Minds & Machines
DOI 10.1007/s11023-010-9213-z

Yes, She Was!

Reply to Ford's "Helen Keller Was Never in a Chinese Room"

William J. Rapaport

Received: 2 August 2010 / Accepted: 13 September 2010

Abstract Ford's "Helen Keller Was Never in a Chinese Room" claims that my argument in "How Helen Keller Used Syntactic Semantics to Escape from a Chinese Room" fails because Searle and I use the terms 'syntax' and 'semantics' differently, hence are at cross purposes. Ford has misunderstood me; this reply clarifies my theory.

Keywords Chinese room argument · Syntactic semantics · Searle

1

Jason Michael Ford's "Helen Keller Was Never in a Chinese Room" (2010) claims that my argument in "How Helen Keller Used Syntactic Semantics to Escape from a Chinese Room" (Rapaport 2006) fails because Searle and I use the terms 'syntax'

W. J. Rapaport (✉)
Department of Computer Science and Engineering, University at Buffalo,
The State University of New York, Buffalo, NY 14260-2000, USA
e-mail: rapaport@buffalo.edu
URL: http://www.cse.buffalo.edu/~rapaport/

W. J. Rapaport
Department of Philosophy, University at Buffalo, The State University of New York,
Buffalo, NY 14260-2000, USA

W. J. Rapaport
Department of Linguistics, University at Buffalo, The State University of New York,
Buffalo, NY 14260-2000, USA

W. J. Rapaport
Center for Cognitive Science, University at Buffalo, The State University of New York,
Buffalo, NY 14260-2000, USA

Published online: 21 October 2010

and ‘semantics’ differently, hence are at cross purposes. I think Ford has misunderstood me, so I am grateful for this opportunity to clarify my theory.

The theory of syntactic semantics (Rapaport 1988) underlies computationalism: the claim that cognition is computable, i.e., that there is an algorithm (or a family of algorithms) that compute cognitive functions (Rapaport 1998). The theory has three parts: First, cognitive agents have direct access only to internal representatives of external objects. As Ray Jackendoff (2002, §10.4) says, a cognitive agent understands the world by “pushing the world into the mind”. Therefore, both words and their meanings (including external objects serving as their referents) are represented internally in a *single* language of thought (LOT). For humans, this LOT is a biological neural network; for computers, it might be some kind of knowledge-representation and reasoning system (such as SNePS; see Shapiro and Rapaport 1987).[1]

Second, it follows that words, their meanings, and semantic relations between them are all syntactic, where syntax is the study of relations *among* members of a *single* set (of signs, or marks, or neurons, etc), and semantics is the study of relations *between two* sets (of signs, marks, neurons, etc., on the one hand, and their meanings, on the other) (cf. Morris 1938). “Pushing” meanings into the same set as symbols for them allows semantics to be done syntactically: It turns semantic relations between two sets (a set of internal marks and a set of (external) meanings) into syntactic relations among the marks of a *single* (internal) LOT. For example, truth tables and formal semantics are both syntactic enterprises, as are the relations between neuron firings representing signs and neuron firings representing external meanings. Consequently, symbol-manipulating *computers* can do semantics by doing syntax.

Finally, understanding is recursive: We understand a *syntactic* domain (call it ‘SYN_1’) *indirectly* by *interpreting it* in terms of a *semantic* domain (call it ‘SEM_1’). But SEM_1 must be *antecedently understood* by considering it as a *syntactic* domain (rename it ‘SYN_2’) interpreted in terms of *yet another semantic* domain, which *also* must be antecedently understood. And so on. But, in order not to make it go on *ad infinitum*, there must be a base case: a domain that is understood directly, i.e., in terms of itself (i.e., not “antecedently”). Such direct understanding is syntactic understanding (Rapaport 1986b) (And perhaps it is holistic understanding; cf. Rapaport 2002).

Thus, the theory of syntactic semantics asserts that syntax suffices for semantic cognition, that cognition is therefore computable, and that computers are hence capable of thinking.

2

Ford claims that my “meld[ing of] the semantics into the syntax, placing all the semantic and syntactical units and their relations into a single set ... effectively

[1] The “marks” or “units” of this LOT (e.g., nodes of a semantic network; terms and predicates of a language; or their biological analogues, etc). need not all be alike, either in “shape” or function. Natural languages, e.g., use a wide variety of letters, numerals, etc.; neurons include afferent, internal, and efferent ones (and the former do much of the internalization or “pushing”). Thanks to Albert Goldfain (personal communication) for emphasizing this last point.

giv[es] up the claim that the syntactic relationships obtain independent of the meanings of the signs, and adding rules that make use of the meanings of the signs" (§2, "The Structure of Rapaport's Argument").

Not quite. In my theory, some "units" (as Ford calls them; I usually refer to them as "marks") are linguistic; others are (let's call them) "conceptual"; the latter are the (internal, hence idiosyncratic or first-personal) meanings of the former. Each set of units has syntactic relations *among* its members. There are also semantic relations *between* the two sets. When the two sets are "melded" (i.e., unioned), we now have three distinguishable classes of relations: the previously syntactic relations *among* the linguistic units, the previously syntactic relations among the conceptual units, and the previously semantic relations *between* the two subsets of units. The last of these three are now also syntactic relations, because they are (now) relations *among* the members of the set that resulted from the melding.

This is all somewhat abstract, so let us consider a more concrete case: Some of my neuron firings are linguistic (encoding linguistic items);[2] others are conceptual (encoding meanings of those linguistic items). Patterns of firings among the former are "syntactic" in my terminology; patterns of firings among the latter are similarly syntactic. But they are all neuron firings in one "melded" biological neural network (i.e., my brain). Patterns of firings between the "linguistic" neurons and the "conceptual" neurons are "semantic" (in my terminology) insofar as they encode relations *between* "linguistic" and "conceptual" neurons; but they are "syntactic" (in my terminology) insofar as they encode relations *among* neurons *tout court*.

What Ford calls "The syntactic relationships [that previously] obtain[ed] independent of the meanings of the signs" have *not* changed, *nor* have any "rules" been "added" "that make use of the meanings of the signs" ("The Structure of Rapaport's Argument"). Nothing has changed (except our perspective), nor have I given anything up.

3

Ford claims that Searle's definitions of 'syntax' and 'semantics' (subscripted with 'S') differ from mine (subscripted with 'R') ("The Structure of Rapaport's Argument"). Mine are, indeed, more general or abstract, but I believe that Searle's fall under them. All of the quotes from Searle that Ford cites are consistent with my definitions. Nor are Searle's definitions "orthogonal" to mine ("The Structure of Rapaport's Argument").

I agree with Ford that "Syntactic$_R$ relations ... could be either syntactic$_S$ or semantic$_S$", but this does not support their being orthogonal. I agree with Ford's statement, because I see Searle's notions as falling under mine and because I see semantic relations (with or without subscripts) as being interpretable as syntactic (with or without subscripts). What enables syntactic relations (relations *among* the members of a single set) to play the role of semantic relations (relations *between* the members of two sets) is the perspective from which the two sets are seen as one.

[2] And, of course, there are many others; see previous note.

Ford cites my example of the syntactic nature of the semantic synonymy of ‘lawyer’ and ‘attorney’. Let us assume for the sake of the argument that these two words are indeed synonymous and that there is such a relation as synonymy; debating those issues is irrelevant to the point. There are two ways to explain the synonymy: (1) We could say that there is an entity (an external referent, an internal concept, a set of neuron firings in my brain—whatever) that is the common “meaning” of both words. Represent this entity by ‘LAW-PRACTITIONER’. The fact that there are semantic interpretation relations between ‘lawyer’ and LAW-PRACTITIONER and between ‘attorney’ and LAW-PRACTITIONER is what makes the two words synonyms. This is a semantic explanation.

(2) Alternatively, we could ignore meanings, referents, concepts, etc., and simply pay attention to the two words. We might note that they are mutually substitutable in all (extensional) contexts, or we might note that they are equally distributed in the language. In other words, we might note that they are synonymous without appealing to meanings. That is, we might consider their synonymy as a syntactic matter. (We might also appeal to a common meaning in order to explain or succinctly describe that syntactic synonymy, but that would be a semantic explanation or description.)

However, there is more to the matter when Ford says that “semantic$_S$ relations ... could be either semantic$_R$ or syntactic$_R$” (“The Structure of Rapaport’s Argument”). This is true insofar as Searle’s notions are special cases of mine (again, special cases are not orthogonal). But there is a notion of semantics, which may be the notion that (Ford says that) Searle has in mind, that *is* distinct from mine, namely, “third-person” or “objective” semantics. What does ‘gold’ mean? Such a question has many answers, even if we only consider referential relationships. ‘Gold’ could refer to the chemical element Au or to any real, physical object composed of Au. Or it could refer to whatever it is that Putnamian experts claim that it refers to (Putnam 1975). Or I could use it to refer to whatever it is that I think is “gold”. Or it could have a “first-person” meaning for me that differs, in minor or major ways, from your “first-person” meaning (this would be a meaning more along the lines of a Fregean sense—rather than a Fregean referent—or a Meinongian object (Rapaport 1981, 1985/1986; Shapiro and Rapaport 1987, 1991), or a particular set of neuron firings in my brain (my concept associated with ‘gold’ as opposed to your concept). And so on.

Note that, not only do we (you, me, Searle-in-the-room, computational cognitive computers, et al.) not have access to external meanings, neither do *I* have access to *your* meanings; they are also external to me (see Rapaport 2003a). So, when Ford later says, “If the description of the CRA *in Rapaport’s theoretical framework* forces us to conclude that Searle-in-the-room will come **to understand the meaning** of the initially uninterpreted Chinese characters by virtue of hand-working a program, then Rapaport will have accomplished his task” (“Translation and Evaluation of the CRA”; my boldface), that means (for me) to associate some (other) concepts (or neuron firings, etc.) with those that arise from the Chinese characters. It does *not* mean to come to know (or believe) what *others* mean by them.

I have argued (in Rapaport 2006 and elsewhere) that, whether or not there is such an “external” or “third-person” meaning, we have no access to it. Are such meanings or semantic$_S$ relations syntactic$_R$ or semantic$_R$? I could say that they are neither (if that means that these notions are orthogonal, so be it). But I could also say that, *from God's (or Nature's)*[3] *perspective*, they are both! The relations between words and such external meanings—knowable only to God or existing in Nature but not (fully) knowable by us—are semantic because they are relations between two sets. But they would be syntactic (*sub specie aeternitatis*)[4] because they are relations *among* a single set: the denizens of Nature.

4

So, is it really the case that "Searle's bumper-sticker slogan, 'Syntax$_S$ is not sufficient for semantics$_S$,' is a very different claim from Rapaport's similar-sounding, 'Syntax$_R$ *is* sufficient for semantics$_R$.' " ("The Structure of Rapaport's Argument")? If 'semantics$_S$' refers to external, third-person semantics, then, not only is syntax$_S$ not sufficient for it, but neither is anything else: *Both computers and humans are in the same boat; neither of us can access such external meanings*.

But, *because* we're both in the same boat, our syntax *does*—indeed, *must*—suffice for the only kind of semantics that we are capable of having: the internal, first-person kind. Yes, our internal concepts (neuron firings, etc.) are, presumably, causally generated by objects and events in the external world, and our beliefs are true insofar as our internal "semantic networks" match the "ontology" of the external world. But we have to take that on faith (or accept pragmatically that all the evidence suggests that they match).

5

Ford suggests that I could "make the following argument: 'If we accept the classical definitions of semantics$_R$ and syntax$_R$, we can show that the CRA fails. Within that conceptual framework, we can generate or discover *semantic content* from the manipulation of *uninterpreted symbols*" ("Helen Keller"). This makes it sound as if I would want to argue that Searle-in-the-room can come to understand, say, everything about hamburgers simply on the basis of reading about them. Not quite. But I would say that Searle-in-the-room can come to understand everything about hamburgers on the basis of syntactic linguistic input (words, sentences) and syntactic non-linguistic input (visual, tactile, olfactory, etc). Why am I calling all these inputs 'syntactic'? Strictly speaking, I suppose, the *inputs*, coming as they do from the external world, are not necessarily syntactic. But they are transduced into marks (for the computer) or neuron firings (for us); and those *are* all syntactic. They are "uninterpreted symbols", though some get interpreted in terms of others of them.

In other words, we must assume the situation of the Robot Reply: Among the members of the set E of experiences that are used to interpret the set C of Chinese characters (see point 3 in §3, "Translation and Evaluation of the CRA") and among

[3] *Deus sive natura.*

[4] Or "viewed from nowhere", to use Nagel's (1986) phrase instead of Spinoza's; cf. Goldstein (2006): 67–68, 276.

the input to Searle-in-the-room are items that can serve as the meanings or interpretations of the members of C. Some of these (e.g., the Chinese character for 'tree', say, and an internal, mental image of a tree) will be input to Searle-in-the-room "simultaneously", in the manner of Helen Keller's experiences of both water and 'w-a-t-e-r' at the well house. But these Robot Reply assumptions are all syntactic.

Ford says, "Here is the key feature of the scenario that Rapaport seems to have neglected: the CRA affords no opportunity for Searle-in-the-room to associate any of the Chinese symbols with any of his experiences or memories of the things that the symbols stand for" ("Helen Keller")—whereas Helen Keller had that opportunity. But my point is precisely that both Helen Keller and Searle-in-the-room (better: Searle-in-the-room + the instruction book, as in the Systems Reply) *did* have this opportunity: Searle-in-the-room should not be limited to linguistic input; the input must include visual, etc., imagery, as in the Robot Reply.

6

And now we come to what *I* see as the "key feature". Ford says, "Unless we can create a computer with sensors that provide the computer *with meaningful experience* (experience that is meaningful *antecedently* to any formal operations performed in accord with the computer's programming) all of Rapaport's subsequent analysis of Keller's case in terms of SNePS is completely irrelevant" ("Helen Keller"). But what could "meaningful experience" mean? As Ford notes, Searle takes "meaningful experience as a biological product" ("Helen Keller"): Intentionality (including, more generally, understanding and cognition) must be biological because only biological systems have the requisite causal properties to produce it. What are those properties? Those that are "causally capable of producing perception, action, understanding, learning, *and other intentional phenomena*" (Searle 1980: 422; my emphasis). But this is either circular (Rapaport 1986a) or empty as an explanation.

Any "units" or marks or neurons that are "antecedently meaningful" can only be so by having been interpreted in terms of other antecedently understood (or antecedently meaningful) units, marks, or neuron firings. And, as I argue, ultimately such items must be understood syntactically (on pain of infinite regress) (Rapaport 1986b).

Ford accuses me of "assum[ing] what [I] need to prove—that symbols produced by sensors interacting with the environment somehow automatically acquire semantic content or meaning" ("Helen Keller"). I don't think I assume it; I think I argue for it: They acquire meaning by association with other symbols. Searle and I start from opposite places, but his starting point (that it's all biological) is empty; I think mine (that it's all internal) is defensible (using Kantian or modern cognitive versions of argument-from-illusion-style arguments; see Rapaport 2000, 2005a).

7

The internal symbols must include "products" of "environmental sensors" ("Helen Keller"). Ford compares a "Keller-like version" of the Robot Reply to what he calls (but I do not recognize as) “Rapaport’s version”. In the former, input comes from “a video screen, speakers, etc.” *together* with Chinese characters (“Helen Keller”). In the latter—his version of “my” version—input comes *only* from the Chinese characters, with “signals to the robot’s effectors [sic][5] ... sent via the Chinese symbols” (“Helen Keller”). Ford continues: “Then, like magic, Searle-driving-the-robot begins to understand Chinese.”

But nowhere do I say that the input to the sensors must be transduced into Chinese characters (i.e., that Chinese characters mediate between the sensory input and motor output). Rather, Ford’s Keller-like version and “Rapaport’s” version are alike. That was precisely my point in Rapaport (2006). The real Rapaport version is no more (but no less) magical than Ford’s Keller-like version.

A more interesting question is whether the original Robot Reply says that the input to the sensors is transduced into *Chinese characters* (i.e., that the external world is represented by Chinese characters, which suggests that it is represented by Chinese *sentences*). Here is the Robot Reply in Searle’s words (“quoting” an imaginary, Yale philosopher, hence my use of double quotes):

> “Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating, drinking—anything you like. The robot would, for example have a television camera attached to it that enabled it to ‘see,’ it would have arms and legs that enabled it to ‘act,’ and all of this would be controlled by its computer ‘brain.’ Such a robot would, unlike Schank’s computer, have genuine understanding and other mental states.” (Searle 1980: 420).

My understanding of this reply is that we are to endow Searle-in-the-room (or Searle-driving-the-robot, as Ford refers to him) with the output of various sensors. What kind of output is this? That depends on the sensor, of course, but it would typically be information that encodes visual, tactile, etc., sensations. How is this encoded? Again, that depends, but it would almost certainly *not* be encoded in the same way that natural-language input would be—i.e., in letters, words, sentences—despite the fact that Searle has some strange ideas about how computational vision systems work:

> A standard computational model of vision will take in information about the visual array on my retina and eventually *print out the sentence*, “There is a car coming toward me.” (Searle 1990: 34–35; my emphasis).

Searle echoes this strangeness in his original response to the Robot Reply:

> ... the answer to the robot reply is that the addition of such “perceptual” ... capacities adds nothing by way of understanding, in particular, or intentionality, in general, to Schank’s original program. To see this, notice that the same thought experiment applies to the robot case. Suppose that instead of the

[5] Does Ford mean external signals sent *from* the external world *to* the robot’s *sensors*? Or does he mean internal signals sent *from* the sensors *to* the effectors?

> computer inside the robot, you put me inside the room and, as in the original Chinese case, you give me *more Chinese symbols* with more instructions in English for matching Chinese symbols to Chinese symbols and feeding back Chinese symbols to the outside. Suppose, unknown to me, *some of the Chinese symbols that come to me come from a television camera attached to the robot* ... (Searle 1980: 420, my emphasis).

As I observe elsewhere (Rapaport 2007: 8), this

> is astounding. ... [I]t is hardly "standard" to have a vision system yield a *sentence* as an output. It might, of course ("Oh, what a pretty red flower."), but, in the case of a car coming at the system, an aversive maneuver would seem to be called for, not a matter-of-fact description.

And such a maneuver would likely be based on some kind of ever-enlarging, internal (symbolic) representation of the oncoming, external car (see Parisien and Thagard 2008 for discussion).

But let's be sympathetic to what I think Searle's real point is: Ultimately, all the input from the external world, be it linguistic or visual or whatever, is encoded in the computer by 0s and 1s and is encoded biologically in us as neuron firings. And, as Searle says, the point is that "all I am doing is manipulating formal symbols" (Searle 1980: 420). But note that, even if the formal symbols are all the same kind (0s and 1s for the computer, neuron firings for us), presumably, they are encoded differently: Some represent language, perhaps via ASCII coding; others represent, say, cars, perhaps via an array of RBG pixels. The key point is that the associations between the former and the latter are *semantic associations, implemented syntactically*.

8

Ford's description of various versions of the Systems Reply is even odder. The Systems Reply says that it is not Searle-in-the-room who understands Chinese but that it is the system consisting of Searle-in-the-room together with the instruction book. (That is, it is neither a static CPU nor a static computer program that understands, but a dynamic *process*: a CPU-executing-the-program; cf. Rapaport 1995).

Ford's Keller-like version ("Helen Keller") begins with "Searle-outside-the-room" (i.e., the system) being "given some Chinese symbols, and somebody puts his other hand in some water". In "gavagai" fashion, Searle-outside-the-room "can narrow down the options: [the symbols might mean] Water[,] Wet[,] Wet and cold[,] Wet hand[,] I now immerse part of you in water ...." (On narrowing down such options, see also Rapaport and Kibby 2007).

Ford's "Rapaport's Version", however, goes quite differently, for reasons that I simply cannot fathom: "Searle-outside-the-room is given *some Chinese symbols*, and somebody finger-spells w-a-t-e-r into his other hand" ("Helen Keller" my emphasis). But why? That doesn't parallel the case at all. I never claimed that the system associates finger-spellings with Chinese characters. My analogy was indeed what Ford calls the ‘Keller-like version’: The system associates symbols (be they finger-spellings or Chinese characters) with real water. To say, as Ford does, that “the robot’s sensors send the information about the environment in the form of Chinese text” (“Helen Keller”) is as misleading as Searle’s own claim (cited above) that computational vision systems output *sentences* of the form “There’s a car coming”.

Later, Ford elaborates: “Searle-in-the-room doesn’t get any pictures or recognizable diagrams[6]—he just gets symbols” (“Helen Keller”). But pictures and diagrams are also symbols, requiring interpretation: Australian aboriginal paintings are typically aerial views, so a Western painting of a horse might be interpreted naively by an Australian aboriginal as a horse lying on its side (David Wilkins, personal communication); cf., also, “puzzle” pictures, such as the one of a black-and-white spotted Dalmatian dog against a black-and-white background of rocks in snow,[7] require an interpretive theory to be recognized.

In my view, “interpretation” comes down to incorporation into a previously existing semantic network (implemented in us by our biological neural network). Incorporation takes the form of associations (links, mappings) between the (transduced) input and the stored (already existing) nodes in the network. Some of those associations will be semantic, relating (linguistic) symbols to their internal, first-person meanings; all of the associations, however, are syntactic—relating internal, first-person marks to other internal, first-person marks.

Ford accuses me of having “smuggled in some meaningful experience, where the real CRA starts with nothing more than uninterpreted symbols” (“Helen Keller”). But uninterpreted symbols is precisely where “meaningful experience” must begin. Actually, there are two possibilities: One is that we begin with antecedently “meaningful” nodes in our mental semantic network, i.e., innate ideas. In that case, all incoming uninterpreted symbols get interpreted by incorporation into this “prior knowledge”.[8] The other is that we begin with a *tabula rasa*. In that case, the first incoming uninterpreted symbols become the prior knowledge in terms of which all subsequent incoming uninterpreted symbols get interpreted by incorporation (Rapaport 1995). And how is the initial “prior knowledge” understood? Syntactically! (Rapaport 1986b).

Given this correction, Ford and I actually agree: “If Searle-in-the-room ... is given *meaningful experiences* ... and the capacity to associate features of those experiences with the Chinese symbols simultaneously presented, then **of course Searle-in-the-room ... will have the potential to learn to understand Chinese**” (“Helen Keller”; my boldface). Again, I’m not sure what “meaningful experiences” are, or how they get their meaning, but if all that Ford has in mind are internal “conceptual” symbols, then we agree.

[6] Who “recognizes” the diagrams?

[7] [http://www.michaelbach.de/ot/cog_dalmatian/].

[8] In general, we understand, or give meaning to, incoming information by embedding it in the context of our prior beliefs. For further elaboration on this idea, as it relates to contextual vocabulary acquisition, see Rapaport 2003b. And for a commentary from a legal standpoint, see *New York Times* 2010.

9

Ford asks, but does not stop to consider, "If everything that we could be concerned with (when investigating the mind and the brain) is syntactic$_R$, why strive to create a semantic$_R$ system?" ("Translation and Evaluation of the CRA"). The answer is simple: Doing so helps us to understand how we understand (Rapaport 1995).

10 Appendix

Ford cites Jahren's (1990) objections to my view. Let me take this opportunity to reply to Jahren, whose critique of my (1988) theory of syntactic understanding and its application to the CRA shows how easy it is in discussing these issues to talk just slightly past one another.

10.1

What, for example, is a natural language, and what does it mean to understand one? For Jahren, a natural-language is "a series of signs used by a system", and "the sine qua non of natural-language understanding ... [is] an ability to take those signs *to stand for something else ... in the world*" (Jahren 1990: 310, my emphasis). But if a natural language is just "a series of signs", it follows that to understand it is to understand the series of signs as used by the system—which is a *syntactic* process. Now, as I urged in "Syntactic Semantics" (Rapaport 1988), to understand, in general, is to map symbols to concepts.[9] Thus, for *me* to understand *you* is for me to map *your* symbols to *my* concepts, which *is*, to use Jahren's phrase, taking "those signs to stand for something else"—but *not* "something in the world" (except in the uninteresting sense that my concepts are things in the world). This is *also* a syntactic process: Insofar as I internalize your symbols and *then* map my internalized representations (or counterparts) of your symbols to *my* concepts, I am doing nothing but internal symbol manipulation (syntax), even though I *am* taking your "signs to stand for something else", namely, my concepts.

How do I understand *my* concepts? Do I take *my* concepts to stand for something else outside me? Yes—I so *take* them, although I only have *indirect* access to the "something else" outside me. The only way I can take *your* symbols "to stand for something in the world" would, pre-theoretically, have to be either directly or else indirectly via *my* symbols (concepts). But *all* of it is indirect, since I can at best take your symbols to stand for the same thing I take mine to stand for, and, in both cases, that's just more symbols (cf. Rapaport 2000).

10.2

Jahren takes me to task for using 'mentality' in a "suprapsychological" sense (citing Flanagan 1984) instead of "in a human sense" (Jahren 1990: 314ff). But what sense

[9] This is the *recursive case* of understanding. In the *base case*, to understand is to be able to manipulate the symbols syntactically. See Rapaport 2006, Thesis 3.

is that? Is it determined by human *behavior* (as in, say, the Turing Test)? If so, then Jahren and I *are* talking about the same thing, since human mental *behavior* might be produced by different processes. Is it determined by the way the human brain does mental processing? But that is too strong for my *computational philosophical* tastes: I am concerned with how mentality, thinking, cognition, understanding—call it what you will—is possible, period. I am not concerned with how *human* mentality, in particular, works; I take that to be the domain of (computational) cognitive *psychology*.[10] However, I *don't* intend (at least, I don't *think* I intend) the *very* weak claim that as long as a computer can simulate human behavior by *any* means, that would be mentality. I *do* want to rule out table look-up or the (superhuman) ability to solve any mathematical problem, without error, in microseconds. The former is too finite (it can't account for productivity); the latter is too perfect (in fact, if viewed as an infinite, God-like ability to know and do everything instantaneously, it, too, is a kind of table look-up that fails to account for productivity; cf. Rapaport 2005b).

Now, having excluded those two extremes, there is still a lot of variety in the middle. So I'll agree with Jahren that, the extreme cases excepted, "a computational system is minded to the extent that the information processing it performs is functionally [that is, input-output, or behaviorally] equivalent to the information processing in a mind" (Jahren 1990: 315)—presumably, a *human* mind. However, Jahren says that two mappings are input-output equivalent "because these mappings themselves can be transformed into one another" (Jahren 1990: 315). This seems to me too restrictive, not to say vague (what does it mean to transform one *mapping* into another?). Jahren gives as an example "solving a matrix equation [which] is said to be equivalent to solving a system of linear equations" (Jahren 1990: 315). But surely two algorithms with the same input-output behavior would be functionally equivalent even if they were not thus transformable. Consider, for instance, two very different algorithms for computing greatest common divisors. They would be functionally equivalent even if there were no way to map parts of one to parts of the other in any way that preserved functional equivalence of the parts.

10.3

Jahren alludes to the symbol-grounding problem: "The semantics$_R$ [that is, the semantics in Rapaport's sense] of a term is given by its position within the entire network" (Jahren 1990: 318). The proper response to this is: 'Yes and no'. *Yes*, in the sense that ultimately all is syntactic, hence holistic, as Jahren observes (cf. Rapaport 2002, 2003a). But *no* in the sense that this misleadingly suggests that nothing in the network represents the external world. For instance, Jahren gives an example of 'red' linked as subclass to 'color' and as property to 'apple', etc. But this omits another, crucial—albeit still internal—link: to a node representing the

[10] On the distinction between computational philosophy and computational psychology, see Shapiro (1992), Rapaport (2003a).

sensation of redness.[11] *Some* parts of the network represent external objects, so an internal analogue of "reference" is possible.

Now, to be fair, Jahren is not unsympathetic to this view:

> ... Rapaport's conception of natural-language understanding does shed some light on how humans work with natural language. For example, my own criterion states that when I use the term 'alligator', I should know that it (qua sign) stands for something else, but let us examine the character of my knowledge. The word 'alligator' might be connected in my mind to visual images of alligators, like the ones I saw sunning themselves at the Denver Zoo some years ago. But imagine a case where I have no idea what an alligator is but have been instructed to take a message about an alligator from one friend to another. Now the types of representations to which the word 'alligator' is connected are vastly different in both cases. In the first, I understand 'alligator' to mean the green, toothy beast that was before me; in the second, I understand it to be only something my friends were talking about. But I would submit that the character of the connection is the same: it is only that in the former case there are richer representations of alligators (qua object) for me to connect to the sign 'alligator'. ... The question ... is whether the computer takes the information it stores in the ... [internal semantic network] to stand for something else. (Jahren 1990: 318–319; cf. Rapaport 1988, n. 16).

Well, the computer does and it doesn't "take the information it stores ... to stand for something else". It *doesn't*, in the sense that it can't directly access that something else (any more—or less—than *we* can). It *does*, in the sense that it assumes that there is an external world. But note that if it represents the external world internally, it's doing so via more nodes! There's no escaping our internal, first-person experience of the world. As Kant might have put it, there's no escape from phenomena, no direct access to noumena.

10.4

I have been avoiding the issue of consciousness and what it "feels like" to understand or to think (though I have something to say about *part* of that problem in Rapaport 2005a). But let me make one observation here, in response to Jahren's description of how we can experience what it is like to be the machine: "in accordance with the Thesis of Functional Equivalence one can be the machine in the only theoretically relevant sense if one performs the same information processing that the machine does" (Jahren 1990: 321). That is, to see if a machine that passes the Turing Test is conscious, we would need to *be* the machine, and, to do that, all we have to do is behave as it does. But just "being" the machine (or the "other mind") isn't sufficient—one would also have to simultaneously be oneself, too, in order to *compare* the two experiences. This seems to be at the core of Searle's Chinese-Room Argument—he *tries* to be himself *and* the computer simultaneously

[11] For a computational implementation of this consistent with the SNePS theory presented in Rapaport (2006) see Shapiro and Bona (2010).

(cf. Cole 1991; Rapaport 1990; Copeland 1993). But he can't use his *own* experiences (or lack of them) to experience his own-qua-computer experiences (or lack of them). That's like *my* sticking a pin into *you* and, failing to feel pain, claiming that *you* don't, either. It is *also* like my *making believe* I'm you, sticking a pin into *me-qua-you*, feeling pain, and concluding that so do *you*. Either one "is" both cognitive agents at the same time, in which case there is no way to distinguish one from the other—the experiences of the one *are* the experiences of the other—or else one is somehow able to separate the two, in which case there is no way for either to know what it is like to be the other. Note, finally, that what holds for me (or Searle) imitating a computer holds for a computer as well: Assume that *we are* conscious, and let a computer simulate us; could the *computer* determine whether *our* consciousness matched *its*? I doubt it.

10.5

Let's return to the syntactic understanding of Searle-in-the-room. Jahren says that Searle-in-the-room does not understand Chinese "because ... [he] cannot distinguish between categories. If everything is in Chinese, how is he to know when something is a proper name, when it is a property, or when it is a class or subclass?" (Jahren 1990: 322). I take it that Jahren is concerned with how Searle-in-the-room can decide of a given input expression whether it is a *name*, or a *noun for* a property, or a *noun for* a class or subclass. In terms of a computational cognitive agent (such as Cassie, discussed in Rapaport 2006), this is the question of how she would "know" that 'Lucy' in 'Lucy is rich' is a proper name (in SNePS terms, how she would "decide" whether to build an `object-propername` case frame or some other case frame) or of how she "knows" that 'rich' expresses a property rather than a class (how she "decides" whether to build an `object-property` case frame rather than a `member-class` case frame; see Rapaport 2006 for details on these SNePS semantic network notions).

In one sense, the answer is straightforward: In Cassie's case, an augmented-transition-network parsing grammar "tells" her. And how does the augmented transition network "know"? Well, of course, we programmed it to know. But in a more realistic case, Cassie would learn her grammar, with some "innate" help, just as we would. In that case, what the arc labels are is absolutely irrelevant. For us programmers, it's convenient to label them with terms that *we* understand. But Cassie has no access to those labels. So, in another sense, she does *not* know, *de dicto*, whether a term is a proper name or expresses a property rather than a class. Only if there were a *node* labeled 'proper name' and appropriately linked to other nodes in such a way that a dictionary definition of 'proper name' could be inferred would Cassie know *de dicto* the linguistic category of a term. Would she know that something was a proper name in *our* sense of 'proper name'? Only if she had a conversation with us and was able to conclude something like, "Oh—what *you* call a 'proper name', I call a___", where the blank is filled in with the appropriate node label.

This is simply the point that native speakers of a language don't have to explicitly understand its grammar in order to understand the language. I once asked (in French) a native French-speaking clerk in a store in France whether a certain noun was masculine or feminine, so that I would know whether to use 'le' or 'la'; the clerk had no idea what I was talking about, but she did volunteer that one said '*le* portefeuille', not '*la* portefeuille'.

Jahren "argue[s] that Searle-in-the-room cannot interpret any of the Chinese terms in the way he understands English terms" (Jahren 1990: 323). But insofar as Searle-in-the-room *is* understanding Chinese, he is *not* understanding English. Neither does Cassie, strictly speaking, understand SNePS networks; rather, she understands natural language, and she uses SNePS networks to do so. Just as a native speaker of English would explicitly understand English grammar only if she had studied it formally, so would Cassie only explicitly understand SNePS networks if she were a SNePS programmer (or a computational cognitive scientist). And, even if she were, the networks she would understand wouldn't be her own—they wouldn't be the ones she was *using* in order to understand the ones she was *programming*. Insofar as Searle-in-the-room *does* understand English *while* he is processing Chinese, he could map the Chinese terms onto his English ones, and thus he would understand Chinese in a sense that even Searle-the-author would have to accept.

References

Cole, D. (1991). Artificial intelligence and personal identity. *Synthese, 88*, 399–417.

Copeland, J. (1993). *Artificial intelligence: A philosophical introduction*. Oxford: Blackwell.

Flanagan, O. J. (1984). *The science of mind*. Cambridge, MA: MIT Press.

Ford, J. M. (in press). Helen Keller was never in a Chinese Room. *Minds and Machines*.

Goldstein, R. N. (2006). *Betraying Spinoza: The Renegade Jew who gave us modernity*. New York: Schocken.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jahren, N. (1990). Can semantics be syntactic? *Synthese, 82*, 309–328.

Morris, C. (1938). *Foundations of the theory of signs*. Chicago: University of Chicago Press.

New York Times (2010, 5 June). Justice Souter's Counsel. editorial, p. A20.

Parisien, C., & Thagard, P. (2008). Robosemantics: How Stanley the Volkswagen represents the world. *Minds and Machines, 18*(2), 169–178.

Putnam, H. (1975). The meaning of 'meaning' ", reprinted in *Mind, language and reality*. Cambridge, UK: Cambridge University Press, pp. 215–271.

Rapaport, W. J. (1981). How to make the world fit our language: An essay in Meinongian semantics. *Grazer Philosophische Studien, 14*, 1–21.

Rapaport, W. J. (1985/1986). Non-existent objects and epistemological ontology. *Grazer Philosophische Studien* 25/26: 61–95.

Rapaport, W. J. (1986a). Philosophy, artificial intelligence, and the Chinese-Room Argument. *Abacus, 3* (Summer), 6–17; correspondence, *Abacus* 4 (Winter 1987): 6–7, *Abacus* 4 (Spring 1987): 5–7. [http://www.cse.buffalo.edu/~rapaport/Papers/abacus.pdf].

Rapaport, W. J. (1986b). Searle's experiments with thought. *Philosophy of Science, 53*, 271–279.

Rapaport, W. J. (1988). Syntactic semantics: Foundations of computational natural-language understanding. In J. H. Fetzer (Ed.), *Aspects of artificial intelligence* (pp. 81–131). Dordrecht, Holland: Kluwer Academic Publishers. (errata online at [http://www.cse.buffalo.edu/~rapaport/Papers/synsem.original.errata.pdf]).

Rapaport, W. J. (1990). Computer processes and virtual persons: Comments on Cole's 'Artificial intelligence and personal identity' ", *Technical Report 90-13*. Buffalo: SUNY Buffalo Department of Computer Science, May 1990); [http://www.cse.buffalo.edu/∼rapaport/Papers/cole.tr.17my90.pdf].

Rapaport, W. J. (1995). Understanding understanding: Syntactic semantics and computational cognition. In J. E. Tomberlin (Ed.), *Philosophical perspectives, Vol. 9: AI, connectionism, and philosophical psychology* (pp. 49–88). Atascadero, CA: Ridgeview.

Rapaport, W. J. (1998). How minds can be computational systems. *Journal of Experimental and Theoretical Artificial Intelligence, 10*, 403–419.

Rapaport, W. J. (2000). How to pass a Turing test: Syntactic semantics, natural-language understanding, and first-person cognition. *Journal of Logic, Language, and Information, 9*(4): 467–490.

Rapaport, W. J. (2002). Holism, conceptual-role semantics, and syntactic semantics. *Minds and Machines, 12*(1), 3–59.

Rapaport, W. J. (2003a). What did you mean by that? Misunderstanding, negotiation, and syntactic semantics. *Minds and Machines, 13*(3), 397–427.

Rapaport, W. J. (2003). What is the 'context' for contextual vocabulary acquisition?. In P. P. Slezak (Ed.), *Proceedings of the 4th joint international conference on cognitive science/7th Australasian society for cognitive science conference (ICCS/ASCS-2003; Sydney, Australia)*. Sydney: University of New South Wales, Vol. 2, pp. 547–552.

Rapaport, W. J. (2005a). Implementation is semantic interpretation: Further thoughts. *Journal of Experimental and Theoretical Artificial Intelligence, 17*(4), 385–417.

Rapaport, W. J. (2005b). The Turing test. In *Encyclopedia of language and linguistics (2nd ed.)* (Vol. 13, pp. 151–159). Oxford: Elsevier.

Rapaport, W. J. (2006). How Helen Keller used syntactic semantics to escape from a Chinese room. *Minds and Machines, 16*(4), 381–436.

Rapaport, W. J. (2007). Searle on brains as computers. *American Philosophical Association Newsletter on Philosophy and Computers, 6*(2) (Spring), 4–9.

Rapaport, W. J., & Kibby, M. W. (2007). Contextual vocabulary acquisition as computational philosophy and as philosophical computation. *Journal of Experimental and Theoretical Artificial Intelligence, 19*(1): 1–17.

Searle, J. R. (1980). Minds, brains, and programs", *Behavioral and brain sciences, 3*, 417–457.

Searle, J. R. (1990). Is the brain a digital computer?. *Proceedings and Addresses of the American Philosophical Association, 64*(3), 21–37.

Shapiro, S. C. (1992). Artificial intelligence. In S. C. Shapiro (Ed.), *Encyclopedia of artificial intelligence (2nd ed.)* (pp. 54–57). New York: Wiley.

Shapiro, S. C.; & Bona, J. P. (2010). The GLAIR cognitive architecture. *International Journal of Machine Consciousness, 2*, 307–332.

Shapiro, S. C., & Rapaport, W. J. (1987). SNePS considered as a fully intensional propositional semantic network. In N. Cercone, & G. McCalla (Eds.), *The knowledge frontier: Essays in the representation of knowledge* (pp. 262–315). New York: Springer.

Shapiro, S. C., & Rapaport, W. J. (1991). Models and minds: Knowledge representation for natural-language competence. In R. Cummins & J. Pollock (Eds.), *Philosophy and AI: Essays at the interface* (pp. 215–259). Cambridge, MA: MIT Press.