Integration without Integrated Models or Theories

Eric Hochstein

Forthcoming in *Synthese*

Abstract: It is traditionally thought that integration in cognitive science requires combining different perspectives, elements, and insights into an integrated model or theory of the target phenomenon. In this paper I argue that this type of integration is frequently not possible in cognitive science due to our reliance on using different idealizing and simplifying assumptions in our models and theories. Despite this, I argue that we can still have integration in cognitive science and attain all the benefits that integrated models would provide, without the need for their construction. Models which make incompatible idealizing assumptions about the target phenomenon can still be integrated by understanding how to draw coherent and compatible inferences across them. I discuss how this is possible, and demonstrate how this supports a different kind of integration. This sense of integration allows us to use collections of contradictory models to develop a consistent, comprehensive and non-contradictory understanding of a single unified phenomenon without the need for a single integrated model or theory.

Keywords: idealization, integration, unification, abstraction, ontic commitments, metaphysical commitments, implicit commitments, models, theories

In an interdisciplinary field like cognitive science, where different scientific domains approach the study of the mind from different perspectives, a lack of integration between these perspectives is often taken to be one of the major stumbling blocks in our study of the mind. Nicholas Cassimatis, for instance, claims that “the lack of integration among cognitive models is a severe impediment to achieving a comprehensive model of human cognition” (2005, p.402). Similarly, Poldrack et al. (2011) argue that a lack of integration among an ever-growing number of scientific theories and models presents “an increasingly critical challenge” to cognitive science (p.1). The importance of integration to cognitive science and the problems it faces are hardly new, and have been highlighted by theorists since the start of the cognitive revolution. As Salvucci notes, “the importance of integration was argued perhaps most famously by [Allen] Newell, who lamented the lack of integration in psychological theories and urged the community to focus on developing unified theories of cognition” (2013, p.830).

But what exactly is required for integration? In the context of cognitive science, one of the most common ways that integration is understood is in terms of whether we can provide a single integrated model or theory of the phenomenon that incorporates all the various insights provided by the different perspectives. For instance, Cassimatis in the passage above stresses that a lack of integration stops us from creating *a comprehensive model of human cognition*. Meanwhile Newell's proposed solution to a lack of integration is to create *integrated theories* of cognition. Salvucci similarly points out that in direct response to Newell, "many computational models have followed an integrated approach to account for various aspects of cognition" (Salvucci 2013, P.830).

In this paper, I argue that if we understand integration as the creation of *integrated models or theories* of a complex phenomenon, then integration quickly becomes unachievable in much of cognitive science due to both practical, and representational, limitations. However, there is a broader sense of integration we can appeal to which can provide all the same benefits. Instead of directly merging our representations into a single model or theory, we can provide a coherent and comprehensive understanding of complex phenomena by learning what sorts of inferences to draw *between* various incompatible models and theories of the same phenomenon. Scientists using models which make contradictory claims can often share certain unspoken or implicit commitments about the phenomenon in the world that is not explicitly reflected in the model itself. By identifying these shared commitments, we can find points of contact between the various models which we can use as a bridge, allowing us to draw inferences across models. This in turn allows the various models to inform, refine, constrain, and build on one another in a coherent manner. As a result, we can appeal to a different kind of integration. This sense of integration allows us to use collections of contradictory models to develop a consistent, comprehensive and non-contradictory understanding of a single unified phenomenon without the need for integrated models or theories.

In order to make this argument, I will start in Section 1 by elaborating on the common assumption that integration in cognitive science involves the creation of integrated models or theories. In Section 2, I demonstrate why this sense of integration is unlikely to be forthcoming in many cases. In Section 3, I explore alternative ways of thinking about integration, and its benefits. In Section 4, I defend a different sense of integration and explain how it is possible to use models which make contradictory claims about a phenomenon to inform, constrain, and contribute to, a non-contradictory understanding of it. In section 5, I provide explicit examples from cognitive science to highlight this alternative account of integration. Lastly, in Section 6, I explore the implications of this for future work in philosophy of science and cognitive science.

1. Traditional Account of Integration in Cognitive Science

What does it mean to "integrate" multiple theories or models in the context of cognitive science? It is commonly thought that this involves combining the different accounts into a single over-arching representation. For a clear expression of this idea, consider Holland, de Regt & Drukarch's claim regarding the neuroscientific study of nerve impulse propagation. They note that it is…

> …argued in current (neuro)scientific literature that views focusing on membrane proteins and membrane lipids should be integrated in a general unifying model. Such a general unifying model is developed to incorporate, integrate and explain all relevant aspects of the nerve impulse by unifying different manifestations of the nerve impulse and the interaction(s) between them. An important argument for developing such a general unifying model is

> to obtain insights in nerve impulse propagation that cannot be acquired using models that focus on only one or a few aspects of the nerve impulse without studying the interactions between them. (2019, p.2)

Or consider Lemerise & Arsenio (2000), who argue that "it is vitally important for developmental psychologists to take a broader view of children's social and cognitive development, and [that] an essential aspect of this broader view involves considering, both theoretically and empirically, how emotional and cognitive processes can be integrated in models of social competence" (p.107). Meanwhile David & Szentagotai (2006) caution that a lack of integration in the field of cognitive behavioural psychotherapy poses a potential threat to the efficacy and effectiveness of such therapies, claiming that this lack of integration is due to a "lack of a coherent and integrative theory" (2006, p.284). Marraffa & Paternoster (2013) even claim that in the pursuit of integration, "future research should mainly be devoted to the development of really integrated models." (p.39).

Even more generalized discussions of integration in cognitive science implicitly presuppose this view. Take Glenn Gunzelmann, who claims that:

> The motivations for integrative models of human cognition have their roots in the origins of cognitive science as a scientific discipline. Even before cognitive psychology emerged, ideas about unifying principles to explain cognition were expressed in the scientific literature, including so-called grand psychological theories proposed during the first half of the 20th century. [...] In the decades since, integrative theories of human cognition have become increasingly prevalent in cognitive science. (2013, p.30)

Or consider Schoelles et al. (2006), who respond to Allen Newell's call for greater integration in the following way:

> Newell also warned us that an unremitting focus on isolated components of cognition would never enable us to see how these components fit together. We argue that Newell was right and that the time to build integrated models of cognitive systems is now. In some sense, our position is neither bold nor novel as there are many examples of other researchers engaged in much the same enterprise. (p.760)

This way of thinking about integration is so common in cognitive science that it is often not explicitly defended or argued for. However, I want to suggest that this is frequently not a helpful way of thinking about integration in cognitive science, and a somewhat different sense of integration is more useful.

2. The Virtues of Incompatibility

The essential role that idealizations and distortions play in scientific modelling has been well documented (for just a small sampling, see: Batterman 2001; Mitchell 2002, 2004; Plutynski 2013; Weisberg 2013; Levy & Bechtel 2013; Longino 2006, 2013; Chirimuuta 2014; Potochnik 2015, 2017; Hochstein 2016a, 2016b; Craver & Kaplan 2018; Holland, de Regt & Drukarch 2019). Put simply, we are often forced to represent the world in idealized ways that we know to be incorrect for different practical and representational purposes. A consequence of our dependence on such idealizing assumptions is that we are frequently left with collections of models that describe the same complex phenomenon in very different and often contradictory ways. Attempts to integrate all these different models into a single representation becomes virtually impossible given the conflicting idealizing assumptions made by the various models. As Angela Potochnik puts it:

> Combining representations of disparate influences into a single unified representation is oftentimes impossible, if only because of the limitations of our modeling techniques and computational powers. […] In our complex world, visits to so many scientists engaged in different pursuits would not culminate in a full explanation of anything. At best, this would yield a hodgepodge of partially related representations, with no way to reconcile their incompatible assumptions and no way to combine their mathematical frameworks. (2017, p.18)

One reason why such integrated models are frequently not possible is because it is often *only* by idealizing or abstracting away from certain features of a complex system that other features of it can come into focus or be represented accurately. In these contexts, to do away with the idealizations in our models would be to undermine their very ability to represent the different aspects of the phenomenon we use them to identify (see: Batterman 2001; Hochstein 2016b; Potochnik 2017).

There are also more mundane reasons for our dependence on idealized models. Even if a single integrated model or theory of a complex cognitive phenomenon were possible in theory, such a model is often not plausible in practice as scientists must always work within time and resources constraints. Thus, even in cases where we can create a single integrated model or representation of a complex phenomenon in principle, it is not always economically or computationally feasible, nor in our best interest, to do so. Instead, we are dependent on using collections of simplified models that each allow for more useful and economical representational tools, but which do not cohere together. Michael Weisberg, who refers to this practice as "Multiple Model Idealization" (MMI), notes that this is common throughout the sciences:

> In ecology, for example, one finds theorists constructing multiple models of phenomena such as predation, each of which contains different idealizing assumptions, approximations, and simplifications. Chemists continue to rely on both the molecular orbital and valence bond models of chemical bonding, which make different, incompatible assumptions. In a dramatic example of MMI, the United States National Weather Service (NWS) employs several different models of global circulation patterns to model the weather. Each of these models contains different idealizing assumptions about the basic physical processes involved in weather formation. Although attempts have been made to build a single model of global weather, the NWS has determined that the best way to make high-fidelity predictions is to employ all three models, despite the considerable expense of doing so. (2013, p.103)

This presents a problem for the traditional account of integration. If integration requires that we can capture the insights of our various representations within a single model or theory, then this is frequently not impossible. Cognitive science is no different in this regard. While we have certainly been able to create integrated models of certain limited kinds of cognitive phenomena, the assumption that such integrative models will always be available, desirable, or that scientific progress in cognitive science depends on them, is not something we should accept uncritically.

Take the study of the action potential of the neuron. The study of the action potential involved the use of a collection of highly idealized models and theories from many different domains, including: electrophysiology, chemistry, biology and physics. This resulted in a collection of models which described different aspects of the action potential in incompatible ways, with no way of merging them all into a single comprehensive representation (see: Trumpler 1997; Hochstein 2016a; Holland, de Regt & Drukarch 2019). Similarly, the study of phenomena like aggression and sexual orientation in cognitive science involved the application of idealized and incompatible models from domains like neurobiology, social psychology, dynamical systems theory, and behavioural genetics (Longino 2006, 2013).[1] All of this means that under the traditional account of integration, our ability to generate integrated models of complex cognitive phenomena is extremely limited. So what does this say about the quest for integration in cognitive science?

3. What is Integration, and What Do We Want From it?

If it is frequently not possible to provide a single coherent and comprehensive model or theory of a complex phenomenon, does this make the project of integration impossible in cognitive science? It all depends on how we understand "integration", and what we hope to gain from it. As William Newell (not to be confused with Allen Newell) points out, the idea of integration is not as straightforward as it may seem, and many scientists are "not even clear on exactly what is *meant* by integration" (2001, p.18). Indeed, scientists talk about integration in very different ways. O'Rourke, Crowley & Gonnerman (2016), for example, argue that integration takes "inputs that are not integrated and puts or brings them together into an integrative relation", yet they point out that what counts as an "integrative relation" is described by scientists as everything from fusing, to melding, to amalgamating, to knitting, to the linking of different kinds of knowledge, to making sense together, to assembling (p.68). Some of these relations (e.g. "amalgamating", "assembling") can often imply the combination of different perspectives and accounts into a single model or representation, yet O'Rourke, Crowley & Gonnerman are quick to point out that such interpretations have often been rejected by scientists on the grounds that they misleadingly imply that various models and representations can be easily fit together, which is not the

[1] Marcin Miłkowski, following Ioannis Votsis, refers to such attempted integrated models in cognitive science as "monstrous". Specifically, a model or theory is considered monstrous "if it contains 'isolated islands' that are confirmationally disconnected, i.e., what these 'islands' imply is completely disjoint." (Miłkowski 2016, p.21).

case (2016, p.68). Meanwhile, other senses of integration (e.g. “making sense together”) do not, at least *prima facie*, seem to require that we construct a single comprehensive representation that includes the insights of the various perspectives. On a similar note, Gabriele Bammer defines integration as…

> …experts from several disciplines plus stakeholders working on a common complex real-world problem in a way that not only brings together their insights but also deals comprehensively with unknowns, all in order to support policy and practice change. (2013, p. 9).

Notice how this definition of integration similarly does not imply that we must develop a single comprehensive representation of some complex phenomenon.

We find a very similar idea advocated by Sandra Mitchell in the context of biology and evolutionary theory, with her notion of “integrative pluralism” (2002, 2004). According to Mitchell, when studying a complex phenomenon, we cannot integrate our various models into a single coherent model that includes all the insights of the different models. However, when we apply our various incompatible models to study different aspects of a single concrete system in the world, the various models will allow us to answer different questions about the same phenomenon, identifying different causal features, and constraints, which altogether allow us to develop a more comprehensive understanding of the system than any of the models in isolation could provide. She ultimately concludes that “the type of integration that can occur in the application of models will itself be piecemeal and, to varying degrees, local” (2002, p.68).

Catherine Stinson (2016) has defended a similar view of integration in the context of cognitive science, rejecting what she calls a “seamless” view of integration and instead adopting a view in which “the integrations we can realistically expect are more partial, patchy, and full of loose threads.” (Stinson 2016, p.1585). Using the study of short-term memory as an example, she argues that cognitive models may identify entities and relations that do not correspond to neural structures and processes identified by neurophysiological models (thus blocking a seamless combination of the two types of models). However, she argues that this does not undermine integration. Instead, she notes that:

> Neuroscientists may not any longer try to find brain parts that correspond to anything like a short-term memory store [identified by many cognitive models of short-term memory], but they do look for processes that are necessary and/or sufficient for short-term memory traces of specific kinds. For example, Zars et al. (2000) is just one of many papers investigating short-term memory in Drosophila. They investigate an enzyme that mediates synaptic plasticity for olfactory learning. This is the sort of integration occurring in cognitive neuroscience. (Stinson 2016, p. 1606)

She proposes that we can integrate the findings of neurophysiological models (identifying the role that certain enzymes play in the mediation of synaptic plasticity), with those of the cognitive models (describing short-term memory in terms of more general processes such as encoding, decoding, and storage) by using the first to identify features or aspects of the neural system that informs our understanding of how short-term memory, as described by the second, is made possible. The various models can mutually inform each other given that "each of these provides a piece of the story about how memory works, highlighting one or more of its important features, but none gives a complete picture." (Stinson 2016, p.1611). Stinson ultimately concludes from these examples that integration should not be about seamless fit between different models, but instead "integration can mean figuring out how several partial models of the same or related phenomenon connect" (Stinson 2016, p.1611).

I think that Stinson is correct, and that she points the way towards what a more realistic theory of integration must do. But the question of how exactly incompatible, or worse *contradictory*, models can be used in this way to contribute to a coherent account of a unified phenomenon, is not something that Stinson provides a comprehensive answer to (as this is not her primary focus in the paper), and so she remains vague on the details. And it is precisely these sorts of details that seem to worry many theorists who emphasize the importance of integrated models in cognitive science. Consider the following passage by Wayne Grey:

> To the extent that the cognitive system is more than a bushel of independent mechanisms, then we need to understand how these mechanisms are integrated to achieve cognitive functionality. Integrated models of cognitive systems are a necessary tool in understanding a cognitive system that is integrated. (2007, p.12)

Or recall Schoelles et al.'s (2006) concern that "an unremitting focus on isolated components of cognition would never enable us to see how these components fit together" (p.760), or Holland, de Regt & Drukarch's claim that "an important argument for developing such a general unifying model is to obtain insights in nerve impulse propagation that cannot be acquired using models that focus on only one or a few aspects of the nerve impulse without studying the interactions between them" (Holland, de Regt & Drukarch 2019, p.2).

The benefit of developing integrated models in cognitive science seem to be that they show us *how* the various distinct models we have developed connect and link together to say something coherent about the system, and how the various entities posited by these models relate to one another in the production of the phenomenon. Such integrated models thereby provide a more consistent and comprehensive understanding of the entire integrated cognitive system than any of the component models can provide.

In the section that follows, I want to demonstrate how we can supplement the insights of Stinson (2016), Mitchell (2002, 2004), and others to address these sorts of worries. More specifically, I provide an account of exactly *how* we are able to take various models that make contradictory claims, and use them to inform, build on, and connect to one another so as to provide a coherent and comprehensive account of the phenomenon without the creation of an integrated representation into which they all must cohere. In this respect, I propose we can attain all the benefits that cognitive scientists seek from integrated models without their construction.

4. Integration Without Integrated Models or Theories

I propose that integrating insights from different scientific models and theories together in a coherent manner does not require the construction of an integrated model or representation. Instead, it only requires understanding how we can draw coherent *inferences* about the same target phenomenon *across* those models. But how is this possible? If we are forced to rely on collections of models which idealize the system in incompatible ways, then it would *prima facie* seem that we cannot draw inferences across these models on penalty of logical contradiction. As a result, each model seems isolated from the others, with each providing information about some restricted aspect of the phenomenon, but without any way of knowing how these different models relate or connect. This would seem to leave us with a kind of radical pluralism, where the various models are completely divorced from one another, used for entirely different purposes. Yet, as Love & Lugar (2013) rightly point out, we should not jump to this conclusion:

> While there is clear warrant for versions of pluralism in response to theory reduction's failure to capture the heterogeneity of reasoning in biology, this has encouraged a neglect of how the coordination of methods, concepts, and other epistemic units occurs. (p.548)

To understand how this is possible, we must understand that our ability to draw inferences across contradictory models requires going *beyond* what is explicitly stated by any model to examine the metaphysical and ontological commitments held by the working scientists who construct and apply such models. As David Danks rightly notes:

> One challenge with computational models of the mind is that it is often unclear just what commitments — metaphysical, epistemological, and methodological — are intended for a particular theory. The mathematics of a cognitive theory are insufficient to know exactly what the theory does and, often just as importantly, does not imply, whether about the world, the cognitive scientist, or future experiments. As a result, we are often left in the position of not knowing whether some empirical data confirm, disconfirm, or are simply irrelevant to a particular computational cognitive theory. (2014, p.13)

Scientific models are not constructed or applied in a vacuum, and working scientists often must start with metaphysical and ontological commitments about the world which are used to frame and structure how their scientific models are built, and what inferences are appropriate to draw from them (for more, see: Hochstein 2019). The implicit ontic and metaphysical commitments of working scientists determine the context of application for their models and regulate which inferences they are warranted in drawing from, and to, those models. Consider that in order to judge some feature of a model *as an idealization*, we must already be committed to certain ontological and metaphysical facts about the way the world is in order to determine what parts of the model deviate from it. As a straightforward example, take the use of infinite population sizes in evolutionary biology. Many evolutionary models frequently invoke infinite population sizes as a way to mathematically determine what sorts of traits would be locally optimal for a particular organism in a particular environment to have. Yet, it would be a mistake to conclude from the success of such models that population sizes must really be infinitely large simply because the model treats them as such. This is not an inference that biologists are warranted in drawing from such models, nor do they. This is because if we look beyond what is explicitly stated in the model, to the implicit ontological and metaphysical commitments held by biologists, we find a commitment to the idea that real populations in nature are never infinite, and so it would be a mistake to draw that inference from such models. Instead, we use such models only to draw inferences about what traits would be locally optimal for an organism in a particular environment to have. Working scientists thus use their frequently unspoken or implicit ontological and metaphysical commitments not just as a guide for when and how to apply our models, but also for which inferences we are licensed to draw from them, and which we are not. Acknowledging and keeping track of these implicit commitments can be the key to understanding how incompatible models connect to one another.

Even when models make contradictory claims about the same system, we can still draw inferences between those models by identifying the shared ontic or metaphysical commitments that are implicit in their creation and application. This allows us to find points of contact between incompatible representations that we can use as a bridge to move between models. In essence, we draw inferences across incompatible models by moving through the shared implicit commitments that working scientists take on board when building and applying these models. By doing this, it gives us a means of allowing different kinds of models, theories, and methodologies to constrain and inform one another, and thus to improve our overall scientific understanding of a single unified phenomenon, without worrying about direct integration of the models or theories themselves. This captures the benefits we seek to attain through the development of integrated models by providing us with a coherent, non-contradictory, and more comprehensive understanding of the phenomenon than can be provided by any model in isolation, yet it does not require a single integrated model to provide this. But how does this apply in the context of cognitive science? Let us consider some examples.

5. Examples from Cognitive Science

5.1 Integrating Models of Place Cells and Grid Cells with Models of Cognitive Maps

It is widely held that the hippocampus and the entorhinal cortex play essential roles in the construction of cognitive maps, spatial representations of the environment that animals use for navigation. The construction of these cognitive maps involves the interaction of (among others) two different kinds of cells: grid cells and place cells. Each of these encode different kinds of spatial information about the organism's environment, which are then combined in the formation of cognitive maps. Now suppose we wish to provide an integrated account of this phenomenon. It turns out our ability to scientifically study this cannot be captured by any single model. This is because numerous contradictory models are required to describe what information exactly is being encoded by the different cells, how the mechanisms of the cells allow them to engage in this task, and how this information is being combined in the construction of cognitive maps. One reason for this is because computational models used to describe the information processing tasks of the individual cells and how they are combined must, by necessity, distort or leave out essential information regarding how the mechanisms of the place cells and grid cells work to produce such computations, as well as how such cells causally interact.

Computational modeling's dependence on idealizations and abstractions have been well documented (see, for example, Sejnowski et al. 1988; Sterratt et al. 2011, p.316; Chirimuuta 2014). This dependence means that computational models of neural circuits cannot be directly integrated with models that describe how the cell physically functions to engage in such computations, which in turn cannot be directly integrated with computational models that describe the way in which this information is combined in the construction of cognitive maps.

To illustrate, suppose we wish to model the physical mechanisms of the grid cell to understand how it engages in information processing. Empirical studies have established that "a neuron's response to activation of a synaptic conductance is determined by its dendritic morphology, and by the identity and subcellular location of its membrane ion channels" (Garden et al. 2008, p.875). In other words, to understand how the neuron will respond to particular inputs, and produce particular outputs, we need to model the dendritic morphology of the neuron. We also know that "neurons important for cognitive function are often classified by their morphology and integrative properties." (Garden et al. 2008, p.875). Thus, place cells and grid cells differ in their morphological characteristics. So to understand how place cells and grid cells compute and interact, we need to model the essential morphological and physiological characteristics of the two cells.

While these models are essential for understanding *how* the different neurons produce their relevant outputs, they do not represent the dynamic electrical properties of the output spike trains themselves. Dynamical models are used to represent these features, and computational models are used to characterize them in terms of information processing. However, when we switch to these models, the essential details of the cells needed to understand how they produce these features must be idealized or ignored.

For instance, Savelli & Knierim (2010) propose a computational model which describes how information is passed from grid cells and place cell, and how this information is combined. However, in describing their model, they note that "both place cells and their grid cell inputs were simulated as spiking units. Place cells of the hippocampus were modeled as integrate-and-fire neurons" (p.3168). The integrate-and-fire model is a highly idealized model of the neuron that distorts a great deal of neuronal structures and dynamics. Likewise, consider Yuan et al.'s Entorhinal-Hippocampal Model for Simultaneous Cognitive Map Building (2015). They propose a computational model for how information from grid cells and place cells can be integrated in the formation of cognitive maps, and demonstrate how a robotic system can use the same principles to navigate its environment. Their model is highly abstract and idealized however, simplifying a great deal of information relevant to how real-world neurons behave. Verena Hafner (2000) also provides a computational model to explain the development of cognitive maps, one that is directly inspired by the discovery of place cells in the hippocampus. The model involves using a neural network to demonstrate how information can be combined by a network system to form maps needed to navigate in an open environment. However, as Hafner notes regarding the model, "the neurons and synapses in this model are idealised and should not be regarded as biologically realistic neurons and synapses" (2000, p.801). Even computational models of the grid cell by itself are highly idealized or abstract, focusing only on certain features of the neuron while ignoring and distorting others (see: Giocomo, Moser & Moser 2011). Both computational and structural models are necessary to provide a more comprehensive understanding of how place cells and grid cells interact to construct cognitive maps, but we cannot fit them into a single coherent representation of the phenomenon.

Does all this suggest that we have no integrated account of how place cells and grid cells interact in the construction of cognitive maps? I propose not. Despite the inability to integrate these various models into a single representation, scientists are able to draw coherent inferences between these various models given shared metaphysical commitments. The first such commitment is that the action potential of the neuron plays an essential role in information processing. As Mazviita Chirimuuta points out, "a central dogma of neuroscience is that the action potential is for information transmission, and that information theory and computational principles are needed to understand why action potentials have particular patterns of generation" (2014, p.131). The second is that the physical structure of the cell, and its computational behaviour, are tightly linked. Philipp Haueis, for instance, describes this second commitment in terms of the often-used slogan "structure determines function" (2022, p.13). For further discussion of this particular metaphysical commitment, see also: Bechtel & Mundale 1999; Eliasmith 2002; Piccinini & Craver 2011; Hochstein 2016b.

Let's consider how these metaphysical commitments allow neuroscientists to draw inferences between various models that describe the interaction between place cells and grid cells in the formation of cognitive maps. Garden et al. (2008) argue explicitly that physiological and structural models of grid cells provide essential constraints on the development and understanding of computational models. Specifically, they claim:

> The dorsal-ventral organization of synaptic integration that we describe is consistent with abstract single-cell and network models for generation of grid-like firing fields. In both cases our data suggest biophysical constraints on how these models might be implemented. (p.887)

Giocomo, Moser & Moser (2011), meanwhile, point out how such coherent inferences can go in the other direction, highlighting how the use of computational models allows us to draw coherent inferences about its underlying physiology. In their words:

> Computational models have been particularly important in the search for mechanisms of grid cells. Theoretical models have for example highlighted the potential role of multiple single-cell properties, such as oscillations and after-spike dynamics, in grid cell formation. With the introduction of in vivo whole-cell patch-clamp and optogenetic methods, the role of these properties can be tested. Direct and controllable manipulation of intrinsic oscillation

> frequencies, the timing of synaptic inputs, or the spiking dynamics of identified grid cells would provide paramount insight into what mechanisms contribute to the formation of spatially responsive neurons. (p.599-600)

These metaphysical commitments even allow inferences from more abstract computational models of cognitive maps to more detailed physiological models of grid and place cells. Hafner, for example, argues that while her computational model of cognitive map formation works with idealized neurons that are not biologically realistic, she also claims that "many aspects of the presented cognitive map model are biologically plausible and extend existing work." (2000). The idea being that we can integrate this model with other more biological plausible models and existing research by understanding *which* parts of the models to draw inferences from, and which we should not. By using biologically plausible models as a guide, it allows Hafner to generate computational models which can build off them, *even if her model itself works with neurons that are not biologically plausible.* We can use morphological or structural models of place and grid cells to understand how the neuron is structured, and thus how it produces its output spike sequences. This information can then be integrated with computational models of the cell by understanding how such spike sequences can be used to encode and decode relevant information about the organism's spatial orientation. We can subsequently integrate *this* information with more large-scale computational models regarding how the information from these neurons are combined and bound together to form cognitive maps. Integration can be achieved without the need for a comprehensive model which combines all these elements together in one representational format.

5.2 Modeling Alzheimer's Disease and its Effects on Cognition

As a second example, consider our attempt to provide an integrated account of Alzheimer's Disease (AD), and how it impairs cognitive functioning. AD is a neurodegenerative disorder in which a buildup of proteins around neuronal cells causes severe damage both to the cell itself, and to its connection to other cells, inhibiting its ability to send and receive signals. This results in a loss of cognitive abilities in everything from memory, to language, to movement. The study of AD has required the use of many different kinds of models, including (but not limited to): neurobiological models of the effected neurons and how they are damaged by the disease (Sompol et al. 2009; Penney, Ralvenius, & Tsai 2020), and computational models regarding how information processing needed to carry out cognitive tasks is impeded by the disease's progression (Horn 1993; Duch 2007; Saraceno et al. 2013). Importantly however, effectively modeling AD must also include network models to highlight how various brain regions are organized and interconnected in order to share information needed to carry out more complex cognitive tasks, and how such organizations are damaged by AD.

We've already discussed how structural or morphological models of the neuron cannot be directly integrated with many computational models given the necessity of invoking distinct idealizing assumptions and abstractions. This problem is further compounded when we introduce network models into the mix. First, it is helpful to provide a general overview of how network models work:

> These consist of graphs that depict a particular network in the biological brain: Nodes correspond to network elements, and edges correspond to network connections. Once a graph has been constructed, its local and global organization – its topology – can be analyzed using the tools and concepts of graph theory. These can be used to identify hub nodes (nodes with a relatively large number of edges) and motifs (local patterns of connectivity that are repeated throughout the network). They can also be used to determine the degree of clustering or modularity in a network (the degree to which nodes are arranged into densely interconnected

> communities of nodes that are sparsely connected to other communities), and for measuring the overall density or degree of randomness of a network's connections. (Zednik 2019, p.27)

Such models often play an essential role in characterizing the way in which large scale neural systems are organized. Network models of this sort play an important role in understanding how AD affects the organization of the brain, and the way in which different regions coordinate action and integrate information (see: de Haan et al. 2012; Hutchison et al. 2013).

The thing to note about network models is that by necessity they must idealize or abstract away from the details of the neural components that make up the network. For instance, Levy & Bechtel point out that "to understand organization, one often needs to abstract from the structural specifics of a mechanism and represent it in a skeletal, coarse-grained manner. […] In this form of modeling, the pattern of causal relations within a system is highlighted, while structural aspects of components are suppressed" (2013, p. 241). Similarly, Carlos Zednik notes that we do not use network models of this sort to represent detailed structural or behavioural features of neural components, since "component operations such as the activity of individual neurons are instead represented in dynamical models of individual neuronal activity" (Zednik 2019, p.35). In this regard, we cannot combine computational models of information processing carried out by individual neurons, models which highlight anatomical details of neurons or brain regions, and network models which identify essential organizational features of the brain, into a single representation. Thus, a single integrated model of the effects of AD on cognition is unattainable. Despite being unable to create a single unified representation, we can still integrate these models by understanding how to draw inferences *across* them to say something coherent and comprehensive about the progression of AD, its effects on the brain, and its influence on cognition.

To demonstrate how this is possible, let us examine how these types of models can inform one another more generally before turning to the case of Alzheimer's Disease in particular. Consider Kitano & Fukai (2007), who highlight that:

> Different wiring topologies give rise to different degrees of clustering among neurons (how densely a population of neurons is mutually connected) and different mean path lengths (on average, how many synapses are present along the shortest path connecting a neuron pair), both of which can influence the temporal structure of synaptic inputs to each neuron in a population and hence that of the output spike sequences. (p.237-238)

If we adopt the same metaphysical commitments as before, then we have a means of understanding how to draw inferences across neuronal morphological models, computational models, and network models. We know that the synaptic inputs, and the output spike sequence of a neuron depends on its structure and morphology. We also know that such processes are how information is encoded and processed by the cell. We likewise know that such input/output relations change depending on their place within a given network. This gives us a point of contact between the various models that allow us to understand how each can relate to the other and contribute to a coherent understanding of the system without there being a single representation into which they fit.

We are now in a better position to see how all these models can be integrated in the study of Alzheimer’s Disease. In AD, a buildup in protein damages not only neuronal cells, but the connections between them. This means that more detailed physiological models of individual neurons affected by the progression of AD can inform our understanding of how the synaptic inputs and spike-train outputs of those neurons will change as a result. This, in turn, influences our understanding of how the information-processing capacities of those neurons is inhibited, allowing us to thereby draw inferences to computational models regarding how cognitive phenomena like memory becomes degraded. For instance, Horn et al. (1993) develops a computational model of memory deterioration based on synaptic deletion and compensation. Likewise, network models provide essential information as to not only how damage to particular parts of the network may lead to a lack of coordination between brain regions during the progression of AD, but also how and why particular neurons are likely to be susceptible to protein buildup in the first place. For instance, Buckner et al. (2009) explicitly note that:

> Maps of cortical hubs, and eventually the detailed paths of fiber tracts supporting them, may provide a means to understand why certain lesions and connectional abnormalities are particularly disruptive. Hubs may also provide insight into Alzheimer's disease (AD) pathology. AD is associated with the pathological accumulation of misfolded proteins, including amyloid-β (Aβ). The identification of cortical hubs may explain why certain regions of cortex show disproportionately high levels of metabolism and, as a result, preferential vulnerability to AD pathology (p.1861)

Note that here, we can integrate physiological, computational, and network models to provide a more comprehensive account of AD's effects on cognition *without* providing a single model or theory in which the insights from all the component models are combined. Instead, we integrate them by understanding *how to move between* the various models that cannot be combined directly. It is this idea that allows us to make sense of the claim by Bressler (1995) regarding the integrative benefits of network models:

> The concept of large-scale cortical networks provides a reasonable framework for integrating results from neuroanatomical, neuropsychological and neurophysiological studies on distributed functioning of the cerebral cortex. (1995, p.299)

Clearly by “integrating results” from various studies, Bressler does not mean we can provide a single comprehensive network model which includes all neuroanatomical, neuropsychological, and neurophysiological information. Instead, a far more plausible interpretation of “integration” here is the idea that we can provide a coherent and comprehensive account of cognitive phenomena by understanding how to draw *relevant inferences between* network models, neuroanatomical models, neuropsychological models, and neurophysiological models.

5.3. Integration and Multiscale Modeling

Recent work on multiscale modeling likewise provides further evidence for the view of integration being defended here. When dealing with complex systems, it is often the case that the behaviour of the system is very different when examined at different temporal and spatial scales, requiring us to employ different mathematical formalisms to represent the various scale-dependent features of the system (see: Green & Batterman 2017; Wilson 2017; Batterman & Green 2021; Haueis 2021, 2022). This dependence on different mathematical formalisms for characterizing different scale-specific behaviours means we cannot provide a single integrated representation which describes all these behaviours simultaneously.

We can, however, integrate these various models by understanding how to draw coherent inferences across these scale-dependent models. This is frequently done by invoking very particular kinds of mathematical techniques. When we move from a macroscale model to a microscale one, for instance, we focus our attention only a restricted part of the microscale system relevant to our pragmatic purposes (this involves invoking boundary conditions on the microscale model). Conversely, when moving from a microscale model to a macroscale one, we often invoke homogenization. With homogenization, we take the widely heterogeneous structures and processes operating at the microscale and treat them as though they were homogenous in nature to identify macroscale behavioural regularities of the system. These mathematical techniques allow us to draw coherent inferences across models at different scales.

While this story further supports the idea that we can integrate these accounts without fitting them all into a single model or representation, one potential concern worth highlighting is that the *method by which this is accomplished* in these particular cases may appear to conflict with the view being defended in this paper. The focus of this paper has been to defend the idea that integration of this sort depends on keeping track of implicit metaphysical commitments held by scientists which link the various models. But it is not apparent that this is what's happening in the case of multiscale modeling of this sort. Instead, the integrative work seems to be done by employing the appropriate epistemic functions of introducing boundary conditions (to restrict the lower-scale modeling domain) and homogenization (converting lower-scale information in a format such that it can correct upper-scale models). These seem to be epistemic, or perhaps pragmatic, solutions to understanding how models relate, which appear neutral in regards to the researchers' background metaphysical or ontic views. This seems to run counter to the idea that identifying such commitments is essential for the relevant inferences between models to occur.[2]

In order to see why these cases in fact the support the view being defended here, it is helpful to explore how boundary conditions and homogenization are applied in the study of cognitive systems. Philipp Haueis (2021, 2022) explores how multi-scale modelling in connectomics (the study of networks in the brain at different scales) works in accordance with these principles. Specifically, he notes that:

[2] Special thanks to a blind referee for emphasizing this objection.

> A multiscale modeling schema in connectomics attempts to describe the relationship between these different scale-dependent features of brain organization. In particular, multiscale models of cortical gradients describe how systematic progressions of network connectivity, circuit architecture and cellular/subcellular densities relate to each other. (2022, p.9)

He identifies three different scales at which the brain is commonly studied: microscale models (which identify cellular and subcellular organizations), mesoscale models (which identifies the organization of neural circuits), and macroscale models (which identify large scale brain networks connecting different anatomic regions). While Haueis agrees that the application of boundary conditions and homogenization do indeed allow us to draw inferences between the models at different scales, he also points out that the application of these mathematical techniques to relate these models depends on what he calls "fundamental presuppositions" that must be adopted by the scientists working at different scales. For instance, he notes that "to gather information relevant to many types of behavior, descriptive models formulate a *fundamental presupposition* about the overall organization of the target system. In connectomics, a fundamental presupposition specifies the relation between neural structure and function in a nervous system or a class of nervous systems" (2022, p.13, emphasis in text). These fundamental presuppositions are required to in order to understand how we can relate the models operating at different scales. For instance:

> In multiscale systems such as the brain, many different features can realize the structure-function relationship presupposed by a connectomics model. Researchers therefore also need *scale-specific modeling assumptions* to link data types to scale-specific features characterized by multiscale gradient modeling. For example: a microscale assumption is that layer-specific differences in histological intensity profiles reflect differences in cytoarchitectonic similarity.

> This assumption links G_{HIST} values to degrees of laminar and cytoarchitectonic differentiation. These features realize the organization presupposed by the structural model, according to which areas with similar cytoarchitecture are more strongly connected than architectonically dissimilar areas. (2021, p.4).

These fundamental assumptions act as exactly the sort of metaphysical commitments that are needed for integration to occur. What this shows is that these sorts of multiscale modeling cases dovetail nicely with the view of integration defended here.

6. Implications and Consequences

One advantage in thinking about integration in this new way is that we need not worry about generating a single integrated model or theory in order attain the virtues of integration. This has the potential to shift our methodological approach in cognitive science. The insistence that "integrated models of cognitive systems are a necessary tool in understanding a cognitive system that is integrated" (Grey 2007, p.12), or that "future research should mainly be devoted to the development of really integrated models" (Marraffa & Paternoster 2013, p.39) should be taken with a grain of salt. The cost, time, and computational resources that have gone into trying to discover and generate a single integrated representation of a complex phenomenon may be more effectively re-allocated to the creation of more simplified and practical models without giving up on the quest for integration or the knowledge that could in principle be gained from integrated models.

If this account of integration is correct, this also means that cognitive scientists need to be far more aware of what ontological and metaphysical commitments they bring to bear on the models they use, and their justifications for them (see: Lombardi 2019). A lack of awareness can lead to problems in which the scientists themselves may be confused as to the ontic and metaphysical commitments that guide their research. As Plutynski notes, "what starts as deliberate simplification may often be confused with actual hypothesis and latter reified into theory" (2013, p.472).

Likewise, scientists may often think that they are studying the same phenomenon, but their implicit metaphysical commitments result in them understanding the phenomenon in different and potentially incompatible ways. This can misleadingly lead to the *appearance* of integration in cases where such integration has not in fact occurred. As a clear example of this, consider the study of aggression. It may seem as though models from different domains are all attempting to study the same phenomenon (namely, *aggression*). The problem is that each domain may have distinct metaphysical commitments regarding what aggression *is* that do not necessarily conform to one another.

In the context of social science, the concept of "aggression" in humans is often implicitly intertwined with psychological traits that are stereotypically considered distinctly masculine, such as being ambitious, assertive, dominant, and confident (for discussion, see: Longino 1983). Those who adopted such a view of aggression found that their theories and models seemed to be vindicated by, and well-integrated with, experimental work done on lab rats regarding the effects of increased testosterone on aggression. Specifically, that increasing testosterone levels in rats seemed to correlate strongly with an increase in aggression. Given that human males, on average, have a greater concentration of testosterone than do females, this experimental work on rats would seem to confirm that men are more aggressive than women, and thus more likely to have traits such as being ambitious, assertive, dominant, and confident (e.g. Money & Ehrhardt 1972; Ehrhardt & Baker 1974). Thus it appeared that we had a clear case where the models and theories of the different domains were nicely integrated. The problem is that the lab experiments did not define "aggression" in the same way the social scientists did, leading to the misleading assumption that the two types of models and theories built on one another. More specifically, in the context of the laboratory studies on rats, "aggression" was operationalized exclusively in terms of the likelihood to engage in fight behaviour. However, when social scientists looked at these results, their implicit gendered assumptions regarding what the phenomenon of aggression entailed meant that they assumed the study confirmed *their* account of the phenomenon, when in fact it said nothing about the relationship of testosterone to any traits beyond fight behaviour. Thus the experimental results did not vindicate the claim that men are more likely than women to be ambitious, assertive, dominant, or confident. As Longino puts it…

> …although aggressivity is identified in the experimental situation with fighting (among caged laboratory rats), when appealed to in social explanations it includes not only combativeness but also such traits as assertiveness, independence, and intelligence. Frequency of fighting is treated as a measure of aggressivity and thus as the measure of other qualities. (1983, p.13).

Here, having different implicit metaphysical commitments regarding how the phenomenon is understood meant that scientists misleadingly assumed they had a clear case of integration when in fact they did not.

What all this means is that in some cases it may appear as though two different models superficially share key metaphysical commitments, when in fact the scientists who employ these models have commitments that put the models at odds with each other, making integration impossible. Conversely, some models may seem to be in direct conflict, but different implicit ontic and metaphysical commitments held by the scientists who employ them can bring them into alignment.

To illustrate, consider the current dispute between representationalists and anti-representationalists in cognitive science. These debates hinge on whether we ought to interpret the brain as a computational system that operates over mental representations, or instead as a complex embodied and embedded dynamical system whose behaviour is best explained via the application of particular kinds of mathematical formalism (with Dynamical Systems Theory being one that is frequently proposed). For additional details on this debate, see: Van Gelder 1995; Van Gelder & Port 1995; Eliasmith 1996; Chemero 2009; Hutto & Myin 2014; Raleigh 2018; Taylor 2022.

How are we to understand the relationship between dynamical models of cognitive behaviour, and computational ones? If those who adopt the computational model are commitment to the metaphysical fact that the brain engages in mental representation, while the dynamical model is committed to the fact that the brain does *not* engage in mental representation, then we seem to have a direct conflict. However, the same models employing the same formalisms can often be interpreted differently when we adopt different sets of metaphysical commitments in their application (for discussion, see: Schwarz 2022).

For instance, some argue that we should understand mental representation merely as a gloss on neurophysiological mechanisms, and not as metaphysical entities in need of explanation (Ramsey 2007; Egan 2010). In which case, models of the brain in terms of computational processes over mental representations can be thought of as something more akin to a phenomenological model, useful mainly for characterizing and predicting all kinds of behavioural capacities and regularities of the system. This sort of information can easily be integrated with dynamical models since such behavioural regularities directly informs our understanding of the dynamics of the system.

Or consider a very different set of metaphysical commitments. Someone can adopt a robust metaphysical commitment to mental representation, while also holding the view that the application of Dynamical Systems Theory does *not* require a metaphysical rejection of such representations. In which case, dynamical models provide essential information as to the dynamics of the system which can inform how the brain constructs and applies mental representations, thereby allowing for integration between the two types of models (for more, see: Botvinick 2012; Eliasmith 2012; Raleigh 2018). Thus integration cannot be determined by looking at the compatibility or incompatibilities of the models themselves, but at what metaphysical commitments one brings to the models.

Settling these disputes will require justifying and re-evaluating the metaphysical commitments that guide how and why we construct the particular models we do, and what sorts of inferences we ought to draw from them. This is no small task, as sometimes working scientists are unaware of the various metaphysical commitments that guide them, while other times the commitments they *report* having may not reflect the commitments that seem to guide the practices that they engage in. This can certainly make the process of integration seem overwhelmingly difficult. However, this does not mean that the task of identifying and understanding what these commitments are is an intractable one.

Numerous projects have appeared in the past few years specifically designed to help facilitate awareness regarding what sorts of commitments guide various kinds of cognitive research, and how these can be used to clarify conflicts. Jacqueline Sullivan (2017) highlights a few of these projects in detail, as well as their current limitations. These initiatives have the potential to help us overcome the daunting obstacles to integration, however their success will depend on a number of factors that we need to be vigilant about.

For instance, certain databases like Neurosynth and Brain-Map are dedicated to providing a standardized taxonomy and set of terminology for cognitive phenomena needed for scientists to effectively communicate across laboratories and domains to find points of agreement. This is accomplished by conducting “automated meta-analyses that integrate neuroimaging data from a diverse array of experiments and represent it in the form of brain activity maps.” (Sullivan 2017, p.134). The problem, as Sullivan rightly points out, is that “when a user inputs a label into a given database to conduct a meta-analysis of the literature, the brain map generated abstracts away from details about the contexts, experimental paradigms and protocols used to produce each data set” (Sullivan 2017, p.134). We can supplement Sullivan’s analysis here by noting that in virtue of abstracting away from all this kind of information, these databases likewise do not provide information as to whether or not scientists working in different labs or different contexts share the same underlying metaphysical commitments when they invoke and apply a given terminology, and so cannot address an important part of the problem.

Other sorts of initiatives are more sensitive to these issues and focus more on fostering coordination between scientists working in different labs and fields. One particularly noteworthy initiative is the Cognitive Atlas:

> One aim of the Cognitive Atlas, a collaborative web-based knowledge-building project, is to promote such coordination (Poldrack et al. 2011; Poldrack and Yarkoni 2016). The Atlas is an ontology that aims to systematically depict: (1) psychological functions currently under study in cognitive neuroscience, (2) the experimental paradigms/tasks/tests used to investigate those functions, and (3) the relationships between (a) different functions and (b) different paradigms used to investigate those functions. General definitions are provided for each hypothetical psychological construct (e.g., “attention”, “working memory”) identified in the Atlas. (Sullivan 2017, p.135).

The Cognitive Atlas is more promising as a means of bringing to light implicit metaphysical commitments of working scientists as “it allows its users to identify and document when they agree or disagree about (a) how a term designating a construct is defined and (b) the asserted relationships between different constructs.” (Sullivan 2017, p.135). However, it too has its own set of concerns we must keep in mind.

First, the success of initiatives like the Cognitive Atlas is dependent on whether enough scientists actually participate in it, and reaching this critical mass is no small task (Sullivan 2017, p.135).

Second, there is a potential worry that Sullivan does not highlight, which is that the sorts of metaphysical commitments that can influence integration are not simply those that deal with our cognitive constructs, and what their relationships are. They can often be baked into conditions surrounding and framing those constructs. For instance, imagine scientists were to agree on a cognitive construct like a "theory of mind mechanism", which we can think of as a mechanism in the brain responsible for interpreting, understanding, and predicting others by attributing thoughts, motivations, and other mental states to them (e.g. Leslie, Friedman & German 2004). One potential problem is that if some scientists have implicit metaphysical commitments about the architecture of the brain being massively modular, while others have a commitment to its being massively interconnected, then this will radically influence whether the models they build that invoke this construct can successfully integrate or not. Here the problem isn't with their metaphysical commitments regarding *what* a theory of mind mechanism is broadly speaking, but about what the architecture of the brain is that frames how they go about *studying* such a mechanism. This means that while the potential for the Cognitive Atlas to help address these obstacles for integration are certainly there, it still has a way to go before it can be the tool we need it to be.

Sullivan notes that some of the most promising initiatives can be found in Psychiatry, with the development of scientific repertoires such as the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRICS) and the Research Domain Criteria (RDoC) Project (2017, p.137-141). These large-scale projects attempt to bring together researchers working in different domains, and to help them coordinate their activities by way of surveys, meetings, and workshops. The goal being to produce an agreed upon set of constructs, research paradigms, and tools to be used to help make explicit potentially conflicting assumptions in the study of psychiatric disorders. Moreover, there is an explicit awareness from those involved in these projects that the cognitive constructs they develop are to be used primarily as heuristics that are constantly under revision (Sullivan 2017, p.140).

This avowed commitment to the tentative nature of their constructs can hopefully help scientists to be more aware of what metaphysical commitments they might bring to the table when invoking them, and to their research more generally. It also encourages a willingness to revise such commitments through the processes of collaborative communication and discovery.

While these scientific repertoires are extremely promising, at the moment they are largely focused on psychiatry and not mainline research in neuroscience and cognitive science more generally. It is for this reason that Sullivan proposes what she calls a “coordinate pluralism”, which involves using each of these different projects to inform and constrain one another. As she puts it, “only a coordinated pluralism that incorporates positive aspects of each initiative and implements them on a broad scale in psychology and neuroscience will have the potential for success” (Sullivan 2017, p.130).

I think that Sullivan provides an encouraging path forward for dealing with the obstacles of integration, just so long as we are careful to make sure these various projects can tease out the relevant implicit metaphysical commitments that guide scientists in different domains and contexts. But there is good reason to be optimistic. While the challenges to integration are not easily overcome, I want to suggest that focusing on these issues, as opposed to whether we can generate a single over-arching integrated model of a complex cognitive phenomena, is a far more productive and helpful way of thinking about integration in cognitive science. The challenges to integration are very real, but none of this suggests that we should give up on integration in cognitive science, so much as reconceptualize what integration entails and what it can give us.

Conclusion

In this paper, I have argued that the traditional interpretation of integration adopted in cognitive science may be problematic and unattainable in many instances. However, this does not mean we should give up the quest for integration. Instead, I propose we adopt a slightly different sense of integration where various models that make contradictory claims can still be used to contribute to a coherent non-contradictory understanding of the same unified phenomenon. This process involves looking beyond the models themselves to the metaphysical and ontological commitments that working scientists implicitly adopt when creating, employing, and drawing inferences from various incompatible models. Identifying and keeping track of such commitments is key to understanding how incompatible or contradictory models can still inform and constrain one another. In this regard, we can have an integrated account of a complex phenomenon even without integrated *models* or *theories*.

Declarations

Conflicts of interest: The authors declare that they have no conflict of interest.

References:

Bammer, G. (2013). Disciplining interdisciplinarity: Integration and implementation sciences for researching complex real-world problems. Canberra: ANU E-Press.

Batterman, R (2001) The devil in the details: asymptotic reasoning in explanation, reduction, and emergence. Oxford University Press, Oxford.

Batterman, R.W., Green, S. (2021). Steel and bone: mesoscale modeling and middle-out strategies in physics and biology. Synthese 199: 1159–1184. https://doi.org/10.1007/s11229-020-02769-y

Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. Philosophy of Science, 66(2), 175-207. http://dx.doi.org/10.1086/392683.

Botvinick, M. (2012). Commentary: Why I am Not a Dynamicist. Topics in Cognitive Science 4: 78-83.

Buckner, R. L. et al. (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J. Neurosci. 29, 1860–1873.

Bressler, S. L. (1995). Large-scale cortical networks and cognition. Brain Research Reviews, 20(3), 288–304.

Cassimatis, N. (2005). Integrating Cognitive Models Based on Different Computational Methods. Proceedings of the Annual Meeting of the Cognitive Science Society 27 (27): 402-407.

Chemero, A. (2009). Radical embodied cognitive science. Cambridge, MA: MIT Press.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. Synthese, 191(2), 127–154.

Craver, C. & Kaplan, D. (2018). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. The British journal for the philosophy of science. DOI: 10.1093/bjps/axy015

Danks, D. (2014). Unifying the Mind: Cognitive Representations as Graphical Models. MIT Press

David, D. and Szentagotai, A. (2006). Cognitions in cognitive-behavioral psychotherapies; toward an integrative model. Clinical Psychology Review 26 (3): 284-298.

de Haan, W., Mott, K., van Straaten, E.C., Scheltens, P. & Stam, C.J. (2012). Activity dependent degeneration explains hub vulnerability in Alzheimer's disease. PLoS Comput. Biol. 8, e1002582.

Duch, W. (2007). Computational models of dementia and neurological problems. Methods Mol. Biol 401: 305–336. doi:10.1007/978-1-59745-520-6_17

Egan, F. (2010). Computational models: A modest role for content. Studies in History and Philosophy of Science, 41, 253–259.

Ehrhardt, A. & Baker, S. (1974) Fetal Androgens, Human Nervous System Differentiation, and Behavior Sex Differences. In Richard Friedman, R. M. Richart, and R. M. Van de Wiele, eds., Sex Differences in Behavior. New York: Wiley. 33-52.

Eliasmith, C. (1996). The third contender: A critical examination of the Dynamicist theory of cognition, Philosophical Psychology, 9:4, 441-463, DOI:10.1080/09515089608573194

Eliasmith, C. (2002). The myth of the turing machine: The failing of functionalism and related theses. Journal of Experimental & Theoretical Artificial Intelligence, 14(1), 1-8. http://dx.doi.org/10.1080/09528130210153514.

Eliasmith, C. (2012). The Complex Systems Approach: Rhetoric or Revolution. Topics in Cognitive Science 4: 72-77.

Garden, D., Dodson, P., O'Donnell, C., White, M., Nolan, M. (2008). Tuning of Synaptic Integration in the Medial Entorhinal Cortex to the Organization of Grid Cell Firing Fields. Neuron 60 (5): 875-889.

Giocomo, L. Moser, M-B, Moser, E. (2011). Computational Models of Grid Cells. Neuron 71 (4): 589-603

Grey, W. (2007). Composition and Control of Integrated Cognitive Systems, in Wayne D. Gray (ed.), Integrated Models of Cognitive Systems. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195189193.003.0001

Green, S. & Batterman, R. (2017). Biology meets physics: Reductionism and multi-scale modeling of morphogenesis. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 61: 20-34

Gunzelmann, G. (2013). Motivations and Goals in Developing Integrative Models of Human Cognition. Proceedings of the Annual Meeting of the Cognitive Science Society 35 (35): 30-31.

Hafner, V. V. (2000). Cognitive maps for navigation in open environments. In Proc. 6th int. conf. on intelligent autonomous systems (IAS-6) (pp. 801–808). Venice: IOS Press.

Haueis, P. (2021). Multiscale modeling of cortical gradients: The role of mesoscale circuits for linking macro- and microscale gradients of cortical organization and hierarchical information processing. NeuroImage 232, 117846.

Haueis, P. (2022). Descriptive multiscale modeling in data-driven neuroscience. Synthese 200: 129.

Hochstein, E. (2016a). One Mechanism, Many Models: A Distributed Theory of Mechanistic Explanation. Synthese 193 (5): 1387-1407.

Hochstein, E. (2016b). Giving up on Convergence and Autonomy: Why the Theories of Psychology and Neuroscience are Codependent as well as Irreconcilable. *Studies in History and Philosophy of Science* 56: 135-144.

Hochstein, E. (2019). How Metaphysical Commitments Shape the Study of Psychological Mechanisms. *Theory & Psychology* 9 (5): 579-600.

Holland L., de Regt H.W., Drukarch B. (2019). Thinking about the nerve impulse: The prospects for the development of a comprehensive account of nerve impulse propagation. Front. Cell. Neurosci.13: 208.

Horn, D., Ruppin, E., Usher, M., and Herrmann, M. (1993). Neural network modeling of memory deterioration in Alzheimer’s disease. Neural Comput. 5: 736–749. doi: 10.1162/neco.1993.5.5.736

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., … Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage 80: 360–378. doi:10.1016/j.neuroimage.2013.05.079

Hutto, D. D. and Myin, E. (2014). ‘Neural Representations Not Needed – no More Pleas, Please,’ Phenomenology and the Cognitive Sciences 13(2), pp. 241–256.

Kitano, K., & Fukai, T. (2007). Variability vs. synchronicity of neuronal activity in local cortical network models with different wiring topologies. Journal of Computational Neuroscience, 23(2), 237–250.

Lemerise, E. & Arsenio, W. (2000) An Integrated Model of Emotion Processes and Cognition in Social Information Processing. Child Development 71 (1): 107-118.

Leslie, A., Friedman, O., & German, T. (2004). Core mechanisms in ‘theory of mind’. Trends in Cognitive Sciences (8)12: 528-533.

Levy, A. & Bechtel, W. (2013) Abstraction and the Organization of Mechanisms. Philosophy of Science 80: 241-261.

Lombardi, V. (2019). Identifying Conflicts Between Models in Science: Examining the Role of Implicit Commitments in Scientific Models. *Major Research Paper for Masters Degree in Philosophy*, University of Victoria.

Longino (1983). Beyond "Bad Science": Skeptical Reflections on the Value-Freedom of Scientific Inquiry. Science, Technology, & Human Values 8 (1): 7-17.

Longino, H. (2006). Theoretical pluralism and the scientific study of behavior. In S. Kellert, H. Longino, & C. K. Waters (Eds.), Scientific pluralism (pp. 102-132). Minneapolis: University of Minnesota Press.

Longino, H. (2013). Studying human behaviour. Chicago: The University of Chicago Press.

Love, A. C., & Lugar, G. L. (2013). Dimensions of integration in interdisciplinary explanations of the origin of evolutionary novelty. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44(4, Part A), 537–550.

MacLeod, M (2018) What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. Synthese 195: 697–720

Marraffa, M., & Paternoster, A. (2013). Functions, levels, and mechanisms: Explanation in cognitive science and its problems. Theory & Psychology, 23(1), 22–45.

Miłkowski, M. (2016). Unification Strategies in Cognitive Science. Studies in Logic, Grammar, and Rhetoric 48 (61): 13-33.

Mitchell, S. (2002) Integrative Pluralism. Biology and Philosophy 17: 55-70

Mitchell, S. (2004) Why Integrative Pluralism? E:CO Special Double Issue Vol. 6 Nos. 1-2 2004 pp. 81-91

Money, J. & Ehrhardt, A. (1972). Man and Woman, Boy and Girl. Baltimore: Johns Hopkins university Press.

Newell, W. H. (2001). A theory of interdisciplinary studies. Issues in Integrative Studies, 19, 1e25.

O'Rourke, M., Crowley, S., & Gonnerman, C. (2016). On the nature of cross-disciplinary integration: A philosophical framework. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 56, 62–70.

Penney, J., Ralvenius, W., and Tsai, L-H. (2020). Modeling Alzheimer's disease with iPSC-derived brain cells. Molecular Psychiatry 25:148–167

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. Synthese, 183(3), 283-311.

Plutynski, A. (2013). Cancer and the goals of integration. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44(4), 466–476.

Poldrack, R., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., . . . Bilder, R. M. (2011). The cognitive atlas: Toward a knowledge foundation for cognitive neuroscience. Frontiers in Neuroinformatics, 5(17), 1–11. doi: 10.3389/fninf.2011.00017

Poldrack, R., and T. Yarkoni. (2016). From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure. Annual Review of Psychology 67: 587–612.

Potochnik, A. (2015). The diverse aims of science. *Studies in History and Philosophy of Science* 53: 71-80

Potochnik, A. (2017). Idealization and the Aims of Science. Chicago: The University of Chicago Press.

Raleigh, T. (2018) Tolerant enactivist cognitive science, Philosophical Explorations, 21 (2): 226-244. DOI: 10.1080/13869795.2018.1477981

Ramsey, W. (2007). Representation reconsidered. Cambridge, NY: Cambridge University Press.

Salvucci, D. (2013). Integration and Reuse in Cognitive Skill Acquisition. Cognitive Science 37: 829–860.

Saraceno, C., Musardo, S., Marcello, E., Pelucchi, S., and Di Luca, M. (2013). Modeling Alzheimer's disease: from past to future. Front. Pharmacol. 4:77. doi: 10.3389/fphar.2013.00077

Savelli, F. & Knierim, J. (2010). Hebbian Analysis of the Transformation of Medial Entorhinal Grid-Cell Inputs to Hippocampal Place Fields. J Neurophysiol103: 3167–3183, 2010.

Schoelles, M.J., Neth, H. Myers, C.W., Gray, W. (2006). Steps towards integrated models of cognitive systems: a levels-of-analysis approach to comparing human performance to model predictions in a complex task environment. Proceedings of the Annual Meeting of the Cognitive Science Society. Red Hook, NY: Curran. 756-761.

Sejnowski, T. J., Churchland, P. S., & Koch, C. (1988). Computational neuroscience. Science, 241, 1299–1306.

Sompol, P., Ittarat, W., Tangpong, J., Chen, Y., Doubinskaia, I., Batinic-Haberle, I., Abdul, H.M., Butterfield, D.A., Clair, D.K. (2008). A neuronal model of Alzheimer's disease: An insight into the mechanisms of oxidative stress–mediated mitochondrial injury. Neuroscience 153 (1): 120-130.

Sterratt, D., Graham, B., Gillies, A., & Willshaw, D. (2011). Principles of computational modelling in neuroscience. Cambridge: Cambridge University Press.

Stinson, C. (2016). Mechanisms in psychology: ripping nature at its seams. Synthese 193: 1585–1614. https://doi.org/10.1007/s11229-015-0871-5

Sullivan, J. (2009). The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. Synthese 167: 511–539.

Sullivan, J. (2017) Coordinated pluralism as a means to facilitate integrative taxonomies of cognition. Philosophical Explorations 20 (2): 129-145. DOI: 10.1080/13869795.2017.1312497

Schwarz, S. (2022). Mental Talk, Model Behavior: Intentional Psychology as Explanatory Model. *Masters Thesis*, Berlin School of Mind and Brain.

Syropoulos, A. (2008). Hypercomputation: Computing beyond the Church-Turing Barrier. Springer.

Taylor, S. (2022). Cognitive Instrumentalism about Mental Representation. Pacific Philosophical Quarterly 103: 518–550. DOI: 10.1111/papq.12383

Trumpler, M. (1997). Techniques of intervention and forms of representation of sodium-channel proteins in nerve cell membranes. *Journal of History of Biology* 30: 55–89.

Van Gelder, T. (1995). What might cognition be if not computation?, Journal of Philosophy, 91, 345-381.

Van Gelder, T. and Port, R. (1995). It's about time: An overview of the dynamical approach to cognition, Mind as motion: Explorations in the dynamics of cognition. (Cambridge, MA, MIT Press).

Weisberg, M. (2013). Simulation and similarity: Using models to understand the world. Oxford: Oxford University Press.

Wilson, M. (2017). Physics Avoidance: Essays in Conceptual Strategy. UK: Oxford University Press.

Yuan, M., Tian, B., Shim, V.A., Tang, H., Li, H. (2015). An Entorhinal-Hippocampal Model for Simultaneous Cognitive Map Building. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Zars, T., Fischer, M., Schulz, R., & Heisenberg, M. (2000). Localization of a short-term memory in Drosophila. Science, 288: 672–675.

Zednik, C. (2019). Models and mechanisms in network neuroscience. Philosophical Psychology 32 (1): 23-51.