2

Representation in classical computational theories: the Standard Interpretation and its problems

In this chapter I reveal what I take to be a popular set of assumptions and tacit attitudes about the explanatory role of representation in the CCTC. I'll suggest that these assumptions and attitudes collectively give rise to an outlook on representation that amounts to a sort of merger between classical computational theory and folk psychology. In other words, the way researchers and especially philosophers have come to regard the importance of representations in the CCTC has been largely determined by their understanding of beliefs and other commonsense notions. This has led to a way of thinking about computational representations that suggests their primary explanatory function is to provide a scientific home for folk notions of mental representations. In other words, symbolic representations in the CCTC have come to be viewed as the scientific analogues for beliefs, desires, ideas, thoughts, and similar representational posits of folk psychology.

This perceived connection between computational representation and folk psychology comprises what I will refer to as the "Standard Interpretation" of the CCTC. While few authors have explicitly stated the Standard Interpretation, at least not in the manner in which it will be presented here, it (or some version of it) has nevertheless played a significant role in shaping the way many people think about the CCTC. After spelling out what I think the Standard Interpretation involves, I'll try to show that it leads us down a path where, despite various claims to the contrary, we wind up wondering whether the symbols of classical models should be viewed as representations at all. This path has been illuminated by two important skeptics of CCTC representationalism, John Searle and Stephen Stich. Searle and Stich both exploit the alleged link between the CCTC and folk psychology to challenge the claim that the CCTC can or should appeal to inner representations. Searle does this by denying that classical symbols can capture what is essential about thoughts, while Stich does this by arguing that beliefs and other folk notions don't belong in a

serious scientific account of cognition. Both criticisms can be seen as ways of making the more general point that if classical symbols are to serve as *reducers* for folk notions, the job description challenge goes largely unanswered.

My own view is that the Standard Interpretation has clouded our understanding of the explanatory work that is *actually* done by the notion of representation in classical computational theory. The mistake is in thinking that computational symbols need to capture ordinary folk psychological notions in order to qualify as representational states. In the next chapter, I'll argue that these problems disappear once we abandon the Standard Interpretation and recognize that the sort of representations invoked by the CCTC are not based directly on folk psychology, but on the kinds of explanatory strategies used by the CCTC. For now, however, my aim is only to present the Standard Interpretation and show how its defense of representationalism runs afoul. To show this, I'll first present some of the basic principles behind the CCTC. Then I'll show how on the Standard Interpretation, computational symbols come to be linked with notions like belief. Finally, I'll present Searle's and Stich's criticism of representationalism, and examine the ways defenders of the Standard Interpretation have responded. We'll see that in the final analysis, the need to treat computational states as representations is left in serious doubt.

2.1 THE CCTC AND THE STANDARD INTERPRETATION

The central doctrine of the CCTC is that cognition is computation, which is itself to be understood as a form of quasi-linguistic symbol-manipulation done in accordance with specifiable rules. John Haugeland captures the basic idea this way: "Reasoning (on the computational model) is the manipulation of meaningful symbols according to rational rules (in an integrated system)" (Haugeland 1985, p. 39). Over the course of the last forty years, there have been many different theories and hypotheses that could be described as part of the CCTC. Many of these differ dramatically in styles of processing, information storage strategies, types of basic algorithms and representational forms. Yet as a group they share the core idea that cognitive systems carry out various tasks by shuffling, connecting, storing, repositioning, comparing, and in other ways maneuvering primitive and complex symbol tokens.

By now, this way of describing the CCTC has become something of an extended slogan, and, like many slogans, it is far from clear just what it all is supposed to mean. For instance, what does it mean to say a computational

system does "symbol manipulation"? Who or what manipulates the symbols and how does it know which manipulations to perform? What determines what the symbols are about, and how do they get their representational content? Moreover, how is such a process supposed to account for cognition?

Some of these questions have been largely ignored by advocates of the CCTC, or at least have been thought to express "mere implementation" matters. For instance, in most computers, symbol manipulations are carried out by some sort of central processing executive. This is typically a sophisticated device that not only keeps track of which operations need to be executed, but actually performs many of the operations as well. Yet the issue of how this is done in actual brains is generally passed over as something to be determined by neuroscience researchers. Other questions are seen to involve uninteresting technical details or even mere terminological quirks. For example, what it means to say that symbols are manipulated is just to say that symbols are erased, written and rewritten in various registers. This erasing and rewriting is what "symbol manipulation" refers to, along with the corresponding vernacular of symbol "shuffling," "combining," "rearranging," and such.[1] Thus, the CCTC claims that there is something like a neural code in the brain and symbols are "written" and "erased" by using this code.

In contrast to these somewhat neglected questions, other matters have been given far more attention and are properly treated as more central to the way the CCTC accounts for mentality. The most significant of these concerns the ways in which a symbol-manipulating paradigm is believed to provide compelling explanations for many of the central features of cognition and intelligence. Indeed, it is with regard to this matter that cognitive science has arguably enjoyed its greatest degree of success. Many would say that the CCTC has provided us, for the first time, with a scientifically robust account of human thought. Fodor describes the central "deep and beautiful idea" underlying the CCTC as "the most important idea about how the mind works that anybody has ever had" (Fodor 1992). Because there are now several superb introductions that describe the basic mechanics of the CCTC framework in considerable detail, what follows is a somewhat truncated synopsis of how the story generally goes.[2]

[1] See Copeland (1993), chapter 4, for an excellent discussion of how the erasing and writing actually get done in computers.

[2] See, for example, Block (1990); Boden (1977); Copeland (1993, 1996); Crane (1995); Harnish (2002); Haugeland (1985); Newell and Simon (1976); and Robinson (1992).

2.1.1 Levels of analysis and mechanized inference

The overarching explanatory picture of cognitive science is one that treats cognition as an input–output process that admits of different levels of analysis. This picture often reflects David Marr's conception of cognition as having three distinct levels of explanation or analysis (Marr 1982). At the top level is the specification of the explanandum – the cognitive task that we are attempting to explain. Marr calls this the "computational" level,[3] where the specification is typically an input–output function expressing the cognitive process or capacity, sometimes called the "task domain." These capacities are functions that can vary from converting numbers into products, sentences into parse trees, playing chess, constructing a medical diagnosis, and so on for a variety of things we think cognitive agents do. It is popular to regard the input as the proximal stimuli to a perceptual system and the output as some sort of motor movement, but the inputs and outputs can be any types of states that define a cognitive operation. In Marr's account, for example, the outputs were often also input to some other early stage of visual processing.

At the next lower level of analysis is what Marr referred to as the "algorithmic" level.[4] In truth, computational systems don't directly convert numbers or arrays of chess pieces. Instead, they convert representations of such things. On a standard computer, for instance, keyboard typings represent the input while the display on a computer screen (or perhaps some sort of print-out) represents the output (Cummins 1989). So at this level, we re-describe the explanandum in representational terms: The new explanandum is not the conversion of numbers into sums, but the conversion of representations of numbers (numerals) into representations of sums. In fact, this is the *general* form of cognitive science explananda, even for non-CCTC accounts like connectionism. What we want explained is how the brain performs this representational conversion. It is at the algorithmic level that theories like the CCTC provide us with a set of operations whereby symbols that designate the relevant inputs are transformed into symbols designating the appropriate output (where appropriateness is determined by the function specified at the top level). This is the level of analysis at which most traditional cognitive theories, like those of the CCTC, are pitched.

[3] This label is somewhat misleading since the sort of operations many regard as computations are not found at this level, but at the lower, algorithmic level.

[4] This same level is also sometimes called the "representational level," or, more controversially, the "syntactic level."

Finally, at the bottom level there is what Marr referred to as the "implementational" level. Here we find descriptions of the actual physical structures and processes that implement the operations occurring at the algorithmic level. It is at this level where physicalism is vindicated because psychological processes are mapped onto actual physical events and processes. In everyday computers, the explanatory posits of this level would be things like wires and circuits and their various causal interactions. In humans and other organisms, the relevant posits would be the actual neurological states, structures and processes that implement the cognitive operations that a given theory proposes.

As noted, in this three-tiered framework it is at the middle, algorithmic level where the CCTC theories attempt to explain the kinds of processes that account for mentality. It is at this level where the critical symbol-manipulations take place. Roughly, the accounts put forth are descriptions of processes in which symbols are shuffled about and combined so that their transformations produce appropriate input–output conversions. The processes are typically assumed to implement an algorithm, whereby complex operations are broken down into simpler and more basic operations – all performed on symbolic tokens. Thus, the explanatory strategy is task-decompositional; to understand a sophisticated cognitive process or capacity, the CCTC posits a series of sub-systems that perform different sub-tasks. We can continue breaking sub-systems down into simpler components until we reach the sort of processes that can be described by implementational elements. The process is familiar to those who build and program computers, but the goal in cognitive science is not to build a smart computer. Instead, the goal is to use our understanding of computers to model and help explain what is alleged to be taking place in our heads when we perform some cognitive chore. Advocates of the CCTC have called our brains "Physical Symbol Systems" (Newell and Simon 1976; Newell 1980) because of the crucial role symbols play in this way of understanding cognitive processes. Formal symbols are essential because they serve as the central elements (or "medium") of the type of algorithmic operations specified by the theory. Orderly symbol manipulations are physically possible because, as computers demonstrate, symbols can be erased, rewritten, stored, combined by purely mechanical processes. As long as the symbols are encoded so that their physical form (sometimes called their "syntactic" properties or their "shape") can be exploited by other elements of the system, these symbolic transformations can give rise to input–output mappings of the sort we are trying to explain. Put simply, the CCTC is the overarching explanatory framework that

attempts to account for cognition by appealing to this kind of symbol-manipulating process.

The explanatory strength of this framework can be seen more clearly when we recognize that physical symbolic operations of this type can correspond to the states and transformation of formal systems. For instance, if we include some connectives and quantifiers (whereby symbols like ">" stand for the material conditional), then the CCTC approach tells us how a physical system can implement a deductive system. Upon receiving the symbolic input "P," a mechanical system can search and retrieve the data structure of the form, "P > Q," and then, on the basis of those tokens and the system's physical architecture, write and register the new token, "Q." In fact, it shows us how formal operations can be mechanized that not only adhere to the basic principles of deductive logic, but more complex formal systems like predicate calculus. It is this feature – their ability to mechanically implement logical processes – that gives actual computers their powerful problem-solving capacities. Computers are often called "number crunchers," but a better description would be "logic implementers." Applying this understanding of computation to psychology, we get at least part of a theory of how our mind/brain works. Cognition is seen as rule-governed symbolic modifications that are possible because the system is sensitive to the physical form of the symbols. Physical systems – including the brain – can thereby use physical symbols to become a sort of inference-machine.

2.1.2 The CCTC, RTM, and the Standard Interpretation

At the end of the last section, I described the purely mechanical symbolic operations put forth by the CCTC as "inferences." However, if we want to account for *actual* inferences of the sort we would recognize as thought processes, then the story, as it stands, won't do. Mechanical operations involving merely formal tokens won't suffice for an account of what we normally think of as *reasoning*. To be an account of reasoning (or inference, decision-making, thinking, etc.), the relevant units of computation must be representational states like beliefs and desires. What's missing from the account just presented are states with full-blown representational status. A critical question then, is how do we come to view formal computational symbols in this way?

The Standard Interpretation suggests an answer to this question. The critical move in constructing a folk psychological interpretation of the computational processes is to exploit the parallelism that is possible

between the causal relations among computational symbols, on the one hand, and the causal and inferential relations assumed to exist between beliefs and other propositional attitudes on the other hand. As we saw in the last chapter, folk psychology treats propositional attitudes as states that play different causal roles. The Standard Interpretation expands on this idea and claims that causal roles distinguish not just propositional attitudes from other mental states, but also different sorts of propositional attitudes. A given representation playing one sort of role is a belief state; the same representation playing a different sort of role is a desire state. The idea that mental processes are mediated by mental representations has been called the "Representational Theory of Mind," or RTM for short. While there are different variants of RTM, most versions regard psychological processes as the interplay of representational states in accordance with formal principles. Some of these principles govern the way primitive representations can be conjoined to form complex representations, whereas others determine the causal relations between representational states. What matters here is that on the Standard Interpretation, RTM helps to wed commonsense psychology to the CCTC. If the inferential relations of the attitudes are casual relations that can be mirrored by the symbols posited by CCTC, then we can treat the symbols as beliefs and explain ordinary inferences and judgments. If we say the symbol "P" represents that it's currently raining while the symbol "Q" represents that I need to drive to work, and ">" represents the material conditional, then the mechanical operation described above – where a token of "P" and a token of "P > Q" led to a tokening of "Q" – can explain how a person decided to drive to work. It is by mirroring the inferential and causal relations of beliefs and the like that the CCTC can provide intuitively plausible accounts of everyday reasoning. If we can do this for a variety of different cognitive capacities and skills, then we have what amounts to a robust theory of ordinary thought.

The picture presented here is often described as one whereby a physical system implements a syntactic engine which drives a semantic engine. It is an appealing model of cognition because it preserves much of what we take for granted about mental states and processes while at the same time showing how these states and processes might be physically realized. Perhaps the best known advocate of this account is Jerry Fodor. Fodor argues that defending the commonsense assumptions about thoughts on the one hand, and, on the other hand, defending the central framework of RTM, and, by extension, the CCTC, all amounts to the same basic enterprise (1987, p. 16). According to Fodor, a theory of cognition provides an endorsement of commonsense psychology if it posits inner states that are

semantically evaluable, have causal powers and abide by the generalizations of folk psychology. The inner symbols of the classical model are the perfect candidate for this explanatory role. As he puts it, "[c]omputers show us to connect semantical with causal properties for *symbols*. So, if having a propositional attitude involves tokening a symbol, then we can get some leverage on connecting semantical properties with causal ones for *thoughts*" (Fodor 1987, p. 18).

The Standard Interpretation, then, offers answers to two important, closely related questions about the nature of representation in the CCTC. The first question is: What type of representational notion is invoked in computational explanations of cognitive processes? The answer proposed is, the same type of representational notions that are invoked by commonsense psychology. Computational processes are seen as a mechanized version of folk psychological reasoning, and this is possible only if the symbols being manipulated are viewed as analogues of familiar commonsense mental representations. The symbols of the CCTC are realizers of beliefs, desires, concepts, and other mental representations put forth by our pre-theoretical framework. The second question is: What is it about the CCTC's explanations of cognition that motivates the positing of representations? And the answer, once again, stems from the Standard Interpretation's assumption that the explanation of cognition is in large measure an explanation of the states and processes put forth by folk psychology. Computational symbols come to be regarded as the representational states of folk psychology so that the explanations provided by the CCTC are pertinent and "make sense."[5] The symbols earn this treatment by virtue of their capacity to causally replicate the intuitive ways we assume thoughts interact to produce mental processes.

As many have noted (Fodor 1987; Pylyshyn 1984; Clark 2001), one of the virtues of this perspective is that it appears to provide a compelling vindication of our commonsense understanding of the mind. With the Standard Interpretation, the threat of eliminative materialism – the view that there are no such things as beliefs and other commonsense mental representations – can be significantly reduced, since the CCTC is seen to be fully committed to propositional attitudes. At the same time, the CCTC also inherits a considerable degree of intuitive plausibility from folk psychology. If we identify the elements of computational systems as the sorts of states we intuitively think are at the heart of cognitive processes, as the

[5] See Pylyshyn (1984).

Standard Interpretation does, then this intuitive understanding of the mind will make the computational perspective seem *prima facie* plausible.

To sum up, on the Standard Interpretation, the CCTC offers an account of cognition that makes use of certain notions of representation. The notions invoked are the same notions of mental representation put forth by commonsense psychology, especially our notion of belief. The symbols serve as representations in this theory by serving as the computational equivalent of these ordinary mental representations. The CCTC thereby provides us with a framework that vindicates folk psychology by providing us with symbolic posits that causally interact in the manner of thoughts and other mental representation. But at the same time, folk psychology helps vindicate the CCTC by providing intuitive support for some of the "folksy" aspects of symbolic representations. The end result is a reading of the CCTC whereby the notion of representation employed is, roughly, the same as our folk notions of mental representation.

2.2 DIFFICULTIES WITH THE STANDARD INTERPRETATION

Despite the overall attractiveness of this perspective, I think it is fundamentally wrong-headed. The Standard Interpretation proposes what I believe to be wrong answers to both of the questions posed earlier: It is not the case that the basic notions of representation employed in the CCTC are folk notions of mental representation, nor is it the case that the motivation for treating computational symbols as representations is so that commonsense psychology can be given a scientific home. In the next chapter, I'll try to provide what I think are the right answers to these questions. But for the remainder of this chapter, I want to explain how the Standard Interpretation leads to trouble. The problems stem from an aspect of the job description challenge presented in the last chapter; in particular, the need to explain how a given structure or state actually serves as a representation. Because the Standard Interpretation claims that computational symbols function as analogues for propositional attitudes, it ignores what is distinctive about CCTC explanations that makes the positing of representations a more integral aspect of the theory. While the symbols can behave in a manner that mirrors the intuitive syntactic and causal nature of mental representations, this is not enough to demonstrate that whatever content they have is explanatorily relevant.

While the specific version of the job description worry I have in mind hasn't received a great deal of attention, something close to it has loomed in the background of more traditional challenges to computational

representationalism. It will help, then, to frame our discussion of these matters by looking at two of these challenges and especially the manner in which defenders of representationalism have responded. One challenge, presented by Searle, is that computational symbols lack the sort of intentional content ascribed to propositional attitudes. According to Searle, this entails that they lack the status of real representations. The second challenge, endorsed by Stich, contends that folk psychological notions of representation are not suited for serious scientific psychology. From this Stich argues that we should adopt a strictly syntactic interpretation of computational psychology and abandon the representational perspective altogether. Both of these challenges exploit the idea that computational representationalism is rooted in folk psychology, that computational symbols qualify as representations by serving as analogues to beliefs and other folk notions of mental representation. While proponents of the Standard Interpretation have vigorously responded to both Searle and Stich, these responses, we'll see, raise as many questions as they answer. In an oft-quoted passage, Haugeland suggests that a motto of the CCTC is, "if you take care of the syntax, *the semantics will take care of itself*" (Haugeland 1985, p. 106). But as these critics point out, it is far from clear how, exactly, the semantics is supposed to simply "take care of itself."

2.2.1 *Searle's criticism of computational representation*

The author who has done the most to argue that in computational systems the semantics will *not* "take care of itself" is Searle (1980). His Chinese Room argument has become such a prominent landmark on the cognitive science terrain that it hardly needs reiteration. A man is locked in a room with a box of Chinese symbols he doesn't understand, but is also given an instruction manual (written in a language that he does understand) that tells him how to manipulate the symbols in various ways. After some Chinese characters are passed along to him, he consults the manual and carries out a number of elaborate symbolic operations, focusing only on the shape of the symbols. Eventually, he hands back through an output slot a different set of Chinese symbols. Unbeknownst to him, he has just been asked some question in Chinese and has responded in a perfectly reasonable manner. How does he do it? By implementing an algorithm that is designed to create the impression that he (or the system) really understands Chinese. But, Searle argues, this is only an illusion; he has no idea what the symbols actually mean. In fact, from the standpoint of the system itself, the symbols have no real representational status – they mean nothing. According to Searle, merely

implementing a formal program is not sufficient for real mentality or understanding; the syntactic operations of the computational model do not provide us with an adequate account of cognition.

As some have noted (Warfield 1998), there are really two distinct conclusions that Searle uses the Chinese Room example to establish. Both involve the claim that running a program is in some way inadequate, but they differ with regard to what, exactly, running a program is inadequate for. The first conclusion is that computational processes are inadequate for instantiating understanding or mentality in general. While the man in the room specifically lacks understanding of the Chinese language, Searle is clearly making a more general point that reaches beyond linguistic knowledge. His intuitive point, and the one that makes the argument so compelling, is that no amount of formal symbol manipulation alone can ever give rise to real comprehension or understanding of anything. The man could instantiate a symbol-manipulating program for any cognitive process and neither the man nor the running program nor the entire system would actually instantiate that cognitive process. This is Searle's rejection of Strong AI, the idea that running the right program is sufficient for having a mind. Searle's Chinese Room can be seen as a counter-example to the metaphysical claim that any system that runs the right program automatically becomes a thinking system.

Searle's other conclusion more directly challenges the representational character of the CCTC. Searle notes that computational systems succeed at their various tasks by manipulating symbols by virtue of the syntactic properties and not by virtue of any sort of semantic content. In his Chinese room, the man simply checks the shape of the characters, checks the instructions, and then shuffles the symbols about by virtue of these two factors. By contrast, *real* mental representations – things like our thoughts and ideas – intuitively interact with one another and produce behavior by virtue of what they are about. This leads Searle to deny that computational symbols are representational states at all, since they lack any real representational content. At most, they have the sort of derived intentionality we discussed in the last chapter – a sort of meaning that is assigned by us, the outside programmers and observers. Since the symbols lack the intrinsic intentionality of real thoughts, there is no sense in which they serve as representations *for* the system itself. In short, Searle argues that it is simply wrong to regard the symbolic structures that mediate computations as representations. As he puts it, "syntax alone is not sufficient for semantics, and digital computers insofar as they are computers have, by definition, a syntax alone" (1984, p. 34).

While both of Searle's conclusions are important,[6] it is the second thesis that directly challenges the Standard Interpretation of the CCTC. It does so by pulling the rug out from under the idea that classical symbols provide a suitable reduction base for the representational posits of folk psychology. Searle's argument is similar to our earlier point that a theory of cognition cannot simply adopt folk notions of representation because those notions involve features, like intrinsic intentionality, that need further explication. You don't get such an explication just by positing states that causally interact in the manner put forth by the CCTC. Searle's claim is not simply that the meaning of computational symbols does no explanatory work, or that the content of the symbols is causally inert. His point is that computational symbols don't have any real content, and consequently computational symbols aren't really representations at all. Searle's argument challenges the very idea that classical computational theory is itself a *representational* theory of cognition. He claims it isn't, because the symbols in classical architectures don't have the original intentionality associated with real thoughts.

Responding to the Chinese Room argument has become a minor industry in the philosophy of psychology, and in the next chapter I'll offer my own rebuttal. But for now the important point to note is how both Searle and the defenders of the CCTC argue from the perspective of the Standard Interpretation and link the explanatory value of computational representations with our commonsense understanding of the mind. Most supporters of the CCTC agree with Searle that the posits of the CCTC need to be sufficiently similar to folk mental representations if they are going to serve as representations. So the debate is not over the *sort of* representational posit put forth by the CCTC. Instead, it is over whether or not the CCTC succeeds in accounting for all of the important aspects of that posit – in particular, whether it succeeds in accounting for the intentional content that we associate with beliefs and other mental representations.

Many popular strategies for responding to Searle concede that, by itself, the CCTC does *not* account for the sort of intentionality needed to treat computational symbols as representations. There is, in other words, a bit of back-pedaling on the idea that "the semantics will take care of itself." If positing representations entails positing belief-like states, and if positing belief-like states entails explaining intentional content, then the CCTC, as such, doesn't actually posit inner representations! But according to

[6] On Searle's own account of intentionality, these conclusions are closely related since he holds that the content of thoughts is closely linked to consciousness and our background understanding.

defenders of the Standard Interpretation, this only means something more needs to be added to convert the inner symbols into real representations. On the Standard Interpretation, while there is nothing about the way the symbol shuffling succeeds in producing appropriate input–output conversions that suggests the symbols actually serve as representations, we nevertheless *can* view them as representations if we supplement the theory with a workable theory of content. For example, a popular response that Searle himself considers – the "Robot Reply" – claims that the CCTC needs to be supplemented with some sort of story that connects the inner symbols to the world via the right sort of causal links. With this proposal, if the inner computational symbols get hooked to the world in the right way, then they actually will have the sort of original intentionality ascribed to folk mental representations, and can thereby serve as tobust representational states.

Before we evaluate this response to Searle's argument, it is worth pausing to consider what it suggests about the explanatory connection between the CCTC and representation. For starters, it would seem to clearly undermine the basic idea that the CCTC explains cognition by invoking inner representations. If it is generally agreed that the computational processes can be fully understood without an appeal to any sort of intentional content, then at most what the CCTC provides is the non-representational chassis or framework whose internal states can be *converted into* representations with some added story about how states get their meaning. That is, the CCTC merely provides the non-representational precursors to (or non-representational vehicles for) the representational states posited by folk psychology. As one author puts it, "... if you let the outside world have some impact on the room, meaning or 'semantics' might begin to get a foothold. But, of course, this response concedes that thinking cannot be simply symbol manipulation. Nothing can think simply by being a computer" (Crane 2003, p. 128). Moreover, if the explanatory role of classical symbols is to provide a vehicle for vindicating the representational posits of folk psychology, then this is certainly an odd sort of vindication. In the last chapter, it was noted that the representational posits of folk psychology are assumed to have their content essentially – beliefs and other propositional attitudes are individuated by virtue of their content. Insofar as computational symbols, *qua* theoretical posits of the CCTC, don't, as such, have any sort of content, then the CCTC fails to capture the most central aspect of what is supposedly being vindicated. Even if the symbols can *acquire* semantic properties through causal links to the world, the need for this is independent of the explanatory framework provided by the CCTC. It's not that the representational character of the CCTC's theoretical posits

establishes a scientific home for beliefs; rather it seems that, at least in part, the desire to find a scientific home for beliefs drives the representational characterization of CCTC theoretical posits.

These considerations allow us to see how the Standard Interpretation of the CCTC leads down a path that gives rise to skepticism about representation in the CCTC. The Standard Interpretation assumes that representations are needed in the CCTC in order to account for the sort of thought processes recognized by folk psychology. It thereby emphasizes the degree to which computational symbols can, by virtue of their syntactic properties, mirror the causal relations of the attitudes. Hence, there is a parallelism between the causal activity of computational data structures and the causal activity commonsense assigns to folk mental representations. But the parallelism alone is not enough to bestow representational status on the symbols; an additional theory of content is needed. Consequently, the symbols' status *as* representations is not, on the Standard Interpretation, something built into the explanatory framework of the CCTC. If positing representations entails positing belief-like states, and if positing belief-like states entails explaining intentional content, then the CCTC, as such, doesn't actually posit inner representations![7]

Let's return to the question of whether or not the standard-plus response to Searle is ultimately successful. Recall that the strategy is to argue that if we can construct an adequate theory of content for computational symbols (one that conforms with our assumptions about ordinary thought content), then we would have all the necessary ingredients for a robust and complete representational theory. Searle denies that such a theory of content has been given, arguing that (for example) the sort of head-world causal links thought to account for content fail to actually do so. But let's set aside the debate over content and focus instead on whether the proposed strategy, if successful, would provide what is needed. On the Standard Interpretation, the CCTC allows us to meet the job description challenge by revealing how something can *function as* mental representations. They function as mental representations by functioning in the way commonsense psychology assumes beliefs and other propositional

[7] Dan Weiskopf has pointed out that if we adopt a functional role semantics, it might be said that the causal roles associated with computational symbols also provide the content for the symbols, so a content story is built into the functional story of the CCTC after all. Yet putting aside the many difficulties associated with functional role semantics, most functional role semanticists concede that syntactic relations cannot be the sole determinant of content for mental representations (Block 1986). Moreover, such an account of content would still leave unexplained what it is about the symbols that makes their role a representational one.

attitudes function in the mind. So if what we want is an account that accommodates our ordinary notion of mental representation, what we wind up with is an account with two parts. One part, provided by the CCTC, is an account of how inner states acquire the causal/functional properties associated with specific sorts of propositional attitudes. The other part, some sort of theory of content, is an account of how the same states come to have the original intentionality associated with mental representations. Put the two parts together and we have, it would seem, an account of the mind that is fully representational and vindicates folk psychology. If having the belief that the Dodgers won is a matter of the mental representation with the content "the Dodgers won" playing a certain functional role (i.e., the believing role), then even though the CCTC does not itself provide a theory of content, it does provide an account of what it would be to serve as a mental representation of this sort. Or does it?

The problem with this picture is that it doesn't allow us to explain what it is for something to serve as a representational state, *simpliciter*. On the Standard Interpretation, the causal-syntactic framework of CCTC leaves that part of the job description unexplained, since it presupposes that data structures playing these different roles are already functioning as a mental representation. What the CCTC shows is how a state serving to represent that the Dodgers won comes to function as a belief or desire. It enables us to explain when the structure is functioning as a belief-type representation and distinguish this from when it is functioning as a desire-type representation. What it doesn't show, on the Standard Interpretation, is how (explanatorily prior to all this) the state in question comes to serve as the representation the Dodgers won. This latter role, whatever it is, would be one that representations serving as different propositional attitudes all *share*. On the Standard Interpretation, the CCTC provides a causal story that allows us to distinguish a believing representation from a desiring representation, etc. It doesn't give us an account that allows us to distinguish inner representations, as such, *from everything else*.

Recall that the job description challenge requires that when a representational posit is invoked in a theory, there is some sort of account of how the state is supposed to function as a representation in the system in question. The problem with the account we've been considering is that, even after the account is supplemented with a theory of content, we still don't have a sense of how inner states are supposed to serve as representations in computational processes. We do get an account of the sorts of causal relations that are involved in making something a representation of a certain sort (analogous to the relations that allow us to distinguish a

hand-held compass from one that is mounted on an automobile's dash). This is provided by the causal/syntactic architecture of CCTC that explains the difference between beliefs, desires and other propositional attitudes. And we can assume we get some sort of account of content-grounding conditions (analogous to the nomic dependency between the needle's position and magnetic north). This might be provided by a theory linking the inner symbols to their intentional objects through some sort of causal relation. But what we don't get is an account of the computational (or physical) conditions that explain why a state or structure should be treated as serving as a representation in the first place (analogous to someone using the compass needle's position to discern the location of magnetic north). What is missing from the Standard Interpretation is a story about how symbolic posits actually serve as representational states in the sort of functional architecture proposed by the CCTC. It isn't provided by the syntactic or causal interactions of the symbols because these causal relations are of the wrong sort. They account for the differences between types of inner representations, but they don't account for the way in which a computational system uses data structures *as* representations. And it isn't provided by proposed content-grounding conditions, such as causal links between the representation and its intentional object, because those conditions generally don't, by themselves, *make* something into a representational state.[8] It is often assumed that the combination of the inner causal relations with the content-bestowing head-world links provides all that is needed. But we can now see that we need something more. The missing feature isn't constituted by these other factors, just as a compass's role as a representational device isn't constituted by its being mounted on the dash and its causal connection to magnetic north. What's missing is an accounr of exactly what it is about the way the brain uses inner symbols that justifies regarding those inner states as representations.

Representations are thus similar to other special science posits that involve multiple dimensions. Consider what it is for some piece of metal to serve as a form of currency. To fully understand a quarter's role as a bit of currency, it would not be enough to describe some of its unique causal relations, such as what it does in a Coke machine. Nor would it be sufficient to explain the process that bestows upon the quarter its value of 25 cents. To be sure, these matters are relevant to our understanding of the coin's status as currency. But one could understand various causal

[8] A similar point is made by Adams and Aizawa (1994) in their criticism of Fodor's account of representation.

interactions and even the process whereby its value is stipulated, and still be ignorant of *how* it actually serves as a unit of exchange (say, by not knowing that it is physically traded for goods). Representations are similar. Besides standing in some sort of content grounding relation and besides participating in a number of other causal relations, they are, more fundamentally, employed to stand for something else. How this happens is left unexplained by the Standard Interpretation.

The problem, as I see it, stems from the way the Standard Interpretation ties representation in the CCTC to commonsense psychology. Because the motivation for treating symbols as representations is connected to the explanatory value of folk psychology, it is seen to stem from considerations that are distinct from the explanatory scheme used by the CCTC. Consequently, there is a blind spot to the way in which that explanatory scheme actually *does* explain how computational symbols serve as representations. The debate prompted by Searle's Chinese Room argument helps reveal that blind spot and show how, on the Standard Interpretation, representation is presented as something that really *isn't* an inherent part of CCTC explanations. In the next chapter, I'll argue that, in fact, when we adopt the correct interpretation of the explanatory role of representations in classical models, we actually *do* get an account that makes the notion of representation a central and indispensable explanatory posit. On the proper treatment of CCTC, the notion of representation invoked can be seen to meet the job description challenge. But before we see how that story goes, it will pay to look at another challenge to computational representationalism.

2.2.2 *Stich's criticism of computational representation*

Stich's challenge to representational computational psychology is somewhat more nuanced than Searle's, but it is perhaps more damaging to the Standard Interpretation. In his 1983 book "From Folk Psychology to Cognitive Science," Stich adopts the Standard Interpretation tenet that treating computational explanations as involving representations amounts to treating computational explanations as committed to the posits of folk psychology. While he thinks this may be possible (unlike Searle), he also thinks it is a bad idea. Folk psychology individuates propositional attitudes by virtue of their content. However, according to Stich, there are a number of reasons for thinking that content-based taxonomies are ill-suited for serious, scientific psychology. Because content-based taxonomies are based upon head-world relations, they will individuate mental states in ways that cross-classify more scientifically respectable taxonomies, such as those

based upon causal powers or causal roles of the internal state. Moreover, Stich maintains that content ascriptions for propositional attitudes are based upon similarity judgments – we ascribe beliefs by imagining what we would say in situations similar to the subject's. Thus, content ascriptions involve a high degree of vagueness, are highly parochial and context sensitive, and fail with subjects who are too dissimilar from ourselves, such as the very young and the mentally ill. In effect, Stich denies the alleged parallelism between the causal/syntactic operations of computational symbols on the one hand, and the content-based descriptions of mental processes presented by folk psychology on the other. Because representational psychology is, for the most part, identified with folk psychology, and because folk psychology makes for lousy science, Stich argues that we need to drop the appeal to representations in CCTC altogether.

In place of a representational cognitive science, Stich argues psychologists should employ a purely causal/physical, or syntactic theory – one that individuates mental states by appeal to their purely non-semantic properties. In other words, Stich agrees with Searle that computational symbols should not be treated as surrogates for folk psychological posits like beliefs. But unlike Searle, Stich does not argue that the CCTC should be abandoned. Instead, it should be re-conceived so that internal symbolic structures are treated as formal tokens individuated by virtue of their purely syntactic properties. Stich's so-called "Syntactic Theory of the Mind" is what he claims computational psychology should look like – a theory that retains the same basic architecture of classical computational theory, but makes no attempt to treat the inner states as representations. The purely syntactic generalizations and taxonomies provided by such an outlook are, according to Stich, much better suited for a scientific psychology because they carve the inner computational machinery at its causally salient joints. So Stich can be seen as making a negative point and a semi-positive point. The negative point is that since a representational perspective relies on a folk psychological framework, and given that such a framework leads to taxonomies that are ill-suited for science, computational psychology should reject representationalism. The semi-positive point is that it is okay for the CCTC to abandon representationalism since a non-representational, syntactic account works just fine in accommodating CCTC explanations.[9]

[9] Though he can hardly be characterized as an anti-representationalist, Fodor has pressed similar points in his well-known paper, "Methodological Solipsism" (Fodor 1980). Fodor endorses what he calls the "formality condition" which states that computational theories should individuate symbolic states without appeal to semantic properties. Here, Fodor appears to join Stich in insisting that treating CCTC symbols as representations is, at best, gratuitous.

Although Searle's argument is clearly very different from Stich's, it is worth pausing to consider the ways in which their views overlap. Both Searle and Stich agree that classical computational accounts of cognition should not be treated as representational. Why not? Because, according to these authors, computational symbols fail to serve as an adequate reduction base for folk notions of representation like belief. Searle concludes from this, "so much the worse for classical computationalism." Stich concludes from this, "so much the worse for folk psychology." Both authors assume (at least tacitly) that the question of whether or not the classical model of cognition provides us with a workable notion of representation is to be answered by focusing on our notions of belief, desire and other propositional attitudes. Searle argues that computational states would need to be more belief-like to qualify as real representations; since they aren't, computationalism is a flawed model of the mind. Stich appears to agree that the representational status of computational states rests on their being belief-like, and since he thinks belief-like states are not scientifically respectable, representationalism is rejected. Despite their criticism of the marriage between computational psychology and folk psychology, both authors share the Standard Interpretation's assumption that positing classical computational representations amounts to positing belief-like states in a computational framework. In fact, they use this assumption to challenge the link between CCTC and RTM.

While defenders of the Standard Interpretation respond to Searle by arguing that it is *possible* to convert symbolic structures into belief-like representations, the challenge posed by Stich forces them to explain why it is *necessary*, or at least beneficial, to do so. This is important for our discussion because it goes right to the heart of the issue of the explanatory value of representations. If Stich's challenge can be answered and it can be shown that our understanding of cognitive processes is significantly enhanced by treating CCTC symbols as propositional attitudes, then the Standard Interpretation would itself be vindicated. We would then have reason to suppose that the folk psychological notion of representation in the CCTC has some degree of explanatory legitimacy. On the other hand, if it should prove quite difficult to answer Stich's challenge and if the benefit of treating CCTC symbols as beliefs and desires is left in doubt, then that can be taken to indicate one of two possibilities. It could be taken to show, as Stich suggests, that CCTC shouldn't be regarded as a representational theory. Or, it could be taken to show that the Standard Interpretation is wrong to equate CCTC representations with folk psychological states, and that a different way of thinking of computational representations is in order.

Because, as we've noted, most authors today endorse the idea that the content of mental representations can be reduced to some other set of naturalistic conditions, many would reject Stich's analysis of content ascription, especially his suggestion that they are based merely on similarity judgments. Most accounts of content determination hold that the semantic content of our thoughts is determined by objective factors (like certain sorts of causal relations) and that a scientific psychology could adopt content-based taxonomies that are neither parochial nor context-sensitive. However, even with these more robust accounts of content ascription, many writers would concede that the syntactic approach offers a more causally accurate picture of the inner workings of the mind. Given that the syntactic properties are what actually determine the causal role of symbols in computational processes (the ones that contribute to the causal powers of the symbols), a syntactic taxonomy would be more descriptive and would have greater predictive power. So why not regard the symbols as simply non-representational tokens that serve as the medium of purely mechanical computational operations? This is the key question Stich poses for CCTC representationalists.

There have been two popular strategies for responding to this challenge and defending representationalism. One strategy is to directly assail Stich's argument for the Syntactic Theory by showing there is a flawed assumption in the argument. The second strategy is to appeal to some explanatory desideratum that is achieved through representationalism and would be lost if we were to adopt a purely syntactic framework. Each can be seen as a way to support the union between the CCTC and folk psychology for those who adhere to the Standard Interpretation.

A common version of the first strategy offers a sort of self-refutation argument against Stich's assumption that taxonomies more finely tuned to the details of the inner workings of some system are superior to those that are more coarse-grained and abstract. The problem with this assumption, goes the response, is that it applies to the syntactic level itself. After all, there are levels of description of brain processing that are purely physicochemical, or perhaps even molecular, that would provide a more detailed analysis and offer more accurate generalizations than the syntactic level. Thus, if Stich's reasoning is sound, we should abandon the syntactic level (and computational psychology altogether) and adopt the physicochemical level of description. But this seems much too drastic.[10] Block calls this argument the "Reductionist Cruncher" and states, "... the physicochemical account will

[10] See, Bickle (2003) for an account that does not treat this as absurd.

be more fine-grained than the syntactic account, just as the syntactic account is more fine-grained that the content account ... if we could refute the content approach by showing that the syntactic approach is more general and more fine-grained than the content approach, then we could also refute the syntactic approach by exhibiting the same deficiency in it relative to a still deeper theory" (Block 1990, p. 280).

While Block's argument offers a valuable lesson for many sorts of reductionism, its application to Stich's challenge is misplaced. Block seems to assume that Stich is advocating a switch from a higher level of analysis to a lower one, where levels of analysis correspond to different levels of organization or composition in physical reality. On this view, the kinds and entities described at a lower level constitute the kinds and entities described at a higher level. Block's complaint is against those who would try to abandon descriptions of higher levels of composition in favor of theories and descriptions at a lower level (abandoning biology, say, for physics). But Stich is concerned with something very different; namely, with determining which properties of the symbolic entities at a single level of organization (the one appropriate for computational psychology) we ought to use when constructing our psychological taxonomies. The transition from a content-based taxonomy to a purely syntactic taxonomy does not involve a transition in levels of analysis or organization. Stich isn't arguing that we should abandon the computational level of analysis at all. Rather, he is challenging the idea that at the algorithmic level of analysis, the posits should be classified by appealing to their alleged semantic properties as opposed to their syntactic properties. The issue is one concerning *types* of taxonomies, not *levels* of taxonomy.

What is really in dispute concerns which properties matter for understanding the sort of job the posits of CCTC actually perform. The debate between Stich and representationalists is about the proper job description for computational symbols. Stich argues they should be viewed as formal tokens that mediate syntactic operations and processes. Block and others argue they should instead be viewed as representations. But because the Standard Interpretation leaves it unclear exactly how data structures are serving as representations in computational processes, it is unclear what it is about computational explanations that warrants a representational job description for the symbols. Some have argued[11] that a syntactic taxonomy will miss generalizations captured by a content-based approach. But it is important to bear in mind that taxonomies and generalizations are cheap. It is, after all,

[11] See, for example, Pylyshyn (1984).

easy to construct various taxonomies for symbolic data structures that are more abstract than those based on semantic properties. If we can vindicate the representational treatment of computational symbols simply by showing that it allows unique generalizations, then we can vindicate *any* taxonomy of computational states since they all allow for unique generalizations. What needs to be shown, then, is not that a content-based approach to computational psychology invokes vocabulary, captures generalizations, and makes classifications, etc. that are *different* from what would be provided by a purely syntactic, non-representational approach. Of course it does. Instead, what needs to be shown is that the vocabulary, generalizations and classifications allowed by the representational approach buys us something worth having. We need some justification for thinking that the generalizations provided by the representational story provide something of significant value that would be missing from a purely syntactic story.

This brings us to the second strategy for responding to Stich's challenge, which is to appeal to some further explanatory desideratum that, allegedly, only the representational picture can provide. When asked for this, the most common response is to invoke some sort of principle of *rationality*. Here it is claimed that the representational picture provides (and the syntactic account leaves out) a framework that allows us to characterize cognitive processes as rational. Perhaps the most explicit expression of this is found in Pylyshyn (1984), so it is worth quoting him at length:

> What I am claiming is that the principle of rationality . . . is a major reason for our belief that a purely functional account will fail to capture certain generalizations, hence, that a distinct new level is required (p. 34) . . . [I]n a cognitive theory, the reason we need to postulate representational content for functional states is to explain the existence of certain distinctions, constraints and regularities in the behavior of at least human cognitive systems, which, in turn, appear to be expressible only in terms of the semantic content of the functional states of these systems. Chief among the constraints is some principle of rationality. (1984, p. 38)

Arguments similar to this have been presented by others.[12] The basic sentiment can be summarized as follows:

(1) Theories of cognition must account for the rationality of cognitive systems.
(2) The only way we can account for the rationality of cognitive systems is by treating internal states as representations.
(3) Therefore, computational theories of cognition must treat inner symbols as representations (and not merely syntactic or functional states).

[12] See, for example, Fodor (1987) and Rudder-Baker (1987).

So here, it seems, we have a clear answer to Stich's challenge. We need to view cognitive systems as rational, and to view them as rational, we need to treat their inner states as representations, which means, on the Standard Interpretation, as commonsense mental representations. Moreover, this could perhaps be used to handle our primary job description concern of explaining how computational symbols actually *serve* as representations. They serve as representations by virtue of serving as not just symbolic tokens, but as states in a system engaged in rational inferences. To explain cognition, we need to account for rationality, and to account for rationality, the symbols posited by the CCTC must function as commonsense mental representations.

Unfortunately for defenders of the Standard Interpretation, this line of reasoning has a number of shortcomings. It is far from obvious exactly what Pylyshyn means by "rational," and there are certainly notions of rationality that could be explained without any appeal to inner representations. For example, if all that is meant by "rational" is that the behavior is appropriate given the circumstances, or that the inner processes provide the right input–output mappings, then it is hard to see why a syntactic model couldn't provide this. In fact, Stich himself offers a point-by-point comparison between, on the one hand, a content-based account of someone being told her building is on fire and subsequently fleeing, and, on the other hand, a purely syntactic story of the same cognitive process and behavior (Stich 1983).[13] Since the syntactic story explains the same life-preserving behavior as the representational account, it seems the syntactic theory can explain rational behavior as well.

Of course, many would argue that to be seen as rational, the behavior itself must be given an intentional description (answering a phone call, fleeing from danger, etc.), and that *this* requires that the behavior arise from inner states characterized in intentional terms. While this is a huge topic that demands of book-length treatment, let me briefly offer some reasons why I reject this perspective. First, many of the arguments for this view rest on dubious claims about analytic definitions for terms used to characterize the relevant behavior. For example, it might be claimed that

[13] Here's a sample of how Stich develops his account: "Why did Mary come running from the building? It's a long story. First, she had a long standing D-state whose syntactic form was that of a conditional, viz. F ⊃ L (corresponding to the desire to leave the building if it is on fire), where F and L are themselves syntactically well-formed strings ... Mary began to inhale smoke ... The direct consequence of inhaling smoke was that Mary came to have a B-state (corresponding to the belief that she was inhaling smoke). From this B-state and the long-standing B-state I ⊃ N (corresponding to the belief that if one is inhaling smoke then there is a fire nearby) she inferred (i.e., was caused to add to her B-store) a token of N ..." (1983, pp. 174–175).

for us to characterize an action as "fleeing from danger," it is necessary for us to assume that the subject believe that there is danger nearby, has a desire to avoid danger, knows how to flee, etc. But it seems clear that our basic understanding of what it is to flee from danger could survive, perhaps with modification, if we dropped representational descriptions of inner causes. In fact, Stich's own syntactic analysis of the woman escaping the burning building shows us how this might go. Second, as Fodor (1987) has noted, there are deep problems associated with the idea that we should individuate behaviors in ways that are sensitive to the content of the underlying psychological states. Insofar as content depends upon relational factors that most would deny are determinants of behavior (such as linguistic affiliation), it is better to classify behaviors in ways that pertain to the causal powers of inner states. That amounts to adopting the syntactic theory when describing the inner causes of behavior. Third, it seems just false that an ascription of rationality requires an intentional characterization of behavior. "Crouching low," for example, is arguably a non-intentional description of behavior, but surely this can be seen as fully rational in certain situations. Finally, we've seen that what cognitive theories typically try to explain is not simply overt behavior, but cognitive capacities characterized as the transformation of input representations into output representations. There is no reason provided by the Standard Interpretation to think the intermediary states need to be representations for these input–output mappings to successfully instantiate some sort of function (like addition). But if these input–output mappings correspond with some cognitive capacity in a systematic way, then it certainly seems they would qualify as rational.

But what about the computational processes themselves? If we want to regard those inner processes as rational inferences, don't we need to treat them as involving inner representational states? Once again, the matter depends on what we mean by "rational." If we take it to simply mean that the processes are in accordance with the rules of deductive and inductive logic, then the answer is clearly "no." We can instead treat the symbols as formal tokens whose causal relations follow the basic rules of logical transformations. After all, logical rules just are rules that govern formal interactions between uninterpreted variables. If the syntactic operations mirror valid logical schema, then it would certainly seem that this notion of rationality doesn't require representational states with content.

However, on another interpretation of "rational" – the one Pylyshyn and others presumably have in mind – a purely formal, syntactic interpretation of computational processes won't do. The reason is that to qualify as rational in this second sense, the relations between inner elements need to

adhere to a different set of rules than those of formal logic. They would need to adhere to the "laws" and generalizations that folk psychology assigns to propositional attitudes and other folk mental states. This is the notion of rationality associated with explanations like, "She left the building because she thought it was on fire" or "He went to the fridge because he wanted a beer and thought there were more left." In other words, to be rational in this sense is to be driven by inner processes that involve states generally recognized as *reasons*. What sort of states are these? Commonsense mental states like beliefs, desires, hopes, fears, and other propositional attitudes, which are, of course, representational states. Consequently, for the computational process to be rational in this sense, we must regard the participating states as mental representations. The argument presented earlier needs to be modified to look more like this:

(1) Theories of cognition must account for the rationality of cognitive systems.
(2) To be rational (in this sense) is to instantiate the processes described by commonsense psychology.
(3) To instantiate the processes described by commonsense psychology is to instantiate processes involving commonsense mental representations like beliefs and desires.
(4) Therefore, theories of cognition must appeal to commonsense mental representations like beliefs and desires.
(5) Therefore, computational theories of cognition must treat inner symbols as commonsense mental representations (and not merely syntactic or functional states).

Initially, this might seem to bring us closer to what the Standard Interpretation needs – an argument that establishes what is gained by treating computational symbols not just as representations, but as representations of the sort recognized by commonsense psychology. CCTC is committed to representational states because it is committed to explaining rationality, and rationality in this context is just defined as processes involving belief, desires and so on. Yet a bit of reflection reveals that this won't work as a response to the sort of challenge Stich is offering. To respond to Stich's syntactic theory, we would need to show what is wrong with a non-representational interpretation of the CCTC, and the suggestion is that it will miss certain "distinctions" and "regularities" associated with a principle of rationality. But when we ask what this means, it turns out that these are just the distinctions and regularities that come along with treating states as commonsense mental representations. In other words, the reason we need to treat computational states as

propositional attitudes is so that we can treat computational processes as the sort of processes that involve propositional attitudes. This is hardly a convincing argument against a skeptical challenge to representationalism. We noted above that the sort of distinctions and generalizations captured by a representational perspective will be different from those offered by a syntactic account. Yet what needs to be shown is not that the representational framework is distinct, but that it is distinct in a way that is substantially superior (in terms of explanatory and predictive power) from the syntactic framework. The suggestion is that representationalism is indeed superior because it allows us to view computational processes as rational. This is what is supposed to be gained by treating the CCTC symbols as belief-type states. But being rational on the current proposal just means treating CCTC symbols as belief-type states! Thus, the argument is circular.

So the defender of the Standard Interpretation who appeals to rationality winds up with the following dilemma. Either we can define what it is for a system to be rational in a way that makes it distinctive, say by appealing to some formal system like deductive logic. But if we go this route, the relevant states can be treated as merely non-representational tokens and the syntactic account works just fine. Or, we can define rationality in a way that makes representations essential – indeed, makes beliefs and desires essential. But then being rational amounts to the same thing as implementing a system with the sort of representational states invoked by folk psychology. If you are a serious skeptic about the value of folk psychology, or of representation in general, rationality in this sense becomes something scientific psychology shouldn't care about. Either way, a principle of rationality doesn't help those who claim that data structures in CCTC systems need to be treated as representations.

In fairness to Pylyshyn and others, it is far from obvious that the appeal to rationality is based upon the same understanding of the explananda of cognitive theories that we have been assuming throughout this discussion. Here, I've characterized a central explanandum of cognitive theories as various cognitive capacities and skills defined in terms of representational input–output functions. We want to know how minds convert representations of chess boards into representations of moves, or how representations of sentences are converted into grammaticality judgments. Under this assumption, it is an open question whether the cognitive machinery responsible for these conversions involves folk psychological states like beliefs and desires – or, for that matter, inner representational states of any sort (apart from the inputs and outputs). It is at least possible, under

this construal of cognition, that a theory of the mind that denies inner representations could prove to be true. However, there are other conceptions of the explananda of cognitive science, and one of them takes it as a given that cognitive processes are of the sort (more-or-less) suggested by folk psychology. On this construal of cognition, one of the things we want explained are ordinary inferences like those mentioned above – e.g., someone deciding to leave a building because she believes it is on fire and wants to avoid injury. Folk psychological processes are, from this perspective, part of what we want theories like the CCTC to explain, and this would include an account of the rationality (and, presumably, irrationality) of these thought processes. To accomplish all this, computational symbols must be treated as realizers of propositional attitudes.

While I think there are a number of problems with the idea that cognitive scientists should assume (or do assume) that folk psychology accurately captures the nature of the mind, and that they should (or do) thereby treat commonsense psychological processes as their explanatory target, I'm willing to grant that sometimes this is the case. Even if this is so, this assumption brings us right back to our earlier point that the CCTC, as understood by the Standard Interpretation, fails to provide the proper reduction base for commonsense notions of mental representations. The reason is that on the Standard Interpretation, there is no account of how computational symbols actually serve as representations in computational processes. While there may be a theory of some sort of content-grounding relation for the symbols (like head-world causal relations) and also an account of the inner causal relations that explain how symbols can play the role of different propositional attitudes (distinguishing the believing role from the desiring role, etc.), we don't get an account of what it is for the symbols to actually function as representations in the type of operations presented by the CCTC.

Since I am ultimately going to argue that there actually *is* an explanatory pay-off in treating CCTC symbols as representations, I won't pursue this matter further here. The critical point is that, once again, the problem is due to the Standard Interpretation of the CCTC. I've suggested that the Standard Interpretation comes with the tacit assumption that we can show how symbols are representations by claiming that they realize or instantiate propositional attitudes. But as should now be clear, this doesn't work. You can't use the fact that A is the proposed reduction base for B to establish that A has all the relevant features of B. That is, you can't make computational symbols function as representational states by proposing that they be the things with which folk mental representations are

identified. Rather, one needs to first establish that computational symbols serve as representations in computational explanations of cognition, irrespective of their role as possible reducing posits of folk representations. Then, once we've established that symbols do indeed play such a role, we can ask whether or not they might be the sort of thing that instantiates beliefs and desires. In the next chapter, I'll argue that, on the proper interpretation of the CCTC, this latter strategy is indeed possible. Once we recognize that the notions of representation at work in CCTC are essential to the type of explanatory framework the CCTC provides, we don't need any *further*, folk psychological justification for treating them as representations. But to see things that way, we need to abandon the Standard Interpretation.

2.3 SUMMARY

In this chapter I've tried to do two things. First, I've presented what I take to be a very popular way of thinking about representation in CCTC that pervades the literature and dominates many discussions and debates. That way – what I've been calling the Standard Interpretation – suggests that CCTC is, by and large, a mechanized version of mental processes as conceived by commonsense psychology. Hence, the notion of representation ascribed to CCTC is seen as the same notion put forth by folk psychology. Second, I've tried to show how this outlook leads to problems for representationalism in classical cognitive science. These include the concession that content (and hence, representation) is not actually an element of computational explanations of cognition, along with a failure to explain what it is for computational symbols to serve as representations (even if an account of content is added). A common slogan is that classical computation requires a medium of representations. But on the Standard Interpretation, this looks like a bluff – there appears to be little reason to think a medium of *representations* is needed, as opposed to a medium of syntactically distinct tokens.

In the next chapter, I'll argue that all of this is due to a faulty understanding of CCTC and the notion of representation it employs. The Standard Interpretation is not the proper way to look at these matters, and when we gain a better understanding of the way computation is thought to explain cognition, we can also see why and how representation is needed. The notions that are needed, however, have little to do, at least directly, with the notions of mental representation found in folk psychology. Instead, they are theoretical posits that are as important and central to

the CCTC as are notions like *algorithm* or *sequential processing*. Hence, I'll argue that CCTC employs notions of representation that allow for a clear explication of what it is to serve as a representation, and one that shows exactly what would be missing from Searle's Chinese Room and Stich's syntactic theory of the mind. We've seen how *not* to understand representation in classical cognitive science; now let's see how it ought to be understood.