

I. Intentionality

In an essay entitled "Intentional Systems", Daniel C. Dennett describes what it means to take an intentional stance toward the behavior of a system.

"One predicts behavior in such a case by ascribing to the system *the possession of certain information* and supposing it to be *directed by certain goals*, and then by working out the most reasonable or appropriate action on the basis of these ascriptions and suppositions."

(Dennett 1978, 6)

Intentionality, Dennett claims, is used as a tool to achieve better predictive power of a system's behavior. One possibility is that this "assumption of rationality" (1978, 6) is useful from an observational standpoint only. A calculator, for example, could possibly be described as rational when it correctly calculates sums, but this is not an example of original intentionality. Instead, the calculator can be said to exhibit *derived* intentionality; its display takes on meaning only because humans designed the display and assign meaning to the input/output relationships. There is no sense in which the calculator possesses the concept of what it means to add two groups of objects. Yet systems we believe to be rational, namely other humans and possibly some higher animals, make these same calculations using internal beliefs about objects they can see, feel, and classify as having common properties. This second possibility, that a theory of intentionality accurately describes the basis for how rational creatures perform mental computations, is a most promising way to begin investigating the nature of mental architecture.

Some justification is required for why a materialistic theory of meaning would want to ascribe intentional properties to mental states. Since this is extensively discussed in the literature and is not the primary focus of this paper, a short synopsis tracing my position on why intentionality is necessary will suffice. In his book entitled *Meaning and Mental Representation* (1989), Robert Cummins argues that theories of meaning fall into two general categories. There are theories that treat meaning as asymmetrical, "a treatment that accords priority ('originality') to mental meaning," and those that instead "hold that mental and nonmental representation are basically the same." (Cummins, 1989, 25) In the above example of a calculator, nonmental meaning was said to be derived from the meaning of mental states. However, I do not generally agree with the claim that there are two distinct types of representation. Mental and non-mental representation are both instances of a single type of representation, and one is not dependent or more privileged than the other. They are different in virtue of their structural characteristics, not an *a priori* distinction in the nature of meaning itself.

The terms original and derived are better understood as capturing differences in the use and flexibility of different representational relationships. Both mental and non-mental representations depend on the detection of covariances by a system with multiple states and input/output relations. For the purposes of this rather non-formal discussion, I use the term covariation not as a technical or philosophical term but only as the best word available for describing the co-occurrence of two natural phenomenon. Yet even this fuzzy notion of covariation must be distinguished from representation, for not every instance of covariation in nature deserves to be called representation.

I take the development of a representation to involve the following things. First, an active representational system, by which I mean things as diverse as planaria, phototropic plants, insects, calculators, humans, and some attempts at artificial intelligence, must have a perceptual apparatus receptive to and capable of resolving the relevant differences in local energy levels. Second, a particular perceptual system must be capable of simply observing a covariation (without necessarily recognizing it as such) before anything else can happen; this involves some type of attentive behavior. In some cases, this is built in to the input/output relationships, as is the case with a calculator. The final step is an encoding of this covariation, which takes place as a result of the physical properties of the representational system. For example, a phototropic flower is different from one that is not by virtue of the fact that its new growth always occurs on the side that is not sunny, causing its stem to bend toward the light. In animals, changes in synaptic connectivity in response to calcium currents during the presentation of a conditioned stimulus leads to a change in the way the animal responds to the stimulus in the future. Some representational systems can even respond to these observed covariations by creating external representations, such as the marking of territory by scent or the creation of maps and signs. This type of non-mental representation is still not fundamentally different, for another representational system coming across this now more explicit characterization of a covariation must reconstruct the nature of this new covariation in the same manner outlined above for this externally constructed representation to actually serve as a representation.

Thus the difference between a covariation and a representation is that something must be interpreted by a representational system to count as a representation. Representations may differ in the extent to which they are unchangeable and how much interpretation is required, but there is only one fundamental type of representation. Although this definition may sound somewhat circular, it is not meant to be a founding pillar upon which I base my discussion. The preceding discussion is included only to give some justification for rejecting the position that nonmental representation is somehow dependent on intentionality. If there is only one type of representation, then this position would have to say that all representations are dependent on (i.e. meaning reduces to) intentionality, including those used by sub-personal agents and calculators. (Cummins, 1989, 23-25)

Instead, I will follow the insight that it is more effective to do the reverse and base intentionality on representation; if a theory proposes that mental computations involve intentionality, it must be shown how the particular representational architecture employed brings about this intentionality. In this way intentionality comes to be what *distinguishes* "rational" representational systems from those that are simpler and do not lead to the development of intentional mental states. It is for this reason that I will investigate the nature of particular representational systems in the next section. However, there are a number of clarifications about what intentionality is and how it has been traditionally treated that must be addressed first.

The question that is of importance for a philosophical theory of meaning is as follows: if it is postulated that there are abstract mental representations used by the brain to make decisions, form beliefs, and generally deal with reality, what is it that makes these representations *mean* what they do? What makes these representations actually *about* one thing (real or imaginary) as opposed to something else, thus distinguishing representations from essentially meaningless placeholders? My claim is that a representation's

meaning cannot be understood in isolation from the whole of the representational system, a position somewhat similar (but only in minor respects) to what Cummins terms "Globalism." (1989, 25) As is often pointed out by functionalist theories of mind, the mental structures and categories that might be used by an information processing system do not necessarily coincide with any identifiable and indivisible structure on a lower level (for example, see Rey, 1996, esp. 176-178). This had led to the development of theories that construct mental architectures independent of biological research. This is the wrong conclusion to draw from their valuable insight, however. Perhaps the functional insight into the nonexistence of individually understandable representations can be better understood as indicating that representational structures only have meaning when viewed in relationship to the *entire* mental state space created during the lifetime of a representational system's development.

This view takes the original question of what makes representations actually mean what they do and makes it answerable only by answering the empirical question, what is the structure of the state space and how does it develop? A complete answer to this question is different for each individual system postulated; a simpler representational system with less flexibility will have a less developed representational space with representations that mean correspondingly less specific things. As an analogy, a mirror telescope with fewer mirrors will not have the resolving power of a telescope with more mirrors. Simpler systems, such as simulated networks, simpler animals and pocket calculators have less representational power and therefore have representations that *mean* less clearly definable concepts. A calculator's system is so simple that it can only represent abstract notions of quantity, lacking even the ability to represent a single quantity separate from the gross computational relationships its program encodes. A simple connectionist network has representations with elementary categorization abilities and is usually trained on a particular type of input. Therefore its representations come to represent relationships in the input set which are applicable to similar types of information, but will produce meaningless garbage if presented with an inappropriate input. Simple invertebrate nervous systems are able to encode relationships between incoming stimuli and motor executions; they develop representations that can separate out the appropriate degree and nature of response based on the degree and nature of the stimulus, but do so quite ignorant of what causes the stimulus.

Humans, however, have the ability to represent (and therefore develop beliefs and desires *about*) specific types of objects, specific instantiations of these types, relationships between objects and groups of objects, and even particular occurrences of the same object or relationship. What type of architecture can allow for this extremely rich detailing of world knowledge, yet remain flexible enough to allow for the learning of *new* concepts and *new* relationships? An example of how this question arises in cognitive science is the productivity of human language - almost every sentence produced by a person is one which has never before been uttered. Even if one ignores the fact that a person's vocabulary is continually growing, humans are able to produce an unending variety of sentences by recombining elements in novel ways using recursive application of structural rules. The success of the generative analysis of grammar to capture this productivity, coupled with the success of computational architectures encoding logical relationships to produce decision algorithms gives credence to the long standing philosophical claim that mental architecture can be understood in terms of logical computations involving symbolic placeholders. The mandate for com-

putational theories that rely on intentionality, then, is to show how these symbolic tokens gain and retain reference to actual objects. If these symbols can be systematically related to instantiations of real world referents, then a theory that details the formal treatment of these symbols has succeeded in describing rational thought.

Attempts to provide a systematic relationship between symbolic representations and their referents have been quite successful at providing structure, evaluative tools, and theoretical support for intentional theories; they have not, however, successfully demonstrated that such a system can indeed have meaningful symbolic representations. From theories that appear elegantly simple to others that are quite complex, each of these theories has difficulties with at least one of several related problems. Misrepresentation, the disjunction problem, the distality problem, and the indeterministic nature of cause and effect relationships are all problems with which intentional theories must contend. To illustrate how these issues relate to intentionality and each other, I will outline the general nature of a simple intentional story as it applies to computational theories of cognition and its subsequent development in response to challenges.

Causal/information theories hold that something is a representation of an event or object in the world if that thing causes a tokening of the representation in the mind. Assuming that the natural relationship between cause and effect holds and that the causal chain of events leading from the presence of an event to the tokening of a symbolic representation can be specified, this simple axiom does a good job of explaining why a representation with a specific content is present in a system. One of the earliest problems seen with causal theories is their apparent inability to provide for the possibility of misrepresentation; an effective intentional theory must also be able to account for errors.

To explain why this is so, it is best to formally state how a causal theory assigns reference. As summarized by Robert Cummins, the central idea of causal theories (and intentional theories in general) is that saying some token " x represents y " in a representational system is the same as saying " x . . . occurs in a percept when, only when, and because . . . [such a system] is confronted by y (whiteness, a cat, whatever)." (1989, 38) According to such a theory, however, a system can not misrepresent a concept. There are instances when, upon seeing a dog, we might say it is a cat. This is in response to the tokening of CAT, where the uppercase letters denote an abstract symbolic representation. However, since causal theories state that a representation is a representation of what causes it, the representation tokened when a dog is seen must be a correct representation of a dog. It appears that this representation must have disjunctive contents; since both cats and dogs can cause the representation, the contents of this representation is CAT OR DOG and there has been no error. Misrepresentation is an incoherent notion in this simple causal theory. (Cummins, 1989, 40)

In order to provide for the possibility of misrepresentation, it must be the case that something which is not in the extension of x (i.e. is not one of its referents) can lead to a tokening of x . A successful theory of content will provide a way for defining the contents of a representation as the simpler CAT over the more complex disjunctive concept CAT OR DOG. If a theory can deal effectively with the disjunction problem, it is capable of misrepresentation; this is indicative of a theory's power to have representations with intentional contents. (Millikan, 1993, 123; Fodor, 1990, 60; Dretske, 1988, 65) To this end, several theorists take the causal account of content and bring additional relationships to bear on the status of a representation.

Although most theoretical frameworks do not neatly fall into exactly one of the following categories, I present them as distinguishable for the salient points they uncover about intentional representations. The three categories addressed are covariation, teleology, and contextual/background information.

The basic tenant of several theories is that the covariation (more strictly defined) between a representation in the head and its referent is satisfactory for giving representations intentional contents, and the disjunction problem is avoided by privileging certain situations that guarantee correct fixation of content. Examples of this tactic include Fodor's asymmetrical dependence theory (1987) and Dretske's treatment of associative learning in his essay entitled "Misrepresentation" (1986). In Fodor's scenario, a disjunction arises when some link between an object in the environment and the tokening of a (wrong) representation in "Mentalese" (Fodor's Language of Thought hypothesis states that mental representations are language like symbols) is "parasitic" on an already established connection between that representation and its correct referent. (Cummins, 1989, 58; Fodor, 1987, 107) When shrews cause mouse representations, it is because mousy properties cause MOUSE and shrews have mousy properties. However, the same relationship does not hold for mice since they are not dependent upon shrews for their ability to cause the tokening of MOUSE representations but only on their own possession of mousy properties. By basing the underlying representational schema on certain causal relationships and attaching instances of misrepresentation to these underlying structures, Fodor hopes to retain the original contents of disjunctive concepts.

This solution to the disjunction problem is not really a solution at all. It eliminates disjunction by idealizing the learning process as something which can develop stable representations amid a host of non-ideal possible situations. Cummins details several of these possible situations in which the fundamental connection that is supposed to obtain between mice and MOUSE cannot be said to be more fundamental and independent of its dependent relationships. (Cummins, 1989, 59) If, to take Cummin's example, a person looking for mice to use in the creation of tribal potion knows that mice and shrews are different *but has only actually seen shrews*, there might be a mouse to MOUSE connection and a shrew to MOUSE connection (and possibly also equivalent connections to a SHREW representation, but this is irrelevant). The shrew to MOUSE connection is dependent on the mouse to MOUSE connection according to asymmetrical dependence, yet the mouse to MOUSE connection, "given the way things are learned, . . . wouldn't exist if it were not for the connection between shrews and |S| (MOUSE)." (1989, 59) This breaks down the alleged asymmetry and produces a case of disjunctive contents once again; MOUSE seems to be expressing the property of being a shrew or a mouse. Another way of phrasing the problem is that the set of properties which are actually useful for distinguishing the correct referent from the incorrect referent (i.e. the set of properties that is somehow missing or ignored when misrepresentation occurs) would also have to be privileged during learning in order for a representational system employing asymmetrical dependence to maintain separate MOUSE and SHREW representations. If these properties are not known then the contents of these representations, even if there are two separate "placeholders" at some level of description, can be described as MOUSE OR SHREW.

Fred Dretske also proposes a method for defining content fixing situations in his essay entitled "Misrepresentation," (1986) an explicit attempt to overcome the disjunction problem in covariance theories.

He states that a certain amount of complexity is required in an information-processing system before it can be expected to have a "clear and unambiguous capacity for misrepresentation . . . and the dependent capacity for belief." (Dretske, 1986, 139) As he defines what kind of complexity is required he comes to the conclusion that a representation can only be assured of having non-disjunctive concepts if the system in which this representation is tokened is capable of "some form of associative learning." (Dretske, 1986, 141) There may be more than one environmental stimuli which signals the presence of a certain event in the environment, and a system that is able to detect more than one proximal stimuli as indicative of this event is more complex than one that can not.

A theory that allows for more than one stimuli to token a representation can no longer say that the presence of such a representation indicates the presence a particular stimuli. If representation R covaries with the presence of either proximal stimulus *a* or proximal stimulus *b* (say that both are indicative of event E having occurred), then a tokening of R does not *mean* that *a* alone is present, or that *b* alone is present. It could be said to mean that *a or b* is present, or it could instead be said that R does not indicate either of these and only reliably indicates the presence of E. Dretske states that this is reason enough to suppose that the representation truly indicates the presence of the event and not the stimuli, yet admits that the contents of the representation may be considered as disjunctive since the representation can be tokened by more than one proximal stimuli. (Dretske, 1986, 141) The contents of the representation will always be the invariant and disjunctive set of proximal stimuli that causes the representation to be tokened.

If the system is capable of associative learning, however, it is able to add to this set of proximal stimuli at any time. Dretske argues that therefore a representation in this system does not indicate the presence of any *fixed* set of stimuli. He claims that the only time-invariant meaning that the representation can have, and hence the only thing it can really indicate, is the actual event. This is one example of a theory that does not rely explicitly on the causal chain of events to fix contents, but instead applies to a more abstract notion of what representations are for or possibly why they developed. The general story is that representations are useful to an organism because they are designed to indicate things in the environment. The environmental event is what the representation is *supposed* to mean, and any failure of the learned stimuli to correlate with the presence of the event does not affect the time-invariant relationship between the representation and the event it is supposed to indicate. (Dretske, 1986, 142) I will further address what is meant by this relationship when I discuss teleological theories, which also rely on a notion of what representations are supposed to do.

Dretske's explanation ends quite abruptly at this point, with almost no explanation of how this associative learning occurs. Much like Fodor's asymmetrical dependence theory, an associative learning mechanism that deals with a new environmental stimuli depends on a reliable covariation between this newly encountered stimulus and the environmental event; it idealizes the learning situation. In reality, the presence of a stimulus does not always indicate the presence of the event (if misrepresentation is ever to happen) and the presence of the event does not always assure that this stimulus is present (or else there would be no reasonable explanation for why a representation could be tokened by any other stimulus and hence no reason for associative learning), meaning that the stimulus does not reliably covary with the environmental

event. Yet Dretske's associative learning mechanism depends on a reliable covariation between a newly recruited stimulus with the desired environmental event.

In later writings Dretske is not completely consistent in his support of what a representation is supposed to indicate (either a more proximal stimuli or more distal event), but he does provide some answer to how this relationship is supposed to hold. He says at one point that the covariation between a stimuli and an event must be correlated at some statistically significant level, making the event the "*maximally* indicated state" (italics mine), in order for the association between the stimuli and the representation to be learned. (Dretske, 1990, 826) In other places, he insists on complete indicator to indicated reliability (i.e. a statistical probability of 1) and that what is indicated is the most distal cause, stating that this is the only way to allow for misrepresentation since it is otherwise "impossible to fool the organism." (Dretske, 1986, 139; Dretske, 1988, 68) The tradeoff is between how successfully a theory allows for misrepresentation and how successfully it allows for associative learning. In either case, misrepresentation occurs when this coextension breaks down and the sign does not covary with the indicated environmental state. These are situations in which the detector of this environmental cue "is no longer capable of indicating what it is supposed to indicate." (Dretske, 1988, 68)

This proximal/distal distinction explored by Dretske is related to what is often called the distality problem: if A reliably causes B and B reliably causes representation R, is there any reason to say that (1) the tokening of R indicates B but not A since B is what actually caused R to be tokened, as opposed to (2) that R actually indicates A since it is the initial cause? The distality problem is a specific example of a larger philosophical problem with causality. Cause and effect relationships are not strictly linear: there are both divergent and convergent causal structures. In addition, there is the possibility that some events do not necessarily have causes but are instead the result of stochastic, random processes. I will refer to these problems with causal stories as the *indeterminate* nature of causality. Some theorists view this point as a form of skepticism that any theory of representation can effectively deal with misrepresentation. However, in the third section of this paper I will instead take the viewpoint that certain types of mental architecture are better suited to dealing with this indeterminacy than others.

Despite their problems, covariance theories do introduce several concepts which seem necessary for developing a powerful intentional theory. If the goal is to develop a successful theory of meaning, the issue of how a system learns and uses learned facts to modulate its performance needs to be addressed. Dretske's use of associative learning emphasizes that tokenings of representations do not occur in temporal isolation. Over a system's lifetime, it encounters a vast set of data about the external world and can bring this explicit contextual information about the regularities seen in natural environments to bear on how it operates in a given situation.

Before addressing theories that explicitly use this type of background information, I will return to the notion of function mentioned in reference to Dretske's associative learning and the wider use of functional relationships found in teleological theories of content. At this point in the development of computational intentional theories, it has proven difficult to fix the content of a representation (i.e. there is no principled reason for believing that the content of a representation is *x* as opposed to *y* or the disjunctive *x or*

y) and allow misrepresentation as long as the work of actually mapping contents to a representation resides only with the representation and its causal relations to events in the world. Perhaps something can be found which fixes the content of a representation by applying only to natural kinds (i.e. categories which reflect actual or "natural" types, such as heat), since natural kinds are non-intentional and the goal of intentional theories is to ground intentional contents in non-intentional terms. As Ruth Millikan puts it, the problem is to understand "what *bare* representation is, and then what being true or false is, over and above bare representation . . . without introducing ad hoc abstract objects, say unanalyzed meanings, senses, propositions, or possible states of affairs, as somehow ingredient in nature." (Millikan, 1993, 123)

Teleological theories (also called teleofunctional theories in some contexts) seek to define a relationship between internal states and the external relationships explicit in natural kinds by positing that something in a system has the function of indicating external things. No longer are internal events merely caused by external events, but specific things internal to the system have the role (or function) to indicate the presence of external things. These theories state that representational systems are able to extract important information about the environment like the time of day, for example, from a natural sign in the environment such as the position of celestial bodies. The relationship between the position of stars and the time of day is one which is guaranteed by the motion of the earth to be correct, yet is only useful to a system which is able to extract such information. Functional theories posit that this requires not only experience and learning, but also the existence of something which has the function of indicating this relationship. The position of a constellation is not a meaningful indicator of time of day since it is not the function of that constellation to indicate time of day, but an internal representation of the time of day based on this relationship requires the presence of something whose purpose it is to extract and represent this information within the system.

Teleological theories of content introduce the idea of natural selection as a possible way to explain the presence and usefulness of particular functions within representational systems. Individual functions develop within and are used by a functional architecture because of their contribution to the welfare and survival of an organism and continuation of a species. Systems that have the ability to detect environmental stimuli are better suited to existing in a particular niche and the genetic traits that produce the mechanisms responsible for detecting these environmental events persist in a species. A trait's usefulness to an organism contributes to the reproductive success of that organism, ensuring that the trait will be more widespread in future generations. This trait is useful because it is responsible for the presence of successful indicators. It is thus the function of this mechanism to indicate the presence of an environmental event; that is what it was *recruited* to do, that is what it is *supposed* to do. The selectional history over many generations of the genetic traits that acquire this function, this representational status, determines whether or not it represents a certain event and therefore determines the contents of a representation.

Ruth Millikan, who uses teleological explanations (though not exclusively) to give an account of the contents of representations, states that natural selection acts upon a system with three distinct parts. (Millikan, 1993, 124-126) First, there is the representation producer. The producer is similar to the mechanisms described by Dretske; it is the mechanism that *produces* a representation as a result of some environmental event. Second, there is the representation itself. This representation is then *consumed* by the

consumer, which performs its function "in, or via mediation of, or despite, [condition] *c*." (Millikan, 1993, 129) Each of these three items acquires a function, and these items develop conjointly via natural selection to work with each other. The producer's function is to "*produce* a representation that indicates" (p. 128) where "to indicate" means that the representation is in a correspondance relation with something in the environment. The representation's function is to represent an environmental event *to* a consumer, and the consumer's function is to produce an effect in a manner modulated by the presence of a representation.

Millikan claims that the introduction of a representation consumer gives her leverage to deal with the indeterminacy that Dretske's theory faces. She points out that the reliability of a representation producer (or any mechanism) cannot be part of the definition of the producer's function, since "no item *effects* [i.e. is a determiner or evaluator of] its own reliability." (Millikan, 1993, 130) Instead of saying that the representation producer produces correct representations a certain statistically significant number of times (i.e. reliably) or mandating complete reliability of indicator/indicated relationships, she states that the function of a representation producer is to produce representations that allow for the proper functioning of the consumers. (p. 130) A producer may fail in its function (either mechanically or because of a breakdown in a natural covariance) and produce a representation that is false, yet the consumer will use the representation as if it were true. The historically defined relationship between the producer and consumer fixes what it is that the consumer expects of the producer; the consumer operates as if on the assumption that the producer produces correct representations, and only in the correct contexts. The representation produced by the producer *misrepresents* the state of affairs *to* the consumer, which is blind as to how and why the representation is tokened. The contents of the representation is fixed by the *very existence of the relationship* between a producer, a representation, and a consumer. This relationship exists because it is useful to an organism and survived the selectional pressures of an environment over many generations.

Although this framework seems to convincingly allow for misrepresentation of some sort, it still leaves something to be desired. The first objection to these teleological theories is that they still don't explicitly deal with what I have termed indeterminacy. Millikan's teleofunctional architecture abstracts away from explicit causal chains and in this respect better captures the abstract notion that representations do not gain meaning purely in virtue of immediate environmental conditions. But at the same time the nature of representation in her theory still depends too strictly on how *individual* representations have acquired meaning through a selectional history and ignores the simultaneous development of *competing* representational structures that results in a unified world representation. All parts of a representational system must be consistent with each other for representations to truly have intentional contents, and Millikan's story about the creation of teleofunctions provides no account for how this cohesiveness is obtained. Second, although I did not state this while outlining her arguments, to a large extent Millikan depends on acquisition through public language as the method for shaping the functions of representation producers and consumers. (Millikan, 1993, 133) A concern I will address more fully in the next section is that basing the formation of mental architecture on an organism's ability to construct or comprehend linguistic architecture both limits the type of representation that can occur and presupposes that linguistic categorization coincides with real categories. In addition, both the positing of functional roles and the creation of mental architecture from lin-

guistic structure seems like a case where representational structure is being grounded in intentional structure, whereas the goal of intentional theories as outlined above is to ground intentionality in representation. Millikan runs the risk here of violating her own mandate to avoid "ad hoc abstract objects" in the development of a theory of bare representation. Finally, the relationship Millikan details between consumer and producer requires extensive and specific reinforcement throughout a selectional history. Although she does address this issue to some degree when she discusses associative learning and rule application, her reliance on natural selection to define functional relationships means that she is building a large amount of innate structure about specific representations into a mental architecture; she has in some sense "hard wired" the roles these agents can play.

Additionally, there is still some doubt that any functional theory can give sufficient criteria for selecting one content (such as MOUSE) over another, coextensional content (like MOUSE OR SHREW in a case where only mice are present in the environment). If in an organism's environment, two things are coextensional, both could equally well be the thing a representation has as its function to indicate. (Fodor, 1990, 73) Dretske and Millikan argue that the content that is correct is the one that describes a connection that actually aided in the selection of the mechanism; a representation described to have the contents MOUSE OR SHREW has no selectional advantage over one that is described as only meaning MOUSE, and it seems ridiculous to suppose that a disjunctive representation has some *a priori* selectional advantage over a non-disjunctive representation in this case. Yet the point is that a representation with the contents MOUSE also has no advantage over one with the contents MOUSE OR SHREW in being selected from generation to generation or during learning since they are coextensional, and there is no principled reason for selecting one coextensional thing over the other. Millikan's reply is that the simplest explanation of a representation's content is correct (in this case MOUSE), but there doesn't seem to be anything systematic about this reasoning. (Millikan, 1993, 221) As Fodor points out in his "Darwin doesn't care" argument, the label under which mice get recognized doesn't matter for natural selection as long as they do get recognized, and thus applying to natural selection doesn't seem to get us anywhere towards solving the disjunction problem. (Fodor, 1990, 73) The important aspect to remember about teleological theories, despite their problems, is that representational systems do have selectional histories, and new aspects of the environment that are important for survival will influence what organisms survive.

Similar to these teleological theories, theories that make use of the extensive background and contextual information available about an intentional representational structure address the nature of learning and development in a clearer fashion than do previous theoretical frameworks. Representational systems have at their disposal information about previous experiences with events in the real world as well as contextual information about a particular event and background information about the methods a system uses to create mental representations. It is possible that misrepresentation occurs in cases when this purely internal knowledge is improperly applied or missing during the production of a particular representation as opposed to when the normal covariances observed in nature break down. Knowledge about a representation's structure and function on a microscopic level (i.e. at the level of simple percepts or patterns of sensory layer activity) in a system may help in defining more rigorously exactly what causes a representation and how they

are structured. Explicit information early enough in the causal chain leading to a tokening of a particular representation may distinguish the actual referent (and hence contents) of a representation as a non-disjunctive concept. For example, many neuroscientists believe that intentionality (inasmuch as philosophical intentionality and what is known in other disciplines as the binding problem are the same thing) is achieved by the matching up of temporally similar rhythmic firing patterns in different parts of the brain. Cummins argues that even if a system's simple percepts may be enough to provide extremely detailed, high resolution information on the nature of an event and the context within which it takes place, deriving the contents of a representation from simple percepts (i.e. proximal stimuli) in a bottom up manner is an extremely difficult and tedious task from both a research and theoretical perspective. (Cummins, 1990, 48-49) Current advances in neuroscience perception research have made this problem seem much less formidable, but there is no assurance that an understanding of all the individual neural events that lead to a tokening of a representation necessarily entail an understanding of a representation's content.

Pursuing this line of argumentation basically amounts to supporting the notion that mental architecture should be investigated in light of its biological implementation. The best way to develop a theory for how a representational system might use contextual and background information is to consider how a representational system stores and encodes background information. Artificial intelligence research has led to the development of a large number of architectures for storing and manipulating this type of information. Its successes and failures support the idea that *certain architectures are better than others* for sustaining the type of representation sought to produce intelligent behavior, despite the fact that all computational architectures are at some level theoretically equivalent.

Throughout this section, the discussion of intentionality in computational mental architectures of representation has taken place without a consideration of the actual structures of the representations themselves beyond the initial assumption that they are symbolic. After this consideration of computational intentional theories in general, it is now necessary to examine the nature of particular representational systems with the belief that this is the best strategy for first detailing the nature of representation and subsequently demonstrating the ability of a mental architecture to develop and support *intentional* representation