in case $\Pr(S|R) > \Pr(S)$. In words: the probability of S obtaining given that R is tokened must be greater than the probability with which S obtains in general.

We can understand a neural state R that probabilistically covaries with a state of affairs S as giving us some information about S (though in a wider sense than that of Dretske (1981), which requires that $\Pr(S|R) = 1$). Knowing that R is tokened tells us more about S than we would have known otherwise. A representation that does not give us any information in this sense is for all practical purposes useless as an indicative representation.

The role of the detection mechanism in Gallistel's explanations is to maintain the covariation of a representation and its corresponding state of affairs. This means, first, that the representation and the state of affairs must covary, at least probabilistically, and second, that the covariation must exist in virtue of the detection mechanism, and more particularly, must exist in virtue of the detection mechanism's establishing a one-way causal link from the state of affairs to the tokening of the representation. This last requirement of causal directionality rules out covariation that exists in virtue of the representation's tokening causing the state of affairs to come about. The reason for the directionality requirement is simply that Gallistel and others' explanations involving inner maps satisfy the requirement. A kind of informatic explanation in which causation goes in the other direction is explored in section 6.

Covariation is exploited when the information supplied by a tokened representation is acted upon by the prosecution mechanism in such a way as to realize the capacity to be explained. The desert ant, for example, exploits the covariation of its internal representation of the direction of and distance to its nest by heading in the represented direction for the represented distance.

In what follows, I will give an account of what I call simple exploitation. More complex forms of exploitation are treated in section 3.3.

I have said that a prosecution mechanism exploits a covariation when the information supplied by a tokened representation is acted upon so as to realize a given capacity. Simple exploitation is exploitation in which the tokening of a given representation is always acted on in the same way. In simple exploitation, then, there is for each representation R what I call an *appropriate action* A. When R is tokened in the right circumstances, A is executed. (The right circumstances are whatever prompts the organism to act on its representation; in the case of the ant, for example, the right circumstances are finding food, or whatever it is that causes the ant to turn for home.)

Most of the present section defines what it is for an action to be appropriate. Suppose that R represents a state of affairs S. A is appropriate just in case executing A improves the chances of the system realizing the capacity to be explained, and does so in virtue of the fact that S obtains. The latter requirement—that the increase in probability occurs in virtue of S—ensures that A's appropriateness is due to its being especially apt when S obtains, rather than just a good move overall.[2] Appropriateness, then, is relative both to a state of affairs and a capacity.

This informal definition can be formalized probabilistically.[3] An action

2. More generally, if a representing mechanism M has states $R_1, \ldots, R_n$ that covary with states of affairs $S_1, \ldots, S_n$, then this covariation is exploited if for each R_i there is an action A_i such that (a) when M enters the state R_i, A_i is caused to occur, and (b) the occurrence of A_i improves the chances of the system realizing its capacity, in virtue of the fact that S_i obtains.

3. The probabilistic idiom may not always seem especially appropriate. A more general definition of simple exploitation would require that, in virtue of the covariation, the action A

A caused by the tokening of a representation R of a state of affairs S is appropriate (relative to a capacity C) just in case the following conditions are satisfied:

1. $\Pr(C|SA) > \Pr(C|S)$; that is, when S obtains, A increases the chances of C being realized.
2. $\Pr(C|SA) > \Pr(C|A)$; that is, the increase in probability occurs at least in part because S obtains.
3. These inequalities hold in virtue of the fact that the execution of A sets in motion a causal mechanism leading to the realization of the capacity C.

As in the case of detection, then, there is, in addition to the formal probabilistic requirement, a requirement that causation occur in a particular direction. Note that the requirement that the tokening of R always cause A can also be weakened to a probabilistic requirement: $\Pr(A|R) > \Pr(A)$, or equivalently, $\Pr(RA) > \Pr(R)\Pr(A)$.

Let me illustrate my definition of what it is to exploit a covariation with another elementary example of simple exploitation. Consider an animal with a neural state that represents the fact that there is a predator close at hand. Suppose that this representation covaries with the presence of a predator. Then the animal may exploit the covariation by running away when the representation is tokened. The appropriate action, then, is running away, which increases the chances of survival when the represented state of affairs obtains, but not when it does not.

2.4 The R-C Model

I have defined a class of explanations in which the explanandum is a capacity C, and in which a representation R of a state of affairs S plays the following role:

---

make a difference to the realization of the capacity to be explained. This definition requires, however, a technical definition of difference making which would take me too far afield in this paper.

1. In virtue of a detection mechanism, *S*'s obtaining raises the probability of *R*'s being tokened, and
2. In virtue of a prosecution mechanism, *R*'s being tokened causes an action that increases the probability, given *S*, of the capacity *C*'s being realized.

These are the *R-C explanations*. My formal characterization of the R-C explanations in this paper is the *R-C model* of explanation.

In an R-C explanation, then, the explanatory role of representations, is to serve as a reliable guide to the world that is consulted so as to determine an action that will, in the circumstances, improve the chances of a creature's realizing the capacity to be explained. This, I propose, is how representations are invoked as "inner maps" to explain animals' navigational capacities.

One may compare this account to elements of Ruth Millikan's theory of mental representation, which turns on the explanatory role of maps in an organism's evolutionary history (Millikan 1984, 1989). Millikan does not, I think, quite succeed in giving a fully explicit account of the role, in evolutionary explanation, played by mappings (and it seems that she perhaps does not intend to do so, leaving the task to philosophers of science). But she does emphasize the role of a representation's "producers" and "consumers", which correspond roughly to my detection and prosecution mechanisms. Correlates of my detection and prosecution mechanisms may also be found, I think, in Dretske (1988).

Let me spell out the elements of an R-C explanation in a little more detail. A good R-C explanation, I propose, contains the following:

1. A specification of a representing mechanism, a detection mechanism, and a prosecution mechanism.
2. A description or demonstration or at least an assertion of the detection mechanism's explanatory virtues, that is, its capacity to establish covariation between the states of the representing mechanism and the states of affairs that they represent.
3. A description or demonstration or at least an assertion of the prosecution mechanism's explanatory virtues, that is, its capacity to exploit the

covariation established by the detection mechanism, more exactly, to produce an appropriate action in response to the tokening of any state of the representing mechanism.

The explanandum—the existence of a capacity—can be deduced from these elements.

The R-C model is well able to capture the form of simple informatic explanations such as the explanation of the ant's homing ability or the unnamed animal's capacity to avoid predators. In the next few sections I ask what other informatic explanations might be treated according to the R-C model. This will involve some additions to the model, especially in sections 3.3, 5, and 6.

These later additions will not, however, much alter the underlying structure of the R-C model. Thus it is possible even at this early stage to note that the R-C model seems to fit well with a causal approach to explanation. It is not obvious that it had to be so: the hallmark of representation is, after all, a semantic, not a causal, relation between the representation and the rest of the world. In turns out, though, that the representations invoked by Gallistel and others to explain animals' navigational capacities do their work in virtue of their central position in a causal mechanism for which the detection and prosecution mechanisms constitute the interface to the world. I ought to note that the R-C model is also compatible with some other accounts of explanation, for example, Hempel's D-N model.

3. Extending the R-C Model

3.1 Covariation and the Representation of Past and Future States of Affairs

It is the responsibility of the detection mechanism to ensure that the corresponding representing mechanism provides an accurate picture of those parts of the world it is supposed to represent. I have characterized accuracy in terms of covariation. Is this too narrow a conception? That question is the subject of sections 3.1 and 3.2.

Covariation is a relation that can hold only between coexisting states of affairs. A representation, then, cannot covary with a future state of affairs. If a certain light covaries absolutely with danger, for example, it covaries with danger now—it lights up for just those periods that count as dangerous.

Even quite simple representations, however, can be of states of affairs in the past or future. Consider the following example. A certain plant (there are many that conform to the description that follows) may begin to produce buds, in early spring, after a warm period lasting a certain number of days. The warm period indicates that spring is nigh, and so that it is reasonably safe to set the budding process in motion. Some period of time—say two weeks—after the warm period, budding begins. The two week delay may be seen as a safety margin.

The mechanism underlying the "decision procedure" is as follows. The warm period changes some underlying state of the plant. Biologists routinely refer to this underlying state as a representing mechanism (in a very light duty sort of way), so I will call the new state R. The tokening of R sets in motion a process that, after a period of two weeks, triggers the budding process. It is natural to talk of R as representing the state of affairs "it is safe to begin budding in two weeks" or "the weather will be tolerably warm starting two weeks from now" or just, for convenience, "spring begins in two weeks". It seems, then, that R represents a future state of affairs. (Never mind whether the plant's R is really a representation; for the sake of the argument, let it stand in for genuine representations of future states of affairs, which representations are, of course, legion.)

Clearly, the representation of the imminent beginning of spring ought to be invoked in any explanation of the plant's capacity to bud at the right time, early but not too early. (Here, by the way, is a case where the probabilistic modeling of capacity realization is entirely apt.) But how can the explanation be subsumed under the R-C model if a representation cannot covary with a future state of affairs?

The answer is really very simple. Think of the relevant state of affairs as a present state of affairs, say, the state of affairs of it being two weeks before the beginning of spring. The accuracy of the representation R can be quantified in terms of its covariation with this state of affairs: the detection mechanism to which *R* is attached is accurate to the extent that *R* is tokened just in case it is two weeks before the beginning of spring.

This use of the notion of covariation in this way is no different, I think, from the use of the notion to quantify the accuracy of a representation of a state of affairs distant in space. A representation meaning "Danger!" need not mean "Danger here!"; it might mean "Danger within 1 mile of my forward line of sight" or "Danger back at the nest!". The covariation of a representation *R* and danger at the nest amounts, as I understand it, to *R*'s being tokened just in case there *is* danger at the nest. In the same way, the plant's *R* is accurate to the extent that it is tokened just in case it is two weeks before the beginning of spring. In this way, the notion of covariation can be used to quantify the accuracy of representations of states of affairs remote in time or space.

In the case of very sophisticated representing mechanisms, such as those in the human mind, it may not ring true to talk as though all representations are of present states of affairs. For example, my belief that, say, it will rain tomorrow, seems to be a belief about tomorrow, not today—it is not the belief that today is a day one day before it rains. Yet the R-C model requires that such representations be understood as corresponding to present states of affairs. Provided that the required correspondence relation is seen as a technical device, related but not identical to the semantic relation of aboutness, I do not see this as a problem for the R-C model. What is important is that the model has a way of handling representations of past and future states of affairs.

3.2 Representations of Permanent States of Affairs

Representations, on the R-C model, always covary to some degree with the states of affairs that they represent. It would seem to follow that a represented state of affairs would have to sometimes obtain and sometimes not obtain. The R-C model, then, cannot account for explanations involving representations of permanent states of affairs, such as a representation of the fact that the star Sirius is ten light years distant.

Is this reasoning correct? As it happens, representations of permanent states of affairs can play a role in explanations of quite simple capacities. An example was mentioned earlier: certain migratory birds learn, during a critical period when they are young, the layout of the stars in the sky, and later in life they navigate by this layout, using their avian constellations (different for different individuals) to locate the celestial pole (Gallistel 1990, 83–88).[4] We know this because experimenters brought up birds in a planetarium where the stars were askew; upon release, the birds flew off in entirely the wrong direction (vivisection for the Information Age).

It turns out, then, that representations of permanent states of affairs require reliable detection and prosecution mechanisms for exactly the same reasons as any other representation. These mechanisms play the usual role in explanations of capacity realization: in the case at hand, they put the bird's brain into one of a set of states correlated with the various possible arrangements of the stars, and whichever state is tokened then causes the bird to fly off in what would be the right direction if the stars were actually arranged that way.

This leaves a technical problem involving the definition of covariation. It would appear that if a state of affairs S is permanent, the probability of S is one, which means that the conditions for covariation with S and exploitation of S (specifically, clause (2) of the definition of an appropriate action in section 2.3) can never be satisfied. To cover the case of permanent states of affairs, we must allow the probabilities that define exploitation and covariation to range over possible states of affairs in which the stars are differently arranged, so that $\Pr(S) < 1$ (where S is the current state of the stars).

3.3 Complex Exploitation

The characterization of the R-C model in section 2.4 allows only for simple exploitation of a covariation, that is, only for exploitation that always produces the same appropriate action in response to the tokening of a given

4. The layout of the stars is not, of course, truly permanent. But compared to the life of a constellation, that of an indigo bunting is very short.

representation, for example, running away in the case of a predator detector and turning a certain angle and scuttling for a certain distance in the case of the desert ant. Simple exploitation is inadequate when the appropriate response to a representation depends on what other representations have been tokened. An important class of examples of this sort are those in which an organism has an internal map of the surrounding terrain that includes more than one feature. The correct action to take, based on the contents of such a map, will be determined not by any one representations, but by the all the representations together.

Consider an example. The frillfin goby, a small fish, makes use of a map of the local geography in finding its way from tidal pools to the sea. From within a tidal pool, other pools and the sea are not visible, and a leap in a random direction is likely to leave the fish on dry land. But if the receding tide leaves a goby cut off from the sea, it is able to jump from pool to pool in a direct route to safety. Controlled experiments show that the gobies can find the sea in this way only if they are familiar with the layout of the pools (a day of swimming around at high tide is sufficient), and thus that their ability is due to their learning the lay of the land (Gallistel 1990, 141–2).

The relevant goby prosecution mechanism, then, must be capable of using the goby's internal map of the tide pools to generate a route between the goby's current location and the sea. The action executed, a series of carefully oriented jumps, will be a function of (a) the representation of the goby's current location, (b) the representations that make up the map, that is, the representations of the positions of all the pools, and (c) the representation of the desired location, that is, the sea.[5] Call these sets of representations R_L, R_M and R_D respectively. The action executed—the jumps made—may be written as a function of these representations, that is, as $A(R_L, R_M, R_D)$, or for short, A_{LMD}.[6] If the goby always calculates correctly, and its internal map

5. Since the reaching the sea is, as far as we know, always a goby's goal, the "desired location" is in this case not a variable, and so will not require a representing mechanism. But a general account of map-based navigation will have to allow for a variable destination.

6. Here I assume, for simplicity's sake, that no other factors influence the goby's choice of route.

is always accurate, then the series of jumps that constitutes A_{LMD} will get it back to the sea every time. More generally, an action A_{LMD} will be appropriate, and so executing that action will count as exploitation of the relevant covariation, if A_{LMD} raises the probability of capacity realization when the represented states of affairs obtain.

Let me put this more formally. Let R_L, R_M, and R_D be, as before, representations of an organism's current position, the surrounding terrain, and its destination, respectively. Let the capacity C to be explained be the organism's ability to reach its destination. Then the covariation between R_L, R_M, R_D and the states of affairs S_L, S_M, S_D that they represent is exploited so as to realize the capacity C if for every combination of R_L, R_M, R_D,

1. $\Pr(C|S_L S_M S_D A_{LMD}) > \Pr(C|S_L S_M S_D)$,
2. $\Pr(C|S_L S_M S_D A_{LMD}) > \Pr(C|A_{LMD})$, and
3. These inequalities hold because the execution of A_{LMD} sets in motion a causal mechanism leading to the realization of the capacity C.

This is of course just the definition of simple exploitation, with A_{LMD} for A and $S_L S_M S_D$ for S. The definition is easily generalized to the case where the Rs are any indicative representations and C is any capacity.

This characterization of complex exploitation is not only unwieldy, however, but is stronger than it needs to be. Each of the representations in R_L, R_M, and R_D may play a role in explaining the capacity C even if some of the actions A_{LMD} are not appropriate, that is, if certain mistakes are systematically made in planning the route to the destination: not a problem for gobies, but common in the realization of navigational capacities less essential to the continuation of life.

A weaker and more concise definition of exploitation is as follows. The covariation of a representation R with a state of affairs S is exploited to realize C just in case

1. $\Pr(C|SR) > \Pr(C|S)$
2. $\Pr(C|SR) > \Pr(C|R)$, and
3. These inequalities hold in virtue of the fact that being in state R at some

stage helps set in motion a causal mechanism leading to the realization of the capacity C.

Thus what is required is that the evidence about S provided by R is somehow utilized at some point, perhaps in the light of the tokening of many other representations, to realize C.

What is missing from this characterization of the explanation of the goby's navigational capacity is, of course, any commitment as to the nature of the mechanisms responsible for maintaining the proper probabilistic relations, that is, any commitment as to the nature of the detection and prosecution mechanisms. This omission is deliberate; my intention is that the R-C model should be compatible with any of a wide range of hypotheses about cognitive architecture. The scheme, then, is intended to capture what a classical computational (Gallistel 1990), a connectionist, and a "dynamic" explanation of the goby's navigational ability all have in common. It does not, of course, apply to explanations that reject the existence of internal representations altogether.

4. Beyond Navigation

4.1 Further Extending the R-C Model

My examples of informatic explanation all involve fairly simple navigational capacities of animals. The R-C model seems to account for the role of representations in the explanation of these capacities well enough, but to what other kinds of informatic explanation might it apply? This question might be broken down into several parts:

1. Does the R-C model apply to the explanation of more complex capacities in more complex animals? To the explanation of human capacities?
2. Does the R-C model apply to psychological explanations with explananda other than capacities?
3. Does the R-C model apply to explanations outside psychology, such as those in neurobiology, or developmental biology?

This section's brief discussion of these questions will lead to, first, a discussion of the informatic explanation of behavior, in section 5, and second, a treatment of the explanatory role of imperative representations, that is, representations of goals, in section 6.

4.2 Other Animals, Other Capacities

Let me begin with the question concerning more complex capacities and more complex animals. Here I am confident that the accuracy of detection mechanisms and the appropriateness of the actions executed by prosecution mechanisms will continue to be a central focus of informatic explanations, and so that the R-C model will continue to be valuable. An example involving communication between monkeys is given at the beginning of section 5.

Despite these successes, there is a very good reason to think that the R-C model cannot possibly accommodate every informatic explanation of animals' capacities to deal with the world. The reason is this: there is no place in the R-C model for imperative representations—representations of *goals*—such as human desires and so on. The representations in an R-C explanation are indicative representations, functioning like beliefs rather than desires. I will suggest a way to extend the R-C model so as to accommodate the explanatory use of imperative representations in section 6.

4.3 Other Kinds of Explananda

There are at least two kinds of explananda other than capacities that receive informatic explanations in the psychological literature. First, and more paradigmatic, is the explanation of individual pieces of behavior. Under this heading I group not only the explanation of physical behavior, but also of mental behavior, such as the tokening of particular beliefs and desires or the formation of particular intentions. In a standard folk psychological explanation, for example, my belief that it is raining and my desire not to get wet explain my forming the intention to pick up my umbrella, a piece of mental behavior, which in turn explains my actually picking up the umbrella, a piece of physical behavior. In section 5, I use a more scientific example to sketch an account of the informatic explanation of behavior. On the view I will advance, the explanation of capacities plays a crucial role in the explanation of behavior, because a piece of behavior is explained as an instance of the realization of a capacity. The view builds on the R-C model, then, rather than replacing it.

The second kind of psychological explanandum I will mention is one much emphasized by Robert Cummins: *computational capacities* (Cummins 1983, 1989). Computational capacities differ from the capacities that are the explananda of R-C explanations in that, whereas an R-C explanation capacity is *world-involving*—its exercise involves the manipulation of or at least interaction with the outside world—a computational capacity is exercised entirely within the mind. Cummins's paradigm of a computational capacity, for example, is the capacity to add numbers. Other, more complex, computational capacities involve such things as the various computations undertaken by the visual cortex that underlie our sense of sight.

Pure computational explanations, it seems, will not involve any appeal to detection or prosecution. Certainly, Cummins's account of the explanation of computational capacities makes no reference to these notions, and I think that his account is essentially correct. Thus there is at least one major class of psychological explanations that is not to subsumed under the R-C model. Because Cummins has dealt with this class of explanations so thoroughly, there is no need for me to discuss it further here.

I should note, however, that a computational capacity may play a critical part in the implementation of a detection or a prosecution mechanism featured in an R-C explanation, and so may itself be a part of an R-C explanation of a world-involving capacity. For example, part of the prosecution mechanism for gobies' navigational capacity (section 3.3) may involve a computation of an appropriate route to the sea from the goby's internal representation of the layout of the surrounding tidal pools (see section 6). The explanation of this particular part of the prosecution mechanism, then, would proceed according to the scheme laid down in Cummins's work. In this respect, the explanation of computational capacities and the explanation of world-involving capacities, such as navigational capacities, may be overlapping endeavors.

4.4 Informatic Explanation outside Psychology

Do informatic explanations outside psychology make the same kind of explanatory use of representation and allied notions as explanations within psychology? To provide a full answer to this question would require a journey into genetics, developmental biology, neurobiology, and other sciences that would belong in a book length treatment of my topic, but which would overwhelm a single article. Nevertheless, let me briefly discuss the prospects of the R-C model outside psychology, with the help of three examples: the "neural code", the "genetic code", and "genetic programs".

The hypothetical neural code is the encoding system that neurons use to encode information in their patterns of firing, often called *spike trains*. No one knows if the same system is used by all neurons, but for simplicity's sake, let me continue to refer to *the* neural code. The neural code stands to full blown mental representation more or less as the genetic code stands to what might be called genetic representation. The genetic code represents amino acid sequences, while genes or gene complexes "represent" things like blood type and body plans. Similarly, the neural code is normally taken to represent numbers, rates, intensities, and other quantitative properties of things, while, say, beliefs represent things like horses and democracy. But the representational ability of genes and beliefs is thought to depend, somehow, to a great extent on the representational power of, respectively, the genetic and neural codes.

Explanations involving the neural code seem to be amenable to treatment by the R-C model for the same reasons as are the psychological explanations considered above. The relevant property of a spike train will covary at least probabilistically with some other, non-neural property, and the spike train will in turn cause some action or other that helps to realize the capacity to be explained. The science is only in its initial stages, but for an example of how such an explanation might look, the reader might consult Rieke et al. (1997).

Perhaps the most obvious objection to my confidence in R-C explanation in the neural realm is that an explanation involving the neural code may not always be an explanation of some particular capacity. The explanations we find in current neuroscience certainly are of capacities, such as memory or perceptual capacities, but it need not always be so. As in my discussion of the explanation of computational capacities, I therefore advocate a pluralistic attitude to the proper form of informatic explanations, while suggesting that R-C explanation will always be very important.

I will now consider the genetic code, meaning the code in virtue of which sequences of nucleotides in DNA and RNA encode sequences of amino acids. It is quite unclear whether the R-C model is of any help in explanations involving the genetic code.

Let me state just one problem. What, in such an explanation, does a nucleotide sequence covary with? What is the relevant "detection" mechanism? One possibility, if we think of the capacity to be explained as having something to do with inheritance, is that the nucleotide sequences are supposed to covary, in the case of DNA, with previous generations of nucleotide sequences, and in the case of RNA, with the nucleotide sequences in DNA.[7] The "detection mechanism", then, includes the process of DNA replication in mitosis, the process of transcription by which RNA is created, and other such processes.

But there is an obvious objection to this suggestion, which is that in many cases, nucleotide sequences do not, and are not supposed to, covary with those of previous generations. The obvious example is sexual reproduction. Another example is bacterial conjugation, in which genetic material is transferred from one bacterium to another.

In answer to this objection, the covariance might be stated at the level of the gene: sequences are supposed to covary with the sequences in the genes from which they are copied. I will not try to work out the details here. Perhaps this is a case where the use of an informatic notion is so simple that the

7. I will suppress, for the sake of the argument, many of the necessary qualifications and complications.

R-C model provides an explanatory surfeit: what is going on in explanations that invoke the genetic code is something much simpler than in a typical R-C explanation, so we end up looking for such things as detection mechanisms where there are none to be found.

There is, of course, a more sophisticated sense in which DNA and other elements of organisms' reproductive systems may be characterized in informatic ways. Genes are often said to encode *programs* for the construction and maintenance of organisms. In this idiom, the genetic material represents the body plan and the instructions for adding appendages such as antennae, eyes, and limbs to the body plan, and so on. It is a matter of some controversy whether this kind of talk is useful for explanatory, or any other, purposes. But suppose that it is here to stay. Might explanations that use this language fall under the R-C model?

Clearly not, unless the R-C model is enhanced so as to cover explanations invoking imperative representations, since on the view of genetic material as program, much of what genes represent takes the form of instructions. Imperative representations of goals will be treated in section 6. Until then, let me make two comments. First, the capacities that serve as explananda in these explanations will be the kinds of capacities that one finds in developmental biology, for example, the capacity humans have to grow neither more nor less than five digits on each hand. Second, some of the same problems that arose in interpreting explanations involving the genetic code as R-C explanations will arise in interpreting explanations involving genetic programs as R-C explanations.

5. The Explanation of Behavior

This section advances an account of the informatic explanation of behavior based on the R-C model. In order to establish firmly the suitability of the R-C model for dealing with my central example, Dorothy Cheney and Robert Seyfarth's well-known study of vervet monkey alarm calls (Cheney and Seyfarth 1990), I will discuss the example at first as though it were a normal case of

R-C explanation. I ask the reader to ignore, for the time being, the fact that the relevant explananda are behaviors as much as capacities.

I have chosen this example not only because the scientists involved have been especially careful in spelling out important aspects of their reasoning, but also because it demonstrates the ability of the R-C model to handle, first, explanations involving primates, and second, explanations in which the representations are public, as in language, rather than private.

Cheney and Seyfarth's main objective is to find out what vervet monkeys' alarm calls represent. They characterize themselves as reasoning by inference to the best explanation: they choose the thesis about the meaning of danger calls that best explains the observed facts, in particular, that best explains monkeys' capacity to avoid predators. By attending to what Cheney and Seyfarth count as evidence, we can see what they count as relevant to an informatic explanation of the monkeys' predator avoidance. (Later I will, as promised, turn to Cheney and Seyfarth's informatic explanations of particular pieces of monkey behavior.)

Cheney and Seyfarth considered two kinds of facts especially relevant to determining what state of affairs the monkeys' alarm calls represent. First, unsurprisingly, is the question as to what causes a monkey to issue a particular warning call. When a monkey spots an eagle, it issues a particular cry, the *eagle alarm*. Cheney and Seyfarth consider two hypotheses about what this alarm represents: the state of affairs *eagle in the vicinity* and the state of affairs *danger from above*. They note that the alarm is given in the presence of eagles but not, say, falling branches, and they conclude that the former hypothesis better explains this pattern of behavior (p. 167–8).

The second kind of evidence Cheney and Seyfarth consider relevant to determining what is represented by alarm calls is the way in which monkeys react to the calls. In response to the same alarm call, different monkeys react in different ways. An eagle alarm, for example, may cause a monkey in a tree to descend to the ground, and a monkey on the ground to take cover in some bushes. Another monkey also on the ground may not react in any way at all. On the whole, each of these responses is appropriate in each monkey's particular circumstances, given that an eagle is in the vicinity. This pattern of behavior is best explained, according to Cheney and Seyfarth, by the hypothesis that the eagle alarm means *eagle in the vicinity* rather than, say, being a signal to other monkeys to imitate the behavior of the alarm-giver:

> . . . the most parsimonious explanation would seem to be that calls denote a type or class of danger, not the caller's behavior, and that an individual's particular circumstances strongly influence the exact nature of his response (Cheney and Seyfarth 1990, 167).

Cheney and Seyfarth's inference to the best explanation, then, seems to proceed in accordance with the following method. Take a capacity, such as predator avoidance. Find out how the detection and prosecution mechanisms work, at the algorithmic level, that is, what triggers alarms and what behaviors are performed in response to the alarms. Infer the existence of the representation that fits best into the resulting explanatory framework, that is, the representation that maximizes the sum of the accuracy of the detection mechanism and the appropriateness of the actions executed by the prosecution mechanism. If this is correct, then Cheney and Seyfarth's view of the explanatory role of alarm calls is in accordance with the R-C model.

Now let me turn to the explanation of behavior. What I suppressed in the above discussion is that Cheney and Seyfarth are concerned primarily with the explanation of the behavior of the monkeys. Specifically, they are concerned with two kinds of behavior:

1. The giving of alarm calls: for example, what explains the fact that a monkey makes, in a certain kind of situation, an eagle alarm call?
2. The reaction to alarm calls: for example, what explains the fact that a monkey reacts to an eagle alarm call, in a certain kind of situation, by taking cover in nearby bushes?

Type (2) explananda are the kind of behaviors that are normally thought of as the explananda of informatic explanations. Typically, a causal account is given of this sort of explanation: the tokening of the representation causes, directly or indirectly, the behavior to be explained. In the case mentioned above, for example, a monkey's hearing an eagle alarm call causes it, by way of some psychological mechanism, to take cover in nearby bushes. Call this the *simple causal account* of informatic explanations of behavior.

Type (1) explananda present a more puzzling prospect. Cheney and Seyfarth propose, as explained above, that what explains the fact that monkeys issue an eagle alarm call when they see an eagle, but not when they see a falling branch, is that the eagle alarm call means "Eagle in the vicinity", not "Danger from above". Clearly, the representation itself cannot play the same role in this sort of explanation that it plays, according to the simple causal account, in the explanation of type (2) explananda, since, on the simple causal account, the explanandum is caused by the tokening of the relevant representation, whereas in this case, the explanandum *is* the tokening of the relevant representation. If the tokening is a cause, it is an Aristotelian final cause. Yet Cheney and Seyfarth treat the two cases quite symmetrically, as though the two kinds of explananda are to be given broadly the same kind of explanation.

In the remainder of this section, I will provide an alternative to the simple causal account that puts type (1) and (2) explananda more or less on a par, and that has as its basis the R-C model of explanation. My proposal is as follows. Standing behind any of Cheney and Seyfarth's behavior explanations, I maintain, is an R-C explanation of a capacity, namely, predator avoidance. The behaviors that serve as Cheney and Seyfarth's explananda are parts of the realization of the capacity. For example, both issuing eagle alarm calls and responding to such calls are parts of the chain of behavior by which monkeys avoid eagles on any particular occasion. Just as a law of nature explains its instances, so the existence of a capacity explains its realizations. Cheney and Seyfarth's explanations of behavior, then, have two steps. The first step is the explanation of a capacity. The second step is the explanation, by the capacity, of some particular realization of the capacity, and so of the pieces of behavior that make up that realization. Representations enter into the explanation directly only in the first step. In so doing, they indirectly explain the behavior, even, as in the case of type (1) behavior, when they do not cause the behavior.

A good deal more ought to be said about the sense in which a capacity explains the pieces of behavior that serve as its realizers on any particular occasion. But this paper is not the place for such a discussion, because the discussion will have nothing in particular to do with the peculiar explanatory power of representations. It is enough for my purposes, here, that the existence of a capacity clearly does, normally, explain the manifestations of that capacity, or even weaker, and still more plausible, that an explanation of the *workings* of a capacity explains the realizations of the capacity.[8] Then, because representations are a part of the explanation of the workings of, say, the vervets' predator avoidance capacity, they are a part of the explanation of the behaviors that realize the capacity.[9]

I have offered an account of the informatic explanation of behavior that is an alternative to the simple causal account. But I have not argued that there cannot be simple causal explanations of behavior. On the contrary, it is a consequence of various positions taken in this paper, both explicit and implicit, that there can be such explanations. Representations, on the R-C model, are among the causes of behavior, and so, on a causal approach to explanation, ought to be able to explain behavior. I deny none of this; my position is that most explanations of behavior in the literature are better understood as based on the R-C model. That is, although we could give simple causal explanations of behavior that invoke representations, normally we give instead explanations that draw on the R-C model.

8. This weakening is intended to sidestep worries about the explanatory power of "dormitive virtues" and the like.

9. Some other topics for discussion ought also to be mentioned. First, one might wonder whether Cheney and Seyfarth's explanations are functional explanations, that is, explanations that invoke the fact that the monkeys' representing and other mental mechanisms are operating in accordance with their biological functions in a typical case of predator avoidance. Second, I am simply assuming that in this sort of case, explanation is transitive in the necessary way: because the representations explain the capacity, and the capacity explains the behavior, the representations explain the behavior. Third, one might wonder how behaviors are to be explained in cases of false alarms. Can these behaviors be counted as "realizations" of the monkeys' capacity to avoid predators? I think they can, in the sense that false positives and false negatives can be understood as a part of an overall strategy of predator avoidance (Godfrey-Smith 1996, part II).

It is an interesting question whether this observation is as true of folk psychology as it is of the scientific literature. That is, are our everyday belief/desire explanations also based on the R-C model, or are they simple causal explanations, as many writers assume? An obvious impediment to the view that folk psychology gives R-C explanations is the need to find a capacity in the explanatory context of which to view any particular piece of behavior. Is this asking too much? Not necessarily: various discussions of, for example, the principle of charity, have argued that we view all human behavior as broadly directed towards certain kinds of goals (Grandy 1973; Dennett 1987). The idea that folk psychological explanation normally requires such a context is not, then, unprecedented.

6. Representations of Goals

In this last section, I suggest a way in which imperative representations, that is, representations of goals, might find a place in the R-C model. The discussion will be rather more schematic than the treatment in sections 2 and 3; I will point the way rather than laying down the law.

An imperative representation might be a desire, an intention, an instruction, a command, or something else. There are, of course, interesting differences between these different kinds of representations, but in what follows, I focus on what they have in common.

On the R-C model, the invocation of an indicative representation or representing mechanism is accompanied, explicitly or otherwise, by an invocation of two other mechanisms: a mechanism responsible for the tokening of the representation (the detection mechanism) and a mechanism by which the tokening of the representation causes some appropriate action (the prosecution mechanism). I propose that imperative representations are also flanked by two mechanisms, one by which a representation is tokened and one by which a tokening causes some other action or series of actions, and each of which plays a distinctive role in explaining whatever capacity is to be explained. I call these mechanisms, respectively, the *planning mechanism* and the *execution mechanism*.

The role of the execution mechanism is to realize the represented goal. In a good R-C explanation that involves an imperative representation of a goal S, then (where S is the "desired" state of affairs), the execution mechanism will increase the probability that S obtains. The importance of the execution mechanism in the mental economy of desires and other imperative representations has long been recognized, as in, for example, Ruth Millikan's definition of an imperative intentional icon (Millikan 1984, 99).

It is, however, the planning mechanism that makes the connection between imperative representations and capacity realization. The role of the planning mechanism is to form goals, the realization of which will promote capacity realization. In a good R-C explanation involving an imperative representation of a goal S, then, the planning mechanism will token the representation in a context where S's obtaining will increase the chances of capacity realization. The role of the execution mechanism is, of course, to bring this capacity-promoting state of affairs about.

As an example, consider once more the goby. We have seen that the goby maintains an inner map of the configuration of local tide pools, and that it uses this map to construct a route that will take it safely back to the sea. It seems reasonable to conjecture that the end product of the route construction is a representation of the route itself. This is a representation of a goal: at the time of construction, what is represented is not yet an actual route, but rather, to put it anthropomorphically, an intended route.

The mechanism that the goby uses to encode its intended route, then, is a representing mechanism for goals. Each state of the mechanism represents a particular route, or more exactly, represents the so far unactualized state of affairs of the goby's taking a particular route back to the sea. The corresponding planning mechanism is the mechanism that constructs the route. When it works well, it will construct a route that does in fact lead safely back to the sea, so allowing the goby to realize its capacity to return to the sea. The execution mechanism is, of course, whatever mechanism translates the representation of the route into a series of jumps that actually trace the route. When it works well, the goby succeeds in taking the represented route.

It is clear that both the planning and the execution mechanism must work reliably if the goby is to reliably return to the sea. Good execution without good planning likely results in the goby accurately tracing a route that does not lead back to the sea. Good planning without good execution likely results in the goby inaccurately tracing a route that does lead back to the sea. Either way, the goby likely fails to reach the sea. An R-C explanation of the goby's navigational ability, then, must include both the planning and execution mechanisms, and must show that each works well, or at any rate, well enough to realize the capacity.

This particular example makes explicit what I noted back in section 2.1: the mechanisms invoked in an R-C explanation may themselves have internal representations. The prosecution mechanism for the goby's inner map, for example—by definition, the mechanism in virtue of which the map causes the goby to make the moves that it does on the way back to the sea—will consist of part of the planning mechanism, and all of the representing and execution mechanisms, for the goby's representation of its intended route. So the prosecution mechanism includes the goby's imperative representation of its intended route. For the same reason, the planning mechanism—by definition, the mechanism in virtue of which the way things are in the world causes the goby to settle upon the route that it does—will consist of part of the prosecution mechanism, and all of the representing and detection mechanisms, for the goby's inner map. So the planning mechanism includes the goby's indicative representation of the surrounding terrain.

In an R-C explanation that invokes both indicative and imperative representations, then, it will be natural to break down the mechanisms into parts. In the case of the goby, there will be five distinct parts: the detection mechanism for the map, the representing mechanism for the map, a route-finding mechanism between the map and the route representation, the representing mechanism for the route, and finally, the execution mechanism for the route, as shown in figure 1.

I should emphasize that the mechanisms invoked by an R-C explanation, with the exception, perhaps, of the representing mechanisms themselves, will likely not be self-contained physical units. Thus there will often be good rea-

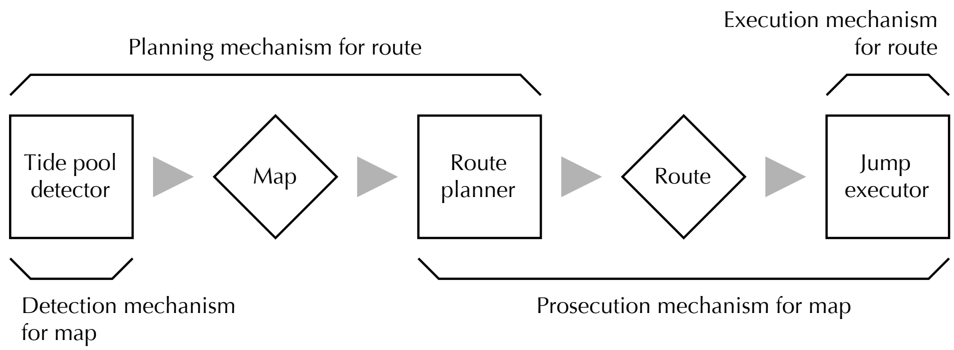

Figure 1: The five components that make up the goby's detection, prosecution, planning, and execution mechanisms. Diamonds stand for representing mechanisms: the map of the tide pools on the left, and the planned route on the right.

son for an R-C explanation to break them down into parts. How they are broken down will depend very much on engineering issues, such as, in the case of psychological explanations, issues concerning cognitive architecture. My claim in this paper is that, despite the very different ways in which representing systems are engineered, one can discern, in many informatic explanations, the outlines of detection, prosecution, planning, and execution mechanisms.

The conditions for the reliable operation of the planning and execution mechanisms can, as in the case of detection and prosecution, be stated more formally in probabilistic terms. The probabilistic condition for the reliable operation of an execution mechanism is

$$\Pr(S|R) > \Pr(S)$$

where R is the representation and S is the corresponding state of affairs. This is identical to the probabilistic condition for reliable detection. The difference is that, whereas in the case of detection, I required that the relation hold in virtue of a causal relation running from S to R, that is, from the outside world to the representation, in the case of execution, I require that the relation hold in virtue of a causal relation running from R to S.

A formal condition for the reliable operation of a planning mechanism is as follows. Call a state of affairs A an *appropriate context* for bringing about a state of affairs S relative to a capacity C if bringing about S when A obtains raises the probability of realizing C. An appropriate context, then, is a context where somehow making it the case that S obtains helps to realize C. Probabilistically, A is an appropriate context for bringing about S relative to C just in case:

1. $\Pr(C|SA) > \Pr(C|S)$,
2. $\Pr(C|SA) > \Pr(C|A)$, and
3. These inequalities hold in virtue of the fact that bringing about S sets in motion a causal mechanism leading to the realization of the capacity C.

A planning mechanism operates reliably if for every imperative representation R of a state of affairs S, there exists an appropriate context A for bringing about S that, when it obtains, causes R to be tokened, or at least raises the probability that R is tokened, so that $\Pr(R|A) > \Pr(R)$, or equivalently, $\Pr(RA) > \Pr(R)\Pr(A)$. A more complex version of this condition can be modeled on the treatment of complex exploitation in section 3.3.

The definition of reliability in an execution mechanism is the same as the definition of reliability in a prosecution mechanism, but with A representing a context rather than an action, and with the causal direction reversed, in the sense that A is required to cause R rather than vice-versa. There is a remarkable symmetry, then, between the explanatory role of indicative and imperative representations on the R-C model. Both kinds of representations must be flanked by mechanisms that establish the same probabilistic relations between the elements R, A, S, and the capacity C:

1. $\Pr(S|R) > \Pr(S)$,
2. $\Pr(C|SA) > \Pr(C|S)$,
3. $\Pr(C|SA) > \Pr(C|A)$, and
4. $\Pr(RA) > \Pr(A)\Pr(R)$.

The difference is in the direction of the causal relations that establish the probabilistic relations. For indicative representations, the flow is from A to R to S to C. For imperative representations, the flow is from S to R to A to C.

References

Cheney, D. L. and R. M. Seyfarth. (1990). *How Monkeys See the World: Inside the Mind of Another Species*. Chicago University Press, Chicago.

Cummins, R. (1983). *The Nature of Psychological Explanation*. MIT Press, Cambridge, MA.

———. (1989). *Meaning and Mental Representation*. MIT Press, Cambridge, MA.

Dennett, D. (1987). *The Intentional Stance*. MIT Press, Cambridge, MA.

Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA.

———. (1988). *Explaining Behavior*. MIT Press, Cambridge, MA.

Gallistel, C. (1990). *The Organization of Behavior.* MIT Press, Cambridge, MA.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge University Press, Cambridge.

Grandy, R. (1973). Reference, meaning, and belief. *Journal of Philosophy* 70.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. MIT Press, Cambridge, MA.

———. (1989). Biosemantics. *Journal of Philosophy* 86:281–297.

Rieke, F., D. Warland, R. de Ruyter van Steveninck, and W. Bialek. (1997). *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA.