

III. Indeterminacy

In section one, I gave a brief definition for what I have termed the indeterminacy of causality. In this section, I will redefine and clarify this notion, giving examples of how indeterminacy in both input and output procedures requires a representational system to incorporate unspecified structure into its mental state space. I will then show how misrepresentation as dealt with in the last section is a coherent concept in connectionist architectures, and outline constraint satisfaction as the method by which connectionist architectures build mental state space that can support unspecified structure. Finally, I will summarize how such a system supports intentional representations.

In her book *Neurophilosophy* (1986), Patricia S. Churchland tries to do something very similar to the goal of this paper; she tries to develop a framework to support the idea that philosophy of mind and neuroscience can be mutually beneficial to each other. As part of this goal, she outlines the development of the philosophical treatment of science. One theme that surfaces repeatedly in this history is the possibility of "Unknowable Reality" (Churchland, 1986, 249) Associated largely with Kant, this idea stems from both philosophical and mathematical concerns about the ability of any theory to completely describe its intended domain. For example, Gödel's Theorem states that for any formal system, there is always something outside of the system which cannot be explained by the formal system's theorems and axioms. For Kant, the primary concern was that "all knowledge is mediated by the mind and its sundry faculties, which impose a structure and an organization on the input." (Churchland, 1986, 248) This concern clearly has consequences for a theory of intentionality; I will argue that it can generally be construed as the same problem I have termed indeterminacy.

The connectionist mental architecture outlined in section two is one which allows for the construction of a mental state space in response to the interplay between continually impinging environmental input and innate structural characteristics. Broadly construed, this view of mental representation agrees with a comment made by Churchland & Sejnowski in *The Computational Brain* (1992): "that is, *brains are world modelers*, and the verifying threads--the minute feed-in points for the brain's voracious intake of world-information--are the neuronal transducers in the various sensory systems." (italics mine, 143) One way to view this modeling process is to see distributed representational systems as proposing and testing theories about regularities in the input. A representational system then faces the same theoretical difficulty when it constructs a model of the world as does a philosopher or mathematician when trying to describe some aspect of reality; a representational system *cannot* construct a model that encompasses all inputs in such a way as to *prove* that its model of the world is correct. Subsets of this model may be absolutely consistent with each other on repeated testing, but no bit of information can ever verify the theoretical framework constructed by a representational system *as a whole*. For the same reasons as Descartes, Kant, and countless others, a representational system will always be at a loss for confirmation that what is hypothesized to be true about the world is an accurate statement of how things are.

This possibility seems to undermine the claim that a representational system can support intentional representations. In one sense it does; but as will be seen, this lack of confirmation doesn't really mat-

ter for a representational system's ability to accurately model its environment. All these related philosophical concerns stem from this same uncertainty. Examples include: the possibility that a different environmental event than the one inferred could be the actual cause for a proximal stimuli, the possibility that what is seen as a moving object is actually a series of still images, the possibility that people see colors differently, and the possibility that an evil demon is deceiving you and you are actually nothing but a figment in someone else's mind. Yet under most circumstances, these concerns do not interfere with the way people interpret their environment. This is because of the vast amount of supporting data that the brain accumulates over a lifetime. Much like theory testing in science, a theory positing a certain mental structure is taken to be valid if it withstands an exhaustive amount of testing, despite the fact that there remains a possibility that the theory might in fact be incorrect.

This testing and revision process takes place at all levels within a mental model's architecture. Long standing, formative (i.e. "older", heirarchically dominant) mental structure can be changed just as can more recent, local mental structure. Examples of these two extremes would include an assumption that surfaces are continuous and that a sound just heard came from the left, respectively. However, the importance of long standing global structure means that it will change only in the face of vastly contradictory evidence; even in this situation, the preferable move is to undermine the contradictory evidence or try to somehow explain it within the existing model. When a formative aspect of mental structure is changed there are large consequences for the interpretation of heirarchically dependent structure; the local relationships between structural elements remain the same, but they take on a completely different interpretation. For example, information that leads to a reassignment of which direction is north will not change the belief that two buildings are next to each other, but it will change a belief about which one is further east.

As is evident from the example above, there must be some way for a mental architecture to encode relationships at lower, more local levels without having to first specify how these local relationships fit into a higher, heirarchichal structure. Similar to the process of completing a jigsaw puzzle by first completing small, unattached sections, it appears that a representational system can fit together coherent pieces of information in the absence of information of how these sections fit into a larger structure. I propose that a representational system has the ability to leave these heirarchichal relationships unspecified under certain situations. Below I will show how this ability can result from the interplay between a set of constraints operating in a constraint satisfaction paradigm implementable in a connectionist architecture.

This ability to have underspecified structure explains how mental architecture is able to deal with indeterminate causality. When the cause of an event is uncertain, either an informed guess is made about the cause (as in a forced labeling situation like that detailed in the last section) or some aspect of the mental model is left unspecified, creating what might be called "dangling" or "unattached" substructure. An example from morphology will serve to demonstrate the link between unspecified causal structure in an input (indeterminacy) and underspecified structure in mental architecture. Many root words show alternations when followed by a suffix, yet during comprehension processes both input forms activate the same semantic representation within mental state space. As was pointed out in section two, no two specific inputs will ever occupy exactly the same points in state space; however, there does develop characteristic attractor

bins in response to the varied contextual distribution of an input. For example, both [veyn] ("vain") and [væn] (in "vanity") activate underlying /vÆn/, where Æ indicates an underlying vowel unspecified for tenseness (from Marlsen-Wilson et al., 1994, 8). A representational system repeatedly exposed to these forms will develop an attractor bin that groups both of these stimuli together so that they function as a unitary semantic concept. The activation of the mental representation for this concept in response to either stimulus reveals the converging (and therefore indeterminate) causal structure of their relationship. A characterization of this relationship relies on the underspecified element Æ, such that *the mental representation is unspecified for tenseness*. The subsequent transformation of this underspecified representation for output procedures must do the mirror of what occurs during the input transformation - it selects one or another surface form depending on contextual information.

Misrepresentation, then, is equivalent to underspecified mental structure. When a cat on a foggy night is believed to be a dog, it is because the subset of features that normally distinguishes cats from dogs is underspecified in a distributed representational system. However, there must be a principled way to distinguish between two cases: those in which a representational system defines a possibly incorrect default structure, and those cases in which the structure remains unspecified for longer periods of time. The motivating factor for this distinction should be how important having specified structure is for subsequent processing; a representational system must be able to differentially recognize the relative urgency of making such decisions.

A connectionist network encoding a violable constraint satisfaction paradigm is capable of weighing these conflicting impulses. Although the constraints I will propose are quite general and not geared towards any one particular constraint satisfaction paradigm, a brief introduction is necessary. Constraint satisfaction networks are designed to find the best of a set of possible solutions to a problem. For a particular problem, a set of constraints are defined which embody the salient aspects of the solution. For example, a possible constraint for a network modeling the phonological patterns of a language states that "Syllables must have onsets (except phrase initially)." (Prince & Smolensky, 1993, 16) A possible output for an input/output mapping is "well formed" if it best meets the constraints defined for the mapping. Constraints are violable and are ranked according to their importance; a possible output that violates only the top ranked constraint and none of the rest is *not* better than one that violates every constraint *except* the top ranked constraint. By specifying all necessary constraints and the order in which they are ranked, a representational system using violable constraint satisfaction can be encoded in the weights of a connectionist network and used to perform an input/output mapping.

Many varied styles of constraint satisfaction have been used for different tasks, but from a theoretical viewpoint the key possibility these systems present is that constraints of all types -- those that are specifically linguistic, those that are involved in motor commands, those that regulate awareness, etc. -- can interact with each other simultaneously. A possible output of a representational system is evaluated according to how well it satisfies *all* the constraints of the problem, with no limits on what the nature of these constraints might be. To address the issues of misrepresentation and indeterminacy as raised above, there are two constraints that will serve to illustrate how aspects of a representational system that are not modality

specific can cooperate with constraints involved in the processing of particular stimuli.

1. Specify all structure completely.

This first constraint is basically a mandate to rely on default mechanisms. In syntax, for example, unspecified structure is costly due to the importance of compressing the highly branched input stream. Such mechanisms as minimal attachment and late closure operate under these default conditions to assign a structure when there is a lack of immediate evidence that could definitively specify one of two possible structures. When such a constraint is obeyed, a representational system assigns a default structure and moves on.

2. Specify only structure that is well supported, even when this may delay processing.

This second constraint leaves structure unspecified in cases where there would be a high cost for making an assumption in subsequent processing. This constraint operates like a cost function to evaluate how important it is to have a particular bit of structure specified. In some cases, it may be better to wait for more evidence or simply move on without completely specifying all hypothesized relationships when subsequent computations do not require absolute specification or are non-essential. Such is the case for many simple observations; if no important decision rests on the resolution of conflicting possibilities, it is usually preferable to *not* make a default assignment and move on.

The operation of these constraints is obviously dependent on a host of other factors that inform a representational system about the urgency of a particular decision. If something like these constraints are hypothesized to exist in actual nervous systems they would obviously operate largely on a subconscious level. If there is the possibility that not making a judgement may cause harm, a default (safe) assignment will be made. However, if the stimulus presents no threat either way, or if further information may lead to a benefit for an organism, a representational system may postpone interpretation of an event. These constraints interact well with a notion used in Smolensky's work on symbolic/ connectionist hybrid networks which he terms *harmony*. Harmony is a property of a connectionist network that encodes how well the constraints of a system are being met at a particular time (Smolensky, 1995). Both 1 and 2 operate to increase harmony, but in some situations they conflict. In these cases, there is a choice between an increase in the local harmony of a system that would result from specifying structure completely and an increase in the global harmony of a system that would result from avoiding a costly assumption.

The use of violable constraint satisfaction gives an intentional representational system the ability to re-evaluate decisions based on their importance. Coupled with extensive background and contextual information compiled over many presentations of stimuli and the subsequent encoding of regularities in the ordering of constraints (through the adjustment of connection weights), such a representational system is able to construct a model of the world that is internally consistent and for most purposes accurate in its representation of real world events. Underspecified structure is either given a default assignment or left unspecified depending on the needs of subsequent processing, and theorized relationships either gain credence or are discarded based on new information.

Coupled with the pertinent aspects of previous treatments of intentionality covered in section one and my investigation of intentionality in connectionist representational structures in section two, this

discussion of indeterminacy gives a fairly accessible account of how intentionality is relevant to cognitive modeling. To clarify, intentional representations result when a representational system is complex enough to allow for the following things:

An assumption that the world is predictable. As the above discussion points out, this point is by no means a given. Yet it is when this assumption is *not* made or comes into question that representational structures appear to be arbitrary and meaningless.

Construction of state space structure consistent with small sets of input. Local circuits in the early stages of perceptual systems develop connections that allow for salient features of stimuli to become prominent in higher processing stages. Simple networks can be trained to perform these mappings with little or no supervision. Previous philosophical frameworks have failed in their attempts to link intentional representations to proximal stimuli because they sought a direct, local connection between these things. Instead, a connectionist system has intentional representations based on its *global* coherence, which is a product of the constant interplay between a system and proximal stimuli that takes place during development.

Construction and constant re-structuring of larger areas of state space out of smaller structural relationships. As opposed to working from large, poorly defined psychological concepts, a theory that depends on the global development of a connectionist representational system to get intentional representations shows how local interactions between existing structure, new information, and hypothesized relationships lead to coherent global structure.

Use and testing of theorized state space structure on new stimuli. Intentional representational systems never stop developing. New data will either conform with and strengthen a representational system's mental architecture or will provide reasons for re-evaluation and possible re-structuring.

These four points illustrate how important the specific nature of a particular representational system is to answering the question of how intentional representation in general is achieved. Despite the theoretical equivalence of computational architectures, only one which remains both globally consistent and flexible *no matter how temporally extended the development period may be* can allow for the development of intentional representations. By giving a glimpse of how performance and learning can be collapsed into a

single process, connectionist architectures have shown how intentional representations can be both globally dependent yet grounded in the local interactions investigated by neuroscientists. Hopefully the successes of connectionist modeling will fuel increased interplay between scientific studies of intelligent systems and theoretical searches for meaning.