A Standard Model of Mind:
AAAI Technical Report FS-17-05

Two Problems Afflicting the Search for a Standard Model of the Mind

Paul Bello
U.S. Naval Research Laboratory
paul.bello@nrl.navy.mil

Selmer Bringsjord
Depts. of Cog Sci & Comp Sci
Rensselaer Polytechnic Institute (RPI)
selmer@rpi.edu

Abstract

We describe two serious problems afflicting the search for a standard model of the mind (SMM), as carried out and prescribed by Laird, Lebiere, and Rosenbloom (LLR). The first problem concerns a glaring omission from SMM, while the second calls into question the evidentiary standards for convergence that motivates the entire SMM agenda. It may well be that neither problem is insuperable, even in the short term. On the other hand, both problems *currently* stand in the way of making any present pronouncements to the effect that a standard model (or substantive portion thereof) exists and can be used as a benchmark against which other researchers might compare their approaches. The pair of problems is offered in a spirit of collaboration, and in the hope that grappling with them will help move the search a bit closer to the sort of undisputed rigor and predictive power afforded by such models in physics. Our order of business in the sequel is straightforward: we present and briefly discuss each of the two problems in turn, and wrap up with some remarks regarding whether or not these problems can be surmounted, and if so, how.

P1: The (Missing) Fundamental Datum

The standard model proposed by LLR is framed explicitly by the authors as an attempt to build consensus on how a *human-like* mind must be furnished in order to support intelligent thought and action. If there is a single fact beyond dispute about human beings, it is that we are conscious for large portions of our total lifetimes. Not only conscious in the sense of being awake, but rather a much wider range of self-consciousness that spans awareness *simpliciter* all the way to consciously representing oneself as a conscious bearer of mental states. For better or worse, this bit of fundamental data about the human condition must be contended with and appropriately treated in anything, schematic or otherwise, that claims to reflect the structure of the mind.[1] Our concern stems from the fact that when one examines either SOAR, or ACT-R, or Sigma, or for that matter their claimed convergence, one finds little explicit mention of consciousness *per se* and nothing whatsoever about self-consciousness.

[1]From this point onward, the term "mind" will be used in the context of human-like minds. On occassions where we wish to address the space of possible minds outside of this narrower scope, we do so explicitly.

In what follows, we argue that even on the most liberal identification of the term "mind" with a project in defining a consensus view of cognitive architecture, consciousness remains a deeply conspicuous omission that flirts with rendering the project incoherent. We open with a rough-and-ready discussion of how consciousness is often characterized in information-processing terms, and its most plausible correlate in LLR's proposed SMM. We show that the introduction of consciousness as a phenomena to be accounted for reveals fault-lines between architectures that are claimed by LLR to be similar enough to warrant the postulation of the SMM. We then consider and rebut a range of objections to both the general requirement of accounting for consciousness within the SMM and to the specifics of our argument. The analysis we offer with respect to consciousness is diagnostic of what we see as a far more insidious and difficult problem to remedy, which will be explored in the second part of this paper.

Conscious Content

One needn't look further than the *Stanford Encyclopedia of Philosophy*, specifically §2.2 of its entry on consciousness (Gulick 20042014), to see that the business of characterizing the various types of conscious content is, put mildly, tricky. The entry gives six different independently specified notions that suggest what it is for a state to be conscious. Rather than getting mired down in the various distinctions among them, we adopt the popular notion that conscious content is *highly integrated* with respect to non-conscious counterpart information. The idea that conscious content is highly integrated was given its most popular expression in the work of Bernard Baars and the development of his Global Workspace Theory (GWT) (Baars 1988; 2002).[2] Roughly, a GWT-style architecture consists of a (usually) large number of processing modules that can work (sometimes) individually but often in tandem to propose

[2]Note that we should not be interpreted as outright affirming Baars's theory *per se*. See e.g. (Bringsjord 1994). High integration is also the (or at least one of the) hallmarks of another prominent theory designed to model concsciousness (particularly phenomenal consciousness), viz. *integrated information theory*, invented long ago and still under very active extension and refinement by Tononi. For a recent overview, see (Tononi 2012; Oizumi, Albantakis, and Tononi 2014).

contents for inclusion in the Global Workspace. The Global Workspace has been characterized as something akin to working memory, although far more fleeting in terms of temporal duration. Once a module or coalition of modules produces output that "wins" entry into the workspace, the winning output gets globally broadcast to all other modules for subsequent elaboration, leading to highly integrated processing. Somewhat less metaphorically, the neural signatures of a process much like what was just described has been identified as is largely considered to be one of the best neural correlates of consciousness (Dehaene et al. 2006). This is not to say, however, that no other processing is occurring across a GWT-style architecture when the workspace is empty, or even after global broadcast. Some modules may not be equipped with algorithms or resources to process the broadcasted content. It doesn't therefore follow that these modules sit idle. They may be processing other information independently of broadcast and producing output that goes into competition for inclusion into the workspace on the next cycle. As our discussion proceeds, we shall refer back to this rough identification of conscious content as integrated, globally available content.

Consciousness, Global Availability, and the SMM

The most obvious objection to be raised against the inclusion of a distinction between conscious and non-conscious content in the SMM is that it is somehow already there. According to the rough and ready definition, conscious content is broadly available integrated content. Insofar as short-term memory stores make their contents broadly available to the rest of the modules in the SMM, it looks as if our worry is misplaced. GWT broadcasting is effectively implemented by short-term memories in the SMM that make their contents globally available.

If we for the sake of argument affirm this claim with respect to both ACT-R and Soar, which are paradigmatic instances of SMM-consistent systems, it seems that ACT-R is much less conscious on any given cycle than Soar is, which strikes us as more than a minor detail to be swept under the rug. ACT-R has a well-known limit of one chunk of information per buffer, leaving the contents of consciousness very sparse indeed. The identification of conscious content with buffer content comes from no less an authority on ACT-R than John Anderson himself (Anderson 2007). Soar, on the other hand, has an unbounded working memory. While something like the latter will strike most as being a little closer to right with respect to phenomenology, it leaves the objective inquirer with questions about whether non-conscious declarative content sufficient for subliminal priming is even possible to represent in Soar. We shall return to this latter point for further elaboration shortly.

Candidate Conscious Contents in ACT-R It may be further objected that the above is a mischaracterization on multiple counts. First, ACT-R's working memory isn't a module, but is actually the set of above-threshold, activated, declarative-memory elements — a set with a cardinality larger than unity. Other researchers have identified the whole of declarative memory as that which can in principle be conscious, and the set of items above a preset activation threshold as those items that comprise the current contents of consciousness, along with whatever is currently in the system's other buffers. At some level, this solves the problem of putting a line of demarcation between conscious and non-conscious facts/memories. But since working memory in ACT-R is treated as highly capacity-limited, identifying the contents of visual awareness with working-memory contents fails to capture the phenomenology of normal visual experience. It certainly seems to neurobiologically normal humans that they are visually aware of much more than the three to five items that many researchers typically identify as capacity limits for working memory. Now it may well be that humans (and human-like SMM instances) are only able to *report* on three to five items in visual awareness, while still having partially processed the rest of the visual scene. If this is case, as we suspect it is, then there is nothing in ACT-R or in the SMM that corresponds to the non-reportable remainder. Simple identification of short-term memory elements with the contents of consciousness seems to be an unworkable strategy.

Candidate Conscious Contents in Soar Moving from ACT-R to Soar, it has been suggested to one of us (PB) that the relevant distinction between conscious and non-conscious processing can be identified in terms of the distinction between working memory as a description of state and the substates that Soar enters during impasse-driven reasoning (Laird, personal communication). On this account, computations in substates correspond to conscious deliberative processing, while computations performed outside of substates are effectively reactive. For the sake of the argument that we are making here, the most pressing problem is that ACT-R has no functional equivalent to Soar substates. So if a distinction is to be made between conscious and non-conscious processing by appeal to substates, the two architectures are incommensurable, which of course can't be the case if they conform to the same SMM. Furthermore, this addendum to the global availability view suggests that our initial concern was indeed valid: global availability is not, on this modified view, the *sine qua non* of conscious content. The deeper and more challenging problem is with an impasse-centric view of consciousness. Such accounts are not particularly new or uncommon. Very recently, Morsella and colleagues have developed and proposed the "Passive Frame Theory," that effectively identifies consciousness with impasse detection and response (Morsella et al. 2016). The theory has come in for heavy criticism (see replies to the previously cited BBS article), some of which focuses on how such theories explain the various other aspects of consciousness outside of conflict situations.

If we are to be as charitable as possible and identify consciousness with both substate processing and the contents of working memory, we *still* are faced with difficulties. On the latest account of working memory in Soar, chunks have activation dynamics similar to those found in ACT-R. There are, however, crucial differences that matter to our discussion of conscious content. Recall that ACT-R divides up declarative memory into working memory (= above-threshold chunks =

contents of consciousness) and latent declarative memory. In its current incarnation, Soar's working memory remains unitary. There is a threshold implemented for forgetting: the removal of insufficiently activated items from working memory. But forgetting in Soar isn't quite the same as limited source activation in ACT. They don't necessarily produce the same sorts of capacity limitations and make wholly different predictions about how much information is conscious information. Approaches to subliminal priming have been developed in a highly modified version of ACT with a natural correlate of awareness associated with an inequality known as the retrieval ratio (van Maanen and van Rijn 2007). It still remains unclear how exactly to address subliminal priming and other forms of non-conscious processing against the backdrop of these modifications to Soar's working memory.

Implications for the SMM Returning to the ACT-R/Soar comparison, it is also true of ACT-R that the individual contents of working memory must be retrieved item-by-item into the retrieval buffer to be matched against the content of procedural memory. If this is so, then we would expect the timecourse of conscious experience to be different for ACT-R than Soar, given that productions in Soar match against the entirety of working memory. So far it looks to us as if ACT-R and Soar differ in both the relative richness and timecourse of conscious experience (using the term loosely). Neither architecture in their unmodified state have much to offer up for an account of non-conscious processing and its influence on behavior, unless we accept the problematic identification of impasse-free production-matching in Soar with non-conscious processing. Perhaps there is less convergence between the two architectures than we thought. Indeed, differences in the structure of their respective working memories plus unqualified claims about global availability reveals sharp inconsistencies between the two approaches.

Beyond Global Availability It should be noted that we have explored the lowest-hanging fruit in terms of a possible correlate for conscious content in the SMM: its identification with globally-available integrated information. In our humble view, this was probably the best shot at locating the what, where, and how of conscious processing in the SMM. Once we move past this conception of consciousness, things get far murkier in a hurry. The availability view we have been discussing is best identified with what the philosopher Ned Block has called "access consciousness," or A-consciousness for short. In a series of famous thought experiments, Block distinguishes A-consciousness from P(henomenal)-consciousness: the "what-it's-likeness" to be in a particular conscious state (Block 1995).

For decades (actually, centuries), 'mind' has been reserved as a term used to denote something that has subjective awareness or phenomenal consciousness: a thing that — to use Thomas Nagel's well-known phrase — there's something it's like to be (Nagel 1974). When you smell a rose, or take a sip of fine, rounded Brunello, or strive with all your mental might to solve an intellectual puzzle for a grade of A in your DiffEQ class, there's something it's like to you to do that. The firm terminological tradition is that seeking a machine with a mind is an activity that marks the searcher as a practitioner of *Strong* AI. Weak AI, in stark contrast, restricts itself to the attempt to build a computational system whose internal and external behaviors meet certain third-person standards (e.g. The Turing Test). In this standard and obvious light, the phrase 'A Standard Model of the Mind' is disturbingly infelicitous; put bluntly, the phrase is entirely inaccurate, especially without a rich account of consciousness that goes beyond global availability.

Why should we be worried about P-consciousness? For one, if we accept (as do most philosophers of mind) that we humans are indeed phenomenally conscious, along with the notion that there is a deep and perhaps unbridgeable explanatory gap that exists between information processing and conscious experience (i.e. the so-called "hard problem of consciousness"), then we are left with unattractive options. Perhaps the most attractive starting point is to resist Block's well-worn distinction between phenomenal and access consciousness, but as of this writing there have been no widely-regarded successes along these lines. On the other end of the spectrum, Christof Koch, Giulio Tononi, David Chalmers, and others have come to the conclusion that consciousness suffuses nature and that the picture of the universe that physics paints for us is incomplete at best and totally wrong at worst. To their credit, Tononi's research group works backward from a phenomenological analysis of experience to a set of postulates about what kind of properties a physical system needs to have in order to account for the essential elements of experience as they have analyzed it — and then they work from the postulates to an increasingly detailed computational and mathematical theory (Oizumi, Albantakis, and Tononi 2014; Tononi 2012). While somewhat non-standard as an approach, reverse-engineering of this sort is often worthwhile and may be a useful tool in helping to refine the current conception of the SMM going forward.

Brief Remarks on Self-Consciousness So far we have only touched upon the distinction between access and phenomenal consciousness. But both discussions assume that there is a *subject*[3] of conscious experience: someone or something that is aware and experiencing. We assume rather conservatively that all mentions of content in an instance of the SMM is self-related. However, self-relation and self-representation are inequivalent. Self-consciousness unquestionably involves the latter, yet cannot be wholly accounted for without the former (Musholt 2013). There is a phenomenological difference between unreflectively driving home from work and noticing your driving upon having attention captured by a motorist who has been pulled over. Once jolted out of automatic behavior, you might say to yourself in inner speech "I should drive extra carefully now." The self is represented explicitly in this and uncounted other less-exciting conscious episodes throughout the day. The ability to explicitly plan, exercise control, autobiograph-

[3]Subjectivity is clearly an obstacle, and it is a prerequisite for a complete account of selfhood: a rich and complicated topic that we have both worked on in the past, but have insufficient space to explore in this short paper (Bello and Guarini 2010; Bringsjord et al. 2015).

ically remember, and engage in dialogue depends critically on self-consciousness of this sort.

In any case, self-consciousness introduces a variety of challenges to be contended with and arguably places further requirements on the SMM. The most oft-discussed property of self-conscious states is that they are immune to error through misidentification (IEM). IEM holds that while self-conscious subjects may be in error about the content of one of their experiences, they are never mistaken about who is having the experience. For example, it may be the case that Paul often mistakes his hunger for irritability. Now he might think to himself "I'm irritated right now," when it is actually the case that he is hungry. Even so, there is no question that it is he himself who is thinking that thought. While this seems both obvious and philosophically picayune, consider that we may learn things about ourselves in the third-person that make a difference for action. If Paul's colleagues are chatting nondescriptly about their preferred choice for the next Center Director and unbeknownst to him, it is he that they refer to, he may do nothing at all. On the other hand, if Paul comes to learn that it is he himself that they speak of, he may try to vigorously talk them out of it.

The relationship between the so-called *de se* contents of self-consciousness and action-production requires a tight connection between knowledge and skills on the one hand, and architectural mechanisms (perhaps involving perspective-taking) on the other. If we are to accept that consciousness ought to fit somewhere in the SMM, it seems at least reasonable to assume that self-consciousness fits as well. If it does, and if the connection between knowledge, skills, and architecture is as necessary as it seems in the case of self-consciousness, then we ought to rethink the sharp divide between them that the SMM as currently developed seems to endorse.

Replies and Rejoinders

We consider a handful of possible replies and offer brief rejoinders.

The No-Consensus Reply "The SMM admits of gaps from the get-go and expects them in places where consensus hasn't been reached. Consciousness is clearly one of those phenomena where consensus hasn't been reached and therefore is not a good candidate for inclusion in the SMM at the present." **Rejoinder**: This strikes us as a double-standard. After all, there is a deep lack of consensus on the nature and function of working memory, which has drastic implications for cognitive architecture (D'Esposito and Postle 2015). Yet working memory (or the equivalent) is central to the SMM. Falling back on an outdated consensus view of working memory would constitute special pleading. Moreover, while we still struggle with the details of working memory (among other things), it is unimpeachably true that there is a distinction between conscious and non-conscious information processing.

The Levels-of-Analysis Reply "The SMM is first and foremost a consensus view on cognitive architecture. The higher powers of mind that supervene on the processes picked out in the SMM, including content-related issues, are outside of the scope of what the project is trying to achieve." **Rejoinder**: As we understand it, there is an implicit sufficiency claim built into the SMM that higher-level cognitive functionality can be expressed in terms of the cognitive architecture it specifies, plus knowledge and skills. As we have already argued, there is insufficiency in the current incarnation of the SMM, and since consciousness is a property of certain mental states and of persons/systems more generally, it cannot be simply identified with either knowledge or skills.

The Who Needs It Anyway Reply "Why care about the distinction between conscious and non-conscious content in the first place? The SMM marks out a space of human-like models and does not exclusively concern itself with cognitive models." **Rejoinder**: If the bounded rationality story that inspired so much of the current incarnation of SMM is to be believed, it has to be well-motivated. The limited bandwith of *conscious* deliberation is often cited as a key driver of bounded rationality. Moreover, if we take the standard line about consciousness being globally available integrated information along with the plausible premise that our beliefs, values, desires, obligations, and dispositions are distributed separately throughout the mind, we see that broadcast allows for the fullest expression of ourselves in decision-making — especially of the moral variety.[4]. Retreating into human-inspiration to avoid giving consciousness its due is a slippery slope. So much of human (and presumably human-like) cognition is predicated upon the distinction between conscious and non-conscious processing that the damage done to any theory of the mind that fails to feature it would be irreparable.

P2: Beyond Boxes and Arrows

The second problem threatening LRR's forthcoming call for, and announced progress toward, SMM is that, put baldly, their perceived alignment of boxes and arrows doesn't constitute verifiable convergence. Yes, LRR *perceive* a convergence between Soar, ACT-R, and Sigma, but absent any rigorous demonstration of such there is no rational reason to attribute this perception to underlying reality, rather than exuberance. Without (a) theorem(s) expressing that, at least with respect to some mental phenomena, initial convergence is in place, it is wishful thinking to aspire to reach, for the mind, what has been reached for the purely physical. Yet there is no such theorem, nor even an antecedent thereof (e.g., a rigorous conjecture), to be had in — the inaugural paper — (Laird, Lebiere, and Rosenbloom forthcoming). We bring this lacuna into focus in a way that both makes it evident, and provides an opportunity for LLR to advance their cause, by supplying either or both of two theorems that would fill in what is painfully missing.

[4]For an excellent book-length treatment of this issue, we recommend (Levy 2014)

Re. Convergence on Symbol Structures for "Memory and Content"

The lowest hanging fruit from which LLR and like-minded thinkers could show convergence, as opposed to merely gesturing toward it suggestively via boxes and arrows, would seem to be in connection with those "symbol structures" used in ACT-R, Soar, and Sigma to represent declarative propositions (which are in turn constitutive of memories, known facts, believed hypotheses, etc. for cognizers). Here, a bit more specifically, and using the authors' own language, is what would need to converge (to quote direction from LRR's paper):

- ACT-R: "chunks with activations and rules with utilities"
- SOAR: "triples with activation and rules with utilities"
- Sigma: "predicates and conditionals with functions"

Now, while a given cognitive architecture could well have a strong suit unmatched by any other architecture (= a domain that it is best at modeling cognition in), every cognitive architecture pitched — to use LRR's own wording — at the "human-level" obviously must model basic arithmetic.[5] More specifically, the axioms of Peano Arithmetic are gently but unmistakably introduced at this grade level. For instance, it must be learned at this grade level that ($\mathcal{I}$) any (natural) number n multiplied by 1 returns n. Let ϕ_A, ϕ_{So}, ϕ_{Si} be three symbol structures couched, resp., in the syntactic machinery of ACT-R, Soar, and Sigma, and each expressing $\mathcal{I}$. We shall leverage standard logico-mathematical tools and view each of these symbol structures as formulae in some formal language.[6] Let $\mathfrak{U}$ be a standard model-theoretic structure, and let $\vdash$ denote (as is standard in AI and mathematical logic) ordinary provability at the first-order level. Then, if, as LRR declare, there is convergence between Soar, ACT-R, and Sigma, one or both of the following two theorems must hold:[7]

Theorem 1 $\mathfrak{U} \models \phi_A$ *iff* $\mathfrak{U} \models \phi_{So}$ *iff* $\mathfrak{U} \models \phi_{Si}$

Theorem 2 $\phi_A \dashv\vdash \phi_{So} \dashv\vdash \phi_{Si}$

The Programming-Language Escape is No Escape One wishing to ignore the challenge of proving Theorem 1 and/or 2 might think that since cognitive architectures are in some sense analogous to, or as flexible as, standard programming languages, LLR and proponents of initial convergence could cogently retort that whatever can be done in a given cognitive architecture or given standard model of the mind can be accomplished in a programming language; hence convergence can be achieved along a different, less formal route. Unfortunately, this reply makes little sense, because if it worked, a group using a knowledge-representation system based on a fragment of FOL could simply harmonize with some group's using a KR system based on full second-order logic (SOL): they could (so the defective story goes) just write computer programs to capture what each group is doing, and convergence could be secured by showing that the programs are functionally equivalent. But of course, mathematically speaking, this is to go from the frying pan into the fire. The reason is that ascertaining whether a program P written in, say, Lisp (or language L), computes the same function f as P' is written in, say, Prolog (or L') is Turing-uncomputable (the problem is Π_2). Moreover, since one formalism for expressing declarative information and its use may allow the expression of problems beyond Σ_0, whereas another formalism may allow the expression of only problems that are fully Turing-solvable, reliance on the programming-language route is just wishful thinking. The upshot is that unless the clear capacity to represent dirt-simple arithmetic facts is stipulated to be outside of SMM, or unless all notions that SMM is inspired and guided by what physics has achieved are — *contra* LRR — summarily dropped, claimed convergence, even just initial convergence, must be accompanied by proofs of Theorems 1 and/or 2. Indeed, there will be many, many theorems in the same spirit that will need to be proved.

Concluding Remarks

Can the pair of problems we have presented be surmounted?

Well, we have already said that one way to instantly surmount the first problem is readily available: viz., simply concede that phenomenal consciousness (which in the human case, as we've explained above, is inseparably part and parcel of thinking), can't be captured in any mechanical format of the type used to describe either a typical cognitive architecture or a typical programming language — or for that matter a standard model of some parts of human cognition. Of course, such a concession means that what LLR are doing is quite different than what scientists like Tononi 2012 are doing, for the latter group is driven front and center by the dream of capturing, in computational process (suitably physically ensconced), phenomenal consciousness. And, the concession really ought to be accompanied by the abandonment of the misleading phrase 'A Standard Model of the *Mind*' (emphasis ours), in favor of something like 'A Standard Model of the Structure and Process of Basic Cognition,' which is of course a mouthful, but accurate (again, given the absence of phenomenal consciousness). The alternative for LLR is to write down, at least at the "boxological" level, some model that shows everyone where and what phenomenal consciousness is. We leave it to the trio to decide their own fate. Neither available move, needless to say, is a cakewalk.

On the other hand, steps can surely be taken in the near term to give a better treatment to both access consciousness

[5] We presume only 4th-grade arithmetic by U.S. Common Core Standards. See e.g. (mcg 2012). Our challenge could be expressed in connection with 2nd-grade arithmetic, because the general axioms associated with operators in Peano Arithmetic are introduced there via declarative statements.

[6] Note that now all mathematical physics corresponds ultimately to theorems expressed as formulae in a formal langauge that is part of a formal logic, and theoretical physics now offers its own axiom systems in which the relevant formulae, in the form of axioms and theorems, cover all of the relevant part of physics. See e.g. (Govindarajalulu, Bringsjord, and Taylor 2015).

[7] Where $\mathfrak{N}$ is the standard interpretation for arithmetic, some readers may be tempted to insist that the three cognitive architectures must also validate $\mathfrak{N} \models \mathcal{I}$, but this isn't needed for our challenge, and we leave it aside.

and certain aspects of self-consciousness in the SMM as it currently stands. A number of candidate psychological processes have been identified as necessary for consciousness, most notably *attention*, which when taken in any degree of detail past the "bottleneck" typically invoked by cognitive architecture researchers, is all but missing from the SMM.

As to the second problem, the situation is quite different, and rather more encouraging. LLR can roll up their sleeves and do the hard work of demonstrating that the declarative languages used in respective cognitive architectures, at least with respect to declarative content that must under any circumstances be part of human-like/level cognizers, conform to one of the key biconditional chains we presented above. To produce these theorems would be to produce the first step toward a standard model that might, at least to a degree, impress a theoretical physicist; and the production thereof might be within the purview of collaborators more than willing to work in the spirit of seeking a standard model of the (human-level) mind.

Acknowledgements

We thank Will Bridewell, John Laird and Gregory Trafton for insightful conversations that sharpened some of the ideas in this paper. This work was made possible through grant N0001415WX01339 from the Office of Naval Research to the first author. The views expressed in this document are solely the authors and do not represent the policies or position of the United States Federal Government or the United States Department of Defense.

References

Anderson, J. R. 2007. *How Can the Human Mind Occur in the Physical Universe?* Oup Usa.

Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge, UK: Cambridge University Press.

Baars, B. 2002. The cognitive access hypothesis: Origins and recent evidence. *Trends in Cognitive Science* 6(1):47–52.

Bello, P., and Guarini, M. 2010. Introspection and mindreading as mental simulation. In *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*, 2022–2028.

Block, N. 1995. On a Confusion About a Function of Consciousness. *Behavioral and Brain Sciences* 18:227–247.

Bringsjord, S.; Licato, J.; Govindarajulu, N. S.; Ghosh, R.; and Sen, A. 2015. Real robots that pass human tests of self-consciousness. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, volume 2015-November, 498–504.

Bringsjord, S. 1994. Baars Falls Prey to the Timidity He Rejects: Commentary on Baars on Contrastic Analysis. *Psyche* 1(10).

Dehaene, S.; Changeux, J. P.; Naccache, L.; Sackur, J.; and Sergent, C. 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences* 10(5):204–211.

D'Esposito, M., and Postle, B. R. 2015. The cognitive neuroscience of working memory. *Annual review of psychology* 66:115–42.

Govindarajalulu, N. S.; Bringsjord, S.; and Taylor, J. 2015. Proof Verification and Proof Discovery for Relativity. *Synthese* 192(7):2077–2094.

Gulick, R. V. 2004/2014. Consciousness. In Zalta, E., ed., *The Stanford Encyclopedia of Philosophy*.

Laird, J.; Lebiere, C.; and Rosenbloom, P. forthcoming. A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*.

Levy, N. 2014. *Consciousness and Moral Responsibility*. Oxford University Press.

2012. *McGraw Hill's Grade 4 Math*. New York, NY: McGraw Hill.

Morsella, E.; Godwin, C. A.; Jantz, T. K.; Krieger, S. C.; and Gazzaley, A. 2016. Homing in on consciousness in the nervous system: An action-based synthesis. *Behavioral and Brain Sciences* 39:e168.

Musholt, K. 2013. Self-consciousness and nonconceptual content. *Philosophical Studies* 163(3):649–672.

Nagel, T. 1974. What Is It Like to Be a Bat? *The Philosophical Review* 83(4):435–450.

Oizumi, M.; Albantakis, L.; and Tononi, G. 2014. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. PLOS *Computational Biology*.

Tononi, G. 2012. *Phi: A Voyage from the Brain to the Soul*. New York, NY: Pantheon.

van Maanen, L., and van Rijn, H. 2007. Accounting for subliminal priming in ACT-R. *8th International Conference on Cognitive Modeling* 1–6.