5-2-2018

Can Global Workspace Theory Solve the Frame Problem?

Katelyn Rivers

CAN GLOBAL WORKSPACE THEORY SOLVE THE FRAME PROBLEM?

by

KATELYN VIOLET RIVERS

Under the Direction of Daniel Weiskopf, PhD

ABSTRACT

The *Frame Problem* originated as an obstacle for classical, symbolic A.I. that was adopted, expanded, and reformulated by philosophers. The version of the problem that I focus on, the *Holism Problem*, points out the difficulty in programming systems to recognize and consider mostly relevant information, given that relevance is context-sensitive. My goal in this thesis is to determine whether the *Global Workspace Theory (GWT)* can solve the holism problem. GWT proposes that distributed parallel processing, global broadcast, and chaotic itinerancy can solve the problem by providing a system with 1) the speed to search through information, 2) access to the information it needs to compute relevance, 3) the ability to synthesize information. I argue that GWT fails to enable a system to recognize *any* relevant information because it inadequately responds to the *Epistemological Holism Problem*, which requires successfully determining the norms by which a system can recognize relevance.

INDEX WORDS: Frame Problem, Relevance, Artificial Intelligence, Global Workspace Theory, Distributed Processing, Modularity

CAN GLOBAL WORKSPACE THEORY SOLVE THE FRAME PROBLEM?

by

KATELYN VIOLET RIVERS

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2018

CAN GLOBAL WORKSPACE THEORY SOLVE THE FRAME PROBLEM?

by

KATELYN VIOLET RIVERS

Committee Chair: Daniel Weiskopf

Committee: Neil Van Leeuwen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2018

ACKNOWLEDGEMENTS

I would like to thank my thesis chair, Dr. Dan Weiskopf, for his patience, support, and commitment to helping me engage in the hardest project I've done thus far. I would also like to thank Dr. Neil Van Leeuwen, the only other member on my thesis committee, for providing me with insightful comments and great readings that helped strengthen this thesis.

Although it has been two years since earning my B.A., I am still in debt to the professors at the University of North Florida who tirelessly dedicated their time to make me the student I am today: Dr. Jon Matheson, Dr. Mitch Haney, Dr. Sarah Mattice, Dr. Paul Carelli, and Dr. Aaron Creller. They all inspired my interests in cognition in unique ways, and for them I am thankful.

And finally, thank you to my peers who listened to me ramble about relevance, and to the conference-goers. I am especially to those who attended the *2017 Mind and Brain Annual Conference* at the University of Michigan-Flint and the *2017 International Conference on Artificial Intelligence and Information* at the University of Porto. They all asked questions and provided comments that helped me engage a wider audience.

Engaging in this high-level research will enable me to approach my future empirical education with confidence and finesse.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

GWT: Global Workspace Theory
GW: Global Workspace
CI: Chaotic Itinerancy

1 INTRODUCTION

The frame problem originated as a problem for classical, symbolic A.I. that eventually became a broader philosophical problem for the creation of intelligent systems. The version of the frame problem that I will focus on, the holism problem, points out the difficulty in programming systems to recognize and consider mostly *relevant* information, given that relevance is a context-sensitive process that requires global assessment of information and the ability to synthesize that information. This problem is critical to the development of general problem-solving machines because recognizing what information and what inferences are relevant are key aspects of solving problems.

The overarching goal of this thesis is to determine whether the *Global Workspace Theory (GWT),* a theory of consciousness created by Bernard Baars (1997, 2005, 2007), can solve the holism problem. To evaluate GWT, I describe Murray Shanahan's and Bernard Baars' work on the topic, which began in 2005. Shanahan and Baars (2005) begin with the narrow claim that parallel processing can solve the problem by providing a system that is capable of handling and searching through a large amount of information. Later, Shanahan realized that merely appealing to parallel processing provided an incomplete solution, so he expanded on the GWT solution by explaining the role of global access and dynamical systems theory in solving the problem (Shanahan 2006, Shanahan 2010). In this thesis, I will argue that there are issues in each step of Shanahan's and Barrs' solutions such that GWT fails to be a complete solution to the holism problem. This thesis aims to propose further questions to those who want to enable a GW system to perform relevance computations—not to knock down the possibility of a GWT system that solves the frame problem. Briefly, the major issue is that Shanahan and Baars interpret the problem as solely a computational one, and thus provide solutions aimed at addressing computational issues. However, the source holism problem “goes as deep as the analysis of rationality" and, accordingly, a complete solution requires a more in-depth analysis of what it takes for a system to *recognize what is and isn’t relevant* (Fodor, 1987, p. 27).

In Section 2, I will describe two versions of the frame problem, how they relate to each other, and why they are such difficult problems. In Section 3, I will describe each step in Shanahan and Baars’ GWT solution, followed by my explanation of why each step falls short. Finally, in Section 4, I will respond to a few objections to my thesis and explain what more is needed to solve the holism problem.

2 THE FRAME PROBLEM AND RELEVANCE

The frame problem points out a difficulty in programming general problem-solving systems (i.e., it is not a problem about perception, nor a problem for encapsulated expert systems) to make decisions without considering all possible information. The frame problem comes in many forms, so I will focus on versions that focus on relevance. In this section, I will describe some terms and concepts that are most important for critical engagement with this work. Namely, in Section 2.1. I will discuss the relevance problem, in section 2.2 I will describe the holism problem, in section 2.3. I will distinguish computational and epistemic issues underlying the holism problem, and in section 2.4 I will discuss two notions of relevance.

2.1 The Relevance Problem

The *Relevance Problem*, which was formulated by Daniel Dennett (1987), asks how we could create an intelligent, open-ended system that can recognize and consider mostly relevant information to update beliefs and plan actions. While systems should be able to generate innumerably many inferences about a given situation, systems also should be able to pick out and consider the *relevant information* to draw *relevant inferences*. The problem, with this, however, is that it's not immediately clear how to program a system to recognize what is relevant to a given task.

To make the problem clearer, Dennett provides a thought experiment (1987, p. 41-42). Imagine that there is a robot, R1, who is locked in a room with both a bomb and a battery on a wagon and tasked with saving only the battery from the bomb. R1 decides that it should pull the wagon out of the room because it would pull the battery out of the room with it. Although R1 knew that the bomb was on the wagon, it was unable to predict that pulling the wagon would pull the bomb with it. So R1D1 was created with the intent that it could deduce the side effects of its actions. R1D1 was able to realize that simply pulling the wagon wasn't enough, but unfortunately, it spent the entire time deducing all the possible side effects of its actions (e.g., that the wall wouldn't change color after pulling the wagon) and never got anything done. So R2D1 was created, and this time it was programmed with rules for ignoring irrelevant side effects. Now, R2D1 was able to realize that the color of the walls was irrelevant, but now it made *even more inferences* about which inferences were and weren't relevant, so it also got nothing done. The robot that solves the frame problem, which Dennett amusingly calls "R2D2," would be able to recognize and consider what is relevant without having to search through its entire set of stored information, yet still be able to make any inference about its environment.

Importantly, I am using the words "recognize" and "consider" in a technical way, and each is a necessary part of the relevance problem as I conceive it. For a system to *recognize* relevant information, it should be able to, when asked about a piece of information, correctly say whether it is relevant to their task at hand. For example, if I ask an artificial system to make me some caramel, it should recognize that the price of slip-on shoes in Guilin, China is not relevant to their task. On the other hand, for a system to *consider* mostly what is relevant, I mean that while making caramel, an artificial system shouldn't think about (and preferably it shouldn't think about whether to think about) the price of slip-on shoes in Guilin, China. It goes without saying that the system needn't be perfect at recognizing and considering *only* relevant information, but intelligent systems should be able to have a reasonable amount of foresight and consider what is relevant at least most of the time.

2.2 The Holism Problem

John Haugeland (1987) complicates the relevance problem by pointing out that the effects of an action are often “situation bound” or “sensitive to the details of the current situation” (p. 83). Any action could have different consequences depending on the current state of the universe (e.g., the consequences of flailing your arm will have different consequences depending on whether there is something next to you). Given that consequences depend on one’s current situation, we cannot properly program certain responses to stimuli as having a set base value because doing so would require the system to determine whether the value should be modified in a situation, which raises a frame problem yet again. For example, one should not program “run” as being *the proper* or *good* response to seeing a bear paw print in the dirt because whether this stimulus indicates danger and the need flee depends on one’s situation (e.g., the print in the dirt could be part of an elaborate role-playing game). The situation-bound nature of the decision-making process entails that “generic or ‘precanned’ responses aren’t reliable” and programming a system with such responses would end up being a hindrance to successful engagement with an environment (Haugeland, 1987, p. 83).

Not only are consequences situation-bound, but also *what information is relevant* to a problem is also situational. This problem was taken up by Hubert Dreyfus (1972, 1992), Michael Wheeler, and many situated cognition theorists. Wheeler (2008) concisely summarizes the context sensitivity of relevance in the following:

> Any system worthy of the epithet ‘intelligent’ must be able to retrieve from its memory just those items or bodies of stored information that are most relevant to its present context, and then decide how to use, update, or weigh that information in contextually appropriate ways, in processes such as belief-fixation and action selection. (p. 424)

Accordingly, the holism problem focuses on how a system recognizes which information is relevant, how it retrieves the relevant information from memory, and how it uses the information, all in contextually dependent ways. Advocates of this problem argue that there is simply no *non-ad hoc* way of determining which information is relevant to the task of getting a drink. For example, while information about the Pyramids of Giza might not be relevant to one's normal drink-getting adventures, it might be important if you must solve a pyramid related trivia problem required to open the fridge door.

Thus, the holism problem complicates the relevance problem in two ways: not only are consequences context sensitive but what information is relevant is also context sensitive. An intelligent system must be open-ended enough to be able to generate innumerably many inferences about action consequences, to recognize what information is and isn't relevant, to consider mostly what is relevant, and it should be able to do these things while being cognizant of its context.

2.3 Wait—What's 'Relevance'?

"When is a conclusion worth deriving?" That is a good question.[1] One might say that a conclusion is worth deriving when it is *relevant*. That's a very reasonable thing to say, and even appears to be the main assumption underlying the relevance problem: to have a functioning system that can solve problems, we need the system to draw relevant conclusions and recognize and consider relevant information.

But what is *relevance*? That is also a good question and will be the question that runs in the background of this entire thesis. There are two notions of relevance that I want

[1] In fact, this question is the focus and title of Henst, Sperber, and Politzer's (2002) paper, "When is a conclusion worth deriving? A relevance-based analysis of indeterminate relational problems."

to distinguish: *relevance as defined by Relevance Theory* and *common-sense relevance*. For convenience, when referring to *relevance as defined by Relevance Theory,* I will denote it relevance*.

A piece of information or inference is relevant* when it produces a measurable cognitive effect. Henst, Sperber, and Politzer's (2002) describe relevance* as follows:

> In relevance theory, relevance is seen as a property of inputs to cognitive processes (e.g., stimuli, utterances, mental representations). An input is relevant to an individual at a certain time if processing this input yields cognitive effects. Examples of cognitive effects are the revision of previous beliefs, or the derivation of contextual conclusions, that is, conclusions that follow from the input taken together with previously available information. Everything else being equal, the greater the cognitive effects achieved by processing an input, the greater it's relevance. On the other hand, the greater the effort involved in processing an input, the lower its relevance.

In other words, an inference is relevant* when it generates a unique conclusion in a context, a conclusion that has cognitive effects such as optimal allocation of energy resources or effort, benefit to the inference maker, etc.

The common-sense notion of relevance, on the other hand, is pertinent to the task at hand, helpful in solving the task at hand, closely connected to the topic at hand, etc. For example, if I ask an artificial system to make me some caramel, the price of slip-on shoes in Guilin, China would be irrelevant because 1) the robot is in Atlanta, which is very far away from slip-on shoes 2) the robot does not need slip-on shoes to make caramel or to buy the ingredients to make caramel 3) there is no economic link between the price of slip-on shoes and the price of the ingredients to make caramel, etc. Thus, this notion of relevance points to conceptual and pragmatic (e.g., do I need x for y?) connections between pieces of information.[2]

In this work, when questioning whether GWT can enable a system to recognize relevance, I am concerned primarily with the common-sense notion. I do discuss relevance* when evaluating GWT, but as we will see that doesn't help solve the problem. Furthermore, while there might be a relationship between relevance and relevance*, I won't go into that here.

2.4 What's the Problem with the Holism Problem?

The holism problem is difficult to solve, according to Dennett (1987), because A.I. "…initially knows nothing at all 'about the world.' The computer is the fabled tabula rasa on which every required item must somehow be impressed, either by the programmer at the outset or via subsequent 'learning' by the system" (p. 47). Underlying this issue is the fact that we're not entirely sure how humans solve the frame problem, which makes successfully performing relevance computations difficult to successfully replicate in a blank-slate system.

Following Sheldon Chow (2013), we can break the holism problem into three distinct worries: the *Generalized Holism Problem*, the *Computational Holism Problem*, and the *Epistemic Holism Problem* (modified from Chow (2013), p. 4-5).

The Generalized Holism Problem is just as described in the previous section: how can a system determine, out of all the possible information, what information is relevant in its specific context.

The Computational Holism Problem focuses on how one computationally delimits

[2] This is not meant to be a complete definition of the common-sense notion of relevance, nor is it meant to pick out necessary or sufficient conditions of being relevant. My goal is to give the reader a general idea of the concept that is at foundation of this paper.

relevant information, or how a system delimits what gets considered during a task. The computational aspects focus on things such as processing speed, time available, amount of information needed to consider, algorithms, etc.

Worries about combinatorial explosion are related to the Computational Holism Problem. One might claim that the problem is difficult because any piece of information, in principle, can be relevant in any given task, which would result in a combinatorial explosion if a system had to search through the entire set of potentially relevant information. Problems of combinatorial explosion occur when a system must consider a huge amount of combinations of information regarding the task at hand, different actions, and different consequences. For any given task, there are many different chains of choices one could make, each of which could result in an even greater number of different consequences, causing the amount of information to explode exponentially. Accordingly, many who respond to the relevance problem attempt to deal with the sheer amount of information that comes into play when recognizing and considering what is and isn't relevant to the performance of a task.

The Epistemic Holism Problem, on the other hand, focuses on how a system recognizes that something is relevant or decides that x is relevant while y is not. These worries are deeply epistemic because, as will be emphasized in later sections of this work, they stem from questions about what relevance is as an epistemic norm and what it means for a piece of information to be relevant. In distinguishing between the Computational and Epistemic holism problems, it should be clear why set up the problem in terms of considering and recognizing relevance: considering relevant things aligns with the Computational Problem and recognizing relevance aligns with the Epistemic Problem.

3 GWT AS A SOLUTION TO THE HOLISM PROBLEM

Bernard Baars and Murray Shanahan advocate that GWT can perform contextually sensitive relevance computations (i.e., it can solve the holism problem).[3] GWT is a cognitive architecture based on the idea that there is a "communications infrastructure" in the brain called the "global workspace" (GW) that is connected to and communicates with a relatively small set of specialist processes. According to GWT, the conscious mental life of an organism is composed of periods of competition and broadcast whereby information collected from the environment is sent to the GW and disseminated to specialist processes, which then use that information to engage in an unconscious competition for access to the GW. The specialist process coalition that wins the competition for the GW is "broadcast" to the rest of the GW infrastructure, becomes the object of the systems' attention, and is used to perform the task at hand. A single, entire specialist process does not win the competition; instead, different combinations of parts of specialist processes—a "coalition"—wins the competition. Importantly, for Shanahan and Baars, the specialist process coalition that wins the competition contains the information that is *relevant* to one's context.

In the following subsections, I will describe, in chronological order, each step in the GWT solution, followed by my objections to the proposed step.

[3] Although they use the term "frame problem" throughout their work, they claim that GWT would be able to provide a system with the means to make relevance computations in different contexts, which is what *I* am calling the "holism problem." This is just a difference in terminology. Along these lines, Shanahan and Baars (2005) write, "What the global workspace architecture has to offer in addition is a model of information flow that explains how an informationally unencapsulated process can draw on just the information that is relevant to the ongoing situation without being swamped by irrelevant rubbish", which is exactly the holism problem as I understand it (pg. 174).

3.1 Parallel Processing as a Solution

In “Applying Global Workspace Theory to the Frame Problem,” Shanahan and Baars (2005) argue that GWT solves the holism problem because it provides a system with what it computationally needs to handle a large amount of information and to resolve issues of combinatorial explosion. According to Shanahan and Baars (2005), what enables GWT to solve the problem is its distributed, massively parallel architecture. In arguing so, Shanahan and Baars locate the source of the problem in the sheer amount of information a system might have to consider in performing relevance computations, i.e., they consider the problem as a computational one.

To concisely summarize the problem, Shanahan and Baars highlight quotes from both Fodor and Carruthers. For example, on the frame problem, Carruthers writes:

> The computational processes that realize human cognition will need to be tractable ones, of course; for they need to operate in real time with limited computational resources… And any processor that had to access the full set of the agent’s background beliefs (or even a significant subset thereof) would be faced with an unmanageable combinatorial explosion. (Carruthers, 2003, p. 23).

and Fodor writes:

> The totality of one’s epistemic commitments is vastly too large a space to have to search if all one’s trying to do is figure out whether, since there are clouds, it would be wise to carry an umbrella. Indeed, the totality of one’s epistemic commitments is vastly too large a space to have to search whatever it is that one is trying to figure out. (Fodor, 2000, p. 31).

In the quotes above, both Carruthers and Fodor note a worry about the amount of information that could come to bear on the task of performing relevance computations. Both worry that a fully unencapsulated cognitive system (a system with access to all beliefs, expectations, etc.) would be unable to perform a relevance computation because either the amount of information would be too much to search through or would result in a combinatorial explosion. Thus, in citing Carruthers and Fodor this way, Shanahan and Baars locate the source of the problem in computational worries.

To provide further clarity to the problem, Shanahan and Baars relate the frame problem worries to computational complexity theory, which distinguishes between tractable, intractable, and undecidable problems. Briefly, a tractable problem is one in which an algorithm could reasonably find a solution because the number of steps it would take to solve the problem, while large, is manageable and one could find a solution using brute force in polynomial-time. For example, figuring out whether a given year is a leap year is a tractable problem because a relatively simple algorithm could solve the problem by brute force. An intractable problem, on the other hand, has no simple algorithmic solution and is computationally infeasible because it can only be solved in exponential time. Even worse, an undecidable problem is a problem that no single algorithm can provide a solution.

Shanahan and Baars ask: when Fodor, Carruthers, and others say that the amount of information one would have to consider is simply too large, do they mean that the problem is intractable or undecidable? To answer this question, they write,

> [T]he real concern over computational feasibility is not an accompaniment to the presumption that first-order logical inference is taking place. The real worry, in the context of the frame problem, seems to be that the set of sentences having a potential bearing on an informationally unencapsulated cognitive process is simply "too big." This suggests that the issue is not really tractability, in the proper sense of the term, for tractability is to do with the rate at which computation time grows with the size of the problem. Rather, the difficulty seems to be the upper bound on n, not the upper bound on T(n), where the n in question is the number of potentially relevant background sentences. (Shanahan & Baars, 2005, p. 163).

In other words, Shanahan and Baars argue that Fodor and Carruthers' frame problem worries do not point out a computationally intractable or undecidable problem because their worries are about the amount of information the system would have to sift through, not the amount of time/number of steps it takes to perform the relevance computation. By definition, worries about computational intractability or undecidability are about the time it takes to solve a problem or the number of steps it would take to solve the problem. However, if the worry is just about the amount of information, the problem is still computationally tractable—not an issue at all.

Even if Fodor and others were suggesting that the holism problem was a computationally intractable or undecidable problem, Shanahan and Baars argue that this poses no *real* threat. First, they argue that neither an intractable nor an undecidable problem, despite their names, are impossible to solve. Intractable and undecidable problems may be inefficient, time-consuming, or require multiple complex algorithms to solve, but they can be solved. For example, one of the most common examples of an intractable problem is the "Traveling Salesman Problem," which points out the difficulty of finding the shortest path between multiple locations using a single search algorithm. This problem becomes efficiently solvable when you use a method other than brute-force search. For example, Google's *Google Maps Directions API* solves the traveling salesman problem by thinking of the problem by making graphs with weighted nodes and then applying statistical methods to find a solution quickly.[4] So, Shanahan and Baars reason that even if the problem is computationally intractable or undecidable, it can still be solved.

Additionally, Shanahan and Baars (2005) argue that Fodor's and Carruthers' reasoning relies on the assumption that the system is not engaging in parallel processing.

---

[4] Google, "The Traveling Salesman Problem."; Google, "Directions API Guide."

In other words, they argue issues of combinatorial explosion assume that there is one central system that serially looks at each potential bit of information, action, and consequence, one at a time, to determine what it system should do. However, Shanahan and Baars argue that, because GWT proposes a distributed, massively parallel cognitive architecture, the amount of information is not too much for a system to manage (p. 160-174). Instead of considering information and performing relevance computations serially, they argue that each specialist process unconsciously runs in parallel and determines whether it has something relevant to offer and if so, it makes itself available to the GW.

By appealing to distributed, parallel processing to solve the problem, Shanahan and Baars deny that the amount of information is too large for a system to search through. Along these lines, they write:

> But why should a large such n [where n is the amount of information one would have to consider to perform relevance computations] be a problem? Recall Fodor's claim that "the totality of one's epistemic commitments is vastly too large a space to have to search." Perhaps the perceived difficulty with a large n is the supposed cost of carrying out an exhaustive search of n items of data. Yet plenty of algorithms exist for search problems that have very favourable computational properties. For example, the task of searching a balanced ordered binary tree for a given item of data is $O(\log_2 n)$. Other data structures for search have even better complexity results. This is why Internet search engines can find every website that mentions a given combination of keywords in a fraction of a second, even though they have records for several billion web pages. (Shanahan & Baars, 2005, p. 163).

Because the distributed specialist processes run in parallel, as opposed to serial non-distributed processes, the system can quickly search through every piece of information without over-burdening the system.

3.1.1 Evaluating Parallel Processing

Now that I had laid out a basic map of how parallel processing was proposed to solve the problem, I will move on to evaluating whether it, in fact, solves the problem. To do so, I will argue that Shanahan and Baars' (2005) interpretation of the problem is too narrow in scope. Then, I will combine the GWT solution with some work by H. Clark Barrett (2005) and Dan Sperber (2005) to describe how specialist processes might perform relevance computations. In doing so, we will hopefully get a better picture of the information flow the global workspace architecture so that it can be properly assessed. Finally, I will argue that appealing to parallel processing does not provide the GWT system with a means to recognize which piece of information is relevant even if one considers the solution in light of the additions from Sperber and Barrett.

3.1.1.1 The Source of the Holism Problem

Shanahan and Baars (2005) interpret the problem as a computational problem because they focus on quotes that imply the problem with the holism problem is that the amount of information a system would have to consider in performing a relevance computation is too large. However, the frame problem is not merely a computational problem. In fact, Sheldon J. Chow (2013) locates five different interpretations of the problem, all of which have slightly different emphases!

Why else—aside from computational troubles—would one worry about programming a system to recognize relevance in the first place? In section 2.3 I began to distinguish between the computational holism problem and the epistemic holism problem. Here, I will go into more depth on the reasons why one should think of the epistemic holism problem as a genuine problem and why a complete solution to the entire holism problem requires solving the epistemic problem as well.

One could argue that what makes relevance especially difficult to program a system to compute is that it has the properties of being what Fodor calls *Quineian* and *isotropic*.

Relevance is Quineian because whether a piece of information is relevant depends on the entirety of one's epistemic commitments. Compare this with "simplicity" which also has the property of being Quineian, according to Fodor. To determine whether a theory is simpler, one must consider the theory in light of the predictions and assumptions it makes, and whether it coheres with other perhaps seemingly unrelated fields of study. Likewise, in considering whether a piece of information is relevant, one must consider the information in light of the predictions it makes, the assumptions the information makes, and whether the information coheres with other things one knows (Fodor, 1983, p. 107-108). Relevance is isotropic in that any piece of information could come to bear on performing a relevance computation (Fodor, 1983, p. 105) but neither we (the programmers) nor the system knows beforehand what that information could be.

Framing the problem in terms of such deep epistemic worries highlights the fact that there the problem has underlying issues regarding the nature of rationality and connection between ideas. While addressing computational issues is necessary to solving the problem, one must address more than just computational issues to provide a complete solution. Accordingly, the Holism Problem that I am concerned with is, at its core, a deeply *epistemic* problem. Chow (2013) also distinguishes between the computational and epistemic frame problem worries. Along these lines, he writes, "epistemological problem does not have to do so much with the computational costs of delimiting what gets considered in a given cognitive task (although this remains an important aspect) as it has to do with considering the *right* things" (p. 6).

A related issue that will crop up later in this work is that it's unclear what Shanahan and Baars mean by “relevance,” which is an open question because the term has been used differently throughout philosophical literature and even in daily discourse. This issue of what is meant by “relevance” is the core of why the problem I am concerned with is an *epistemic issue*. As Chow (2013) notes (p. 6, footnote 5), we mean something different from “epistemic frame problem” than Shanahan (2009) does:

> To summarize, it is possible to discern an epistemological frame problem, and to distinguish it from a computational counterpart. The epistemological problem is this: How is it possible for holistic, open-ended, context-sensitive relevance to be captured by a set of propositional, language-like representations of the sort used in classical AI? The computational counterpart to the epistemological problem is this. How could an inference process tractably be confined to just what is relevant, given that relevance is holistic, open-ended, and context-sensitive? (Shanahan, 2009, section 3).

Shanahan’s interpretation of the epistemological problem is that there is a difficulty translating relevance computations into symbolic representations and finding the proper rules of transformations amongst the symbolic representations so that classical AI can generate relevant inferences. However, this isn’t the issue I’m concerned with here. What I’m calling the epistemic issue is focused on figuring out and applying the rational norms that govern relevance computations.

To be clear, when I refer to the *Computational Holism Problem* I am referring to worries about the amount of information a system would have to consider in performing a relevance computation and how to delimit specific sets of information without considering each piece of information in the set. The *Epistemological Holism Problem*, on the other hand, refers to determining the norms by which a system can recognize relevance and programming a system to apply those norms with a reasonable level of success.

Through the fact that Shanahan and Baars only interpret the frame problem as a computational complexity issue, they set out to describe a system that can deal with computational complexity issues with ease. However, this only addresses the computational part of the problem, and *you cannot solve the problem without addressing the epistemic worries*. Thus, Shanahan and Baars are already on shaky grounds in appealing to GWT as a solution to the holism problem.

3.1.1.2 Expanding Upon GWT

Interpreting the frame problem is incredibly difficult; there are anthologies dedicated to this very task. So, one shouldn't stop at saying that an interpretation of the problem is dubious—instead, it's best to also question, given the way Shanahan and Baars interpret the frame problem, whether GWT would solve the problem they think it does. To answer this question, I will expand upon *how* the specialist processes make their relevance computations. The closest Shanahan and Baars get to describing how the specialist processes perform their relevance computations comes from the following quote:

> Consider the computational processes that might underlie the likening of a Rorschach inkblot to, say, an elephant. Fodor's argument hints at a centralised process that poses a series of questions one-at-a-time—is it a face? is it a butterfly? is it a vulva? and so on—until it finally arrives at the idea of an elephant. Instead, the global workspace model posits a specialist, parallel process that is always on the lookout for elephantine shapes. This process is aroused by the presence of an inkblot that actually resembles an elephant, and it responds by announcing its findings. The urgency with which this process commends itself means that the information it has to offer makes its way into the global workspace, and is thereby broadcast back to all the other specialist processes. (Shanahan & Baars, 2005, p 168).

The suggestion that there are certain processes on the lookout for certain kinds of stimuli and that automatically activate when that stimulus is present is not an uncommon one. For example, one can also find this reasoning "Enzymatic Computation and Cognitive Modularity," in which Barrett (2005) describes an enzyme-like modular system whereby information is sent to a "bulletin board" that is monitored by all modules, and the information is used by whichever module can process it. The information in the bulletin board is available for all modules to use, and if the information is taken up by a module and unable to be used it is sent back to the bulletin board for reuse by other modules. Importantly, for Barrett, relevance computations are made possible by global access to all potential information and by the lock-and-key method of bonding—if any information does not "fit into" a given module, then it is marked irrelevant.

This lock-and-key method of bonding could be how the GWT specialist processes realize what is and isn't relevant. Recall that GWT proposes that the specialist processes gain access to information about the system's current situation and environment, *and then* each specialist process runs in parallel, using the broadcasted information, to determine whether it, itself, is relevant. If the information about the environment is broadcast to all modules, each of which is on the look-out for information they can process, they can automatically activate via the lock-and-key method once the information is available. Compare this to Fodor's (1983) idea of mandatory processing in modular systems, whereby information sent to a module must be processed when (and only when) the information is available. Bolstering this idea, Shanahan (2005) writes that if the specialist process determines it has something relevant to contribute, "it wakes up and starts to lobby for access to the global workspace" (p. 57).

Even if we understand the specialist processes as mandatorily activating, via the lock and key method, in response to presented stimuli, one still might question the "terms" of the competition—i.e., what defines the parameters for a winning and losing competition? To imagine what this might look like,[5] we can look "Modularity and Relevance: How Can a Massively Modular Mind Be Flexible and Context Sensitive" in which Dan Sperber (2005) argues against the claim that modular systems[6] cannot be context sensitive by proposing one way in which a modular system could achieve context-sensitivity.

Briefly, Sperber argues that a modular system, like GWT, could be context-sensitive through an attention-focusing competition over energy resources, and that the system can decide what is most relevant to focus on by optimally allocating energy resources. In other words, Sperber's work suggests that relevance is determined in specific contexts by an optimal allocation of energy resources (presumably the natural resources of the brain—blood oxygen, glucose, etc.).

The first major step in Sperber's explanation is to argue that context-sensitivity is enabled through the differential allocation of energy resources amongst modules; certain aspects of one's environment might be more salient, and thus more energy will be allocated to process that aspect. Sperber explains this using Simons and Chabris' (1999) experiment on inattentional blindness as an example. In Simons and Chabris' (1999), participants were asked to count the number of times a basketball was passed by actors

[5] I am not saying that Shanahan and Baars are committed to this view. I simply bring Sperber's work into the picture as one way in which Shanahan and Baars might respond to my objections so that I can show why this too does not help their case.
[6] One may wonder: Why does Sperber use the term "module" where Shanahan uses the term "specialty process" and are these supposed to be interpreted as different things? Shanahan considers himself to be doing "post-reflective philosophical" work that shuns disembodied metaphysical reasoning, and in doing so, he also shuns metaphysically laden terms like "module." In doing so, he simply avoids terms like "module," "belief," etc. Furthermore, even though Sperber uses the term "module" he distances himself from the "Fodor module," i.e. modules with the properties Fodor describes in *Modularity of Mind*. For our purposes, we can think of cognitive modules and specialty processes as referring to basically the same thing.

in a video while a person in a gorilla suit unexpectedly danced through the scene, at one point passing *directly in the middle of the group of people*. Simons and Chabris found that about half of their participants failed to notice the unexpected gorilla, a very salient but unexpected event, while engaged in the other cognitive task (pg. 1069). Although the animal was clearly present in one's visual field, about half of the participants' animal detection module didn't process the information. Instead of all modules activating for every stimulus, modules perform their operations when the stimulus is salient enough for a perceiver to fixate on it (Sperber, 2005, p. 60). Once a perceiver fixates on something, the module(s) responsible for processing that information is/are mandatorily activated, and more energy is allocated to the module(s). Sperber argues that this must be the case because, in a massively modular system, there are innumerable inputs that are presented to modules, and the brain simply cannot process every input.

The brain does not fixate on certain stimuli or allocate energy randomly, however; rather, Sperber argues that the brain allocates energy beneficially, a way that attempts to maximize the benefits for the perceiver. One should naturally question *how* the brain comes to allocate energy beneficially—how does the brain know which inputs are the right inputs to focus on? This is the holism problem. In short, Sperber argues that the brain will allocate energy to the process that makes the most relevant,[7] most beneficial, and least costly inference.

[7] Sperber is using this term in a technical way, and, importantly, in a way that is different from the way the holism problem uses the term. The kind of relevance the holism problem refers to is the mundane, every day kind of relevance whereby the information that is relevant in a context is the most pertinent, on-topic, and helpful toward one's goal. Thus, unlike Shanahan, Sperber can avoid the circularity worry because he is not using the very same concept to respond to the holism problem. To avoid conclusion, when referencing Sperber's technical term, I will denote it as "relevance*".

According to relevance theory (Sperber et al.1995, Sperber & Wilson 2004, Sperber 2005), an inference is relevant* when it generates a unique conclusion in a specific context, a conclusion it would not generate if the context were not taken into consideration. In other words, a context-sensitive modular system, in part, decides relevance by allocating energy resources to the process that generates unique, beneficial inferences. Furthermore, Sperber and Deirdre Wilson (1996) argue that needs of efficiency also constrain cognitive systems. Specifically, a balance between expected effort and benefit (which are likely not to be calculated and explicitly represented, but expressed by “physiological indicators (Sperber, 2005, p. 65)) also constrains the competition over energy resources. In short, Sperber argues that modules can determine relevance by 1) engaging in a competition for energy resources whereby the winner of the competition provides a unique inference when considering a piece of information and the context and 2) the competition is constrained by cognitive efficiency (a balance between effort and benefit).

3.1.1.3 Objection: Recognizing Relevance

In the previous section, I expanded upon Shanahan and Baars’ picture of the global workspace theory so that we see exactly how the GWT system might perform relevance computations. Briefly, the solution so far is that the GW system solves the holism problem because its massively parallel architecture provides it with the computational power needed to sift through information and isolate what is relevant. I added to this by describing how the specialist processes might handle information: the specialist process automatically processes the relevant information via the lock-and-key analogy and what determines whether something is relevant is whether it generates a beneficial and relevant* (as defined by relevance theory) inference. I expanded the solution in this way because Shanahan and Baars (2005) seemed to hint that the GW system works in this way and because it would have been difficult to evaluate the solution without discussing how the specialist processes might perform relevance computations.

In this subsection, I will argue that an appeal to parallel processing—even if the GWT solution is expanded in the way I have proposed—pushes the problem back rather than provides a solution.

Objection to Shanahan and Baars. My main objection to Shanahan and Baars' solution is that merely appealing to parallel processing will not enable a system to recognize when a piece of information is relevant because it does not provide a system with the *epistemic tools* needed to recognize what is relevant.

First, Shanahan and Baars' (2005) solution does not provide the system with any satisficing conditions—any conditions to tell the system when to stop its competition. In failing to do so, Shanahan and Baars fail to provide the system with a means for a system to recognize what information is and isn't relevant. In other words, because Shanahan and Baars fail to satisfyingly address what it means for information to be relevant and the conditions under which a piece of information is relevant, the current GW system would not terminate the competition and thus would not output a relevance computation.

In addition to this, merely appealing to parallel processing fails to address the fact that relevance appears to require a global assessment of information. Recall that one cannot perform relevance computations in isolation from other information because relevance is sensitive to global properties of entire belief systems. The system simply needs access to more than just the information at hand for the coalitions to recognize when they have something relevant to offer. In other words, the information must be integrated in such a way that the specialist processes have access to the information they need to perform relevance computations.

OBJECTION TO SPERBER. Even if we appeal to the expanded version of GWT, the holism problem is not solved. In particular, Sperber's account pushes the problem back because it, too, does not satisfyingly explain how the system recognizes relevance. Sperber argues that a system solves the holism problem through competition for energy resources whereby the most relevant*, most beneficial inference that requires the least amount of cognitive effort to produce will win. Given that relevance* (and regular old relevance), what is most beneficial, and how much effort something requires are all context-dependent, there cannot be pre-assigned base values for each of these properties. Thus, the problem arises as to how a cognitive system ought to assign values to these properties: how does a system decide what is most relevant*, what is most beneficial, and what requires least effort to produce?

Does the cognitive system have to search through all the inferences and assign values to every inference and then select the one that meets the parameters? If so, we have the holism problem again. The solution to the problem shouldn't require a system to consider all the irrelevant information to find the relevant information; the system needn't be perfect and consider *only* what is relevant, but it certainly shouldn't consider all or too much irrelevant information. Sperber and Wilson (1996) note that there simply is information that we don't consider (pg. 530), so it appears they wouldn't prefer this method either.

I see two options: either these values are pre-wired in some sense (the holism problem challenges the possibility of this), or there must be something that tunes these values. If there is another mechanism tuning these values, we need an explanation of how this mechanism recognizes relevance. Even if there isn't a system that calculates expected values, per se, but rather physiological indicators set the values, there still needs to be a mechanism for tuning the physiological indicators. I am not arguing that values of expected relevance*, benefit, and effort must be explicitly represented for Sperber's account to work, but that there must be something that tunes and applies these properties. Whatever does this is not free from frame problems.

3.2 Global Broadcast as a Solution

Shanahan (2006) also realized that appealing to parallel processing was not enough. The issue Shanahan saw with his previous work with Baars (2005) was that the specialist processes were too informationally encapsulated, thus, he needed to describe how the specialist processes get access to the information they need to make relevance computations.

To address this problem, Shanahan (2006) added that the global broadcast mechanism should give the system access to the information it needs to perform relevance computations. In particular, *information from the environment* is disseminated to the specialist processes to use in performing their relevance computations. For further clarification of the view, see the diagram created by Shanahan and Baars (2005) that describes the relevance computations in GWT:

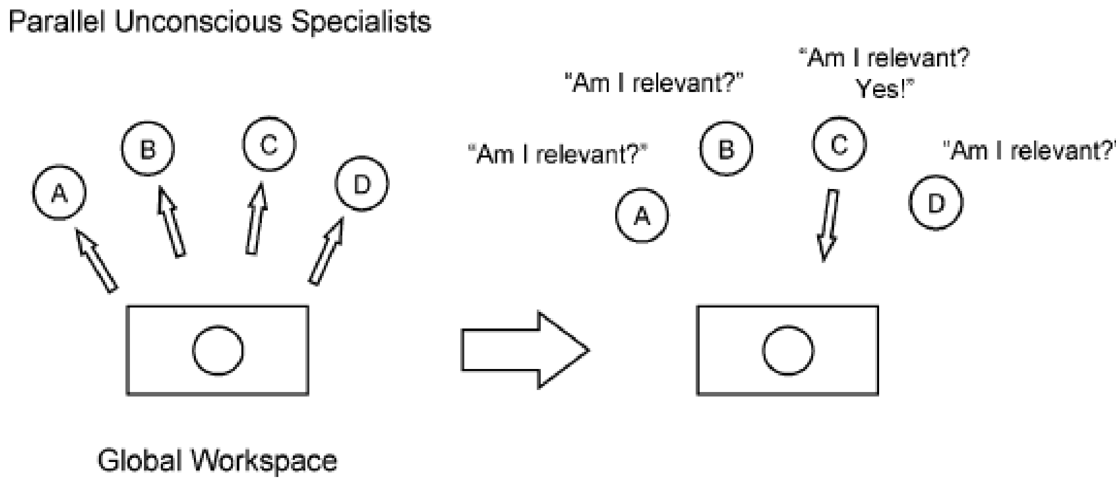

Figure 1: Information flow and relevance computations in the GW architecture. Photo Credit: Shanahan & Baars (2005) p. 169.

According to this addition, the GWT strategy involves employing global access to information *and* distributed, parallel processing to solve the holism problem. Accordingly, 0each specialist processes gain access to information about the system's current situation and environment (left side of the image), and then each specialist process runs in parallel, using the broadcasted information, to determine whether it, itself, is relevant (right side of the image). Once the specialist process determines it has something relevant to contribute, "it wakes up and starts to lobby for access to the global workspace" (Shanahan 2005 p. 65-57).

GWT is not the only architecture that appeals to global access for relevance computations. Barrett (2005) also describes an architecture that relies on global access to information via a "blackboard":

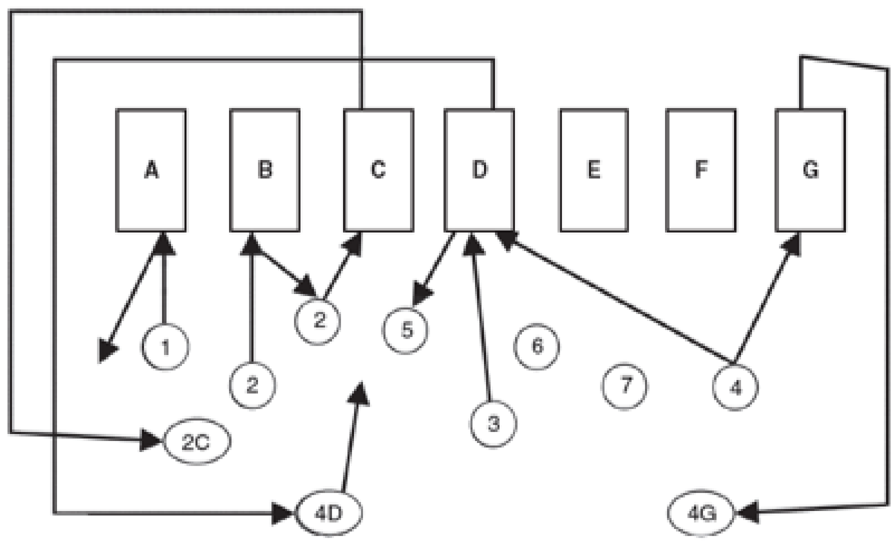

Figure 2: Barrett's EMS architecture.
Photo Credit: Barrett (2005) pg. 267.

The information in Barrett's enzyme-like modular system (henceforth EMS) is sent to a "bulletin board" that is monitored by all modules, and the information is used by whichever module can process it. The information in the bulletin board is available for all modules to use, and if the information is taken up by a module and unable to be used it is sent back to the bulletin board for reuse by other modules. Importantly, for Barrett, relevance computations are made possible by access to all potential information and by the lock-and-key method of bonding—if any information does not "fit into" a given module, then it is marked irrelevant.

Like Barrett, GWT proposes that global access to information via the global broadcast mechanism can help perform relevance computations. Global broadcasting, according to GWT, involves sending information from the environment to the GW, which is then sent to all specialist processes, and each process sifts through the information to determine what is and isn't relevant. If the broadcast information is something the module can't process, then the information is marked irrelevant. Shanahan and Baars (2005) describe how this might work:

> Consider the computational processes that might underlie the likening of a Rorschach inkblot to, say, an elephant. … the global workspace model posits a specialist, parallel process that is always on the lookout for elephantine shapes. This process is aroused by the presence of an inkblot that actually resembles an elephant, and it responds by announcing its findings. The urgency with which this process commends itself means that the information it has to offer makes its way into the global workspace, and is thereby broadcast back to all the other specialist processes.
>
> The Rorschach inkblot example is especially pertinent, since it is closely related to analogical reasoning. As Fodor emphasizes, reasoning by analogy is the epitome of an informationally unencapsulated cognitive process because it "depends precisely upon the transfer of information among cognitive domains previously assumed to be irrelevant" (Fodor, 1983, p. 105). Moreover, analogical reasoning is arguably central to human cognitive prowess (Gentner, 2003). So we should be able to offer a prima facie case that the global workspace architecture relieves the computational burden allegedly brought on by the informational unencapsulation of analogical reasoning. The Rorschach blot example is a start. (Shanahan and Baars, 2005, p. 168).

In this quote, like Barrett, Shanahan and Baars suggest that something like a lock-and-key method of search through broadcast information can determine what information is relevant. Importantly, broadcasting information about the environment to the GW is supposed to provide each specialist process with the contextual information it would need to perform relevance computations.

3.2.1 Evaluating Global Broadcast

While global broadcast does provide each specialist process with more information than mere parallel processing provided, I argue that an appeal to global broadcast does not aid in solving the relevance problem for two reasons.

OBJECTION 1: THE GOLDILOCKS PROBLEM OF BROADCAST. First, I argue that the GWT system does not broadcast the information needed to perform relevance computations. In a GW system, information *from the environment* is broadcast to the specialist processes, which might include the problem at hand and some other base conditions. To see why broadcasting information readily available in the environment isn't enough, consider the following example, "**STAYING DRY:"**

> **Problem:** it is raining outside, I don't want to be soaked by the rain, but I don't have an umbrella.
> **Base conditions:** I have a newspaper a poncho in my backpack.
> **Task:** How will I know to take out my poncho?

Something as simple as figuring out how to protect oneself from the rain turns out to be an extremely nuanced task because more facts than the ones presented in the environment are needed to perform relevance computations. In this case, for me to recognize that information about the poncho is relevant, I would have to know 1) that it's not appropriate to avoid getting my clothes wet by removing them, 2) that I should isolate *my own* newspaper and ponchos as potential water protection devices (i.e., it's not appropriate to snatch someone's umbrella), 3) while also selectively retrieving information from memory regarding how each item reacts with water (e.g., newspapers are not water-resistant, but ponchos are). In short, there exists information about societal norms and about physics that are wholly relevant to the situation at hand, but also not readily apparent through one's sensations of the environment.

Accordingly, instead of posing a solution to the problem, it appears that appealing to global broadcast or global access simply calls for a slight reformulation of the problem. A global broadcasting mechanism makes information available to specialist processes for access, but how does the global broadcasting system isolate which information is important for making the relevance computations? Not just *any* information will do, so the broadcasting mechanism ought to be able to distinguish which information gets broadcast and which does not. There is more than just one layer of explicit information in the highly complex process of reasoning—systems must know that certain information (i.e., societal information) is sometimes relevant in making other relevance computations. The question at hand, then, is how *this information* is made available to the specialist processes so that they can make their relevance computations.

OBJECTION 2: RECOGNIZING RELEVANCE IN FAMILIAR SITUATIONS. Shanahan (2010) argues that global broadcast enable a system to recognize relevant information in familiar situations. He argues this presumably because, if one previously made the correct relevance computation in the past, one would not have to go through the process of searching through information to find a relevance computation again in a similar situation—instead, one would just have to retrieve from memory the correct relevance computation. However, Shanahan concedes that global broadcast and parallel processing alone would not enable a system to recognize what is relevant in unfamiliar situations, which ultimately is his motivation for adding the third part of the solution that I will discuss in section 3.3. Shanahan argues that the GWT system thus far would not be able to recognize relevance in unfamiliar situations because, if information is sent to the modules using the lock-and-key method, then unfamiliar situations wouldn't be calibrated to activate certain specialist processes automatically.

In this objection, I want to assess whether GWT has what it takes to recognize what is relevant even in *familiar situations*.

Referring to my **STAYING DRY** example, Shanahan would argue that if I successfully realized that information regarding my poncho was relevant and I encountered a similar scenario in the future, I should be able to recall my poncho-related-information automatically. However, in the subsequent exposure to a similar scenario, to automatically recall the correct relevance computation, I would have to *recognize the* *scenario I'm in as familiar*. What, then, does it take to recognize a subsequent similar scenario as familiar? In short, recognizing a subsequent similar scenario as familiar requires the ability to perform relevance computations in the first place! It requires being able to pick out just the right situational information to compare to a prior situation and picking out the right prior situation with which to compare the relevant situational information. For example, in **STAYING DRY**, I would have to be able to isolate information related to the problem, base conditions, and tasks and recognize them as relevant objects of comparison to previous experiences I have had—and *not* focus on irrelevant information such as whether there are any squirrels present in both scenarios.

This issue appears to arise because two situations can be familiar or similar while not having the same properties. **STAYING DRY** 1 and **STAYING DRY** 2 could differ in the intensity of rain, location, what I'm wearing, who's around, etc., but still be similar. The task of recognizing **STAYING DRY** 2 as familiar and similar to **STAYING DRY** 1 thus requires ignoring the irrelevant different properties. However, because the claim that a GWT system can perform a relevance computation is so far under-motivated, we don't yet have reason to believe that parallel processing and global broadcast will even enable a system to perform relevance computations in familiar situations.

3.3 Chaotic Itinerancy as a Solution

In the previous sections, I worked through the first two steps in the GWT solution to the holism problem. First, parallel processing was supposed to solve the problem by providing the system with the computational power it needed to perform relevance computations. However, this solution failed on three fronts: 1) it accepted too narrow an interpretation of the problem, 2) it failed to explain how the specialist processes could recognize what was relevant in a given situation because 3) the information in the processes were not integrated in such a way that promoted global assessment of information. Thus, Shanahan (2006) added global broadcast to the solution by arguing the broadcast mechanism gave the specialist processes the information they needed to perform relevance computations. However, this again failed to provide a complete solution because 1) GWT proposes that environmental information is broadcast and made available to the specialist processes for their use in performing relevance computations, but that is not the right kind of information needed to perform relevance computations, and 2) recognizing that a situation is familiar—i.e., similar to a situation previously experienced—requires the ability to perform a relevance computation, an ability that has not yet been substantiated in the GWT system.

Relatedly, Shanahan saw a gap in the system's ability to integrate and *work together* to figure out what is relevant. In other words, Shanahan realized he needed to explain how specialist processes could communicate with each other and have access to any potential combination of specialist processes (e.g., he realized he needed to explain the synthesis of information in the GWT system).

Recognizing these flaws, Shanahan (2010 and 2012) proposed that the dynamical systems theory (DST) concept of "chaotic itinerancy"[8] (CI) could enable a system to integrate information in such a way that would fill in the final missing pieces of the GWT solution puzzle (p. 142). First, to define cognitive integration Shanahan (2012) writes, "perfect cognitive integration is achieved when the animal brings the totality of what it

---

[8] Shanahan is not just pulling this concept out of nowhere; the concept has been demonstrated to apply in different areas of cognition, including associative memory (Tsuda et al., 1987), olfaction in rats (Freeman, 1987), etc.

knows to bear on the ongoing situation—its grasp of the sensorimotor contingencies of multiple domains and its understanding of the associated affordances, plus the full contents of both its long-term (episodic-like) memory and its short-term (working) memory" (p. 2704). This definition of cognitive integration is important because it that explaining how cognitive integration could occur would address the seeming quineian and isotropic properties of relevance.

First, Shanahan (2012) argues that the GW system must engage in a "winner-takes-all competition" for access to the GW because the GW system's connective core can only contain a limited set of information (it has "limited bandwidth," p. 2709). Shanahan (2010) argues that the chaotically itinerant neuro-dynamics allows the system to engage in the winner-takes-all competition successfully. Briefly, Shanahan argues that chaotic itinerancy explains how specialist process coalitions are formed, broken up, and new combinations are formed "yielding a serial procession of global metastable states punctuated by transients" (Shanahan, 2012, p. 2710). Chaotic itinerancy (CI) is the mathematical pattern by which the GW system can integrate information in such a way that it results in a perfect combination of parallel processing and a serial procession of items of attention. In response to this, in this section, I argue that Shanahan fails to ensure that the winning coalition tracks the epistemic norm of *relevance* as opposed to something else entirely.

In the simplest terms, we can think of CI systems (e.g., a brain)[9] as systems that fluctuate from state to state via varying degrees of chaos, i.e., between states of chaos and ordered behavior (Shanahan, 2010, p. 141-143; Kaneko & Tsuda, 2003). CI occurs

---

[9] Brains are not the only systems claim to exhibit chaotic itinerancy. The concept is applied in physics and other non-cognitive and non-neural domains as well (Tsuda, 2013).

at the levels of individual neurons, neuronal networks, and at the level of a brain structure itself; it is hypothesized that the itinerancy at the higher levels within the brain is inherited from the CI at lower levels (Tsuda 1991). Tsuda (1991) describes the ordered behavior in a CI system as “history-dependent” activity whereby exposure to new stimuli initially causes chaotic activity within neurons, but the neurons eventually come to regularly follow certain, regular axonal pathways after learning occurs (p. 174). To make this clear, Tsuda (2013) describes CI in rat olfaction: “After the learning of several odor inputs, the activity of the olfactory bulb becomes a chaotic wandering among learned states; if an input is new, however, the activity converges to one of the learned states, i.e., the activity is represented by a limit-cycle attractor if an input has been learned previously [Freeman, 1987; Skarda and Freeman, 1987].”

Applying this to GWT and the holism problem, Shanahan (2010) argues that chaotic itinerant wandering between brain states is the mathematical pattern by which the serial procession of information contained in the GW is determined. Importantly, CI systems can quickly revisit the same state and move to multiple different successor states, which helps the GW system engage in an off-line exploration of possible outcomes and affordances within the environment (Shanahan, 2010, p. 142-144). To solve the holism problem, adding CI to the picture allows the GW system to foresee possible outcomes and search through information quickly and efficiently (p. 179). Along these lines, Shanahan (2010) writes:

> In order to search this [decision] tree effectively, it must be possible both to foresee the outcome of several actions or behaviors chained together and to investigate multiple possibilities branching away from a single state… To search through a combinatorially structured space of behavioral sequences, the system must be capable of revisiting a state it has already seen and generating a *different* successor. (pg. 142).

This is exactly what chaotic itinerancy brings to the table: the ability to generate new states from what was previously experienced or learned. Thus, it appears that, in addition to aiding memory and learning, Shanahan is arguing that chaotic itinerancy aids in conceptual blending and insight in such a way that it can enable to a system to sift between what is and isn't relevant in unfamiliar situations.

3.3.1 Evaluating Chaotic Itinerancy

In a personal correspondence with Shanahan, he concisely presents what he thinks chaotic itinerancy does in performing relevance computations:

> My hypothesis is that this kind of dynamics [chaotic itinerancy] allows all the active brain processes to enter into a kind of competitive / co-operative negotiation with each other to form the dominant coalition in the brain. There might be an ebb and flow between many transient coalitions on the way, with temporary alliances of brain processes forming and breaking apart, but eventually one will emerge as dominant and determine what the animal / person actually does. The key idea is that any process could potentially enter into a coalition with any other, thereby addressing the combinatorial issue here [the combinatorial issue that arises in recognizing relevance in unfamiliar situations].

In other words, Shanahan argues that the fact that chaotic itinerancy enables conceptual blending and enables an ebb and flow of coalitional formation entails that the system will eventually determine what is relevant in the process of this chaotic dynamics.

The issue with this is that appealing to CI merely describes the formal or mathematical dynamics of the system—how the systems move from state to state so that eventually a single process is isolated—but nothing about this means *relevance* would be the property being tracked. Shanahan assumes that CI will allow the system to eventually settle on a single coalition of specialist processes, that this coalition will win the competition and thus will be relevant, but this solution does not mean that the system could track *relevance.*

One might suggest that the very idea of a winner of a competition entails that the winner is relevant. Shanahan (2005) in fact takes this line of reasoning when he writes, “for the unfolding contents of a global workspace to make sense—for it to exercise any form of coherent influence on the world—the parallel specialist processes that contribute to it must serve a common remit. Even though these processes compete for access to the global workspace, the very idea of a winner only makes sense if there is a universal criterion for success, a criterion that selects the process or coalition of processes that is, in some sense, most relevant at any given time to an ongoing situation” (p. 62). However, this reasoning amounts to the claim that simply winning the competition *ipso facto* makes the winner relevant—which simply hasn’t been substantiated. It *would* make a lot of sense and be convenient if the winning coalition was the most relevant coalition, but the mechanisms underlying how this would occur has yet to be explained. While the very concept of winner might imply that there is *some* criterion for success, nothing in this implies that relevance is the criterion or precludes the option that there is an additional criterion for success that is not sensitive to nor does it track *relevance*.

The frame problem is first and foremost an epistemic problem. Creating a system that computationally can handle a large amount of information is necessary to solve the problem, but it certainly is not sufficient. As Chow (2013) puts it,

> Fodor rightly points out that (what I am calling here) the Epistemological Relevance Problem [what *I* call the holism problem] “goes as deep as the analysis of rationality'' (Fodor, 1987, p. 27). We can see why: the demands to be satisfied in determining what is relevant just are the demands of rationality. To repeat, humans appear to be largely rational—in Fodor's words, “witness our not all being dead'' (Fodor, 2008, p. 116) and determining what is relevant is part and parcel of rationality. (Chow, 2013, p. 12).

Shanahan’s line of reasoning lacks the thoughtful consideration of epistemic concerns about rationality that is required to solve the problem. Winning the competition doesn't mean relevance is the thing being tracked—i.e., the process that wins the competition might be tracking something else—so the system might be computing what is "relevant" in some technical sense but be missing the norms of relevance as they apply in everyday reasoning. Also, claiming that what wins the GW competition will be relevant seems to imply that everything the system focuses on is relevant, but people focus on irrelevant things all the time. If Shanahan and Baars want to claim that GWT is the realistic way that human brains solve the frame problem, then they need to explain how a system can focus on relevant things while still allowing for the possibility that the system might sometimes focus on irrelevant things.

The reason Shanahan added chaotic itinerancy to the picture was that Parallel Processing and Global Broadcasting did not explain how the specialist processes could create any potential, novel combinations of coalitions so that the system could generate insights and novel relevance computations. Chaotic itinerancy might certainly allow novel coalitions to form through its chaotic behavior, the ability to revisit a previous state, and the ability to generate long-lasting patterns of activation; however, the issue comes in when considering whether chaotic itinerancy is enough to bring about the *right* coalitions, the *relevant* coalitions. Neither CI, nor merely winning a competition for access to the global workspace, means that relevance will be tracked at all.

Make no mistake—I am not arguing that the cognitive architecture that solves the holism problem will guarantee that a system always forms the correct relevance computation or that the system is always focused on what is relevant. Instead, I am saying that the person who designs the architecture needs to clearly and accurately pin down that relevance is, as an epistemic norm, and then create an architecture that can allow a system to track that epistemic norm at least enough as humans do. It’s true that humans aren’t perfect at always considering only what is relevant in their immediate situation, so a system that solves the holism problem will not have to be perfect either.

4 OBJECTIONS & CONCLUSIONS

Shanahan and Baars began with arguing that parallel processing was enough to solve the problem, but over the years, they add more factors to their solution. In evaluating GWT, I analyzed each step along Shanahan and Baars made along the way to their final solution, namely, that a combination of parallel processing, global broadcast, and chaotic itinerancy solve the holism problem. After explaining GWT and the proposed solution, I expanded Shanahan and Baars' (2005) work by imagining how specialist processes might engage in their competition using work from Sperber (2005) and Barrett (2005). I argued that, if a GW competition determines relevance, then there needs to be a stopping point for the competition. However, an appeal to distributed, parallel processing alone (even when combined with the work done by Sperber and Barrett) failed to provide stopping conditions for the competition. Next, Shanahan reasoned that broadcasting information about the environment is supposed to help aid the specialist processes to determine whether they are relevant to the ongoing situation. However, I argued that this too pushes the problem back because simply broadcasting information from the environment leaves out key information needed to perform relevance computations—information that is not present in the environment alone. Hence, I argued that GWT's version of global broadcast does not do the job. Finally, Shanahan argued that he needed to appeal to chaotic itinerancy to fill in the remaining gaps in his view. Along these lines, chaotic itinerancy enables a system to generate new states from what was previously experienced or learned. I objected to this on account that a wandering between brain states and eventually settling on a single coalition following the mathematical dynamics of chaotic itinerancy fails to ensure that the winning coalition tracks the epistemic norm of *relevance* as opposed to something else entirely.

Next, I will consider a few objections to my conclusion:

OBJECTION 1: HEURISTICS AND EVOLUTION. One might suggest that a set of goals and needs, in addition to dynamic coupling with an environment, evolution, and learning, are enough to tune a system to recognize what is and isn't relevant. For example, if a system has the goal "don't get eaten by a bear," eventually evolution, trial-and-error, learning, etc. will teach a system that they should be on the lookout for bear sized and shaped paw-prints. I take this response as saying that evolution and learning, etc., can provide a system with *heuristics* that can either directly tell a system what is relevant in its situation or can at least significantly delimit (in Shanahan's (2010) words, "scaffold" (p. 177)) what the system can consider.

A response to OBJECTION 1 comes from *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, in which Fodor argues that any system that implements any number of heuristics is going to have some way of deciding which heuristic to implement over others. If evolution and interaction with the environment were to program heuristics into a cognitive system, the system would then have to decide which heuristics were relevant to its current needs. Using the example from above, the system would not have only one heuristic, namely "if there are bear shaped and sized paw prints, a bear is probably nearby;" instead, it would have a plethora of heuristics that are irrelevant to its current need. For example, a system might have other heuristics such as "if there's smoke, there's fire" or "if there's thunder, there's lightning" and so on. The system would then have to know which heuristics are relevant to its current needs and situation. These heuristics couldn't be mandatorily implemented in a system because there are conceivably many scenarios that present the bear paw-print, smoke, or thunder stimuli yet we don't run away in fear. People don't take out their umbrella and put in their rain boots if the movie they're watching includes thunderstorm stimuli, so there must be a way to override the heuristics—and thus its implementation can't be mandatory. Thus, even an appeal to evolutionarily programmed heuristics pushes the problem back.

Setting aside the negative, GWT is an important architecture because it attempts to get rid of central, serial control while still giving a mechanism (the GW) potential access to important information. On the surface, it seems like GWT provides what is needed to solve the holism problem. First, it is true that a system would need a way to search through information quickly, and parallel processing does provide a system with a way to quickly search through information through enabling a system to consider more than one piece of information at a time. Parallel processing is both an empirically substantiated view of how the brain works (Baars, 2002; Dehaene & Naccache, 2001) and is employed as a strategy in computing. Second, the global broadcast mechanism does seem to be a good way to give information to specialist processes (assuming that the mechanism first knows how to isolate the correct information to broadcast). Next, adding chaotic itinerancy to GWT does add something interesting and unique to the picture: the ability to "transition between a large repertoire of network configurations, allowing an exploratory cognitive state and the efficient response to changing external events" (Hellyer et al., 2014, p. 451). Shanahan (2010) suggests that CI could enable cognitive behavior such as establishing regular patterns of information flow, enabling the ability to revisit previous thoughts to come to different conclusions later, and even the generation of novel ideas. And finally, GWT—as a theory that postulates that parallel processing, a broadcast mechanism, and

CI are features of the brain—does have some empirical support and coheres with other empirical work.

I am not saying that GWT is neurophysiologically implausible; rather, if the brain solves the holism problem, then it is likely not in virtue of this version of GWT as currently presented by Shanahan and Baars. In other words, GWT in its current form is an incomplete project—but that doesn't mean it's a total failure. The core issue is that GWT simply has yet to provide a system with a way to reliably recognize any *relevant* information because it fails to address the epistemic issues that are a necessary part of finding a solution to the problem. The issue with the GWT solution as it currently stands is not that this kind of architecture is doomed never to perform a relevance computation, but that its founders failed to provide successful measures by which a system could recognize that something is relevant. In other words, Shanahan, Baars, and others have not provided successful, non-question-begging standards by which a system can determine that an item of information is relevant with a reasonable level of foresight and success.

REFERENCES

Baars, B. J. (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4*(4), 292-309.

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in brain research*, *150*, 45-53.

Baars, B. J. (2007). The global workspace theory of consciousness. *The Blackwell companion to consciousness*, 227-242.

Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind & language*, *20*(3), 259-287.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79, 1–37.

Dennett, D. (1987). Cognitive Wheels: The frame problem of AI. In Z. W. Pylyshyn (Ed.) *The Robots Dilemma: The Frame Problem in Artificial Intelligence (Theoretical Issues in Cognitive Science)* (p. 41-64). Norwood, NJ: Ablex Publishing Corporation.

Dreyfus, H. L. (1979). *What computers can't do: The limits of artificial intelligence* (Vol. 1972). New York: Harper & Row.

Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason*. MIT press.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Z. W. Pylyshyn (Ed.) *The Robots Dilemma: The Frame Problem in Artificial Intelligence (Theoretical Issues in Cognitive Science)* (p. 41-64). Norwood, NJ: Ablex Publishing Corporation.

Franklin, S. (2003). An autonomous software agent for navy personnel work: A case study. In *Human interaction with autonomous systems in complex environments: Papers from 2003 AAAI spring symposium* (pp. 60-65). Palo Alto: AAAI.

Franklin, S., Strain, S., Snaider, J., McCall, R., & Faghihi, U. (2012). Global workspace theory, its LIDA model and the underlying neuroscience. *Biologically Inspired Cognitive Architectures*, 1, 32-43. Chicago.

Freeman, W. J. (1987). Simulation of chaotic EEG patterns with a dynamic model of the olfactory system Biological Cybernetics, 56, 139-150.

Freeman, W. J. (2003). Evidence from human scalp electroencephalograms of global chaotic itinerancy. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *13*(3), 1067-1077.

Google, "The Traveling Salesman Problem." Last updated *October 13, 2017.*

https://developers.google.com/optimization/routing/tsp/tsp

Google, “Directions API Guide.” Last updated June 17, 2017. https://developers.google.com/maps/documentation/directions/start

Haugeland, J. (1985). *Artificial intelligence: The very idea*. MIT press.

Haugeland, J. (1987). An overview of the frame problem. In K. M. Ford & Z. W. Pylyshyn (Eds.) *The Robots Dilemma Revisited: The Frame Problem in Artificial Intelligence (Theoretical Issues in Cognitive Science)* (p. 61-88). Norwood, NJ: Ablex Publishing Corporation.

Hellyer, P. J., Shanahan, M., Scott, G., Wise, R. J., Sharp, D. J., & Leech, R. (2014). The control of global brain dynamics: opposing actions of frontoparietal control and default mode networks on attention. *Journal of Neuroscience*, 34(2), 451-461.

Henst, J. B. V. D., Sperber, D., & Politzer, G. (2002). When is a conclusion worth deriving? A relevance-based analysis of indeterminate relational problems. *Thinking & Reasoning*, *8*(1), 1-20.

Horgan, T., & Tienson, J. (1996). *Connectionism and the Philosophy of Psychology*. MIT Press.

Kaneko, K., & Tsuda, I. (2003). Chaotic itinerancy. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *13*(3), 926-936.

Lormand, E. (1996). The Holorobophobe’s Dilemma. In K. M. Ford & Z. W. Pylyshyn (Eds.) *The Robots Dilemma Revisited: The Frame Problem in Artificial Intelligence (Theoretical Issues in Cognitive Science)* (p. 61-88). Norwood, NJ: Ablex Publishing Corporation.

Shanahan, M. (2005). Global access, embodiment and the conscious subject. *Journal of Consciousness Studies*, *12*(12), 46-66.

Shanahan, M. (2006). A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and cognition*, *15*(2), 433-449.

Shanahan, M. (2009). The Frame Problem. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2016/entries/frame-problem/>.

Shanahan, M. (2010). *Embodiment and the inner life: Cognition and Consciousness in the Space of Possible Minds*. Oxford University Press, USA.

Shanahan, M. (2012). The brain's connective core and its role in animal cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1603), 2704-2714.

Shanahan, M., & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition*, *98*(2), 157-176.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059-1074.

Sperber, D. (2005). Modularity and relevance. *The innate mind: Structure and contents*, *1*, 53-68.

Sperber, D., & Wilson, D. (1996). Fodor's frame problem and relevance theory (reply to Chiappe & Kukla). *Behavioral and brain sciences*, *19*(3), 530-532.

Sperber, D., & Wilson, D. (2004). Relevance theory. *Handbook of Pragmatics. Oxford: Blackwell*, 607-632.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge University Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*(1), 31-95.

Tsuda, I., Koerner, E., & Shimizu, H. (1987). Memory dynamics in asynchronous neural networks. *Progress of Theoretical Physics*, *78*(1), 51-71.

Tsuda, I. (2013). Chaotic itinerancy. *Scholarpedia*, *8*(1), 4459.

Wheeler, M. (2008). Cognition in context: phenomenology, situated robotics and the frame problem. *International Journal of Philosophical Studies*, *16*(3), 323-349.