From Ockham to Turing --- and Back Again

Michael Rescorla

Abstract: Beginning with Turing himself, many researchers have suggested that mental processes are Turing-style computations. Proponents typically develop this picture in conjunction with *the formal-syntactic conception of computation* (FSC), which holds that computation manipulates formal syntactic items without regard to their representational or semantic properties. I explore an alternative *semantically permeated* approach, on which many core mental computations are composed from inherently representational elements. The mental symbols over which the computations operate, and hence the computations themselves, have natures inextricably tied to their representational import. We cannot factor out this representational import to generate an explanatorily significant formal syntactic remainder. I argue that the Turing formalism provides no support for FSC over the semantically permeated alternative. I then critique various popular arguments for FSC.

§1. Computation as formal syntactic manipulation?

Turing (1936) helped launch the computer revolution by advancing the Turing machine as an analysis of symbolic computation. Many authors, including Turing himself (1947, p. 111), have since proposed that the Turing machine or some similar computational formalism might provide a good model for mental activity. This proposal is now sometimes called *the classical computational theory of mind* (CTM). Putnam (1967) introduced philosophers to CTM. Fodor (1975) advanced CTM as a foundation for cognitive science. CTM proved controversial, with many philosophers vigorously dissenting (Dreyfus, 1992), (Searle, 1980). Researchers have also proposed various rival foundations, including connectionism (Smolensky, 1988) and dynamical systems theory (van Gelder, 1995). Nevertheless, CTM retains prominent advocates among both philosophers and cognitive scientists.[1]

I want to discuss how CTM bears upon the traditional picture of the mind as a representational organ. Many notable mental states are *about* a subject matter: my belief *that Barack Obama is president* is about Barack Obama; my desire *that I drink some water* is about water; my perceptual experience *as of a red sphere standing before me* is about redness and sphericality; and so on. Historically, most philosophers have assigned a crucial role to "aboutness" (or *intentionality*) when elucidating reasoning, decision-making, perception, and other paradigmatic mental activities. Hence the proliferation of theories trafficking in reference, truth-conditions, representational content, propositions, etc. All these theories prioritize *intentional descriptions*, which individuate mental states at least partly through their representational or semantic properties.

What is the relation between computational modeling and intentional description?[2]

[1] Ironically, Putnam (1988) has become one of CTM's harshest critics. Fodor (2000, 2008) also now rejects CTM as a theory of cognition *in general*, although he still holds that it well describes many important mental processes (such as perception and language comprehension).

[2] In addressing this question, I restrict attention to the Turing machine and kindred computational formalisms. I do not consider computation by neural networks, because I am concerned solely with *classical* versions of the doctrine that the mind is a computing system. For purposes of this paper, "computation" means "Turing-style computation." See (Gallistel and King, 2009) for a recent, detailed case that Turing-style models of the mind offer important advantages over neural network models.

According to current orthodoxy, computation manipulates formal syntactic items without regard to their representational or semantic properties. I will call this *the formal-syntactic conception of computation* (FSC). Fodor (1981, pp. 226-227) offers a classic statement: computational processes "are formal because they apply to representations in virtue of (roughly) the *syntax* of the representations… What makes syntactic operations a species of formal operations is that being syntactic is a way of *not* being semantic. Formal operations are the ones that are specified without reference to such semantic properties as, for example, truth, reference, and meaning." Other proponents of FSC include Field (2001), Gallistel and King (2009), Haugeland (1985), Pylyshyn (1984), and Stich (1983). All these authors combine FSC with CTM. According to CTM+FSC, mental activity manipulates formal syntactic items without regard to their representational or semantic properties. Perhaps mental states *have* representational properties. But we should delineate *non-intentional syntactic descriptions* that leave those properties unmentioned.

Stich (1983) espouses an extreme version of the formal syntactic approach. He advises cognitive science to describe the mind through purely syntactic models that ignore representational import. He recommends that scientific psychology jettison mental content altogether. Few proponents of CTM+FSC condone this extreme rejection of mental content. Most proponents try to secure a central explanatory role for formal mental syntax while *also* preserving a central role for representation (Fodor, 1987, 2008). All proponents agree that cognitive science should include a level of description that characterizes mental states in syntactic, non-semantic terms.

I think that FSC fits many computations quite well, including computations executed by standard personal computers. However, I reject FSC as a theory of computation *in general*.

There is no sound reason to hold that all computation manipulates formal syntactic items. (Rescorla, 2012b) introduces an alternative *semantically permeated* conception that integrates representation much more thoroughly into computational modeling. On the semantically permeated conception, certain computational models individuate computational states through their representational properties *as opposed to* any alleged formal syntactic properties. Specifically, computational models of mental activity can type-identify mental states in representational terms rather than formal syntactic terms. We can model the mind computationally without postulating formal mental syntax.[3]

In §2, I present basic elements of the semantically permeated conception. In §3, I argue that the Turing machine formalism is quite hospitable to semantically permeated computation. In §4, I discuss how current explanatory practice within cognitive science bears upon the contrast between formal-syntactic computation and semantically permeated computation. In §5, I critique some popular philosophical arguments for FSC. I will not argue that the semantically permeated conception is superior to FSC. Nor will I develop the semantically permeated conception in full detail. My discussion is programmatic. I seek only to convince you that current philosophical discussion has precipitously embraced a formal-syntactic picture of mental computation, at the expense of an equally appealing semantically permeated alternative.

§2. Individuating mental symbols

Let us begin by considering Fodor's version of CTM. Fodor advocates *the representational theory of mind* (RTM), which postulates *mental representations* comprising *the language of thought* (or *Mentalese*). Mentalese contains primitive symbols and compounding

[3] Burge (2010, pp. 95-101) and Peacocke (1994) propose somewhat similar treatments of computation. For critical discussion of these and other neighboring positions, see (Rescorla, 2012a).

devices for generating complex expressions. It has a *compositional semantics*: the meaning of a complex Mentalese expression is determined by the meanings of its parts and the way those parts are combined. As Fodor (1987) emphasizes, RTM explains two crucial phenomena: *systematicity* (there are systematic relations among which thoughts a thinker can entertain) and *productivity* (even though the mind is finite, one can entertain a potential infinity of thoughts). For instance, we explain productivity by positing a finite base of primitive Mentalese symbols, combinable through compounding devices into complex expressions. Iterated application of compounding devices generates a potential infinity of expressions.

According to CTM+RTM, mental activity instantiates Turing-style computation over the language of thought. Mental computation stores Mentalese expressions in memory locations, manipulating those expressions in accord with mechanical rules. To delineate a computational model of a mental process, we specify the Mentalese expressions manipulated by the process, and we isolate mechanical rules governing how the process manipulates those expressions.

I assume that CTM+RTM is correct. I focus on the following key question: when we construct computational models of mental activity, to what extent should representational properties inform how we individuate Mentalese expressions?

§2.1 Semantic neutrality, indeterminacy, and permeation

According to Fodor, "mental representations… have both formal and semantic properties," and "mental representations have their causal roles in virtue of the formal properties" (1981, p. 26). Crucially, formal properties underdetermine semantic properties: "mental representations can differ in content without differing in their intrinsic, formal, nonrelational, nonsemantic properties" (1991, p. 298). To illustrate, consider Putnam's (1975)

Twin Earth thought experiment. Oscar's mental states represent water, while Twin Oscar's corresponding mental states represent twater (the substance on Twin Earth). This is a *semantic* difference between Oscar's mental states and Twin Oscar's corresponding mental states. But Fodor holds that Oscar's mental representations have the same *formal syntactic* properties as Twin Oscar's corresponding mental representations.

On Fodor's approach, computational models of the mind should type-identify Mentalese symbols through their formal syntactic properties rather than their semantic properties. Phrased in more ontologically loaded terms, Fodor postulates an array of *formal syntactic types* to serve as the items manipulated by mental computation. For example, he postulates a formal syntactic type WATER that could denote either water or twater, depending on the thinker's causal relations to the external world. Initially, Fodor (1981, pp. 225-253) held that formal syntactic type *constrains* meaning while leaving meaning underdetermined. WATER could denote water or twater, but it could not denote dogs. Fodor's later work (1994, 2008) suggests a stronger indeterminacy thesis: a Mentalese syntactic type could have had an *arbitrarily* different meaning, had it figured differently in the thinker's psychology or her causal interactions with the world. WATER could denote dogs, or Bill Clinton, or anything else. Many researchers explicitly endorse the stronger indeterminacy thesis (Egan, 1992, p. 446), (Field, 2001, p. 58), (Harnad, 1994, p. 386), (Haugeland, 1985, p. 91, pp. 117-123), (Pylyshyn, 1984, p. 50).

I say that an entity is *semantically indeterminate* when it does not have its meaning essentially. A semantically indeterminate entity could have had different meaning without any change in its fundamental nature, identity, or essence. I say that an entity is *semantically neutral* when it bears an arbitrary relation to its meaning (assuming it even has meaning). A semantically neutral entity could have had *arbitrarily different* meaning, or no meaning at all, without any change in its fundamental nature, identity, or essence. Semantic neutrality entails semantic indeterminacy, but not vice-versa: semantic indeterminacy entails only that the entity could have had *some* different meaning, while semantic neutrality entails that the entity could have had *any* different meaning. Egan, Field, and Haugeland hold that Mentalese syntactic types are semantically neutral, as do most other contemporary advocates of CTM. Fodor's early work treats Mentalese syntactic types as semantically indeterminate, while his later work seems to treat them as semantically neutral.

I will explore computational models that type-identify Mentalese symbols at least partly through their semantic properties. In this spirit, I say that an entity is *semantically permeated* when we cannot change its meaning while holding fixed its fundamental identity, nature, or essence. A semantically permeated symbol is not a piece of formal syntax requiring an interpretation. Rather, the semantics of the symbol is built into the symbol's inherent nature. The symbol "comes with its meaning attached." I propose that we postulate an array of semantically permeated Mentalese symbols (or mental representations). For example, we can posit a Mentalese word DOG that necessarily denotes dogs, a Mentalese word SQUARE that necessarily denotes the property of being square, a Mentalese word AND that necessarily expresses conjunction, and so on.

To understand my proposal, one must keep the type/token distinction firmly in mind. I propose that we postulate semantically permeated Mentalese symbol-*types*, not that we postulate semantically permeated *tokens*. If we postulate a Mentalese symbol-type DOG that has its meaning essentially, we do not thereby claim that tokens of this type have their meanings essentially. Any possible token of DOG denotes dogs *so long as it is a token of DOG*. A given token of DOG might not have denoted dogs, but then it would not have been a token of DOG.

Although semantically permeated Mentalese individuation departs significantly from contemporary orthodoxy, it has a strong historical basis. Going back all the way to the 14th century, consider William of Ockham. Like most philosophers, Ockham holds that the connection between a natural language word and what it represents is arbitrary, since we can change the word's meaning as we please by changing our linguistic conventions. The English word "dog" could just as easily have denoted cats. Ockham also postulates a mental language whose elements have fixed, unchangeable denotations. He describes the contrast between natural language and mental language as follows (*Summa Logicae*, I.1):

> A concept or mental impression signifies naturally whatever it does signify; a spoken or written term, on the other hand, does not signify anything except by free convention. From this follows another difference. We can change the designation of the spoken or written term at will, but the designation of the conceptual term is not to be changed at anybody's will.

In my terminology, Ockham holds that natural language words are semantically neutral but that Mentalese words are semantically permeated. Ockham offers no hint that we can "hive off" the denotation of a Mentalese word, leaving behind a non-semantic syntactic residue. Whereas Fodor posits formal mental syntax subject to varying interpretations, Ockham posits a mental language whose nature fixes a unique interpretation.

There are many other precedents for my proposal. As Burge notes, a traditional view holds that "concepts' identities are inseparable from their specific intentional properties or functions. Thus a concept of an eclipse could not be the concept that it is if it did not represent, or if it were not about, eclipses." (2007, p. 292). In my terminology, the traditional view holds that concepts are semantically permeated. Rather than speak of "concepts," I speak of

"Mentalese symbols" or "mental representations." I choose these locutions so as to accommodate *non-conceptual representations* (if such there be). For example, Burge thinks that perception is non-conceptual. He postulates *perceptual attributives*, which are perceptual analogs to predicates or concepts. He holds that "perceptual attributives determine, or *specify*, the attributes they attribute. They are not only *as of* the attributes; they are as of the same attribute in every context of use and with regard to any possible situation" (2010, p. 76). In my terminology, Burge holds that perceptual attributives are semantically permeated. My formulations are general enough to encompass perceptual attributives and other putatively non-conceptual mental representations.

§2.2 Developing the semantically permeated viewpoint

I characterized "semantically permeation" in fairly sketchy terms. In particular, I used the rather nebulous term "meaning," rather than some more precise technical term. One might develop the semantically permeated viewpoint in various directions, depending on how one glosses the crucial term "meaning." For example, one might type-identify Mentalese expressions by citing Russellian propositions, or sets of possible worlds, or Fregean senses.[4] There are many other options. A complete semantically permeated theory must choose among these options. I remain neutral regarding such details, which will not affect my argumentation. For present

[4] For representative modern treatments of the Russellian, possible worlds, and Fregean approaches, see (Salmon, 1986), (Stalnaker, 1984), and (Peacocke, 1992) respectively. In (Rescorla, 2012b), I develop a broadly Fregean version of semantically permeated CTM+RTM. Schneider (2011, p. 100) rejects a semantically permeated individuative scheme for Mentalese, partly because she thinks that such a scheme cannot handle *Frege cases*, i.e. cases where a thinker represents the same entity under different modes of presentation. Schneider mistakenly assumes that a semantically permeated scheme must type-identify mental symbols in Russellian fashion. She does not even consider an alternative Fregean approach that type-identifies mental symbols by citing externalistically individuated modes of presentation.

purposes, what matters is that a semantically permeated individuative scheme classifies mental states *at least partly* through their representational import.[5]

The Twin Earth thought experiment shows that representational import does not always supervene upon internal neurophysiology. Thus, semantic permeation induces an *externalist* approach to Mentalese symbol individuation. For example, semantically permeated theorists will postulate a Mentalese word WATER that necessarily denotes water. When Oscar wants to drink water, he stands in a certain cognitive relation to the Mentalese word WATER. Twin Oscar's mental states do not represent water, so he does not stand in any significant cognitive relation to WATER. Instead, he stands in a cognitive relation to a type-distinct Mentalese word TWATER. Even though Oscar and Twin Oscar are neurophysiological duplicates, they entertain different Mentalese word-types. Externally determined denotation plays a crucial role in type-identifying the relevant words. Oscar and Twin Oscar have different mental languages.

Some philosophers respond to the Twin Earth thought experiment by recommending that we replace *wide content* (which does not supervene upon internal neurophysiology) with *narrow content* (which does so supervene). On Fodor's (1981, p. 227, p. 240) early view, each formal syntactic type determines a unique narrow content but not a unique wide content. So formal syntactic type underdetermines vital aspects of a mental state's representational import (e.g. WATER might denote either water or twater). I count these Fodorian formal syntactic types as semantically indeterminate, even though each such type allegedly has a certain kind of content

[5] Subtle issues arise concerning the "compounding devices" that generate complex Mentalese expressions. To ensure that complex Mentalese expressions are semantically permeated, we must isolate compounding devices with fixed compositional import. For preliminary discussion, see (Rescorla, 2012b). This paper focuses on issues raised by the individuation of primitive Mentalese words.

(narrow content) essentially. Semantic permeation requires that an entity have its meaning essentially *in some sense of "meaning" correlative with "representational import."*[6]

Phenomena such as context-sensitivity and reference failure might lead one to refine, qualify, or weaken the semantically permeated viewpoint in various ways. There are many complexities here that I must skirt.[7] This paper emphasizes abstract issues that arise however exactly one develops the semantically permeated viewpoint.

§2.3 Explanation and taxonomization

Critics often lambast semantically permeated entities as mysterious or obscure. Putnam complains that "[n]one of the methods of representation we know about has the property that the representations *intrinsically* refer to whatever it is that they are used to refer to" (1988, p. 21). He warns us to be "highly suspicious of theories that postulate a realm of 'representations' with *such* unlikely properties" (p. 22). He hints, without asserting, that positing such representations is tantamount to positing entities with "magical" powers.

I find Putnam's objection unconvincing. A key point here is that semantically permeated types are *types*, whose primary role in our discourse is to taxonomize tokens. Any psychological theory must adopt a taxonomic scheme for categorizing token mental states, processes, and events. We reify the categories by positing a collection of types. The types are abstract entities

[6] Around the mid-1990s, Fodor abandons narrow content. A constant element in his position is his emphasis upon formal syntactic types that underdetermine representational import. Aydede (2005) proposes that we type-identify Mentalese symbols partly by what he calls their "semantic properties." However, he seems reluctant to develop this proposal in an externalist direction (p. 203, fn. 26). He instead inclines towards permeation by some kind of narrow content. Ultimately, then, Aydede's proposal seems closer to Fodor's (1981) view than to my view.

[7] For example, one might postulate a Mentalese demonstrative THAT that can only denote some demonstrated entity but whose *particular* denotation depends upon context. More specifically, one might propose that THAT does not have its *denotation* essentially but does have something like its *character* in the sense of (Kaplan, 1989) essentially. This proposal individuates THAT partly by *context-insensitive* aspects of its representational import but not by *context-sensitive* aspects. Does the proposal classify THAT as semantically permeated? To answer the question, one requires a more precise definition of "semantic permeation" than I have provided.

corresponding to our classificatory procedures. Semantically indeterminate types correspond to a taxonomic scheme that underdetermines meaning. Semantically neutral types correspond to a taxonomic scheme that leaves meaning completely unconstrained. Semantically permeated types correspond to a taxonomic scheme that takes meaning into account. As Burge (2007, p. 302) notes, there is nothing obscure or magical about the latter taxonomic scheme. On the contrary, it figures centrally in ordinary mentalistic discourse. Setting aside generalized skepticism about abstract entities, I see no *metaphysical* problem about semantically permeated Mentalese types. For instance, there is nothing supernatural about a Mentalese symbol-type DOG that refers to dogs by its essential nature. Semantically permeated Mentalese symbol-types are mere "ontological correlates" to a taxonomic scheme that type-identifies mental states, events, or processes partly through their semantic properties.

Individuation serves explanation. When we debate the proper individuation of Mentalese symbols, we are ultimately debating the proper format for psychological explanation. The central issue here is not ontological but explanatory. How do our best psychological explanations type-identify mental states, events, and processes? Should we employ a taxonomic scheme that cites representational properties of mental states? Or should we employ a taxonomic scheme that leaves representational properties underdetermined?

One might find room in one's theorizing for *both* kinds of taxonomic scheme. For example, Fodor (1998) devotes an entire book to entities that he calls "concepts." A Fodorian concept determines a unique denotation (1998, pp. 20-21), (2008, p. 70), so it is semantically permeated. Thus, Fodor postulates both semantically permeated types (concepts) *and* semantically indeterminate types (formal syntax).

§2.4 Formal-syntactic computation versus semantically permeated computation

CTM+FSC embraces formal syntactic taxonomization. Proponents view the mind as a Turing-style device that manipulates semantically indeterminate syntactic types. Mental computation operates over these types, without regard to representational properties. In that sense, the mind is a "syntactic engine."

One must distinguish between *syntactic* description and *neurophysiological* description. Beginning with Putnam (1967), advocates of CTM have repeatedly stressed that computational models are *multiply realizable*: systems with wildly heterogeneous physical properties can instantiate the same computational properties. In particular, formal syntactic description is supposed to be far more abstract than neurophysiological description, citing properties shareable by diverse physical systems: carbon-based, silicon-based, and so on. Fodor (2008, p. 91), Gallistel and King (2009, pp. 137-148), Haugeland (1985, p. 5), Stich (1983, p. 151), and virtually all other proponents of CTM+FSC tout multiple realizability as a prime benefit of their approach. They commend syntactic description for ignoring incidental neurophysiological details. Thus, CTM+FSC prioritizes formal syntactic descriptions that abstract away from both semantic properties *and* neurophysiological properties. Proponents hold that mental computation operates over semantically indeterminate, multiply realizable syntactic types.

On the alternative picture that I will explore, many notable mental computations operate over semantically permeated Mentalese expressions without intercession by formal mental syntax. These computations store semantically permeated Mentalese symbols in memory locations, manipulating the symbols according to mechanical rules. The symbols lack formal syntactic properties of any theoretical interest. On this picture, many core mental computations are composed from inherently representational building blocks. The mental symbols over which the computations operate, and hence the computations themselves, have natures inextricably tied to their representational import. We cannot factor out representational import to generate an explanatorily significant formal syntactic remainder.

The alternative picture has a positive component and a negative component. The positive component is that we promote good psychological explanation of various core mental phenomena by delineating Turing-style models defined over semantically permeated Mentalese symbols. The negative component is that formal syntactic taxonomization adds no explanatory value to theorizing about those same phenomena. I assume that fruitful explanation is our best guide to underlying natures. Given this assumption, the positive and negative components jointly indicate that certain core mental computations manipulate mental symbols whose natures are inextricably tied to their representational import.

A satisfying scientific psychology will certainly include *non-representational neurophysiological* descriptions. But will it include *non-representational syntactic* descriptions? Given multiple realizability, neurophysiological description is fundamentally distinct from formal syntactic description. If we allow a sufficiently disjunctive or gerrymandered taxonomic scheme, it may be that every mental representation has a semantically indeterminate, multiply realizable syntactic type. What I doubt is that each mental representation has an *explanatorily significant* semantically indeterminate, multiply realizable syntactic type. There are indefinitely many ways to type-identify mental states. Most taxonomic schemes hold no interest for us (e.g. a scheme that type-identifies mental states by citing the current temperature on Mars). Only certain special taxonomic schemes serve our explanatory ends. Does a formal syntactic taxonomic scheme for mental representations serve psychological explanation? Or is formal mental syntax a gratuitous theoretical posit? The issue here is not whether we *can* individuate mental representations in formal syntactic fashion. The issue is whether we *should*.

§3. Turing computation over semantically permeated types

Philosophers and cognitive scientists often suggest that formal syntax is an integral component of computational modeling. In this spirit, Fodor (1981, p. 241) writes: "Computations just *are* processes in which representations have their causal consequences in virtue of their form." Similarly, Gallistel and King (2009, p. 107) say that the symbols manipulated by Turing machines "are to be regarded as purely arbitrary symbols (really data), having no more intrinsic reference than magnetic patterns," while Haugeland (1985) defines a computer as "a *symbol-manipulating machine*" (p. 106), where "the meanings of symbols (e.g. words) are *arbitrary*… in the sense that there is no intrinsic reason for them to be one way rather than another" (p. 91). These passages, and many similar passages found throughout the literature, are offered as uncontroversial remarks that everyone should accept.

I believe that philosophical discussion of computation should ground itself in the mathematical theory of computation, construed broadly to include recursion theory, complexity theory, and theoretical computer science. If we take mathematical computation theory as our guide, then there is no reason to insist that computational modeling requires formal syntax. Abstract mathematical models of computation are indifferent between semantic neutrality, indeterminacy, and permeation. Nothing about the *mathematics* of computation mandates explanatorily significant formal syntactic types.

To see why, let us examine the Turing machine formalism. Philosophers commonly describe Turing machines along the following lines:

(1) A Turing machine consists of a scanner and infinite paper tape, divided into cells. The scanner manipulates strings of strokes inscribed upon the tape. The scanner can erase a stroke, write a stroke, or move to the left or right. A machine table enshrines routine, determinist rules governing these manipulations. The scanner's action is determined entirely by its current "internal state" and by the tape's configuration at the scanner's current location.

A string of stokes does not have essential meaning or content. We could interpret it however we please. Thus, (1) enshrines the orthodox conception of computation as defined over semantically indeterminate entities.

Heuristic descriptions such as (1) have undeniable pedagogical value. However, they do not capture the notion of Turing machine in full generality. Turing machines can operate over entities other than strings of strokes. A proper description should invoke some general notion of "symbol," rather than restricting attention to strokes. Another problem with (1) is its talk about a "scanner" that "moves" along a "tape." Although Turing indulges in such talk, we can only construe it metaphorically. Why restrict attention to literal tapes rather than other physical embodiments, such as silicon chips? Finally, how can a rigorous mathematical theory deploy informal notions such as "motion"?

A proper formulation must divest (1) of all picturesque embellishments. Once we jettison inessential metaphors, the following core notion emerges:

(2) A Turing machine contains a central processor and an infinite linear array of memory locations. The scanner can access one memory location at each stage of computation. There is a finite set of primitive symbols, any of which can be inscribed at a memory location. The processor can erase a symbol from the currently accessed memory location, write a symbol to the currently accessed memory location, or access the previous or next location in the memory array. A machine table enshrines routine, deterministic rules governing these manipulations. The central processor's action is determined entirely by its current "internal state" and by the contents of the currently accessed memory location.

I submit that (2) provides a much more accurate characterization of Turing computation than (1).

Once formulated with proper generality, Turing's conception enshrines no bias towards semantic indeterminacy. A semantically permeated symbol can be inscribed or erased at a memory location. Routine deterministic instructions can mention semantically permeated symbols. Thus, there is no obvious bar to a Turing machine defined over semantically permeated symbols. Such a machine contains an infinite, linearly structured array of memory locations (the "cells" of the "tape"). It contains a central processor (the "scanner"), which can access one memory location at a time. It performs the same elementary operations as a Turing machine defined over semantically indeterminate items: accessing the next memory location in the linear memory array ("moving to the right"); accessing the previous memory location in the linear memory array ("moving to left"); inscribing a symbol in a memory location; erasing a symbol from a memory location. It merely performs these operations upon semantically permeated rather than semantically indeterminate items.

Somewhat more formally, consider how modern computation theory codifies Turing's conception in mathematically rigorous fashion. Taylor's definition (1998, p. 73), which is representative of the modern literature, runs roughly as follows:

(3) A Turing machine *M* is an ordered quadruple $<Q, \Sigma, q_{init}, \delta>$, where Q is a nonempty finite set (the set of *states* of *M*); Σ is a nonempty finite *alphabet of symbols*, including the null symbol (*blank space*); $q_{init} \in Q$ is a privileged *initial state*; and δ is a partial function (the *transition function* of *M*) from $\Sigma \times Q$ to $(\Sigma \cup \{L, R\}) \times Q$.

(3) mentions a finite alphabet of symbols Σ but say nothing about the nature of those symbols. As far as the definition goes, the symbols could be semantically neutral, semantically indeterminate, or semantically permeated. So the abstract modern definition of Turing machine is entirely congenial to a semantically permeated individuative scheme for Mentalese, including a scheme that takes wide content into account. The abstract definition provides no reason to associate each Mentalese word with a formal syntactic type.

To illustrate, suppose that the transition function δ maps $<r_1, q_1>$ to $<r_2, q_2>$, where r_1, $r_2 \in \Sigma$ and q_1, $q_2 \in Q$. This corresponds to the following mechanical rule:

(R) If the central processor is in state q_1, and if it accesses a memory location that contains symbol r_1, then replace r_1 with r_2 and shift to central processor state q_2.

Neither rule R nor our formalization through δ requires that r_1 and r_2 be semantically indeterminate. Rule R applies just as well to symbols individuated through their representational properties as to symbols individuated through formal syntactic properties. For example, r_1 might be a Mentalese word WATER that necessarily denotes water.

We should reject the popular philosophical view that all Turing machines are defined over semantically indeterminate items. This popular view is tempting only if one confines attention to overly picturesque descriptions such as (1). The rigorous, abstract notion of Turing machine, as encapsulated by (2) and formalized by (3), generates no impetus towards formal syntactic computational vehicles.[8] We can therefore model mental activity as Turing-style computation over semantically permeated Mentalese types, without postulating formal mental

[8] In (Rescorla, 2012b), I offer a similar diagnosis for other mathematical models of computation, including the *register machine* and the *lambda calculus*. In each case, I argue that the relevant computational formalism is hospitable to semantically permeated individuation.

syntax. Genuinely computational models can describe mechanical manipulation of mental representations *individuated partly through their representational import*. Despite what many researchers suggest, the mathematical study of computation does not favor FSC.

Some readers may worry that semantically permeated computation requires an "inner homunculus" who interprets Mentalese expressions. Consider Rule R, and suppose that the types r_1 and r_2 are semantically permeated. For mental computation to conform reliably to R, surely it must evaluate whether a given token has type r_1. And surely that requires evaluating whether the token has appropriate semantic properties. More specifically, suppose r_1 is a Mentalese word WATER that necessarily denotes water. In manipulating this word, doesn't the system first need to check whether the word denotes water versus twater? Wouldn't that require deciding whether the system is located on Earth versus Twin Earth?

These worries are misplaced. A computational model defined over Mentalese expressions taxonomizes mental states and operations by citing mental symbol-types. In any normal case, the symbol-types are not *objects* of cognition or awareness. Rather, they are abstract types that *we theorists* cite to taxonomize mental states, events, and processes. Mental computation does not normally represent mental symbols. It *tokens* the symbols. Conformity to Rule R does not require that a computational system evaluate whether tokens have types r_1 or r_2. Conformity simply requires that the system move appropriately between mental states that have types r_1 and r_2. In the special case where r_1 is semantically permeated, the system need not evaluate representational properties of mental states, because it need not evaluate whether states have type r_1. It need merely move appropriately between states with appropriate representational properties. For example, suppose r_1 is a Mentalese word WATER that necessarily denotes water. In manipulating this word, the system need not evaluate whether tokens represents water versus twater, or whether the system's location is Earth or Twin Earth. The system need merely transit in the right way among mental states, some of which represent water. As long as the system so transits, it conforms to rule R.

Implementing a semantically permeated Turing-style model does not require *evaluating* representational properties of mental states. It requires *transiting* appropriately between mental states with appropriate representational properties. Thus, there is no reason to suspect that semantically permeated computation requires an inner homunculus.

§4. Modeling mental computation

I have presented two opposing conceptions of computation: the formal-syntactic conception and the semantically permeated conception. How do the two conceptions apply to computation in physical systems? In my view, the answer depends upon the physical system. Computer science routinely offers semantically indeterminate models of artificial computing systems (e.g. personal computers). The explanatory and pragmatic success of these models provides strong evidence that the relevant systems compute by manipulating formal syntactic items. However, we cannot immediately infer that *minds* likewise compute by manipulating formal syntactic items. There are many differences between minds and artificial computing systems. Our best models of the former may differ markedly from our best models of the latter.

In §2.4, I suggested that certain core mental phenomena are best handled through semantically permeated rather than semantically indeterminate computational modeling. I intend this suggestion in a tentative, conjectural spirit. I also allow that some important mental phenomena are best described through semantically indeterminate computational models, or through models that contain a mixture of semantically indeterminate and semantically permeated symbols.[9] At present, current cognitive science does not offer anything resembling well-confirmed Turing-style models of specific mental phenomena. As well-confirmed Turing-style models emerge, we can assess the extent to which they are semantically neutral, indeterminate, or permeated. Until then, we should keep all theoretical options open.

Fodor (1975, 1981) and his allies frequently claim that cognitive science postulates mental computation over semantically indeterminate formal syntactic types. They claim that formal syntactic description figures essentially within our best scientific theories of mental activity. I think that such claims vastly overstate the centrality of formal mental syntax to contemporary scientific practice. Although current cognitive science may describe *certain* mental phenomena in formal syntactic terms, it eschews formal syntactic description when explaining numerous core mental phenomena. It describes numerous mental processes in representational terms *as opposed to* formal syntactic terms. I will not defend my assessment here. But I will illustrate by considering one important mental process: *perception*.

The perceptual system reliably estimates distal properties (e.g shapes, sizes, and distances) based upon proximal sensory stimulations (e.g. retinal stimulations). For example, the visual system estimates the distance of a perceived body based upon numerous visual cues, including convergence, binocular disparity, linear perspective, motion parallax, and so on (Palmer, 1999, pp. 200-253). The visual system also consults distance cues when estimating distal size: if bodies *A* and *B* subtend the same retinal angle but distances cues indicate that *A* is farther away, then the perceptual system will typically estimate that *A* is larger than *B* (Palmer, 1999, pp. 314-327). Perceptual psychology studies such phenomena. It provides detailed

[9] For example, Gallistel and King (2009) mount a compelling case that *dead reckoning* manipulates symbols inscribed in read/write memory. (Rescorla, 2013b) suggests that current science describes certain cases of invertebrate dead reckoning in non-representational terms. So these may be cases where formal-syntactic computational description is more apt than semantically permeated computational description.

psychological models that explain how the perceptual system estimates distal properties based upon proximal sensory stimulations (Feldman, forthcoming), (Knill and Richards, 1996), (Vilares and Körding, 2011).

As Burge (2010) and Peacocke (1994) emphasize, the science routinely describes perceptual activity in representational terms. It individuates perceptual states through representational relations to specific shapes, sizes, distances, and so on. For example, models of distance perception type-identify perceptual states as estimates of specific distances, while models of size perception type-identify perceptual states as estimates of specific sizes. Perceptual psychology does *not* attribute formal syntactic properties to perceptual states. As Burge puts it, "there is no explanatory level in the actual science at which any states are described as purely or primitively syntactical, or purely or primitively formal. One will search textbooks and articles in perceptual psychology in vain to find mention of purely syntactical structures" (2010, p. 96). Taking perceptual psychology as our guide, it is sheer fantasy to postulate that perceptual activity manipulates formal syntactic items. The science describes how proximal sensory input *as characterized neurophysiologically* determines a perceptual estimate *as characterized representationally*. Formal mental syntax plays no role (Rescorla, forthcoming a).

Any adequate theory of perception must, among other things, illuminate the *neural mechanisms* that underlie perceptual estimation. Perceptual psychology offers some theories in this vein (Pouget, Beck, Ma, and Latham, 2013). The theories tend to be rather more speculative than theories couched solely at the representational level. However, we can all agree that a completed perceptual psychology will contain non-representational *neural* descriptions. The question is whether it should include non-representational *syntactic* descriptions. Should we postulate multiply realizable, semantically indeterminate types manipulated during perceptual processing? Current science postulates no such types. It describes perceptual activity in *representational* terms and *neurophysiological* terms, but never in *formal syntactic* terms.

One might recommend that we *supplement* perceptual psychology with formal syntactic descriptions. But that recommendation needs sound backing. Nothing about current perceptual psychology taken in itself suggests any significant explanatory role for formal syntactic computational modeling.

Beginning with Quine (1960), some philosophers have insisted that intentionality should not figure in mature scientific psychology. They argue that intentional discourse is unclear, interest-relative, explanatorily bankrupt, or otherwise unscientific (Churchland, 1981), (Field, 2001), (Stich, 1983). This attitude prompts Stich (1983) to recommend that we replace intentional psychology with purely syntactic computational modeling of the mind. Even philosophers more sympathetic to intentionality, such as Fodor (1987, pp. 16-26), often suggest that formal syntactic modeling provides intentional psychology with a secure scientific grounding that it would otherwise lack.

I dispute all these assessments. There is nothing unscientific about intentional discourse per se. If we consider how current science actually works, rather than how various philosophers think that it should work, then representation looks like a thoroughly legitimate theoretical construct. In particular, it has repeatedly proved its explanatory value within perceptual psychology. The diverse arguments offered by Churchland, Field, Quine, and Stich against the scientific credentials of intentional discourse are notably less compelling than the intentional explanations routinely offered by perceptual psychology. Moreover, I see no clear respect in which formal syntactic modeling is more scientifically respectable than intentional psychology.

Representation rather than formal mental syntax occupies center stage within our current best scientific theories of perception.

Over the past few decades, philosophers have defended FSC through various arguments that formal syntactic description offers decisive advantages over representational description. I cannot rebut all these arguments here. But I will now critique several of the most prominent arguments.

§5. The mechanisms of cognition

A natural thought is that semantically permeated computational description does not seem *mechanical* enough. Shouldn't a truly mechanical account "bottom out" in mindless responses to formal syntactic items? Consider rule R from §3, and suppose again that the relevant types r_1 and r_2 are semantically permeated. Some readers will insist that a physical system conforms to R only by virtue of conforming to the formal syntactic rule:

(R*) If the central processor is in state q_1, and if it accesses a memory location that contains symbol r_1*, then replace r_1* with r_2* and shift to central processor state q_2,

where r_1* and r_2* are formal syntactic types. Even if intentional description is useful for certain purposes, formal syntax plays an essential causal or explanatory role.

The argument as stated is highly schematic. Let us consider how one might develop it less schematically.

§5.1 The causal relevance of content

An extreme version of the argument maintains that representational properties are *causally irrelevant* to mental activity. Suppose we grant this premise. Then it becomes natural to pursue semantically indeterminate models, so as to secure causal theories of mental computation. Egan (2003), Gallistel and King (2009), and Haugeland (1985) argue along these lines.

Luckily, there are good reasons to reject the crucial premise that representational properties are causally irrelevant. These issues have been thoroughly litigated over the past few decades, so I will just briefly highlight four key points:

(1) We have a strong pre-theoretic intuition that mental content is causally relevant. For example, whether I want to drink water versus orange juice seems causally relevant to whether I walk to the sink versus the refrigerator. To deny pre-theoretic intuitions along these lines is to embrace radical epiphenomenalism.

(2) As emphasized in §5, cognitive science offers numerous explanations that type-identify mental states in representational terms. The explanations certainly look like causal explanations (Burge, 2007, pp. 344-362). For example, current theories of sensorimotor control describe how an intention to move my finger to a certain egocentric location causes certain muscle activations (Bays and Wolpert, 2007). Taken at face value, these theories depict how representational properties of my intention causally influence my muscle activations.

(3) In (Rescorla, 2014), I provide a detailed argument that mental content --- including wide content ---- is causally relevant to mental computation. In particular, I argue that representational properties can causally influence elementary computational operations. Thus, genuinely causal explanatory generalizations can individuate computational states representationally.

(4) Various widely circulated arguments maintain that mental content --- especially wide content --- is causally irrelevant to mental activity. I agree with Burge (2007, pp. 316-

> 382) that these arguments are serious flawed. I critique some of them in (Rescorla, 2014).

Given (1)-(4), we may with clear conscience pursue intentional causal explanations of mental activity. Semantically permeated versions of CTM instantiate this explanatory strategy.

I now elaborate upon (4) by critiquing two representative arguments that mental content is causally irrelevant.

Haugeland observes that “meanings (whatever exactly they are) don’t exert physical forces” (1985, p. 39). If we say that meanings “affect the operation of the mechanism,” then “all the old embarrassments return about exerting forces without having any mass, electric charge, etc.: meanings as such simply cannot affect a physical mechanism” (p. 40). He concludes that “meanings don’t matter” to computational operations (p. 44). On that basis, he urges us to model mental activity as formal manipulation of semantically neutral symbols.

I agree with Haugeland that meanings do not exert physical forces. Meanings are abstract entities, so they do not participate in causal interactions. It does not follow that intentional properties are causally irrelevant. Intentional *properties* can be causally relevant even though intentional *contents* do not enter into causal transactions. To adopt a well-worn analogy: numbers are abstract entities, so they cannot causally interact with physical objects; yet an object’s mass, as measured by some number, is causally relevant to physical processes. One can specify causally relevant properties by citing abstract entities. Thus, Haugeland’s argument does not establish that genuinely causal explanations should ignore intentional content.

Haugeland offers a second argument that intentional properties do not “matter” to mental computation. He suggests that meanings “matter” only if there is an inner homunculus capable of “‘reading’ the symbols in [the thinker’s] mind, figuring out what they mean, looking up rules of reason, deciding which ones to apply, and then applying them correctly" (p. 41). He denies that such a homunculus exists. In a similar vein, Gallistel and King write: "[f]or many purposes, we need not consider what the symbols refer to, because they have no effect on how a Turing machine operates. The machine does not know what messages the symbols it is reading and writing designate (refer to)" (2009, p. 108).

I agree that no inner homunculus interprets Mentalese symbols. I agree that a typical Turing machine does not "know" the semantics of symbols it manipulates. We should not infer that representational properties do not "matter" or that they have "no effect" on computation. If I throw a baseball at a window, then the window does not contain a homunculus that inspects baseball momentum. The window does not know the baseball's momentum. Nevertheless, the baseball's momentum is causally relevant to whether the window breaks. The momentum "matters" to the window's breaking. Likewise, representational properties can "matter" to mental computation even though no inner homunculus inspects a symbol's meaning. There is no cogent inference to the causal irrelevance of representational properties from the non-existence of an inner homunculus, or from the fact that typical Turing machines lack semantic knowledge.

I have critiqued two arguments that mental content is causally irrelevant. The literature offers many additional arguments along the same lines, some geared towards mental content in general, some geared more specifically towards wide content. I refer readers to (Rescorla, 2014) for further discussion of these matters.

§5.2 Fodor on implementation mechanisms

Let us now assume that representational properties are causally relevant. Even so, intentional causation may seem rather mysterious. Perhaps there are genuinely causal generalizations that cite representational properties, but surely that is not the end of the matter. Surely we should ground intentional causal generalizations in *non-intentional* mechanisms. For example, whether a mental state represents water depends upon complex causal-historical relations to the environment. Shouldn't we isolate underlying computational mechanisms that prescind from all such causal-historical relations, citing only "local" properties of mental states?

Fodor develops this viewpoint. He holds that intentional content should figure in laws offered by scientific psychology. He also espouses a nomological theory of causal relevance, so that appropriate participation in intentional laws ensures the causal relevance of intentional properties (1990, p. 148). At the same time, he insists that intentional laws require an *implementation mechanism*: "it's got to be possible to tell the whole story about mental causation (the whole story about the implementation of the generalizations that belief/desire explanations articulate) *without referring to the intentional properties of the mental states that such generalizations subsume*" (1987, p. 139). He proposes that formal syntactic mechanisms implement intentional laws. On the resulting picture, *syntactic* mechanisms ensure the causal relevance of *intentional* properties.

To motivate his approach, Fodor (1987, p. 19) observes that the computer revolution enables us to build a machine such that

> The operations of the machine consist entirely of transformations of symbols; in the course of performing those operations, the machine is sensitive solely to syntactic properties of the symbols; and the operations that the machine performs on the symbols are entirely confined to altering their shapes. Yet the machine is so devised that it will transform one symbol into another if and only if the propositions expressed by the

> symbols that are so transformed stand in certain *semantic* relations --- e.g. the relation that the premises bear to the conclusion in an argument.

By positing semantically indeterminate mental syntax, we explain how mental activity respects semantic relations among mental states: "*if* the mind is a sort of computer, we begin to see how you can have a theory of mental processes… which explains how there could be nonarbitrary content relations among causally related thoughts" (1987, p. 19). CTM+FSC, unlike all rival theories, shows how mental states can participate in causal processes that "track" their meanings.

Fodor often suggests that we should individuate syntactic types by their *shapes*: "to all intents and purposes, syntax reduces to shape" (1987, p. 18), and "formal operations apply in terms of the, as it were, shapes of the objects in their domains" (1981, p. 227). Taken literally, such passages are unacceptable. As numerous commentators have observed (Bermúdez, 1995b, p. 364), (Block, 1983, pp. 521-522), internal states of an ordinary personal computer do not have shapes that are relevant to their syntax. Computation *can* be defined over syntactic types individuated partly by their shapes, but it *need not* be.

In other passages, Fodor recognizes that talk about shape is misleading (1987, p. 156, fn. 5). He envisages some more general notion of "form," not specifically tied to *geometric* form. Unfortunately, Fodor systematically equivocates between two very different positions. The first position holds that Mentalese syntactic types are individuated by *physical* or *neurophysiological* properties: "[t]okens of primitive Mentalese formulas are of different types when they differ in the (presumably physical) properties to which mental processes are sensitive" (2008, p. 79). The second position holds that Mentalese syntactic types are individuated *functionally,* so that "computational states and processes are multiply realized by neurological states (or whatever)" (2008, p. 91). On the first position, all tokens of a primitive Mentalese type share some salient physical or neurological property. On the second position, "[w]e can't take for granted that computationally homogeneous primitive Mentalese expressions *ipso facto* have neurologically homogeneous implementations; indeed, we had better take for granted that they often don't" (2008, p. 90). These two positions are inconsistent. Moreover, the first position blatantly flouts multiple realizability --- one of the main virtues of CTM advertised by Fodor (1975) himself.[10]

Setting aside these equivocations, the key point here is that Fodor's *arguments* do not support multiply realizable, semantically indeterminate Mentalese types. To see why, consider an ordinary personal computer. Computational states in the computer are realized by electromagnetic states. If we program the computer appropriately, then causal interactions among electromagnetic states "track" semantic relations among corresponding computational states. For example, we can program the computer so that it carries premises only to conclusions logically entailed by those premises. To explain why computational activity respects semantic relations among computational states, we can mention correlations between electromagnetic states and semantic properties. We thereby describe the machine in *electromagnetic* terms and *representational* terms. We can also describe the machine in formal syntactic terms, but doing so is not necessary for explaining non-arbitrary content relations among the machine's causally related states. Correlations between electromagnetic states and representational states already suffice for a satisfying explanation.

A similar point applies to mental activity. Let us grant that a complete theory of mental causation must isolate a non-intentional implementation mechanism. There are at least two options:

(1) Intentional generalizations are implemented by neurophysiological processes.

[10] For related criticisms of Fodor, see (Aydede, 1999), (Bermúdez, 1995a), (Prinz, 2011), and (Tye and Sainsbury, 2012, pp. 85-87).

(2) Intentional generalizations are implemented by formal syntactic processes.

(1) and (2) are compatible, but (2) goes beyond (1). To vindicate (1), we correlate intentional states with neural states, and we describe how transitions among neural states track transitions among intentional states. To vindicate (2), we must do more. We must introduce a formal syntactic description that applies not only to humans but also to diverse possible physically heterogeneous creatures. Perhaps we *can* isolate such a formal syntactic description. But why *should* we? Fodor has provided no sound argument that a good theory of implementation mechanisms requires (2) rather than (1).

According to Fodor, "[i]t is central to a computational psychology that the effects of semantic identities and differences on mental processes must always be mediated by 'local' properties of mental representations, hence by their nonsemantic properties assuming that semantics is externalist" (1994, p. 107). I agree that local properties of mental representations mediate the effects of semantic identities and differences. Transitions among representational mental states are not magical. Mental states are realized by neural states, and transitions among mental states are implemented by neural processes. Ultimately, the brain is just responding to local brain states. However, these uncontroversial observations do not favor (2) over (1). The "local non-semantic properties" to which a computational system responds may be neural rather than syntactic. As long as the system instantiates reliable correlations between neural and semantic properties, causal interactions among its internal states can "track" meanings in a non-arbitrary way.

Fodor wants to establish (2). But his appeal to non-intentional implementation mechanisms does not support (2) in addition to, or instead of, (1). In effect, Fodor's exposition bridges this argumentative gap by systematically equivocating between (1) and (2).

§5.3 Chalmers on causal topology

Unlike Fodor, Chalmers carefully maintains the distinction between syntactic and neurophysiological description. He argues that syntactic description "yields a sweet spot of being detailed enough that a fully specified mechanism is provided, while at the same time providing the minimal level of detail needed for such a mechanism," where a "fully specified mechanism" is one that provides "a recipe that could be copied to yield a system that performs the [cognitive or behavioral] function in question" (2012, p. 245). Representational description does not fully specify a mechanism, because it does not provide an explicit recipe that one can readily convert into a physical machine. Neural explanation fully specifies a mechanism, but it includes undesirable neural details. Formal syntactic description is genuinely mechanistic (unlike representational description), and it also offers a desirable level of generality that eludes neurophysiological description.

I reply that extra generality does not necessarily promote good explanation. Suppose we want to explain why John failed the test. We might note that

John did not study all semester.

Alternatively, we might note that

John did not study all semester *or* John was seriously ill.

There is a clear sense in which the second explanation is more general than first. Nevertheless, it does not seem superior. This simple example illustrates a widely recognized *problem of irrelevant disjunction*: if one boosts generality by appending irrelevant disjuncts, then no explanatory gain results (Williamson, 2000). Thus, the mere fact that formal syntactic explanations offer greater generality than neuroscientific explanations does not show that they yield any explanatory advance. They may achieve greater generality only by citing surreptitiously disjunctive or gerrymandered types.

These worries exert particular force against Chalmers's approach to computational modeling. Chalmers's account hinges upon two key definitions:

- The *causal topology* of a system is "the pattern of interaction among parts of the system, abstracted away from the make-up of individual parts and from the way the causal connections are implemented" (2011, p. 337).
- A property *P* is *organizationally invariant* just in case "any change to the system that preserves the causal topology preserves *P*" (2011, p. 337).

According to Chalmers, a computational model individuates computational states by citing organizationally invariant properties. For that reason, computational explanation is much more general than neural explanation.[11]

Chalmers's notion of *causal topology* provides a widely applicable procedure for converting scientific explanations into more general explanations. Given a theory T_1 of some physical system (e.g. the digestive system), one extracts the causal topology attributed by T_1 to the system. One then constructs a new theory T_2 that describes this causal topology in organizationally invariant terms. T_1 may mention various non-organizationally-invariant properties (e.g. enzymes in the digestive system), but T_2 ignores such properties. In general, we would not regard T_2 as constituting any explanatory advance. For example, we would not hold that an organizationally invariant description of the digestive system offers any special insight

[11] Chalmers (2011) combines his analysis with a systematic theory of the *physical realization relation* between physical systems and abstract computational models. The theory leaves no room for physical realization of semantically permeated models. In (Rescorla, 2013a), I criticize Chalmers on this score by citing specific examples drawn from CS. In (Rescorla, forthcoming b), I propose an alternative theory of the physical realization relation. My alternative theory applies equally well to semantically indeterminate computational models and semantically permeated computational models.

into digestion. Chalmers insists that, in the special case of cognition, organizationally invariant description yields an explanatory advance (2012, p. 245). To validate this assessment, he must cite features that distinguish cognition from other phenomena (such as digestion) into which organizationally invariant description offers no special insight. Why suspect that causal topology deserves a more prominent role in the science of cognition than the science of digestion?

I raised this objection in (Rescorla, 2012b, pp. 8-10). Chalmers (2012, p. 246) responds that my objection

> ignores the crucial difference between cognition and digestion: the former is an organizational invariant (setting externalism aside for now) while the latter is not. Causal topology does not suffice for digestion, so no adequate explanation of digestion wholly in terms of causal topology can be adequate. But causal topology suffices for cognition, so we can expect an explanation of cognition in terms of causal topology to be adequate. Such an explanation has the potential to cut at the joints that matter where a mechanistic explanation of cognition is concerned.

Chalmers defends his assessment by citing Lewis's (1972) functionalist analysis of the mind, according to which "[p]sychological properties… are effectively defined by their role within an overall causal system: it is the pattern of interaction between different states that is definitive of a system's psychological properties" (Chalmers, 2011, p. 339). Assuming this functionalist analysis, "[s]ystems with the same causal topology will… share their psychological properties (as long as their relation to the environment is appropriate)" (Chalmers, 2011, p. 339). That, says Chalmers, is why organizationally invariant description illuminates cognition but not digestion.

Lewis proposed his functionalist approach as a piece of *conceptual analysis*. He sought to analyze ordinary psychological concepts: *belief*, *desire*, and so on. As Burge (2007, p. 376) and

Putnam (1992) complain, analytic functionalism remains very poorly developed, lacking anything like the specificity one normally expects from a conceptual analysis. Advocates have yet to correlate a single mentalistic concept with a clearly defined causal or functional role. It seems unlikely that any clear, compelling example will ever emerge. Despite what Lewis claims, talk about causal or functional roles does not seem to capture the meaning of ordinary psychological discourse. Thus, Chalmers's argument rests upon an unsupported and implausible conceptual analysis of mentalistic terms.

Content externalism casts further doubt on Chalmers's argument. As Burge (2007) and Putnam (1988) emphasize, externalism raises serious difficulties for many versions of functionalism. Applied to Chalmers's theory, the core difficulty is that mere causal topology does not suffice for a physical system to instantiate desired representational properties. Suitable relations to the external physical or social environment are also required. Through various hedges and caveats, Chalmers acknowledges the threat posed by externalism. In my opinion, however, he does not fully recognize the threat's magnitude.

Consider Chalmers's claim that organizationally invariant description of a cognitive or behavioral function provides "a recipe that could be copied to yield a system that performs the function in question." Assuming that we specify the cognitive or behavioral function in externalist terms, this claim is false. For example, Burge argues that perceptual psychology routinely individuates perceptual states in externalist terms (e.g. through representational relations to specific distal shapes, sizes, or distances). Assuming that Burge is correct, a computational system that replicates the perceptual system's causal topology need not replicate relevant representational properties of perceptual states (e.g. representational relations to specific distal properties). I can concede that "causal topology suffices for cognition," and hence that an organizationally invariant duplicate of the perceptual system instantiates *some* mental activity. But this activity may differ significantly from mental activity *as described in the externalist terms employed by perceptual psychology*. Our duplicate need not perform the representational functions executed by the perceptual system (e.g. estimation of specific distal properties).

Chalmers may reply that causal topology *plus suitable embedding in the environment* suffices for desired mental activity. This reply is plausible. Also plausible is that causal topology *plus suitable embedding in the human body* suffices for digestion. In neither case do we have any solid reason to believe that causal topology *taken on its own* yields a valuable level of description. The human digestive system and a causal-topological duplicate thereof need not be type-identical in any scientifically important respect. Why believe that the human mind and a causal-topological duplicate thereof are type-identical in some scientifically important respect?

Chalmers claims that we can factor out externalist elements of mental activity, leaving behind an explanatorily significant organizationally invariant remainder. He provides no reason to believe this claim except Lewis's conceptual analysis of mentalistic discourse. Once we reject Lewis's analysis, the claim no longer seems compelling. We have no reason to think that organizationally invariant description illuminates the essential natures of representational mental states, any more than it illuminates the essential natures of digestive states. For example, we have isolated no explanatorily significant respect in which states of the human perceptual system are type-identical to corresponding states of a causal-topological duplicate. Both the perceptual system and the digestive system *have* a causal topology. In neither case do we gain any evident insight by describing this causal topology, rather than *specific* causal interactions among *specific non-functional states* that instantiate the topology.

§5.4 Formal syntactic mechanisms?

Proponents of CTM+RTM usually endorse two distinct theses. First, a Turing-style model of the mind decomposes mental processes into iterated elementary operations over symbols, conforming to precise, routine rules. Second, mental operations over the symbols are sensitive only to formal syntactic properties, not to semantic properties. I have disentangled the two theses by offering a view that endorses the first thesis but not the second. A Turing-style model of the mind must isolate elementary operations over mental symbols, and it must delineate rules governing how those operations are applied. But we have found no reason to assign formal syntax a privileged causal or explanatory role when describing the symbols, the operations, or the rules. For example, we have found no reason to think that rule R (as defined over semantically permeated mental symbols) requires supplementation or replacement by rule R* (as defined over semantically indeterminate syntactic types).

Quite plausibly, one must specify a non-representational implementation mechanism for a rule such as R. One must explain how the brain reliably conforms to R. But how does formal mental syntax advance this enterprise? What explanatory value does formal syntactic description contribute to a psychological theory that already contains appropriate semantically permeated computational descriptions *and* appropriate neurophysiological descriptions? Even if formal syntactic description of mental activity is possible, it may be an explanatorily idle abstraction from representational or neural description, just as organizationally invariant description of digestion would be an explanatorily idle abstraction from enzymatic description.

§6. Turing's legacy in the philosophy of mind

Widespread commitment to FSC reflects a cluster of interrelated factors: overly picturesque expositions of mathematical computation theory; distorted analyses of explanatory practice within scientific psychology; ill-motivated Quinean skepticism about intentionality; fallacious arguments that representational content is epiphenomenal; hasty appeals to implementing mechanisms for intentional laws; indiscriminate invocation of explanatory generality; underdeveloped functionalist analyses of mentalistic locutions; and so on. Once we reject these flawed arguments, we see that CTM+FSC is not well-grounded. A semantically permeated version of CTM becomes vastly more attractive than current philosophical discussion recognizes. The semantically permeated approach is not committed to supernatural entities, spooky action-at-a-distance, inner homunculi, or other mysterious posits. It simply type-identifies mental computations through their representational properties, as opposed to any alleged formal syntactic properties.

According to Fodor (2000, pp. 1-22), Turing's main contribution to cognitive science was showing how formal syntactic manipulations by a machine can respect semantic properties. I agree that this was a pivotal contribution *to computer science and Artificial Intelligence*. I think that its importance *to scientific psychology and philosophy of mind* remains undemonstrated. I would instead locate Turing's fundamental contribution to philosophy of mind elsewhere. Turing showed that iterated elementary symbolic manipulations conforming to precise, routine rules can yield astonishingly sophisticated computational activity. He thereby enabled the first mechanical models even remotely suited to accommodate paradigmatic mental processes such as reasoning, decision-making, perception, and so on. These developments conferred unprecedented substance and plausibility upon the ancient doctrine that the mind is a machine. I have shown how we can extricate this fundamental contribution from the orthodox emphasis upon formal syntactic manipulation. By doing so, we may hope to achieve a more satisfying reconciliation of two enduring pictures: *mind as machine* and *mind as representational organ*.

Acknowledgments: I am indebted to audiences at UCLA, Columbia University, and the Southern Society for Philosophy and Psychology for comments when I presented earlier versions of this material. I also thank José Luis Bermúdez, Tyler Burge, Peter Carruthers, Frances Egan, Kevin Falvey, Juliet Floyd, Christopher Gauker, Mark Greenberg, Christopher Peacocke, Gualtiero Piccinini, and Daniel Weiskopf for helpful feedback.

Works Cited

Aydede, M. (1999). On the type/token relation of mental representations. *Acta Analytica*, 2, 23-50.

---. (2005). Computationalism and functionalism: syntactic theory of mind revisited. In G. Irzik and G. Güzeldere (Eds.), *Turkish Studies in the History and Philosophy of Science*. Dordrecht: Springer.

Bays, P., and Wolpert, D. (2007). Computational principles of sensorimotor control that minimize uncertainty and variability. *Journal of Physiology*, 578, 387-396.

Bermúdez, J. L. (1995a). Nonconceptual content: from perceptual experience to subpersonal computational states. *Mind and Language*, 10, 333-369.

---. (1995b). Syntax, semantics, and levels of explanation. *Philosophical Quarterly*, 45, 361-367.

Block, N. (1983). Mental pictures and cognitive science. *Philosophical Review*, 92, 499-541.

Burge, T. (2007). *Foundations of Mind*. Oxford: Oxford University Press.

---. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

Chalmers, D. (2011). A computational foundation for the study of cognition. *The Journal of Cognitive Science*, 12, 323-357.

---. (2012). The varieties of computation: a reply. *The Journal of Cognitive Science*, 13, 213-248.

Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78, 67-90.

Dreyfus, H. (1992). *What Computers Still Can't Do*. Cambridge: MIT Press.

Egan, F. (1992). Individualism, computation, and perceptual content. *Mind*, 101, 443-459.

---. (2003). Naturalistic inquiry: where does mental representation fit in?. In L. Antony and N. Hornstein (Eds.), *Chomsky and His Critics*. Malden: Blackwell.

Feldman, J. (Forthcoming). Bayesian models of perceptual organization. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization*. Oxford: Oxford University Press.

Field, H. (2001). *Truth and the Absence of Fact*. Oxford: Clarendon Press.

Fodor, J. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

---. (1981). *Representations*. Cambridge: MIT Press.

---. (1987). *Psychosemantics*. Cambridge: MIT Press.

---. (1990). *A Theory of Content and Other Essays*. Cambridge: MIT Press.

---. (1991). Replies. In B. Loewer and G. Rey (Eds.), *Meaning in Mind*. Cambridge: Blackwell.

---. (1994). *The Elm and the Expert*. Cambridge: MIT Press.

---. (1998). *Concepts*. Oxford: Clarendon Press.

---. (2000). *The Mind Doesn't Work That Way*. Cambridge: MIT Press.

---. (2008). *LOT2*. Oxford: Clarendon Press.

Gallistel, R. and King, A. (2009). *Memory and the Computational Brain*. Malden: Wiley-Blackwell.

Harnad, S. (1994). Computation is just interpretable symbol manipulation; cognition isn't. *Minds and Machines*, 4, 379-90.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press.

Kaplan, D. (1989). "Demonstratives." In J. Almog, J. Perry, and H. Wettstein (Eds.), *Themes from Kaplan*. Oxford: Oxford University Press.

Knill, D. and Richards, W. (Eds.). (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249-58.

Ockham, W. (1957). *Summa Logicae*, in his *Philosophical Writings, A Selection*, Ed. and Trans. P. Boehner. London: Nelson.

Palmer, S. (1999). *Vision Science*. Cambridge: MIT Press.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge: MIT Press.

---. (1994). Content, computation, and externalism. *Mind and Language*, 9, 303-335.

Pouget, A., Beck, J., Ma, W. J., and Latham, P. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16, 1170-1178.

Prinz, J. (2011). Has Mentalese earned its keep? On Jerry Fodor's *LOT 2*. *Mind*, 120, 485-501.

Putnam, H. (1967). Psychophysical predicates. In W. Capitan and D. Merrill (Eds.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.

---. (1975). *Mind, Language, and Reality: Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press.

---. (1988). *Representation and Reality*. Cambridge: MIT Press.

---. (1992). Why functionalism failed. In J. Earman (Ed.), *Inference, Explanation and Other Philosophical Frustrations*. Berkeley: University of California Press.

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge: MIT Press.

Quine, W. V. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Rescorla, M. (2012a). Are computational transitions sensitive to semantics?. *Australasian Journal of Philosophy*, 90, 703-721.

---. (2012b). How to integrate representation into computational modeling, and why we should. *Journal of Cognitive Science*, 13, 1-38.

---. (2013a). Against structuralist theories of computational implementation. *British Journal for the Philosophy of Science*, 64, 681-707.

---. (2013b). Millikan on honeybee navigation and communication. In D. Ryder, J. Kingsbury, and K. Williford (Eds.), *Millikan and Her Critics*. Malden: Wiley-Blackwell.

---. (2014). The causal relevance of content to computation. *Philosophy and Phenomenological Research*, 88, 173-208.

---. (Forthcoming a). Bayesian perceptual psychology. In M. Matthen (Ed.), *Oxford Handbook of* *the Philosophy of Perception*. Oxford: Oxford University Press.

---. (Forthcoming b). A theory of computational implementation. *Synthese*.

Salmon, N. (1986). *Frege's Puzzle*. Cambridge: MIT Press.

Schneider, S. (2011). *The Language of Thought: A New Philosophical Direction*. Cambridge: MIT Press.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417-424.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-74.

Stalnaker, R. (1984). *Inquiry*. Cambridge: MIT Press.

Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge: MIT Press.

Taylor, R. G. (1998). *Models of Computation and Formal Languages*. Oxford: Oxford University Press.

Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42, 230-265.

---. (1947). Lecture to the London Mathematical Society on 20 February 1947. In D. Ince (Ed.), *Mechanical Intelligence*. Amsterdam: North-Holland.

Tye, M. and Sainsbury, M. (2012). *Seven Puzzles of Thought*. Oxford: Oxford University Press.

van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, 7, 345-381.

Vilares, I. and Körding, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, 1224, 22-39.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.