Belief in The Singularity is Fideistic

Selmer Bringsjord[1] • Alexander Bringsjord[2] • Paul Bello[3]

[1]Department of Computer Science
Department of Cognitive Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
brings623@gmail.com

[2]Motalen Inc.
Troy NY 12180 USA

[3]Human and Bioengineered Systems, Code 341
Office of Naval Research
875 N Randolph St
Arlington VA 22203 USA
paul.bello@navy.mil

Contact: Selmer.Bringsjord@gmail.com

January 25B 2012

Abstract

We deploy a framework for classifying the bases for belief in a category of events marked by being at once *weighty, unseen, and temporally removed* (*wutr*, for short). While the primary source of wutr events in Occidental philosophy is the list of miracle claims of credal Christianity, we apply the framework to belief in The Singularity, surely — whether or not religious in nature — a wutr event. We conclude from this application, and the failure of fit with both rationalist and empiricist argument schemas in support of this belief, not that The Singularity won't come to pass, but rather that regardless of what the future holds, believers in the "machine intelligence explosion" are simply fideists. While it's true that fideists have been taken seriously in the realm of religion (e.g., Kierkegaard in the case of some quarters of Christendom), even in that domain the likes of orthodox believers like Descartes, Pascal, Leibniz, and Paley find fideism to be little more than wishful, irrational thinking — and at any rate it's rather doubtful that fideists should be taken seriously in the realm of science and engineering.

1   Introduction; Plan

We deploy a framework for classifying the bases for belief in a category of events marked by being at once *weighty, unseen, and temporally removed* (= *wutr*). While the primary source in Occidental philosophy of such events is credal (= orthodox) Christianity,[1] we follow Dennett (2007) in viewing philosophizing as equally applicable to religion and science, and apply this framework to the dominant basis ($\mathcal{A}$) for belief in The Singularity, surely — whether or not itself religious in nature — a wutr event. We conclude from this application not that The Singularity won't come to pass, but rather that regardless of what the future holds, the failure of a fit between $\mathcal{A}$ and either rationalist or empiricist argument schemas in support of this belief implies that believers in the "machine intelligence explosion" are simply fideists. While it's true that fideists have been taken seriously in the realm of religion (e.g., Kierkegaard 1986 in the case of some quarters of Christendom), even in that domain the likes of believers like Descartes, Pascal, Leibniz, and Paley,[2] in line as they are with Christian orthodoxy and hence rationalism, find fideism to be little more than wishful, irrational thinking — and at any rate it's rather doubtful that fideists should be taken seriously in the realm of science and engineering.

1.1   Preliminaries

To make the situation a bit more tidy before we begin in earnest, we take a series of preliminary steps.

First, we acknowledge the initial oddness of speaking of belief *in an event*. Traditionally, of course, the targets of belief (and knowledge) are propositions — though we certainly do say such things as that "Knox believes in General Washington," in which case we are pointing to belief in a *person*. The situation before us is easily and quickly made cleaner: When we say that some believe in The Singularity, where this is an event, we simply mean

---

[1]Which shouldn't be confused with the denomination known as 'Greek Orthodox' — a denomination that does though happen to itself be orthodox/credal in our sense. An elegant characterization of orthodox Christianity is provided by Chesterton (2009). Along the same lines, and no doubt paying homage to his intellectual and spiritual hero, is Lewis's (1960) *Mere Christianity*. A more mechanical and modern characterization is obtained by simply following Swinburne (1981) in identifying orthodox Christianity with the union of the propositional claims in its ancient creeds (e.g., Apostle's, Nicene, Athanasian), which then declaratively speaking within this limited scope harmonizes Catholicism and Protestantism.

[2]And — see note 1 — Chesterton, Lewis, and Swinburne.

that some believe that The Singularity will *occur*. In a parallel that will form a persistent theme in the inquiry herein, when someone says that they believe in The Resurrection, they are reporting their belief that the event in question (Jesus rising from the dead) happened. Note that belief in an event is thus paired with belief that a certain proposition is true. The same kind of association is in play in the case of belief in a person, since Knox clearly believes such things as that Washington is competent.[3]

In a second preliminary step, note that the properties *being weighty*, *being unseen*, and *being temporally removed* are here applied to *events*; and we invoke the already-seen abbreviation of this three-part adjective: *wutr*. We assume that the property of being temporally removed is clear enough to obviate any sustained analysis. This property applies to an event if it's purported occurrence is either beyond the recent past or immediate future. Hence, the aforementioned Resurrection is temporally removed. So is WWI, the American Revolutionary War, the death of Adolf Hitler, the falling of a vast part of California into the Pacific Ocean due to a major earthquake, the Second Coming, and the arrival in 2020 of aliens superior in intelligence to most currently alive Norwegians. Clearly, The Singularity is temporally removed. We devote section 2 to characterizing the first two properties in 'wutr.'

In a third preliminary move, we denote by $\mathcal{S}$ the event associated with The Singularity (the arrival on Earth of computing machines more intelligent, indeed *vastly* more intelligent, than human persons,[4]) and denote by **S** the corresponding *proposition* that this event will in the near future come to pass. By 'near future' we mean to encompass any length of time short of a century; hence we charitably adopt a temporally latitudinarian stance with respect to those confident that $\mathcal{S}$ will occur. On this stance, we are of course allowing much more time than any reasonable interpretation of 'foreseeable future,' and this is a phrase often used to frame predictions that advanced computing-machine intelligence will or will not arrive. For instance, Turing (1950) famously declared that he could foresee a time when humans not only routinely ascribed intelligence and other mental attributes to computers,

---

[3]We are happy to agree that believing in a person includes more than mere propositional belief, but this topic isn't germane to our objectives herein.

[4]We recognize that The Singularity has now come to be associated with a *group* of events (e.g., the group often is taken to include the ability of human persons to exist in forms that are not bio-embodied), but to maintain a reasonable scope in the present paper we identify $\mathcal{S}$ with only the "smart-machine" prediction, which is quite in line with e.g. the sub-title of the highly influential (Kurzweil 2000): "When Computers Exceed Human Intelligence." There is also in alignment with the *locus classicus*: (Good 1965).

but also when his test (the so-called 'Turing Test') would be passed; indeed he specifically predicted that by 2000 a level of such intelligence on par, linguistically speaking, with that possessed by humans would arrive.[5] This prediction turned out to fall completely flat, as we all know by now.[6] In a second example, this one falling on the side of pessimism, Floridi (2005) has argued that a certain ingenious test of self-consciousness cannot in the *foreseeable future* be passed by a computing machine.[7] Finally, note that Chalmers (2010) recounts a number of the time-indexed predictions about when $\mathcal{S}$ will supposedly occur; our allowing a full century is in this context hyper-charitable. As confirmation of this, consider: Good (1965): $\mathcal{S}$ by 2000; Vinge (1993): the explosion between 2005–2030; Yudkowsky (1996): 2021; and Kurzweil (2000): 2030. On the other hand, Chalmers himself appears to believe that $\mathcal{S}$ will occur within *centuries* (note the plural). Since we wish to retain the concept of a foreseeable future, this is too large a range for us to use herein.

1.2 Plan for the Remainder

With these preliminary matters settled, we announce our planned sequence for the remainder: In the next section (§2), our exposition aided by consideration of the claims of credal Christianity,[8] we briefly characterize the category of the weighty and unseen, into which, as will be seen, surely The Singularity falls. Then (§3) we briefly summarize the three main epistemic positions of *empiricism*, *rationalism*, and *fideism*. These positions are sketched with help from the basic but eminently sensible and non-partisan epistemological framework erected by Chisholm (1977). Next (§4), we present our proof-by-cases argument for the claim that belief in **S** is fideistic — an argument that

---

[5]We recognize that Turing's optimism was constrained by certain conditions regarding how long a computing machine's prowess on his test would last, but such niceties can be safely left aside.

[6]As a matter of fact, Turing, like — as we shall see — those predicting the coming $\mathcal{S}$, would seem to be guilty of the same fatal sin: failing to give a rationalist (or even an empiricist) argument for the prediction in question. One of us rather long ago happily conceded that the Turing Test will be passed (Bringsjord 1992), but this concession was not accompanied by any timeline whatsoever — and if there *had* been a timeline, it would have been an exceedingly conservative one.

[7]A counter-argument can be found in (Bringsjord 2010).

[8]This is as good a spot as any to say that we could mine the supernatural event-claims of Islam and Judaism instead of those in credal Christianity, but we aren't that familiar with these other two monotheistic religions, and Western philosophy, for better or for worse, has certainly focused on the event-claims of Christianity of the other two historical monotheistic religions.

will in turn require at least some study of the dominant basis for believing that **S** holds; that is, some study of the aforementioned $\mathcal{A}$, a basis due originally to (1965), and ably modernized recently by Chalmers (2010). A brief conclusion wraps up the paper.

2   The World of the Weighty, Unseen, and Temporally Removed

We've already commented on the property *temporally removed*. What is meant by 'w' and by 'u' in the composite adjective 'wutr'? We haven't the space here to give a rigorous definition, and such a thing isn't needed anyway, because illuminating examples abound in philosophy, especially in the philosophy of (again, Occidental) religion, which typically relates to such things as whether God exists, and whether he really has intervened, and will intervene, directly in our world. Philosophy of religion typically targets those things which are in turn the targets of faith, and as such, things which are at once weighty and unseen. Here is the writer of Hebrews (11:1) in the New Testament: "Now faith is being sure of what we hope for and certain of what we do not see." The context of this passage indicates that what is believed in faith targets things both weighty enough to be earnestly hoped for, and invisible — things that, in short, are miraculous. As philosophical treatments of miracles indicate, miracles are by definition weighty and (to nearly all, anyway) unseen. For example, as is noted by McGrew (2010), we would hardly count as a miracle, or even a purported miracle, some stray, minor deviation from physical laws in a remote corner of the inanimate universe.

Likewise, the context of sustained historic treatments of faith and reason, such as Leibniz's (1998) *Theodicy*, point to events both weighty and unseen; namely, what Leibniz calls the "oracles" of God; that is, the "major" miracles claimed by the creeds of Christianity. These events are paradigmatic examples of profundity and (at least from the perspective of generations living long after the times at which they are to have occurred) invisibility. And the same source of an ostensive definition of w-&-u is found in contemporary treatments, for instance in the work of Oxford philosopher Richard Swinburne (1981, 2010) — work explicitly devoted to substantiating the credal claims of orthodox Christianity. So, these are the examples we provide to clarify the weighty-and-unseen: the Resurrection, the virgin birth, and so on. We are in no way saying that The Singularity is supernatural in nature; we are saying that in *structure*, and specifically with respect to wutr, The

Singularity ($= \mathcal{S}$) parallels the — to again use the Leibnizian term — oracles of Christianity.

Put in terms of propositions (i.e., the underlying content of declarative sentences traditionally signaled in English by "that" phrases), and generalizing to some degree, we can say that *propositions* are weighty-and-unseen when they directly and immediately entail the existence of some being(s), and/or the occurrence of some event, which is at once by nearly any metric such that were it to obtain, or were it known to be arriving in the future, (i) would cause rational agents to significantly alter their beliefs and their behavior, and (ii) involves beings as of yet invisible. The proposition **S** that The Singularity will come to pass within a century certainly seems to qualify as w-&-u with flying colors, for this proposition makes reference to a profound event, and to a being or beings (immeasurably smarter-than-human computing machines) that are invisible as of now, and perhaps invisible even after they arrive on the scene. So condition (ii) is satisfied. What about condition (i)? Anyone who knows even a smidgeon of the literature on The Singularity knows those who expect it often adjust their "cognitive maps." They consider for instance how best to prepare for and perhaps to a degree manage $\mathcal{S}$. Chalmers (2010) is an example of such level-headed cerebration.

3   The Tripartite Framework

Now, what is the framework we have available? By our lights, the basis for believing some wutr proposition P conforms to only one of three normative views: namely, *rationalism*, *empiricism*, or *fideism*. In order to flesh out these bases, we turn to a discrete continuum of epistemic "strength" provided by Chisholm (1977). There are of course any number of ways to unpack the trio, but it's safe to say that Chisholm's scheme is eminently reasonable, and that the result that we obtain (belief in **S** is fideistic) would be generated by *any* epistemologically sensible unpacking of the three concepts in question.[9]

Chisholm's spectrum of the strength of a proposition for a rational human mind is a nine-point one, and ranges from 'certainly false' to 'certain.' At the halfway point are propositions said to be *counterbalanced*. There are

---

[9]For example, our conclusion about believers in The Singularity would be obtained by turning instead to (Pollock 1974). This is as good a place as any to mention that both Chisholm's scheme, and Pollock's, are "computing-machine friendly." One of us has made use of Chisholm's strength-factor scheme to ground software for engineering argumentation; see (Bringsjord, Taylor, Shilliday, Clark & Arkoudas 2008). And Pollock himself built an artificial agent on the basis of his epistemology; see for example (Pollock 1989, Pollock 1995, **?**).

then four positive strength factors working up from there: first *probable*, then *beyond reasonable doubt*, then *evident*, and finally *certain*. Certain propositions include the indubitable truths of formal logic (e.g., *modus ponens*, $1 = 1$, etc.), and presumably "Cartesian" truths such as "I exist," and "It seems to me that I'm sad." What kind of thing is evident? For the most part, the evident would be populated by those propositions we affirm on the strength of sense perception. For example, that there is a computer screen in front of you when you are typing out a sentence such as the present one is evident. This proposition isn't certain: you might be hallucinating, after all; but it's — as we might say — *close* to certain. You wouldn't want to say, for example, while spying a coffee cup in front of you, in perfect health and having not ingested any mind-altering drugs, that the proposition that there's a cup in front of you is merely beyond reasonable doubt: you want to say, instead, that you are well within your epistemic "rights" in holding that it's *extremely* likely that there's a cup before you. This, again, is the category of the evident.

But moving down another Chisholmian notch in strength, we do in fact hit *beyond reasonable doubt* — which of course famously coincides roughly with what it takes in the United States to legally convict someone of murder. That is, to convict someone of this kind of crime the evidence must make some such proposition as Jones is guilty beyond reasonable doubt. Finally, note that to convict on this standard, it's not sufficient to know that it's merely *probable* that Jones did it. Some proposition P being probable is the last notch before we reach *counterbalanced*, which of course means that a purely rational agent wouldn't bet in favor of P, and wouldn't bet against it. A perfectly rational agent who is agnostic about some proposition P would regard P to be counterbalanced.

What about the "negative" side of Chisholm's continuum? Since neither the empiricist not the rationalist, if abiding by their respective programs for belief fixation, would assent to propositions on negative side of *counterbalanced*, we have no need here to explore this epistemic terrain. Of course, all bets are off when it comes to the fideist. Kierkegaard even went so far as to recommend embracing the logically incoherent; that is, to recommend embracing certain propositions that are, viewed intellectually, certainly false (such as that Abraham was obligated to refrain from killing Isaac, and obligated to kill him). But we have no need to discuss the four notches of strength on the negative side in any detail.

Armed with Chisholm's spectrum, we can now offer encapsulation of the three main standards for belief to be applied to belief in **S**:

- **Rationalism**: The view that belief in a wutr proposition P must be supported by deductive proofs or arguments, where the inferences in this reasoning are each formally valid, and the premises are at least probable.

- **Empiricism**: The view that belief in a wutr proposition P must be supported by direct, neurobiologically normal sense perception of the constituents (i.e., of the being or event in question) of the propositions in question (making P, as noted above, evident), perhaps augmented from there by *some* formally valid deductive proof or argument.

- **Fideism**: The view that one ought to believe a wutr proposition P despite having little or no evidence for P (i.e., put in terms of arguments, every argument for P has at least one proposition at or below the level of counterbalanced).

Each of these doctrines are partitioned in our comprehensive breakdown into at least a *strong*, *moderate*, and *weak* sub-forms. This more fine-grained breakdown is beyond our needs in the present essay, but we do need hear a significant portion of the breakdown for rationalism (for reasons that will soon become clear). Accordingly, we note here that *strong rationalism* is the view (and as it happens, *our* view) that any human person believing some wutr P ought to have on hand at least one outright proof of P; that is, have on hand a formally valid chain of deductive inference originating from premises that are each certain.[10] The doctrine of *moderate rationalism* holds that if Jones abides by this doctrine and believes P, then Jones must have on hand at least one formally valid argument for P whose premises $P_1, P_2, \ldots, P_n$ are each at least evident, where each P_i is evident. And following suit we can say that *weak rationalism* requires only that the premises involved in deductive reasoning for the wutr P in question are at least probable. Readers will no doubt get the driving idea from the foregoing; the story would continue on, all the way through not only a more fine-grained ontology of rationalism,[11] but empiricism and fideism. In the case of the latter,

---

[10]Some readers will inevitably ask: "Is there any such thing?!" We are of course well aware of the fact that even some axioms in some axiomatic set theories are controversial, and hence perhaps not certain. (Even the power set axiom in ZFC has its detractors.) Nonetheless, whatever one can deduce in deductively valid fashion from, say, $1 = 1$, would be certain, and one would be well-advised to believe such a consequence. For instance, $1 = 1 \lor Q$, for any proposition Q, would be an acceptable disjunction for even a strong rationalist to believe.

[11]For example, we could distinguish between the strength of inferential links in the argument for wutr P.

the "bravest" fideists are those who believe self-contradictory propositions; Kierkegaard, as noted above, is known for commending the absurd, or certainly false, for assent. On the other hand, the most "timid" fideists would be those who believe a wutr P despite the fact that one or more premises are counter-balanced. In this case, under the "weakest-link principle," there is still wishful thinking.[12]

It's important to note that the above R-E-F framework is erected under the assumption that the human beings we are talking about are neurobiologically normal (and indeed alert readers will have noticed that we employed this condition in our definition of empiricist belief) and have had sufficient nurturing and training to be able to reason at the level of first-order logic. This assumption does idealize the situation to some degree, but we have known since the experiments of Piaget and colleagues that such human beings are certainly among us (e.g., see Inhelder & Piaget 1958), and indeed you no doubt are one of them.[13]

We conclude this section by pointing out that a nice testbed for understanding and contrasting the three different schemes for belief in wutr events and propositions can be found in the case of the credal Christian miracles. McGrew (2010) provides a thorough, readable discussion of the various forms of argument in favor of the veridicality of the credal miracles, some of which (in connection, e.g., with the Resurrection) are rationalist (e.g. Paley 2010), and some of which are empiricist (e.g. Habermas 1984). In addition, Swinburne (2010) has recently provided a formidable empiricist argument for the miracle of the Incarnation.

4   Belief in The Singularity is Fideistic

We now articulate and defend our claim that belief in The Singularity is fideistic, and hence that such belief, while perhaps acceptable in the realm of religion, is not acceptable in the realm of science, where rationalism and empiricism together reign justifiably supreme. The basic line of reasoning

---

[12]Barbarically put, the principle states that an argument for Q is only as strong, overall, as the weakest inferential link in that argument. We leave aside the fascinating subject of fideism "forced" by decision-theoretic considerations. One who for example agrees with Pascal's Wager may decide to believe even if the best propositional evidence is counter-balanced, just because the potential disutility of not believing is infinitely large.

[13]That there are such humans in no way is inconsistent with results (e.g., those produced by the ingenious experimentation of Johnson-Laird 2000) showing that most humans fail to reason at the level of FOL. For additional evidence that some people are pretty darn good at deductive reasoning that coincides with FOL, see (Rips 1994).

in the argument is quickly and easily stated: We examine the main line of serious argument in support of **S**, and observe that by rationalist and empiricist standards this reasoning fails to fall under either umbrella. By disjunctive syllogism, the proponent of **S** is a fideist.

Without further ado, then, what is the argument? It's the one alluded to above, first given by Good (1965), and polished considerably by Chalmers (2010). The kernel of the argument, expressed in prose:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make. (Good 1965)

Chalmers reasonably takes Good to be here arguing for the second premise, that is, (P2), in the following overarching argument ($\mathcal{A}$). In this argument, 'HI' is human intelligence, 'AI' is artificial intelligence at the level of human persons, 'AI$^+$' is artifical intelligence above the level of human persons, and 'AI^{++}' refers to super-intelligence constitutive of $\mathcal{S}$. Note that we have labeled the conclusion in line with previously introduced notation.

$\mathcal{A}$:
(P1)   There will be AI (created by HI).
(P2)   If there is AI, there will be AI$^+$ (created by AI).
(P3)   If there is AI$^+$, there will be AI^{++} (created by AI$^+$).
∴   **S**   There will be AI^{++} ($= \mathcal{S}$ will occur).

Of course, $\mathcal{A}$ is deductive in form, and formally valid. Unfortunately, that's about where the good news ends for the proponent of The Singularity. To see this, we reason as follows. In order for belief that **S** to qualify as rationalist, the premises in question must be in Chisholm's continuum either probable, beyond reasonable doubt, certain, or evident. There can be no denying that (P1) isn't certain; in fact, all of us can be quite certain that (P1) isn't certain. Our certainty in the lack of certainty here can be established by showing, formally, that the denial of (P1) is consistent, since if not-(P1) is consistent, it follows that (P1) doesn't follow from any of the axioms of classical logic and mathematics (for example, from a standard axiomatic set theory, such as ZF). How then do we show that not-(P1) is consistent? We derive it from a set of premises which are themselves consistent. To do this, suppose that human persons are information-processing machines more powerful than standard Turing machines, for instance the

infinite-time Turing machines specified and explored by Hamkins & Lewis (2000), that AI (as referred to in $\mathcal{A}$) is based on standard Turing-level information processing, and that the process of creating the artificial intelligent machines is itself at the level of Turing-computable functions. Under these jointly consistent mathematical suppositions, it can be easily proved that AI can *never* reach the level of human persons (and motivated readers with a modicum of understanding of the mathematics of computer science are encouraged to carry out the proof). So, we know that (P1) isn't certain.

But as a matter of fact the reasoning we have just summarized suffices to show that (P1), and for that matter (P2) and (P3) as well, cannot be classified as *beyond reasonable doubt* or *evident*. Why? The answer is straightforward, and water-tight: It's not beyond reasonable doubt that those who hold that the human mind processes information in a manner above the Turing Limit are wrong. This point is made for example in the brief (Bringsjord & van Heuveln 2003), and made again in the sustained, book-length (Bringsjord & Zenzen 2003).

But there are also out-of-the-armchair reasons why (P1) isn't evident. Recall that we said evident propositions are typically those recommended by direct sense perception. But what is it that we perceive which provides reason to believe that human-level machine intelligence is coming, on the strength of human engineering? The answer is: "Nothing." For the fact of the matter is that a sharp toddler of today makes a mockery of any computing machine with designs on natural-language communication. And even if we leave natural-language communication out of the picture, and refer instead to human-level problem solving specifically in areas that would seem to be positively ideal for computing machines, we perceive not the steady advance of computing machines, but their paralysis when stacked against the capability of humans. For example, consider automatic programming, which is one of the original dreams of AI. Today, in 2012, we quite literally have no computing machines that can, having been supplied with a standard mathematical specification of an arbitrary number-theoretic function f (from the natural numbers to the natural numbers), supply as output a new computing machine that computes f — even when the input functions are as simple as those given to students in introductory programming classes!

We come then to the last possible escape from fideism available, at least in principle, to the believer in $\mathcal{S}$: weak rationalism, in the specific instantiation of this doctrine consisting in the claim that (P1)–(P3) are merely each probable. This move makes for an epistemic humility that we haven't seen among those proclaiming the arrival of superintelligent machines. Nonetheless, the point is that it seems to be a move available to the believer in $\mathcal{S}$.

In addition, there is no denying that while in philosophy of religion the vast majority of cases made for the propositions central to the Christian brand of monotheism accept the burden of strong or at least moderate rationalism, there are instances of *weak* rationalism. Swinburne (1991), for example, argues only that the existence of God is more probable than not.

So, is the trio (P1)–(P3) probable? We don't think so. AI and the computational conception of mind, following Glymour (1992), can be said to have begun over two millennia ago with Aristotle's knowledge representation and reasoning frameworks; and yet, again, here we are, with hardware that moves information in silicon at a rate that makes the transmission speed of the brain seem as slow as a disoriented caterpillar by comparison, and we still don't have a machine that can problem-solve, even in highly formal domains like computer programming, at the level of a mediocre novice. It seems to us that at this point it's looking highly *un*likely that HI will produce AI, and moreover we have no reason to think that AI would be able to produce AI^+ at any rate. We concede that this isn't much of an argument. Is there any more principled philosophical reason for holding that one or more of the trio are less than probable, and hence that believing in The Singularity is to slip into fideism? Yes.

We give an argument based, first, on the observation that $\mathcal{A}$ is itself based on the concept of ever-increasing intelligence. More specifically, we note that it follows deductively from the trio in question that, where $L(M)$ yields the level of intelligence of a machine (or class of machines; we assume for the sake of argument that humans are bio/carbon-based computing machines), $L(\text{HI}) < L(\text{AI}^+)$. In fact, it follows deductively from the three propositions in question, and the defensive move that we have invoked on behalf of the proponent of **S** seeking to avoid falling into fideism, that

> **(P4)** Proponents of the case $\mathcal{A}$ for **S** at this stage in the present dialectic know that it's probable that machines AI^+ will arrive such that $L(\text{HI}) < L(\text{AI}^+)$.

But if the proponents of the case in question know this, then surely they must know what the difference in intelligence between HI and AI^+ consists in. If they don't know what the difference consists in, then they aren't within their epistemic rights in asserting (P4). In fact, in that state of ignorance, asserting (P4) certainly has the look and feel of the core spirit of fideism, which is to forge ahead and believe, in the absence of the normal prerequisites. Do those who believe in The Singularity understand what the difference in question is? Apparently not. We have scoured the writings of pro-$\mathcal{S}$ thinkers for even an atom of an account of the difference, and have come up utterly empty.

In fact, these writings, merged, yield self-refutation. For example, Chalmers (2010) understands that mere processing speed of hardware, ever increasing in conformity with Moore's Law, *contra* Kurzweil (2000), is insufficient to support the claim that super-human machine intelligence will probably arrive. As Chalmers notes, speed is one thing; that which is computed *by* that speed is quite another. What the proponent of $\mathcal{S}$ thus needs to supply in order to dodge the descent into fideism is the difference between HI and AI^+ *cashed out in a differential between the respective functions computable by each class of machine.* These details cannot be found in the literature — anywhere. Of course, the proponent of The Singularity could retort that ultraintelligent computing machines have super-human intelligence because, for example, they can: play better chess than any humans, or push further into complexity-intractable spaces (e.g., can solve in a reasonable amount of time more of the space of problems in the general propositional satisfiability problem) than humans have managed, or out-score any human on general tests of intelligence, and so on.[14]

Unfortunately, there are two fatal problems with this response. First, the response runs afoul of Chalmers' observation that speed in and of itself is in the end nothing worth writing home about. We've known for a long time, for example, that we have an algorithm for playing perfect chess. So if outerspace aliens landed tomorrow and proudly proclaimed that they can play invincible chess via this algorithm, because they have faster hardware than ours (perhaps implanted in their bodies), we really wouldn't be that impressed — and in analogy we wouldn't be inclined to say that if these aliens are just computing machines, ultraintelligent machines had arrived on our planet. The second fatal problem is that the current upward march of AI research is gradually producing precisely the sort of machines touted by

---

[14]In our experience, the concept of intelligence as it's used in communication between those believing in $\mathcal{S}$ comes at least close to be conflated with the concept of *power*, or more precisely, *information-acquisition* power, conjoined with processing speed a la Moore's Law. Once this conflation occurs, the notion that machines of the future will be ultraintelligent quickly arrives on scene. Why? The point can be put in sci-fi terms: We imagine a *Terminator 3*-like event in which unmanned machines hooked into all digital information on the planet suddenly break through any and all privacy restrictions on use of this data, and proceed to exploit it. These machines are now able to do things that are unprecedentedly "intelligent." For example, the machines may now be able to prevent human crimes before they happen. (E.g., machines with access to everyone's email, and the processing power to check them for plans of foul play, could thwart criminals.) Needless to say, while this notion of information-theoretic super-intelligence is coherent, and may in fact even be likely to materialize, no fundamentally new functionality is in play, and hence, while in our interaction with believers in The Singularity we witness the conflation in question, the case for **S** isn't insulated from our counter-argumentation.

the believer in $\mathcal{S}$ in the rejoinder under review at the moment; but these machines aren't in any way regarded to be a quantum leap beyond HI. We expect that soon enough computing machines will be able to process in real time all data relevant to the coördinated automated driving of every vehicle on our planet. Do these machines deserve to be called 'ultraintelligent'? No. They are just fast processors; their core functionality is rather trivial.

5   Conclusion

We conclude, then, that proponents of **S** are indeed fideists. This in no way implies that **S** is false. We have friends, and suspect you do as well, who assert the wutr propositions of this and that religion in the absence of a rationalist or empiricist basis for such assertions, and we wisely resist declaring that therefore they have put their faith in falsehoods. The most that can be said (unless of course disproofs of the propositions in question are on hand) is that faith of the fideistic sort is certainly in operation, and this is so whether or not the targets of that faith are real. So it is as well for the believer in The Singularity at the close of the investigation carried out herein.

References

Bringsjord, S. (1992), *What Robots Can and Can't Be*, Kluwer, Dordrecht, The Netherlands.

Bringsjord, S. (2010), 'Meeting Floridi's Challenge to Artificial Intelligence from the Knowledge-Game Test for Self-Consciousness', *Metaphilosophy* **41**(3), 292–312.
URL: *http://kryten.mm.rpi.edu/sb_on_floridi_offprint.pdf*

Bringsjord, S., Taylor, J., Shilliday, A., Clark, M. & Arkoudas, K. (2008), Slate: An Argument-Centered Intelligent Assistant to Human Reasoners, *in* F. Grasso, N. Green, R. Kibble & C. Reed, eds, 'Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8)', Patras, Greece, pp. 1–10.
URL: *http://kryten.mm.rpi.edu/Bringsjord_etal_Slate_cmna_crc_061708.pdf*

Bringsjord, S. & van Heuveln, B. (2003), 'The mental eye defense of an infinitized version of Yablo's paradox', *Analysis* **63**(1), 61–70.

Bringsjord, S. & Zenzen, M. (2003), *Superminds: People Harness Hyper-computation, and More*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Chalmers, D. (2010), 'The Singularity: A Philosophical Analysis', *Journal of Consciousness Studies* **17**, 7–65.

Chesterton, G. (2009), *Orthodoxy*, Moody Publishers, Chicago, IL. First published 1908.

Chisholm, R. (1977), *Theory of Knowledge 2nd ed*, Prentice-Hall, Englewood Cliffs, NJ.

Dennett, D. (2007), *Breaking the Spell: Religion as a Natural Phenomenon*, Penguin, New York, NY.

Floridi, L. (2005), 'Consciousness, agents and the knowledge game', *Minds and Machines* **15**(3-4), 415–444.
URL: *http://www.philosophyofinformation.net/publications/pdf/caatkg.pdf*

Glymour, C. (1992), *Thinking Things Through*, MIT Press, Cambridge, MA.

Good, I. J. (1965), Speculations Concerning the First Ultraintelligent Machines, *in* F. Alt & M. Rubinoff, eds, 'Advances in Computing', Vol. 6, Academic Press, New York, NY, pp. 31–38.

Habermas, G. (1984), *The Resurrection of Jesus: An Apologetic*, University Press of America, Lanham, MD.

Hamkins, J. D. & Lewis, A. (2000), 'Infinite Time Turing Machines', *Journal of Symbolic Logic* **65**(2), 567–604.

Inhelder, B. & Piaget, J. (1958), *The Growth of Logical Thinking from Childhood to Adolescence*, Basic Books, New York, NY.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V. & Legrenzi, M. S. (2000), 'Illusions in reasoning about consistency', *Science* **288**, 531–532.

Kierkegaard, S. (1986), *Fear and Trembling*, Penguin, New York, NY.

Kurzweil, R. (2000), *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Penguin USA, New York, NY.

Leibniz, G. (1998), *Theodicy*, Open Court, Chicago, IL.

Lewis, C. S. (1960), *Mere Christianity*, Macmillan, New York, NY.

McGrew, T. (2010), 'Miracles', *Stanford Encyclopedia of Philosophy* .
    URL: *http://plato.stanford.edu/entries/miracles*

Paley, W. (2010), *Evidence of Christianity*, Qontro Classic Books. Paley's apology was first published in 1794. The book is available through Project Gutenberg. The full official title of the book is *A View of the Evidences of Christianity.*

Pollock, J. (1974), *Knowledge and Justification*, Princeton University Press, Princeton, NJ.

Pollock, J. (1989), *How to Build a Person: A Prolegomenon*, MIT Press, Cambridge, MA.

Pollock, J. (1995), *Cognitive Carpentry: A Blueprint for How to Build a Person*, MIT Press, Cambridge, MA.

Rips, L. (1994), *The Psychology of Proof*, MIT Press, Cambridge, MA.

Swinburne, R. (1981), *Faith and Reason*, Clarendon Press, Oxford, UK.

Swinburne, R. (1991), *The Existence of God*, Oxford University Press, Oxford, UK.

Swinburne, R. (2010), *Was Jesus God?*, Oxford University Press, Oxford, UK.

Turing, A. (1950), 'Computing Machinery and Intelligence', *Mind* **LIX (59)**(236), 433–460.

Vinge, V. (1993), 'The Coming Technological Singularity: How to Survive in the Post-Human Era', *Whole Earth Review* .

Yudkowsky, E. (1996), 'Starting into the Singularity'.
    URL: *http://yudkowsky.net/obsolete/singularity.html*