Analysis and Metaphysics
Volume 11, 2012, pp. 106–132, ISSN 1584-8574

DE SE PUZZLES, THE KNOWLEDGE ARGUMENT, AND THE FORMATION OF INTERNAL KNOWLEDGE

ERICH RAST
erich@snafu.de
Institute for the Philosophy of Language
Universidade Nova de Lisboa

> You could not step twice into the same river; for other waters are ever flowing on to you. To those entering the same river, other and still other waters flow.
> – Heraclitus, Fragments 41 & 42

ABSTRACT. Thought experiments about de se attitudes and Jackson's original Knowledge Argument are compared with each other and discussed from the perspective of a computational theory of mind. It is argued that internal knowledge, i.e. knowledge formed on the basis of signals that encode aspects of their own processing rather than being intentionally directed towards external objects, suffices for explaining the seminal puzzles without resorting to acquaintance or phenomenal character as primitive notions. Since computationalism is ontologically neutral, the account also explains why neither Lewis's two gods nor Mary's surprise in the Knowledge Argument violate physicalism.

Keywords: phenomenal character, computationalism, de se attitudes, physicalism, acquaintance, the Knowledge Argument

Introduction

In this article a theory of internal knowledge is proposed that provides an explanation of de se puzzles like those of H. N. Castañeda (1967, 1975, 1989), Perry (1977, 1979, 1989), Lewis (1989), and Stalnaker (1981, 2004, 2008) as well as allowing for an alternative way of solving Jackson's Knowledge Argument. While the view defended in this article bears many similarities with the acquaintance hypothesis of Conee (1994), Bigelow & Pargetter (1990, 2006), and Tye (1999, 2000, 2009) it differs from these accounts in various

aspects, most notably in assuming computationalism. As will be argued below, computationalism gives rise to a form of ontological neutrality that makes the solution to the Knowledge Argument (almost) trivial. In a nutshell, the following theses will be defended: (i) A certain kind of knowledge arising from sensory inputs or from reflecting on one's own cognition is factual and also propositional in a sense that will be explicated in more detail below. (ii) Internal knowledge cannot be reduced to external knowledge about the physical world, but this irreducibility arises trivially from the way it is formed. (iii) Despite the previous points the position remains compatible with physicalism.

So according to the way Nida-Rümelin (2009) puts it an instance of the New Knowledge/New Fact thesis will be defended that does not violate physicalism. As the view is widespread that such a combination must in one way or another be incompatible with physicalism, a section on its own will be devoted to a defense of the proposal against possible criticisms. The account will also be briefly compared to the acquaintance view by which it has been inspired.

1. Some Puzzles

The following example will serve as a point of departure for the subsequent discussion of similar thought experiments from the literature:

> Suppose you have never eaten a truffle omelet in your life. When you eat one for the first time, you taste something that you have never tasted before. Given certain ceteris paribus clauses—your olfactory sense is working in the normal way and you're not seriously ill—it is safe to assume that you already have had the ability to experience the taste of a truffle omelet prior to having tasted one. Given, again, certain ceteris paribus clauses—the meal wasn't violently forced down your throat, the surrounding air was not polluted with toxic waste, and so on—after having eaten a particular truffle omelet you will have learned how *that particular* omelet tasted and be able to recall certain, though not all, aspects of this experience from memory. And after your second or third truffle omelet you will likely have learned how truffle omelets taste in general, although all of them differ from each other a little bit in taste.

One might think that something as innocuous as tasting a new dish would not stir up many philosophical controversies. Yet it seems that scenarios very similar to the omelet example have caused a decent amount of bewilderment among contemporary philosophers of mind and language.

To start with a well-known example from contemporary epistemology, Jackson (1986) devised the following scenario to show that physicalism does not hold:

> Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon this, including of course functional roles. If physicalism is true, she knows all there is to know. [...] It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. (Jackson 1986, p. 291)

In the Philosophy of Language similar examples have been considered starting with Castañeda (1967) and Perry (1977, 1979). These have been invented for a different purpose, though, namely for showing that I-thoughts in thinking cannot be reduced to third-person ways of referring to oneself without loosing some explanatory power that is essential for the explication of behavior. The Two Gods thought experiment of Lewis (1979) brings these puzzles to the point by formulating a de se puzzle in a possible worlds setting with agents that are omniscient about the external world:

> Consider the case of two gods. They inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still I can imagine them to suffer ignorance: Neither one knows which of the two he is. They are not exactly alike. One lives on the top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. (Lewis 1979, pp. 520–1)

The similarities and dissimilarities between these scenarios will be explored in the following sections. For reasons of space, closely related issues like the qualia debate or the logical modeling of attitudes will only be grazed.

2. Similarities and Dissimilarities

The similarity of the omelet example to the Knowledge Argument is fairly obvious. Both seeing something red and tasting an omelet are sensory experiences. In both cases the person hasn't made the respective experience before. In both cases, a purporter of qualia would claim that a particular red

experience or a particular truffle omelet experience respectively feels a particular way—a way that cannot be substituted by having another experience or emulated or imagined on the basis of theoretical knowledge alone. The key difference between the examples is that Mary is theoretically omniscient about the physics of her own neurophysiology. As unreasonable as this may be (see below why) this assumption may be also added to the omelet example. Why, then, has the omelet example been chosen as a starting point instead of the original Knowledge Argument? There are two reasons: First, the assumption that the agent in the thought experiment is omniscient regarding his own physiology—or, in a strengthened version of the thought experiment, about the physical universe as a whole—is only needed for arguing against physicalism. For investigating what internal knowledge might be it is not needed and should be avoided, as it is rather implausible. The compatibilist conclusion will turn out to be a side effect of the proposed view about internal knowledge formation, as it follows trivially from the computationalist assumption, but is not the main motivation for coming up with the notion of internal knowledge. We believe that the internal knowledge thesis is appealing independently of the question whether it is compatible with physicalism or not. Second, people that are not blind may find it easier to imagine never having tasted a truffle omelet before than never having seen a color. The omelet example illustrates the fact that except for the strong assumption of her omniscience about color vision there is nothing special at all about the Mary scenario. We often encounter new tastes and odors and a philosophical account of phenomenal character has to be compatible with and give some explanation for this mundane phenomenon.

It takes a bit more efforts to explain what the omelet example has in common with de se puzzles. In response to Perry (1977, 1979) and Castañeda (1967), Lewis devised his thought experiment to show that de re attitudes, knowledge and belief in particular, modeled on the basis of possible world semantics principally do not suffice for representing epistemic states of agents. In a 'standard' setting based on normal modal logic an agent's epistemic state is represented by the set of possible worlds reachable from the actual world by the agent's accessibility relation for the respective modality. In case of belief this set may be called the agent's belief set. The more the agent learns, the more this set shrinks. If, for example, someone learns that the Bradypus pygmaeus is a three-toed sloth then all worlds in which the Bradypus pygmaeus is not a three-toed sloth are eliminated from his belief set. The belief set of a supposedly omniscient agent like that of one of the gods in Lewis's example only contains one world, the actual world.[1] Suppose one of the gods is called 'Zeus' and the other 'Yahweh.' In Lewis's example the gods must still realize I am Zeus and I am Yahweh respectively, I am living on the coldest mountain, I am throwing down thunderbolts, and

so on. This lack of insight seems to conflict with their assumed omniscience, which has lead various researchers to believe that de se attitudes are irreducibly internal (see e.g. Castañeda 1989).

From a purely technical perspective, only regarding the logical representation of the agents' epistemic states, solving these puzzles is not very hard. Solutions range from early suggestions such as property-ascription theory and centered possible worlds (Lewis 1979) and Stalnaker's recent pragmatic variant thereof (Stalnaker 2011), impossible worlds (Hintikka 1975), over structured propositions (Richard 1983; von Stechow 1984; Cresswell 1985), situation theory (Barwise & Perry 1983), to hyper-intensional logics (Thomason 1980; Bealer & Mönnich 1989; Moschovakis 1994, 2006; Muskens 2005). All of these accounts are based on the idea of using entities that are more fine-grained than possible worlds as a representation of semantic content. However, none of these solutions give any substantial insights into the question what exactly Zeus learns when he realizes that he* himself is Zeus.[2]

With respect to that question, de se puzzles and the Knowledge Argument are parallel to each other. Just like in the Mary example it seems as if physicalism could not be true if Zeus and Yahweh were still able to learn who they are after they have already learned everything there is to know about the physical world. Being confronted with such a perhaps unwanted conclusion one might doubt whether a truly omniscient agent would not be able to realize who he is on the basis of his vast knowledge about the external world. After all, he also knows all there is to know about his body and his brain.

Dennett (1991) has given an analogous reply to the Mary example. In his opinion, the thought experiment is based on unfruitful intuition-pumping and does not refute physicalism. Mary might not be surprised at all when she sees red for the first time and would not learn anything new when she sees red. Many philosophers, however, have found this reply unsatisfactory and this strategy of explaining away the problem has not been very popular in case of de se attitudes either. At least some people have a strong intuition that no ensemble of physical facts about color recognition or taste perception can result in knowing how it feels to see red or knowing how a truffle omelet tastes respectively. Correspondingly, thinkers like Wittgenstein, Chisholm, and Castañeda had strong intuitions that genuine I-thoughts are not reducible to third-person thoughts about oneself.

Let us stay with the Knowledge Argument for a moment, though. A vast number of alternative replies have been given to the Knowledge Argument by people that would essentially like to retain physicalism or at least do not believe Jackson's argument refutes it decisively. Addressing all of them would go beyond the scope of this article, but two prominent strands of replies need to be mentioned. According to the ability hypothesis Mary does not gain

new knowledge but an ability. Different variants of this thesis have been defended by Lewis (1983, 1988/1990), Nemirow (1980, 1990), and Churchland (1985, 1989). It has been criticized extensively, see for example Conee (1994) and Coleman (2009), and does not seem to be viable in general. In a nutshell, the main criticism of this position is based on the fact that an agent can have and often will have the respective ability in the first place. For example, there seems to be no reason why Mary could not have had the ability to recognize red objects before having seen something red for the first time. Stipulating, on the other hand, that she gains the ability to recognize how it feels to see something red clearly begs the question; for the physicalist will claim that she already has this ability by virtue of her omniscience about color vision. More would have to be said about these arguments, but laying them out in more detail is not the purpose of this article. For what it's worth we will follow the above mentioned critique here and not further pursue this line of thought in what follows.

The second position to be mentioned is commonly labeled acquaintance hypothesis. According to this view, acquaintance with certain properties like redness or sensory experiences of them cannot be explained solely in terms of knowing-that or abilities. This thesis has been defended by Conee (1994), Bigelow & Pargetter (1990, 2006), and Tye (2009).The position to be laid out below is similar to theirs and a discussion of the differences and similarities between the present account and the acquaintance view will be postponed until Section 5.

Before going on, let us take a closer look at the parallels between the Knowledge Argument and de se puzzles. These are really just two sides of the same coin, though the connection between the two sides has only been explored sporadically (see e.g. Stalnaker (2004)). First of all, it seems striking that Lewis's two gods can be taken as a counter-argument to physicalism just as well as the Mary example. If the gods are omniscient about the physical world and each of them still learns something when he realizes that he* himself is the F-er (for whatever property F is under discussion), then, so it seems, their belief sets must contain something non-physical that is eliminated by the learning process. This is at least so under the additional premise of logical atomism and modest versions of metaphysical realism according to which by 'physical' any constellation of matter and forces acting upon it is meant and it is assumed that logical languages describe such constellations of matter and forces, and nothing else if physicalism is true.[3]

Less radical de se puzzles like Perry's (1979) supermarket example or Rudolf Lingens, who is lost in the Stanford library, in Perry (1977) cannot be considered directly parallel to the Mary example, because the agents in them are not omniscient. The purpose of these examples is to show that there is, either at a linguistic level or at the level of cognition, an irreducible first-

person perspective. Only when the agent has an appropriate I-thought, which cannot be 'reduced' to any 3rd-person way of the agent thinking about himself, will he realize that he himself is the F-er and elicit the corresponding I-behavior like cleaning up a trail of sugar in the supermarket after having realized that it was he* himself who caused it. Despite all these differences to the Knowledge Argument, these examples seem to illustrate a point that the Mary example seems to show as well. If Mary indeed learns something new, whatever that is, what she learns seems to be intrinsically tied to her experiences and her perspective. No third person observer can learn the same at pain of making what is learned 3rd-person reducible and a fortiori already known to him by assumption. So both sorts of puzzles stipulate some irreducible perspectivity centered on the experiencer (viz., thinker).

3. Computationalism

As an assumption in what follows a computationalist view of the mind will be presumed. The idea hereby is that a fruitful solution to any of the above puzzles requires a form of computationalism, yet no detailed arguments shall be given for this assumption here. It shall only be briefly motivated and in case of doubt should be taken as a genuine assumption. There are two main reasons for taking computationalism as a basis for considerations about the puzzles mentioned above: First, non-computationalist theories of the mind like for example non-computational functionalism in psychology are generally not specific enough to tackle the problem at hand in a satisfying manner. These theories might very well be suitable for formulating other philosophical positions and play an important role in the research of cognition in general, but they simply leave to many loopholes open for other solutions to the above puzzles that could be correct, but are not the ones we would like to discuss in what follows. The goal of this article is to lay out one among many possible coherent explanations of internal knowledge and theory pluralism prohibits us from regarding it as the only one. Second, adopting computationalism will allow us to talk about mental phenomena in familiar computational terms that hopefully make it easier to identify the position, for the differences between internal and external knowledge that will be introduced in the next section are subtle and may easily become obfuscated by a more liberal terminology. The question what actually hinges on computationalism will be taken up again in Section 6.

First of all, the term 'computationalism' must be clarified. Sometimes 'computationalism' is used more or less synonymously with 'symbolic, hard AI.' This is not the way the term is to be understood here, though. Computationalism is also sometimes erroneously taken to imply physicalism, which is simply not the case. As it is understood here, computationalism about the

mind means roughly speaking no more than the hypothesis that the brain or the mind solely works on the basis of processes that are representable by computable functions as long as we abstract from issues concerning input and output in a suitable way.

With some caveat discussed below, a computation will from now on be understood as a process that can be expressed by applying the conversion rules of the untyped λ-calculus (α-, β-, and η-conversion) to a term of the untyped λ-calculus. The input of a computation consists of arguments to an unreduced 'redex' term of untyped λ-calculus to which the conversion rules of the calculus may be applied. The result of a computation is a term in canonical form, if there is such a form. (As it is possible to express infinite loops in the untyped λ-calculus some computations might not halt.) A computer is an open system that is receiving inputs, storing and processing them as signals, and occasionally produces results (outputs), where (a) the processing of the inputs is computational and (b) the system as a whole is not a hypercomputer. The first part of the condition is obvious, but something needs to be said about the second one. For example, a source of randomness attached to a computer could theoretically increase the computational power of the device beyond the power of a Turing machine, although it is hard to see how this additional expressive power could be harnessed in practice. There are many other devices in the hierarchy of hypercomputers that could theoretically be produced by extending an open computational system. For example, a Turing machine with the ability to go through an infinite list in finite time would be a (theoretical) hypercomputer. Likewise, a PC with the sublime ability to perform every subsequent step in a program in ½ of the time of the previous step, i.e. an implementation of a so-called 'Accelerated Turing Machine', would be a hypercomputer. By limiting the following considerations to computers in the narrow sense, not including hypercomputers, a minimal explanation of the puzzles and a minimal account of internal knowledge will be obtained. Software that runs on an ordinary computer can also be adapted to run on a hypercomputer, provided that the hypercomputer is based on an extension of the computational model. So if ordinary computationalism suffices for explaining a phenomenon, then hypercomputers would not add anything substantial to this explanation and should be avoided.[4] On the other hand, if they existed, hypercomputers could perform a variety of 'magical' tricks an ordinary computer could not perform, and such a stipulation ought to be avoided. Hence, no hypercomputers are allowed.

Another potential source of worry needs to be addressed. Although reducing terms of λ-calculus is mostly a sequential business, it could be argued that in light of connectionist models of the brain we should take into consideration additional forms of parallel processing. Using λ-calculus as a basis may indeed be a simplification; perhaps this presumption needs to be

adjusted. λ-calculus has the advantage of providing a canonical theoretical model for sequential computation, the same of which cannot be said about theoretical models of parallel computation such as the Actor Model (Hewitt, Bishop & Steiger 1973) or the π-calculus (Milner, Parrow & Walker 1992). However, the arguments of the following paragraphs would not be affected in any substantial way by allowing forms of concurrency that are not expressible in λ-calculus. Connectionist models are explicitly taken to fall under the label 'computationalism' in what follows and if they require genuine parallelism the above definitions would have to be adjusted.[5]

A third point deserves some emphasis: As already mentioned above, a computationalist position about mental phenomena does not imply any particular ontological position like monism, dualism, physicalism, or epiphenomenalism. Computationalism only concerns the way information in a system is represented and processed, namely by computable functions and their application in a computer, and is neutral about underlying substance principles. Although computationalism is often combined with physicalism, a Cartesian dualist could just as well defend a computational theory of the mind. In fact, computationalism of the mind is more than an attractive position for the dualist, as it is prima facie rather unclear what non-mystical alternatives there are. This is also the reason why a broader functionalist perspective has not been adopted. Remaining agnostic about the underlying processing principles might serve well for doing empirical psychology, but can hardly be taken as a viable philosophical stance. After all, there are not so many options on the table:

(A) Is the mind/brain hypercomputational? Yes/No.
(B) Is the mind/brain computational but not hypercomputational? Yes/No.

Answering 'yes' to A has been excluded on independent, methodological grounds (minimality). Answering 'no' to B does not for itself refute a theory based on a positive answer to B and accounts that reject computationalism altogether tend to be shrouded in mistery. Notice, however, that once option B is chosen, multiple realizability does follow from the above notion of computation. Consequently, if it can be made plausible that internal knowledge compatible with physicalism may be formed by a machine, then the explanation can be transferred to a computational model of human internal knowledge formation provided it is general enough.[6]

A final issue shall be mentioned for clarification. Sometimes people claim that computationalism is a metaphor. To them, it is the idea that the mind or brain of humans works like a computer. This thesis seems to be based on an unnecessarily narrow definition of what a computer is, and is decidedly not what is meant here. Computationalism is the thesis that the human mind or brain *is* a computer in the sense laid out above.[7]

4. Internal Knowledge

Now that some background assumptions have been fixed, let us return to the omelet example and address the question what would happen when a computer, say a robot built for that purpose, tastes a truffle omelet. The answer appears to be clear. The robot receives sensory input caused by the truffle omelet, the resulting signals are processed according to the current state of the robot (the 'program') and the circuits of the robot (another aspect of the 'program'–the distinction between hardware and software is irrelevant in the present context), are possibly stored, and the robot possibly produces some motor output. While input and output are external to the robot, the signals representing inputs, the state of the robot, and the actual processing are different aspects of the ongoing computation. The central thesis of this article is that the Mary example and de se puzzles can be explained in much the same way by looking at the state of the brain and the role the respective signals play within the organism as a whole.

Here is what happens when Mary sees a red object for the first time, as described from a computational perspective: Mary (i) receives a particular sensory input (presumably caused by a red object, but this does not matter) and (ii) performs computations based on this signal in her brain or mind (taken as a computer). These computations cannot be substituted—under any reasonable way of understanding 'to substitute' in this context—by (iii) states and computations of the computer representing physical knowledge or any other knowledge whatsoever. For suppose they could be substituted. From a computational point of view this would amount to the claim that one λ-term A responsible for the processing of color vision could be substituted within a larger term C by another λ-term B responsible for the storing of a vast amount of physical knowledge about color vision to yield term C' and both C and C' would for any input arguments reduce to exactly the same canonical term. Obviously, intuitions break down in this scenario, since the λ-terms in question would have to be insanely complex to represent even just a small aspect of how the brain or mind works, but that is not the point. The point is that two programs that serve a completely different purpose can by definition not replace each other, because their purpose must be defined in terms of input, output pairs. If their input, output pairs differ they cannot have exactly the same purpose and they fulfill different roles as subprograms in the computational system as a whole.

Finally, knowledge is formed when (iv) aspects of an input signal and its processing are stored in a way that they may be retrieved later. In the jargon of the mind this means that Mary might remember her first red experience and she might be able to compare it with prior and future color experiences. Comparison with prior color experiences explains her surprise

when she sees a red object for the first time despite having exhaustive physical knowledge. Mary's surprise comes from the trivial fact that she did not process a signal presenting something red before, even though she has acquired all physical knowledge about her processing of those signals. Comparison with future red signals allow her to learn the concept RED by iterative abstraction and generalization from particular red inputs under varying light conditions.

This way of looking at the example does not seem to be far apart from how Dennett or even Quine would describe it. So one might ask where the phenomenal character of experiences enters into this picture. But this question is not very hard to answer from a computational point of view either. The signal presenting a certain sensory input in combination with the state of the computer and the way it is being processed within the computer at a given time is the phenomenal character—for the computer, at the given time. The signal is fine-grained enough and processed in a certain way—in a way no other signal is processed by the computer at the given time.

Now there is seemingly a problem with this point of view that must be addressed before going on. Of course, one computer cannot just take a signal within another one and implant it into his own processing to check that a certain signal presents a particular phenomenal character in that system. First, doing this would likely be physically impossible. Second, the state of the second computer will likely differ from that of the first one. And third, because of the second point, there can be no guarantee that the signal would actually be processed in the same way even if it could be transferred from one machine to the other. This is not really a problem, though. The fact that one computer cannot check that a certain signal directly presents a certain phenomenal character within the computations of another computer does not invalidate the explanation of how phenomenal character arises in the first place. In the way they have been understood so far, signals have all the properties a fruitful explanation of phenomenal character should have. A signal presenting a particular shade of red within a computation is not the same as the object causing it and is not the same as light of a certain wavelength, and so forth. A signal feels a certain way because it is processed in a certain way and fixes possible future states of the system in a way that only this signal can do. And signals are fine-grained enough, since they would otherwise not be able to (re-)present the explanandum.

The idea that there are presentations of experiences in the mind that are to the thinker in actual cognition what they represent is certainly not new. Amongst others, Castañeda (1975, 1989) held this view and developed his own trope theory to account for it. Against this idea one might argue that, since we cannot check the way presentations of others work within others (think about inverted spectrum examples), we simply cannot know whether

there are presentations that are fine-grained enough, and since we cannot know this we should instead adopt an externalist 'semantics of thinking', one that we understand better because it is more closely tied to natural language meaning. How do we know, for example, that Mary doesn't represent every red object in a very coarse-grained manner, as a representation that directly corresponds to the meaning of a natural language sentences like 'There is something that Mary Smith experiences at 5 o'clock in front of her', or 'Mary sees a red ball'?

The answer to this worry is to a vast extent empirical in nature: As a matter of fact humans do not generally remember sensory experiences in a coarse-grained way. They practically always store a more fine-grained signal that also represents other aspects of the computational system and aspects of how the signal was processed at that time. Knowledge formed on the basis of the sensory experience is to a large extent internal, because it involves internal aspects of cognition as well. Mary stores much more than what English phrase 'being red' means when she sees something red for the first time and much of what she stores concerns her own cognition rather than the tomato in front of her. In other words, there is nothing wrong in principle with an externalist view about the 'objects of thought' except that it is empirically wrong when being used as a description of actual cognition —in contrast to natural language meaning, which is to a large extent externalist because languages are public means of communication.

Some defenders of qualia will likely find the whole computational explanation unconvincing, as they feel that it still leaves too many questions unanswered. For example, it seems as if a computational explanation cannot give a satisfying answer to the question why a particular kind of signal and not another one presents a particular sensory experience. A few words have to be said about this alleged explanatory gap.[8] First, computationalism not only allows for, but also correctly predicts this gap. As laid out above, one computation cannot check the role a signal plays in another computation without becoming functionally equivalent with the other one, in case of which no checking would be involved. Second, there is a more general argument why there must be a gap. Having a certain sensory input and (possibly) later recalling certain aspects of it differs both from an external and from an internal perspective from the processing of an explanation of how the respective input is processed, stored, and later retrieved. This ought not come as a surprise. Any explanans differs from the explanandum, or else we would not speak of an explanation in the first place. A particular shade of red has a particular phenomenal character to Mary at a given time because she is in a particular state at that time and this particular sensory input plays a particular role for the evolution of future computations. Trivially, an explanation of this process cannot substitute her processing of the sensory input.

Let us now turn in more detail to the question how and where knowledge enters this picture. A computer forms knowledge when it actually stores aspects of a signal that directly presents a particular state of the computer within the computations of the respective device and at the meantime has the ability to later retrieve those aspects of the signal. As laid out above almost by definition neither the signal processing nor the resulting knowledge can be substituted for other forms of knowledge. If forming knowledge essentially means that aspects of signals are stored, then the fine-grainedness of the signals imposes an upper limit on the fine-grainedness of the knowledge representation. However, since aspects of signals in current parlance may also include any aspects of their internal processing, in theory a knowledge representation of this kind can be almost arbitrarily fine-grained. For example, every signal representing red could be stored in combination with an internal state representing the subjective flow of time, making all of the representations internally distinguishable from each other even if the sensory apparatus has a certain minimal resolution. So for all what its worth knowledge representations can be as fine-grained from a computational perspective as the limitations of the computer allow.

Having said all that, what distinguishes 'ordinary' knowledge from internal one? The answer is this: the subject matter. Internal knowledge is about aspects of cognition itself whereas ordinary knowledge is, roughly speaking, about the world. In lack of a viable theory of cognition—computational or otherwise—this distinction will inevitably lack some precision, but for the present purpose it seems to suffice to base it on some form of intentionality. Internal knowledge, as it is understood here, involves knowledge formed from presentations of objects in cognition that are themselves part of cognition and present aspects of it. In computational jargon: Internal knowledge is knowledge resulting from the storing of aspects of a signal that represent part of its own processing. In contrast to this, external knowledge involves intentional representations of the external world. In computational terms: Mostly aspects of a signal are stored that represent a respective sensory input. There is one terminological issue that must be addressed in this context. Authors in the Meinongian tradition such as Priest (2005) use 'intentional' in a much broader sense according to which we can think about round squares and the corresponding mental representation is intentional towards a round square—whatever that means. In contrast to this, here the adjective 'intentional' is meant in the narrower, but nevertheless rather loose sense of 'pointing/being directed towards an external object.' In other words, according to the present way of talking someone's thought about a round square would not intentional towards a round square, but rather be classified as an internal presentation of an impossible object that is used in cognition to muse about round squares. Knowledge formed on the basis of such a presentation

is in the suggested terminology internal and not external. In contrast to this, a signal whose main purpose in an ongoing computation is to represent aspects of a physical object as they are given by sensory inputs would be intentional in current parlance.

Closely related to this distinction is another one, that between presenting a state of the device within its computation and representing states of the external world, which has been presumed silently so far. A particular red-experience must be presented in someone's mind in one way or another. In computational parlance this means that a certain signal in the computer presents a red sensory input at the given time to the machine within the current computation. Insofar as the computations are concerned the signal is for the computer that particular shade of red at the given time. In contrast to this, when a signal merely represents something red this does not imply that the system is actually processing the redness of that thing. The representation is in this case of such a kind that within the computation the representational sign represents something else that is not present in the computation at the given time.[9] To give a perhaps more intuitively accessible example, suppose the number 3 is stored in a memory device by setting the bits of two binary registers. Then this constitutes a rudimentary form of internal knowledge of 3 (for instance, as part of the storage of a more complex signal representing the presence of three tomatoes within the visual field) by virtue of the fact that the machine may later retrieve the state of the registers in a way that makes the subsequent computation one that involves the number 3. On the other hand, a purely representational signal such as a graphical representation '3' could only serve for subsequent computations in the same way via an additional translation step from '3' to the state 11 of the binary register. External knowledge is formed on the basis of representations that are intentional, whereas internal knowledge is formed on the basis of presentations that may or may not be intentional.

These distinctions are admittedly sketchy, but not much more can be expected at the current state of research. They are further complicated by the fact that in actual cognition rarely just only one aspect of a signal is stored. Although the philosophical use of the notion does not suggest this, it is by far not clear whether intentionality does not, perhaps, come to a degree and signals might be more or less directed towards the world. If so, then knowledge would also come to a varying degree of 'externality.' Consider, for example semantic externalism. Representations of meanings of natural language sentences are mostly or entirely external according to this position. Nevertheless, they must be represented in some way in the brain or mind of a speaker when he 'grasps' a corresponding thought and in that respect are also internal or linked to internal presentations. Take, on the other hand, Mary's first red experience. Perhaps the signal stores a particular

shade of red plus some aspects of its processing such as Mary's surprise about it. In this case the knowledge in question will be mostly internal, because the subsequent processing when the memory is retrieved will be one involving presentations of the original signal and the accompanying surprise about it rather than mere intentional representations directed at, say, the tomato that was present at the time in front of her. Mary may form the internal knowledge without knowing what a tomato is or what redness is in general. Yet there does not seem to be any principal reason why such a signal could not, at the same time, also represent external aspects of the tomato. So perhaps the external/internal distinction is a matter of degree. This does not, however, make it less fundamental. We may speak of mixed signals that have external or internal aspects, but will continue to talk about internal versus external knowledge as if there was a clearcut dichotomy. This is meant to be a loose way of talking about 'mostly internal' and 'mostly external' knowledge.

Equipped with this rough conceptual framework we are able to address the irreducibility of internal knowledge to external knowledge. It is the same sort of trivial irreducibility as has been laid out above for signals in general. While a particular past input or another aspect of a past state of the computer may in principle be represented by different signals—both from the perspective of types of signals and that of tokens thereof—these signals all represent the respective aspects of the past signal. (If not, they represent something else and we have classified the signal incorrectly.) Let 'actualization' be taken as a shortcut for the transition from dispositional knowledge that p to the episodic being aware that p is the case with high certainty during a certain time period. Then external forms of knowledge that are actualized when a computer accesses its memory lead to signals that represent something else than the signals resulting from the actualization of internal knowledge. Consider, for example, again the internal knowledge Mary may form on the basis of her first experience of a red object in comparison to her external knowledge about her processing of sensory input caused by red objects. If activation of these different kinds of knowledge lead to the same sort of signals then internal and external knowledge would be the same—but they are not!

Being able to remember, however vaguely or precise, a past red experience does not require one to have knowledge about neurophysiology, and as the Mary example shows the opposite does not hold either. Certain signals are not 'substitutable' for each other, simply because different sorts of signals in a computer play different roles in the evolution of future states of the device.

To recapitulate, Mary gains internal knowledge when she first sees something red, and this knowledge is genuine fresh knowledge of a particular kind, and having this knowledge can be explained in terms of the workings

of a computer. Whether a device is computational or not, on the other hand, is not a question of ontology but a question of what the term 'computation' means, and this question can be answered precisely on the basis of seminal work by Church, Turing, and many others.

5. Omniscience and De Se Puzzles in General

Of course, not all computers are limited to storing and processing particular sensory inputs in a peculiar way; some may also store and process aspects of their current state whether or not these have resulted directly from any input. Bearing that in mind, one may ask what happens in the brain or mind of Zeus when he realizes that he* himself is Zeus. What exactly happens in someone's brain or mind in this case is not so well understood yet, but there is no doubt that something happens. If so, and if Zeus is learning something when he has the respective insight as purporters of de se puzzles claim, then the computer 'brain or mind of Zeus' changes from one state before the insight to another state after the insight. Zeus forms internal knowledge: Aspects of the signals presenting the insight 'Oh, it's me who throws thunderbolts!' are stored for later retrieval and subsequent processing. If this were not the case then we would not say that Zeus has learned something.

From these considerations it follows that the thought experiment is misleading by asserting that the two gods are omniscient. Once any of the gods has the respective insight, his internal state has changed and he invariably lacks physical knowledge of this state change. Now any newly acquired knowledge of the first state change must result in a second state change, and so on for any higher-order knowledge about the agent's own state. Hence, an agent can in principle not obtain complete knowledge of his own physical (or mental) state. This conclusion can be drawn independently from the question whether physicalism is true or nor; again, only computationalism is assumed.

Apart from this flaw of the thought experiment the actual formation of internal knowledge in case of the two gods is no less mysterious than Mary's first color experience or tasting a truffle omelet. For example, Zeus may think at some time: I am Zeus. In order to constitute internal knowledge relevant aspects of this thought, including the fact that it has occurred in actual cognition in a sort of assertoric mode of thinking, must be stored and be ready for later retrieval.[10] Likewise, in other de se puzzles like Perry's supermarket example a respective thought 'I am the F-er' is stored in a way that allows some of its aspects to be retrieved later. In case of typical de se puzzles these aspects are usually those that are relevant for the explanation of corresponding externally observable I-behavior. For example, in Perry's (1979) supermarket example John Perry does not realize that a damaged

package of sugar is leaving a trail behind his shopping cart. He tries to find the one who is making a mess. Once he realizes that he* himself is the one whose cart produces the trail of sugar, he will start to clean it up. The key difference of these examples to the Mary scenario is that de se puzzles prima facie concern internal knowledge based on signals that occur within the computational system without being linked to particular sensory inputs in any direct and obvious way. Apart from that, they work the same way and can be explained in similar terms on the basis of the external vs. internal knowledge distinction.

6. Comparison to the Acquaintance Thesis

It seems striking that the present suggestion bears similarities with the acquaintance accounts of Conee (1994), Bigelow & Pargetter (1990, 2006) and Tye (2009) mentioned in the beginning. It also differs from their accounts, though. One major difference to Conee's position is that in his view knowledge by acquaintance is not factual, propositional knowing-that. He writes:

> The learning is a matter of Mary's becoming acquainted with the visual experience that ordinarily results from seeing something red, and this acquaintance consists in Mary's experiencing phenomenal redness. She experiences the quality, and that teaches her what seeing red things is like. She does not learn any new fact. (Conee 1994, pp. 140-1; emphasis added)

A similar point of view is also defended by Bigelow & Pargetter (1990) on the basis of a possible worlds semantics for belief with a special indexical acquaintance relation that is also supposed to account for de se puzzles. In both views, knowledge by acquaintance is a special form of knowledge besides knowing-how and knowing-that and is not reducible to one of the others. However, this cannot be quite right. There is no reason to believe that Mary's experiencing of a particular shade of red is not a fact. Consider the parallel case of de se puzzles again. There is different behavior associated with an agent having had a thought of the form John Perry is making a mess than one of the form I am making a mess, although to the two-dimensional semantic externalist the corresponding sentences 'John Perry is making a mess' and 'I am making a mess' express the same semantic content. In the present view the two thoughts differ from each other, as they are tied to different future behavior, and the formation of corresponding internal knowledge involves aspects of these thoughts that account for their difference. (If this weren't the case, there would not be anything puzzling about de se puzzles.) Correspondingly, if taken seriously the two gods example shows that two different facts are learned when for example Zeus has the insight

Zeus is throwing down thunderbolts versus I am throwing down thunderbolts. What Conee seems to mean in the above passage is that Mary does not learn an additional fact about the external world. However, strictly speaking not even this can be right, since Mary also learns that she had a certain sensory experience. If this weren't a fact about the external world, then physicalism would be false, but why should one beg the question in this way when the same example can also be explained in more neutral terms? According to our view internal knowledge is an instance of propositional knowing-that in the same sense as de se belief is an instance of propositional believing-that. It is propositional in the sense that propositions are sets of doxastic alternatives from which an agent removes alternatives when he learns something and it is this sense of 'propositional' that is relevant in the context of discussing the puzzles. Both Mary and robot Mary learn a new fact when they see a red object for the first time. The knowledge in question is an instance of knowing-that because after having seen a red object for the first time Mary and robot Mary have gained knowledge that they have experienced something red, where this knowledge does not imply that the object that caused their experience was red. This position remains compatible with physicalism not because internal knowledge is not propositional or not factive knowledge, but because Mary's physical state changes when she experiences a red object if physicalism is true. In mental parlance, the knowledge in question is knowledge that a certain mental phenomenon occurred. In computational parlance, presuming computationalism, this corresponds to the fact that aspects of signals presenting the processing of a red object within a given computational system are stored and may be retrieved later by the system.[11] There is nothing special about the ontological status of these signals; nothing can be inferred from their existence about the Mind/Body problem, because they can be described in purely computational terms without presupposing a particular substratum in which the computation takes place. Within a computer any experiences, including 'inner experiences' caused by cognitive processes, are represented by particular sorts of signals. Each signal corresponds to a term of the λ-calculus, or a suitable process calculus if parallelism is needed, in the corresponding abstract representation of the computer and its state.

Tye (2009) has recently also defended the acquaintance thesis at great length. In his view direct acquaintance with an object through the senses gives rise to a special form of knowledge by acquaintance that is not an instance of knowing-that. Again, the knowledge in question is not factual or propositional:

> The fact is that there need be no fact I know in knowing the color. I can know a thing simply by being conscious of it. (Tye 2009, p. 99)

The same critique as above applies to this position. Moreover, Tye also defends a direct reference account of perceptual content, the 'Singular (When Filled) Thesis', according to which the content of a perception 'contains' an object when the perception is veridical and is gappy in case of hallucinations and misperception. The content is nonconceptual, i.e. it does presume the existence of phenomenal concepts.

In our view such a theory of perceptual content does not work. The nature of perceptual content is to a large extent determined by sensory physiology and a vast array of optical illusions suggests that perceptual judgments and perceptual content cannot be separated from each other. Generally speaking, representations of perceptual content based on Russellian propositions which presume 'objecthood' of some of the entities involved seem to be problematic. In the Philosophy of Language from which Tye borrows the notion gappy content has turned out to be a double-edged sword. On one hand, it seems to offer a solution to the direct reference theorist for the case when a referent is missing. On the other hand, whenever cognitive significance comes into play different missing objects require different gaps, and so in the end these gaps are not gaps at all. Regarding perceptual content, different hallucinations and misperceptions lead to different behavior, and differences in behavior in this case must be explainable by differences in perceptual content. So gappy content alone does not suffice.

Having said that, it is worth mentioning that the present suggestion may very well be compatible the acquaintance-based theories laid out above. The main concern is rather that these positions rely on notions like acquaintance, perceptual content, or non-propositional knowledge that are not only not needed, but also seem to point into the wrong direction. Apart from that, our position shares many similarities with them. According to both views the ability hypothesis is rejected as an explanation of phenomenal experience. Mary does not need to gain a new ability, and even if she did, it would not explain her learning process. Both accounts are supposed to be compatible with physicalism. Both Tye and Conee also seem to agree with us that having a particular kind of experience is a precondition for the formation of corresponding knowledge and that having the experience trivially changes the agent's state. So the main difference between acquaintance views the one proposed here seems to be that the former rely on some form of nonfactual knowledge to solve the respective puzzles, whereas in our view the distinction between internal and external knowledge suffices for that purpose.

7. Objections and Replies

Objection 1: The Zombie Argument. Chalmers (1996; 2005) devised the famous Zombie argument against materialism. Does this argument also speak

against computationalism? The idea would be that because of multiple realizability Zombies could implement the same abstract computations as humans in the actual world independently of which ontological position is chosen regarding human cognition. Since Zombies would in this scenario process, store, and retrieve signals exactly the same way as we do, and according to the argument's premise they would nevertheless not know what it feels like to see something red, computationalism about human thinking would be refuted if at the meantime it were maintained that humans can know what it feels like to see something red. Much has been said about the original argument and the criticisms of it also apply to this version. In a nutshell, they are as follows: Zombies are perhaps not positively conceivable; they might only appear to be conceivable because we do not yet know enough about the kind of computations that give rise to consciousness and conscious experience. This is basically the reply of Dennett and also the one favored here. Recall the checking problem. It is possible that certain computations give rise to, say, qualia or consciousness without any third-person observer ever being able to verify that this is the case, much in the same way as each of us is not able to strictly speaking verify that solipsism is false. Moreover, if metaphysical possibility is to make any sense (which is not to say it does), then conceivability ought better not entail it. For it seems perfectly reasonable to assume that, say, it is positively conceivable that it rains tomorrow albeit being metaphysically impossible because God has other plans.[12] Yet another response also concerns metaphysical possibility. Perhaps Chalmer's assumption that materialism must hold necessarily is too strong and the thought experiment is ultimately a strawman. Suppose materialism holds in the actual world. Suppose Zombies are conceivable and metaphysically possible. Since what is possible is not the case, one might argue that materialism based on what is the case is more than enough. Ironically, from this point of view Chalmers would come along as an old-fashioned materialist rather than a property dualist. There is no room here to further delve into these details, though. The bottom line of the reply is that the Zombie argument is not sound.

Objection 2: Any instance of the New Knowledge/New Fact View is incompatible with physicalism.

Another worry about the proposal is that one might have a strong intuition that it cannot work under any circumstances. The belief that any instance of the New Knowledge/New Fact View must be incompatible with physicalism seems to be ingrained by many and something has to be said about it. In assessing the puzzles it is important to bear in mind that once time is taken into account it becomes unrealistic to expect, even just for a thought experiment, that an agent may be omniscient about all of his future states. But even if this possibility is granted, there are still infinitely many things a

seemingly omniscient agent cannot know at a given time. For example, an agent a cannot know at time t_1 that he has had a certain experience e at time t_2, where $t_1 < t_2$. He may know at t_1 that he will have had e at a time t_2. Let us call this knowledge attribution K_1. Let further K_2 be the knowledge attribution that a knows at t_3 that he has had experience e at t_2, where $t_2 < t_3$. These attributions not only have differing truth-conditions, they also imply different states of a provided that physicalism is true. If K_1 is true, a might be in state S_1 at t_1. This state is not a result of having had e. On the other hand, If K_2 is true a will be in another state, say S_2, at t_2 that is the result of having had e. Obviously, the agent cannot know at t_1 that he is in state S_2 at t_1 because he isn't. But it is the state itself, via the storing of aspects of signals, that gives rise to internal knowledge. Consequently, K_1 cannot be internal knowledge of e while K_2 might or might not amount to genuine internal knowledge. So the reply to the above worry is that once time and an agent's state of his mind or brain are taken into account, the New Knowledge/ New Fact position becomes much less problematic than it might seem at first glance, since for the physicalist the new internal fact will supervene on a new external fact about the respective agent's brain, yet learning the external fact would not amount to learning the internal fact since thinking a thought_i is different from thinking an explanation of it_i.

Objection 3: Presuming computationalism is begging the question. Philosophers just love to accuse other philosophers of begging the question because anyone who argues in a deductively valid way can be accused of that. However, in this particular case one might ask what question would actually be begged. On the basis of assuming computationalism it was argued that de se puzzles and the Knowledge Argument are compatible with physicalism, where the main tenet was to lay out the external versus internal knowledge distinction. Assuming computationalism of the mind, or more commonly of the brain, seems to be the right thing to do. The burden of proof in this matter is on the side of the anti-computationalists. Suppose, however, there were good reasons to reject computationalism and instead opting for (non-mystical?) alternatives such as descendants of the Penrose-Hameroff view. Even then the above position could perhaps be reformulated in more traditional terms of the philosophy of mind, but in any case a rejection of computationalism would not be so devastating. It was argued that if p, then q; clearly, if someone argues not p this is much less of a concern than if someone were to argue p and not q.

Objection 4: Talking about signals is a way of introducing phenomenal concepts in disguise.
The idea of this objection is that the way we have implicitly defended what Chalmer's calls Type-B materialism based on phenomenal concepts, i.e. the

position that there are phenomenal concepts that are epistemically distinct from physical concepts but ontologically identical to or supervening on corresponding physical concepts. Talking about internal signals is from this point of view just a (perhaps undesirable) way of introducing phenomenal concepts. The reply to this objection is twofold. First of all, since computationalism is ontologically neutral only a weaker thesis could have been defended: The distinctions that were made suffice for accounting for explaining the phenomenal character of experiences and de se attitudes without violating physicalism. Second, the notion of a signal, as it has been used throughout this article, does not presuppose the existence of any kind of concept. Knowledge can be formed on the basis of a single signal, occurring once during computation, by storing it in an appropriate way. A concept, on the other hand, is a classificatory device that allows the computer to recognize several signal tokens as belonging to a certain type of signal, i.e. it is an ability and, more specifically, a special kind of algorithm for classifying signal tokens.[13] But internal knowledge can be formed without possessing a corresponding concept. For example, Mary might be able to remember a red experience without possessing the concept RED. Moreover, in reply to the ability thesis it was already pointed out that it seems likely and is possible that Mary already possesses the concept RED prior to seeing anything red for the first time. In this point of view particular signals that present a certain sensory input within cognition and corresponding phenomenal concepts are different things, albeit being closely related to each other. Phenomenal concepts are generally not fine-grained enough to account for all cases of internal knowledge formation.

Objection 5: The talk about signals and 'internal knowledge' is too unclear and not well-defined.
This final objection must be granted to some extent. As mentioned earlier, we do not have a viable computational theory of cognition yet. The way internal knowledge has been distinguished from external one in Section 4 and 5 leaves indeed many things to desire. 'Signal', as it has been understood here, is not a very clearcut term either. Signals do not only present sense data but may be much more complex. They may comprise various aspects of the current computation, including implementations of algorithms, instructions for memory access, and so on. However, insofar as only an abstract model of computation is concerned it may be assumed that signals are represented by terms of the untyped λ-calculus and their reductions. In this abstract sense signals are fairly well-defined. Finally, some remark is in place about the term 'internal knowledge.' This name has been chosen instead of 'introspective knowledge' for the following reason. Introspection has a long history in the Brentano-Husserl tradition and is commonly associated with phenomenological methodology. While the formation of internal knowledge

could be understood as involving a process of introspection like it is understood in Computer Science, where 'introspection' is used more or less synonymously with 'reflection' to characterize the ability of a programming language or program to reflect about its own characteristics, implementation, or state, it would be too easy to confuse this term with the equivocal one from the Brentano-Husserl tradition. Likewise, speaking of 'reflexive knowledge' would have created conflicts with existing uses of 'reflexive' in modal logics and mathematics. 'Internal knowledge' ought to be understood as a technical term and not too much should be read into the internal-external distinction. In particular, internal knowledge does not imply any first-person authority over what is known. It is possible that a given computer stores aspects of a signal in a more lossy way than is possible when the signal is measured from the outside and stored in another, more reliable storage device. Correspondingly, another person might very well know more about what is going on in one's own cognition than oneself.

8. Summary and Conclusions

The internal knowledge thesis is based on the idea that in actual cognition thoughts and sensory inputs are signals that are used in actual computations. Particular sorts of signals determine possible future states of the computer and other sorts of signals determine other possible future states of the computer. Trivially, under this standard functional role view the former cannot be substituted by the latter. Signals that present particular sensory inputs or features of an ongoing computation within the ongoing computation have been labeled 'internal' whereas intentional signals representing objects not present to the computation have been labeled 'external.' Externality/internality comes to a degree, because in reality almost all signals are mixed. When knowledge is formed by storing mostly internal signals for later retrieval it is mostly internal knowledge.

From the perspective of this position the Mary example and Lewis's two gods are explained as follows. When Mary learns what it's like to be red the learning itself involves a state change rendering her previous knowledge incomplete. The same happens when Zeus has the insight that he himself* is Zeus. Neither can Zeus know that he just thought I am Zeus without having had the thought in the first place nor can Mary know how it feels like to see something red before having seen something red. Thoughts arising from theoretical knowledge about one's color vision and the external world in general are realized by different signals than thoughts arising from internal knowledge. The two forms of knowledge are trivially irreducible to each other, because they are based on different sorts of signals. Computationalism is ontologically neutral, hence the given explanation of de se attitudes and

phenomenal experience are compatible with any ontological position about the mental including physicalism.

One positive lesson to draw from these considerations is that the epistemic states of agents with internal knowledge must be represented by entities more fine-grained than sets of possible worlds. Many philosophers of language, but perhaps not so many epistemologists, already make this assumption. Regarding the problem of describing a 'logic of cognition' the above considerations seem to favor Neo-Fregean hyperfine-grained accounts, but arguing for a specific one was not the goal of this article and is left for another occasion.

NOTES

1. If this world is not the actual world, the agent has learned one or more false propositions; if the belief set is empty, he has attempted to learn a contradiction but one might want to safeguard against this deviant case.

2. The * marks what Castañeda (1967, 1989) calls a 'quasi-indicator', a non-anaphoric first-person use of the third person pronoun that indicates the occurrence of an irreducible I-thought in thinking with a corresponding de se attitude.

3. In some sense Lewis's example is stronger than it has been described above, since logic does not prevent us from talking about non-physical things. No matter which description we choose the gods always learn 'something more.' If we allow this reading there are ways to cast doubts on Lewis's initial premises. However, for the current purpose it suffices to assume the much weaker interpretation according to which every world is a maximal truth-maker constituted out of atomic physical positive and negative facts—which is not to say that this position of logical atomism is metaphysically unproblematic.

4. See (Siegelmann & Sontag 1994) for a comprehensive overview of hyper-computation.

5. The λ-calculus can be fully encoded by the π-calculus (Milner 1990), but not vice versa, and the same holds for other process calculi. While it is on one hand easy to come up with artificial examples that require parallel computation, for example parallel-OR will terminate even if one of its processes does not terminate whereas a sequential implementation of OR might not terminate in this case, it is on the other hand not at all obvious whether connectionism as a philosophical position requires genuine parallelism. Many actual implementations of neural networks are sequential and not every program running on a parallel computer is parallel in the sense of not being translatable to a functionally-equivalent sequential program of similar complexity.

6. This argumentation scheme, of course, does not show that human internal knowledge formation actually works that way; it only provides us with one way to counter the Knowledge Argument and explain de se puzzles. How humans form a particular sort of knowledge is an empirical question.

7. For stylistic purposes 'computational system', 'system', and in context sometimes also 'machine' will be used synonymously with 'computer' in what follows. Neither of these uses is supposed to imply that the implementation is physical.

8. N.B.: The explanatory gap in question above is not one regarding consciousness in general but only regarding the phenomenal character of experiences. Perhaps similar considerations could be made about consciousness—for example, consciousness could be an intrinsic property of some computations but we might never be able to find that out by an analysis of these computations only—, but this topic is left for another occasion.

9. Langer (1951) has made a very similar distinction between discursive and presentational symbols in the different context of developing a cultural semiotics, and the present view is inspired by her work. Similar distinctions can also be found throughout Castañeda's work.

10. 'Assertoric mode of thinking' is a metaphor. What is meant is that the agent must believe that the thought is true, as opposed to only considering it hypothetically.

11. Just to make this clear: The object could be imaginary or an optical illusion, the result of a damage in the brain or someone stimulating parts of it, of some telepathic influence on the mind, and so on. It is irrelevant in the present context whether the signal has the cause it is commonly thought to have. Intimate knowledge can also be formed when there is no corresponding red object.

12. Bringing divine entities into play obviously poses problems regarding the interplay with well-understood logical possibility. But of course the same argument could be made without recurring to God. What is metaphysically possible could even be a subset of what is physically possible.

13. This view is essentially that of Tichý (1988) and Moschovakis (1994, 2006).

REFERENCES

Bealer, G., and Mönnich, U. (1989), "Property Theory," in Dov Gabbay (ed.), *Handbook of Philosophical Logic*. Dordrecht: Kluwer, 133–251.

Bigelow, J., and Pargetter, R. (1990), "Acquaintance with Qualia", *Theoria* 56: 129–147.

Bigelow, J., and Pargetter, R. (2006), "Re-acquaintance with Qualia", *Australasian Journal of Philosophy* 84: 353–378.

Castañeda, H.-N. (1967), "Indicators and Quasi-Indicators," *American Philosophical Quarterly* 4: 85–100.

Castañeda, H.-N. (1975), *Thinking and Doing*. Dordrecht: D. Reidel.

Castañeda, H.-N. (1989), *Thinking, Language, Experience*. Minneapolis, MN: University of Minnesota Press.

Churchland, P. (1985), "Reduction, Qualia and the Direct Introspection of Brain States," *Journal of Philosophy* 82: 8–28.

Churchland, P. (1989), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA: MIT Press.

Coleman, S. (2009), "Why the Ability Hypothesis is Best Forgotten," *Journal of Consciousness Studies* 16 (2/3): 74–97.

Conee, E. (1994), "Phenomenal Knowledge," *Australasian Journal of Philosophy* 72: 136–50.

Cresswell, M. J. (1985), *Structured Meanings*. Cambridge, MA: MIT Press.

Dennett, D. C. (1991), "Epiphenomenal Qualia?," in *Consciousness Explained*. Boston, MA: Little, Brown and Company.

Dennett, D. C. (2005), *Sweet Dreams*. Cambridge, MA: MIT Press.

Hewitt, C., Bishop, P., and Steiger, R. (1973), *A Universal Modular Actor Formalism for Artificial Intelligence*. IJCAI.

Hintikka, J. (1975), "Impossible Possible Worlds Vindicated," *Journal of Philosophical Logic* 4(3): 475–484.

Jackson, F. (1982), "Epiphenomenal Qualia," *The Philosophical Quarterly* 32 (127): 127–136.

Jackson, F. (1986), "What Mary Didn't Know," *The Journal of Philosophy* 83(5): 291–295.

Lokhorst, G.-J. C. (2000), "Why I am Not a Super Turing Machine," Manuscript of a talk given at University College London on 24. May 2000, URL http://homepages.ipact.nl/~lokhorst/hypercomputationUCL.pdf.

Langer, S. (1951), *Philosophy in a New Key*. Cambridge, MA: Harvard University Press.

Lewis, D. K. (1979), "Attitudes De Dicto and De Se," *Philosophical Review* 88(4): 513–543.

Lewis, D. K. (1983), "Postscript to 'Mad Pain and Martian Pain," in David Lewis (ed.), *Philosophical Papers*, Vol. I. New York: Oxford University Press, 130–32.

Lewis, D. K. (1988), "What Experience Teaches," in *Proceedings of the Russellian Society*. Sidney: University of Sidney.

Lewis, D. K. (1990), "What Experience Teaches," in William Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 499–518. (reprint of Lewis 1988)

Milner, R. (1990), "Functions as Processes," in Michael Paterson (ed.), *Automata, Languages and Programming*, Lecture Notes in Computer Science, Vol. 443. Berlin-Heidelberg: Springer, 167–180.

Milner, R., Parrow, J., and Walker, D. (1992), "A Calculus of Mobile Processes Pt. 1," *Information and Computation* 100(1): 1–40.

Moschovakis, Y. N. (2006), "A Logical Calculus of Meaning and Synonymy," *Linguistics and Philosophy* 29: 27–89.

Moschovakis, Y. N. (1994), "Sense and Denotation as Algorithm and Value," in Väänänen, J. and Oikkonen, J. (eds.), *Logic Colloquium '90* (Association for Symbolic Logic) 2: 210–249.

Nemirow, L. (1980), "Review of *Mortal Questions*, by Thomas Nagel," *Philosophical Review* 89: 473–77.

Nemirow, L. (1990), "Physicalism and the Cognitive Role of Acquaintance," in William Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 490–99.

Nida-Rümelin, Martine (2009), "Qualia: The Knowledge Argument," in Zalta, Edward (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/qualia-knowledge/ (version of 2010-02-12).

Penrose, R. (1989), *The Emperor's New Mind*. Oxford-New York: Oxford University Press.

Penrose, R. (1994), *Shadows of the Mind*. Oxford-New York: Oxford University Press.

Perry, J. (1977), "Frege on Demonstratives," *Philosophical Review* 86: 474–497.

Perry, J. (1979), "The Problem of the Essential Indexical," *Noûs* 13: 3–21.

Priest, Graham (2005): *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford: Clarendon.

Quine, W. V. O. (1956), "Quantifiers and Propositional Attitudes," *The Journal of Philosophy* LIII (5: March): 177–187.

Richard, M. (1983), "Direct Reference and Ascriptions of Belief," *Journal of Philosophical Logic* 12: 425–452.

Siegelmann, H. T., and E. D. Sontag (1994), "Analog Computation via Neural Networks," *Theoretical Computer Science* 131: 331–360.

Stalnaker, R. (2004), "Knowing Where We Are, And What It Is Like," manuscript of a talk given at NYU's La Pietra Conference on Consciousness, Florence 2004, URL http://www.nyu.edu/gsas/dept/philo/ faculty/block/lapietra/Stalnaker.pdf.

Stalnaker, R. (2008), *Our Knowledge of the Internal World*. Oxford-New York: Oxford University Press.

Stalnaker, R. (2011), "The Essential Contextual," in Brown, J. and Cappelen H. (eds.), *Assertion: New Philosophical Essays*. Oxford: Oxford University Press, 137–151.

Thomason, R. (1980), "A Model Theory for Propositional Attitudes," *Linguistics and Philosophy* 4: 47–70.

Tichý, P. (1988), *The Foundations of Frege's Logic*. Berlin-New York: De Gruyter.

Tye, M. (2009), *Consciousness Revisited*. Cambridge, MA: MIT Press.

von Stechow, A. (1982), "Structured Propositions," technical report of the SFB 99, Universität Konstanz.

Analysis and Metaphysics
Volume 11, 2012, pp. 133–160, ISSN 1584-8574

THE *WHEEL OF SAMSARA* AS DESCRIPTIVE DYSFUNCTIONAL ORGANIZATIONAL TYPOLOGIES

MURRAY HUNTER
murrayhunter58@gmail.com
University Malaysia Perlis

ABSTRACT. Organization theory has been dominated by occidental paradigms with *'Asian'* philosophies making minimal contribution. This article descriptively explains the phases and realms within the *Wheel of Samsara* from Buddhist Dharma. These phases and realms are reframed and represented as a descriptive model of dysfunctional organizational typologies. This article is an initial step in merging oriental philosophy into organizational theory.

Keywords: Buddhism, Dharma, emotions, organization, organizational development, Oriental philosophy

1. Introduction

Unlike psychology, organizational theory is dominated by the occidental paradigm with minimal oriental thought and philosophical influence. Even the rush to understand Japanese management in the 1980s took an instrumental and positivist viewpoint, rather than a cultural and philosophical perspective, which explained the phenomenon without context. Contemporary writing on Asian organization and management is focused on the marketing, strategic, and socio-political factors, providing readers with an action orientation, without delving into too much philosophy from the region. Although we know the *how*, *where*, and *why* of Asian business, we have not gained very much new input or insights of new knowledge into the metaphorical sea of management philosophy. Even the rising numbers of Asian management academics residing in universities within the west take an occidental viewpoint, while Asian academics within the region import management theory rather than tap into the vast array of local philosophy that has potential applicability to organization and management.