Cognitive psychology does not reduce to neuroscience

Lincoln J. Colling (lincoln.colling@maccs.mq.edu.au)
Macquarie Centre for Cognitive Science, Macquarie University
Sydney, New South Wales 2109 Australia

Reece P. Roberts (r.roberts@auckland.ac.nz)
Research Centre for Cognitive Neuroscience, Department of Psychology, University of Auckland
Private Bag 92019, Auckland 1142 New Zealand

Abstract

Contemporary scientific investigations of the mind have increasingly looked towards the brain in order to explain intelligent behavior. This is most evident with the rise of cognitive neuro-imaging. This approach has, however, been met with mixed reactions. On the one hand, classical cognitive scientists—in the computationalist–functionalist tradition—have argued that cognitive neuro-imaging does not, and cannot, answer questions about the cognitive mechanisms that are responsible for creating intelligent behavior; it is limited to questions about neural function or the neural basis of cognition. On the other hand, there are those who argue that an understanding of intelligent behavior can only be gained through study of the brain. We suggest that both views are misguided. We will present a third option: That neuroscience, properly used, can be employed in the development of cognitive theory, but that cognitive science does not reduce to neuroscience, because intelligent behavior can only be understood by studying how the brain interacts with the body and the brain-body with the environment.

Keywords: extended mind; reductionism; neuroscience; embodied cognition

Introduction

Classical cognitive scientists have argued that the goal of cognitive science is to uncover facts about mental information-processing systems responsible for producing intelligent behavior. For this reason, facts about the brain are seen as irrelevant because they have no bearing on the nature of information processing. To use the words of one prominent advocate of this position, "[n]o amount of knowledge about the hardware of a computer will tell you anything serious about the nature of the software that the computer runs. In the same way, no facts about the activity of the brain could be used to confirm or refute some information-processing model of cognition" (Coltheart, 2004, p. 22). Neuroscientists too have sought to understand intelligent behavior, with some in the field adhering to an extreme philosophical position which largely "refuse[s] the psychological any genuine causal explanatory role" (Bickle, 2003, p. 115), reduces mind to brain, and affords higher level theories and methods only a "heuristic role… in the search for lower mechanisms and reductions" (Bickle, 2003, p. 130)[1]. These two approaches—cognitive science and

---

[1]Bickle claims that the reductionism he favors is the reductionism found in actual neuroscientific practice. This is a controversial claim. However, it will be Bickle's reductionism that is primary focus of the present paper.

neuroscience—use different methods and vocabularies and, while some have argued that they are pitched at different levels of explanation and have distinct explanatory targets (e.g., Coltheart, 2004), others have argued for a single explanatory target and have suggested that psychology should be reduced to neuroscience (e.g., Bickle, 2003).

The aim of this paper is to examine these two viewpoints from the perspective of the cognitive psychologist. In the first half, we will examine the role that neuroscience and neuroscientific reasoning can play in theorizing about cognition. We will suggest that studying the brain can and does tell us important facts about information-processing systems, and that neuroscience can generate novel behavioral predictions that can be incorporated into cognitive theories. In the second section, we aim to defend cognitive psychology from reduction to neuroscience by examining how intelligent behavior relies crucially on extra-neural structures and the environment in which the organism is embedded.

The role of neuroscience

As seen above, classical cognitive science has often been dismissive of neuroscience. This view stems, in part, from a philosophical commitment to the principle of multiple realizability. We will first examine the claim of multiple realizability to see how it is used to protect psychology from reduction. Following this, we will examine some of the problems with multiple-realizability. In doing so, we aim to show that once a commitment to multiple realizability is abandoned the clear division between structure and function begins to fade.

Multiple Realizability

Multiple realizability has been one of the most prominent defenses against reduction of the special sciences (any science above physics). Those who have favored reductionism have either had to explicitly deal with multiple realizability within their reductionist program or have had to question its truth (e.g., Bickle, 2003). However, a commitment to multiple realizability within a non-reductionist program might do more than simply ward off reductionism; it might also create an artificially strong divide between implementation and function thereby obscuring important functional consequences that arise directly out of particular details of physical implementation.

Multiple realizability forms a central part of the functionalist tradition of philosophy of mind. Functionalism is the idea that mental states (or rather, cognitive processes) should be defined by the functional role that they play rather than how they are physically realized. For example, the cognitive process of reading aloud might be defined as the cognitive process that has the function of transforming printed letter strings into vocalizations of the words that correspond to those printed letter strings. Multiple realizability, simply stated, is the idea that one function, such as word reading, can be realized, or implemented, by many different, physically distinct, systems. If two functionally isomorphic systems (maintaining a particular set of input-output relations) can be physically distinct, then nothing useful can be learnt about the function itself by merely studying the physical makeup; doing so will only give details about how the function is implemented.

To understand the problem with multiple realizability, it is necessary, following Shapiro (2000), to introduce the concepts of grain of analysis and of causally relevant properties. These terms can be made clear by way of an example. Two mousetraps, functionally isomorphic, serve the function of catching mice. That is, they take mice as their input and by a particular set of internal state transitions produce caught mice as their output. One mousetrap has a spring-loaded bar made from copper while the other has a spring-loaded bar made from steel. Here then are two physically distinct systems that both perform the function of catching mice. Do they count as two distinct realizations of a single functional kind? We would argue that, in any meaningful sense, they do not. The metal, the copper or the steel, is not a causally relevant property of a mousetrap. That is, it is a physical property that, when changed, does not alter how the mousetrap performs its function. What is causally relevant about the bar is that it is rigid. To be sure, steal and copper are distinct realizations of the property of rigidity, but, as Shapiro (2000) suggests, these differences are not relevant to their function in the mousetrap. We can screen off these differences and they need not enter our explanation of the mousetrap. When a causally relevant property is changed it brings about a change in how the mousetrap performs its function. For example, one mousetrap could use a spring-loaded bar that snaps shut to trap and kill the mouse. If we were to physically alter this mechanism and replace it with a box with a trap door that closes thereby trapping the live mouse instead, we have altered a causally relevant property of the trapping mechanism as we have altered how the mousetrap performs its function. However, again there is a question of whether the two mousetraps should count as distinct physical realizations of a single functional kind. On one grain of analysis the answer would be yes. Both mousetraps realize the function of catching mice. On a finer grain of analysis, however, the answer is not so clear. One mousetrap catches and kills mice, while the other catches mice alive. These two mousetraps could be realizations of two distinct functional kinds—a live mouse catcher and a mouse killer.

Thus for two mousetraps (or brains or computers) to be distinct realizations of a single functional kind, they would need to differ in terms of their causally relevant parts while remaining functionally isomorphic at the grain of analysis of interest.

Fine-grained analysis The grain of analysis problem is clearly evident when contrasting two different approaches to artificial intelligence: good old fashioned artificial intelligence (GOFAI) with non-GOFAI approaches, including connectionism. The GOFAI approach to artificial intelligence has its roots in the work of Newell and Simon (e.g., Newell & Simon, 1997) and, in particular, the law of qualitative structure for symbol systems. This law states, "a physical symbol system has the necessary and sufficient means for general intelligence" (p. 87). A symbol system is a type of computer that might be realized as a desktop computer or a biological brain. What matters for intelligent behavior is not whether the system is made from silicon-stuff or brain-stuff, but rather that it is a physical symbol system. But what is a physical symbol system? Simply put, a physical symbol system is a physical system consisting of symbols (tokens) that can be joined together to form symbol structures such that these symbol structures can be manipulated through processes such as, for example, concatenation, separation, creation, and duplication, according to the various rules (or syntax) of the system. Thus, a simple binary adder might be considered a form of a physical symbol system because it takes symbols (binary ones and zeros) and by joining them together using a simple logical rule (for example, logical AND) it produces the "intelligent" behavior of adding. Similarly, the human brain might also be a physical symbol system because it contains symbols (brain states) that form symbolic structures manipulated according to syntactic rules. Therefore, it is claimed that what matters for cognitive science is that human brains are physical symbol systems and whether the symbols are particular physico-chemical brain states or binary digits realized by electromagnetic switches is not relevant to our understanding of intelligent behavior.

To be sure, there have been many successes within the GOFAI paradigm, and it certainly seems that at a coarse grain of analysis some aspects of human intelligence may be approximated by a physical symbol system. But is this also the case at a finer grain?

Approaches to artificial intelligence such as connectionism differ from GOFAI in a number of key respects. Primary among these is a lack of emphasis on symbolic manipulation. Instead, connectionist modelers place a greater emphasis on network structure, and they may refer to the nodes in their network being neuron-like or their network models being neurally plausible. However, to say that connectionism emphasizes structure over function would be a misrepresentation of the true value of connectionism. As mentioned earlier in our discussion of mousetraps, causally relevant physical properties can have functional consequences, and this is certainly the case when

comparing biological brains and connectionist networks on the one hand with GOFAI approaches on the other. These functional differences are most clear at a finer grain of analysis.

A complete treatment of connectionism is outside of the scope of the current paper; however, for our current purposes it will be sufficient to highlight a few areas in which a connectionist structure has microfunctional consequences—that is, how implementation constrains function in cognitive systems at a fine grain of analysis.

Human brains are biological systems, and human minds are the products of biological systems. Viewing the mind through the lens of the biological brain puts greater emphasis on certain functional features of minds that might otherwise be ignored by an abiological cognitive science. It has been said that "[n]othing in biology makes sense except in the light of evolution" (Dobzhansky, 1973, p. 125). Why should it be that biological cognition is different? We claim that it is not. There are a number of constraints that restrict the nature of possible computational solutions that can be employed to problems routinely faced by biological organisms. These constraints may result in microfunctional (or fine-grained functional) differences between biological (or biologically inspired) cognition and non-biological cognition. Furthermore, the way evolutionary processes work also means that the final solution to some complex computational problem is constrained by the more basic capacities upon which it is built. That is, microfunctional properties of human cognition may arise from the fact that complex capacities are built out of simpler solutions to more evolutionarily basic problems. Other constraints faced by biological cognitive systems might include how the system behaves when damaged, how the system copes under conditions of degraded input, or the ability to employ its cognitive resources under a wide range of circumstances (Clark, 1989). These constraints are brought into sharp focus by considering cognition as a biological phenomenon.

Primary among the constraints faced by biological cognition is time. That is, processing must be able to take place on a time scale that is appropriate for a particular biological organism in order for that organism to survive in its environment. Thagard (1986) argues that there is a selective pressure in favor of fast computation and that if our minds had operated substantially slower then "our species would never have evolved" (p. 305). Thus, the temporal properties of a cognitive system should enter into our functional description. Approaches to artificial intelligence based on parallel computation recognize this constraint while those that favor serial computation (such as GOFAI approaches) do not (Thagard, 1986). Now, if it is the case that only machines built using parallel architectures (that is, architectures that are brain-like) are capable of meeting the time constraints imposed on biological cognition then it is the case that structure can be a causally relevant factor in determining microfunctional properties. Thagard claims that this is the case and that these temporal properties arise as a direct result of the network structure favored by connectionism.

Another microfunctional property to arise from connectionist networks is crosstalk. Crosstalk refers to the errors that arise due to partial activation of units corresponding to similar representations. For example, in the CID model of word perception (McClelland, 1985), presenting two words simultaneously to the network results in the activation patterns from both words being combined. This can cause interference between the activation patterns of the two words and can cause partial activation of units corresponding to other words that share letters with the two original words. This interference causes a characteristic pattern of errors. So for example, if the network is shown the words SAND on the left and LANE on the right and then "asked" to report the word on the left the network may respond with SANE. This migration error, where the E from LANE migrates to SAND, arises directly out of the structure of the CID model. Originally, McCelland thought that this might be a drawback of the CID model; however, far from being a drawback of the model it is actually a positive aspect because this pattern of errors mimics those produced by humans (e.g., Mozer, 1983). The important point here is that this microfunctional property arises directly out of the architecture of the network. It is important, however, not to overstress the match between the architecture of CID networks and real brains, but this example serves to show that there is not a clear, sharp division between function and implementation. Rather, these lines are blurred and they will continue to blur as more is learnt about neuroscience and as more is learnt about the microfunctional features of cognition.

Science does not happen in a theoretical vacuum. The theoretical baggage that the cognitive scientist carries into an experiment not only affects how they might interpret data, but it also affects which observations might count as usable data and which might be discarded as noise or as irrelevant (Lakatos, 1970). Studying how it is that brains implement the functions they do serves to refocus the cognitive scientist's attention on aspects of the data that might otherwise be neglected or ignored. For example, time constraints on biological cognition might be ignored out of a philosophical commitment to multiple realizability and observations of particular error patterns might be regarded as noise rather than valuable sources of data. What the preceding examples aim to show is that facts about the brain can yield scientific progress in developing models of cognition and can produce novel behavioral predictions. Models that cannot account for speed of computation, computation under natural environmental conditions, and various other microfunctional properties of natural human cognition, are not satisfactory models of human cognition. In our discussion of connectionism it is not our intention to give the impression that connectionism is the final word on cognitive modeling. To be sure, connectionism is only one set of tools that can be employed in order to understand better the functional consequences that result from particular

structural implementations. More detailed, and more biologically realistic models of neural computation—possibly incorporating cellular and molecular details—are needed and it would be premature to set an *a priori* limit below which structural details may be screened off as functionally irrelevant for explaining intelligent behavior. Neuroscience, and a biological view of cognition, here serves the purpose of highlighting those physical details (e.g., network-like structure) that might be important for understanding functional properties of cognitive processes and microfunctional features of intelligent behavior.

The role of brain imaging

Not only can "brain-like" computation produce novel behavioral predictions, so too can techniques that measure actual brain functioning in behaving humans. The rise of brain imaging has been met with much skepticism within certain segments of the cognitive science community. Much of this criticism is for good reason, but to draw conclusions about brain imaging's in principle inability to answer questions about the mind from its methodological failings would be too hasty. Coltheart (2004) identifies two goals of neuroimaging: localization of cognitive functions and using neuroimaging data "to make theoretical decisions entirely at the cognitive level" (p. 21). As an example, included in this second goal Coltheart lists using neuroimaging to judge between different cognitive models of some cognitive system, a goal which Coltheart believes imaging has (so far) utterly failed in achieving. However, there is another possible use for neuroimaging: we suggest that the real power of neuroimaging lies in using neuroscientific data for *generating* and developing cognitive models which, stripped of neuroscientific language, can be pitched solely at the functional level[2].

The discovery of mirror neurons is an example of both the poor use of neuroscientific data and the valuable role neuroscientific data can play in generating theories about cognition. Mirror neurons, which were originally discovered in monkey premotor cortex, have been implicated in many cognitive tasks ranging from facilitating observational learning to understanding the goals and intentions of others (for a brief review, see Giacomo Rizzolatti, 2005). Many of these proposals have been very speculative and have gone well beyond what the data suggest; however, not all proposals have fallen into this trap. The real value of the discovery of mirror neurons has been that is has stimulated new ways, or at least reinvigorated old ways, of thinking about cognition, by emphasizing the intimate links between perception, action, and cognition. Action-centered cognition has a long history going as far back, at least, as William James (1905 [1890]), and action was also at the core of the ecological psychology of Gibson (1979). What mirror neurons have added to the debate is that they provide direct evidence of the links between perception, action, and cognition; this marks a clear shift from the amodal, action-neutral, symbolic approach to cognition found in classical cognitive science.

An example of how mirror neuron theory can be turned into a (purely) cognitive theory can be found in the emulator theory of action prediction put forward by Wilson and Knoblich (2005). Wilson and Knoblich suggest that the mirror neuron system provides a means by which an observer can predict the actions of an observed agent by internally simulating, or emulating, the agent's actions. That is, they propose that a system (consisting of mirror neurons) exists that has the purpose of generating real-time simulations of the movement dynamics of observed agents. These simulations can stand in for information from the external world, thereby bypassing the delay that would be introduced by waiting for actual sensory input, and these predictions can then be used to drive behavior.

In developing the emulator theory of action prediction, Wilson and Knoblich (2005) point to the properties of mirror neurons that provide the basis for their ideas. For example, they cite evidence from fMRI (e.g., see Buccino et al., 2001) which shows that action observation in humans results in somatatopic activation of premotor areas that are also active during action execution (this is consistent with single unit recording studies in monkeys). Stated functionally, this is evidence that action execution and action perception share a common neural code. Furthermore, Wilson and Knoblich (2005) point to evidence from PET (Stevens et al., 2000) suggesting motor and parietal regions (known to contain mirror neurons) are preferentially active for biologically possible movements over biologically impossible movements. In addition, Wilson and Knoblich also point to evidence from fMRI that shows mirror neuron activity when observing skilled movements is contingent on the observer being similarly skilled (e.g., see Calvo-Merino et al., 2004). Stated in cognitive psychological terms, this suggests that observers represent the actions of observed agents by mapping the agent's actions onto their own motor repertoire. Last, they point to evidence from single unit recording studies in monkeys (the nature of these experiments make them difficult to replicate in humans) that show mirror neuron activity to be predictive in nature, with neurons that fire during observation of grasping firing even when the final movement is obscured from view, suggesting that dynamic action-centered representations of other agents exhibit a type of representational momentum. That is, they extrapolate action beyond what is actually perceived. Taken together, these functional descriptions can be used to generate a purely functional/cognitive theory which generates behavioral predictions (which might not otherwise have been thought of) that can be tested with behavioral data. For example, it can be predicted that if observers predict the actions of others by mapping observed actions onto their own action repertoires, then it may be possible to detect the effects of the observer's action repertoire on the

---

[2] Although we have stripped the neuroscientific language from the example outlined below, it is still up for debate whether neuroscientific language *should* be stripped from cognitive explanations—we choose to avoid this debate in the present paper.

predictions they generate. In particular, action predictions should be more accurate when observed actions are more easily mapped onto the observer's action repertoires. That is, predictions should be more accurate when there is a better match between the observer's and the observed agent's action systems. This prediction has been tested in several experiments using a self/other prediction paradigm. (The logic behind this paradigm is that the self-prediction condition provides a perfect match between the observer's and the observed agent's action repertoire while the other-prediction condition does not). Evidence from behavioral studies confirms this prediction of the emulator theory (e.g., see Colling, Thompson, & Sutton, 2009; Flach, Knoblich, & Prinz, 2003). As more details about properties of mirror neurons come to light, these can be incorporated into the (neuro) cognitive emulator model and further novel behavioral predictions can be generated. For science to progress it is not enough to only falsify old theories with new empirical observations and to develop new theories that can account for the excess empirical content. Development of new theories that lead to the prediction of novel, or unexpected, facts (that are later corroborated) is also vitally important (Lakatos, 1970). Thus for neuroimaging to inform our theorizing about mind and behavior it need not (only) be used to falsify cognitive theories, but it can also lead to scientific progress when it predicts novel behavioral effects.

A non-reductive role for neuroscience In the preceding section, we highlighted two examples of how data from neuroscience can be used to inform cognitive theorizing. First, we showed that in certain cases structural properties of the brain can have functional consequences for cognition. Viewing cognition through the lens of biology can help to highlight those functional properties that might be missed by an abiological cognitive science. Second, we demonstrated that novel behavioral predictions can be generated based on findings from the neuroimaging literature, suggesting that the demands which Coltheart imposes on neuroimaging—that of distinguishing between two competing cognitive theories—fail to appreciate neuroimaging's value to the discipline of cognitive psychology. By viewing neuroimaging in this light, it places constraints on the possible information processing models for a particular cognitive function proposed by cognitive psychologists and serves to highlight those aspects of behavioral data that may be important. This, however, does not amount to a reduction to neuroscience as it allows for a conservation of the terminology, language and, most importantly, the methods of cognitive science. Some have argued that it is not enough to explain intelligent behavior with cognitive laws and mechanisms that are informed by neuroscience, but that cognitive laws/mechanisms must be replaced by neuroscientific laws/mechanisms. In the final section, we suggest that neuroscientific laws/mechanisms *cannot* replace cognitive laws and mechanisms (or psychological laws). Neuroscientific mechanisms fail to be reductions of cognitive mechanisms because they fail to be a

"more comprehensive theory [that] contains [the] explanatory and predictive resources that parallel those of the reduced theory" (Bickle, 1998, p. 28).

An extended mind argument against reduction

One of the standard defenses against reductionism has been multiple realizability. However, for the reasons outlined in the preceding section, the truth of multiple realizability seems doubtful and, therefore, a defense against reductionism based on multiple realizability seems untenable. Instead, we propose an extended mind argument against reduction, which holds that the cognitive sciences cannot reduce to neuroscience because to generate models of intelligent behavior we need to know about more than just what is happening in the brain. Neural explanations cannot replace cognitive explanations because they fail to parallel the predictive power, and richness, of cognitive explanations. In the following section, we will present this argument with the aim of demonstrating that the neural-centric view fails because it ignores the fundamental role that the organism's body and environment play in explaining intelligent behavior; it is often the case that the (cognitive) mechanisms underlying intelligent behavior reside not solely in the brain, but in the rich interaction between brains, bodies and environments.

This position lies in stark contrast to that of Bickle (2003), who puts forward a defense of what he calls "ruthless reductionism": the reduction of cognitive psychology not to cognitive neuroscience but rather to molecular and cellular neuroscience. Bickle's favoured example of such an approach—his "parade case for current reductionistic neuroscience" (p. 114)—is the reduction of long-term memory consolidation to molecular mechanisms of long-term potentiation (LTP), and he does not spare any details in his thorough treatment of the molecular mechanisms responsible for LTP. We commend this detailed molecular neuroscientific investigation just as we have commended connectionism and functional neuroimaging. Cellular and molecular neuroscience have an important part to play if they give us a more fine-grained account of those functional properties of neurons and neural networks that are causally relevant for the production of intelligent behavior. Molecular neuroscience (especially studies of LTP) can, for example, provide us with information about changes in synaptic efficacy (connection weights) and their microfunctional consequences. Where we part company with Bickle, however, is in his conclusion that a detailed exposition of neural mechanisms naturally leads to reductionism. An examination of the extended mind argument will show why this conclusion is flawed.

The extended mind argument, as put forward by Clark and Chalmers (1998), simply stated, is that cognition is more than what happens in the head. On this view, those parts of the external environment on which certain aspects of cognition rely may be properly counted as part of the cognitive system producing intelligent behavior. At the heart of the argument is the parity principle. In explaining

the parity principle, Clark and Chalmers ask the reader to consider two situations: the first involves a subject performing some task entirely "in the head" through the mental manipulation of some internal symbol or representation; the second involves the same task being completed but this time with the previously internal symbols realized as some external symbol, representation, or artifact. They argue that if the first case of mental manipulation is to be considered cognitive then the second case, where physical manipulation of a physical object fills the same role as mental manipulation, should also be considered cognitive. It is important to note, however, that the extended mind argument that we endorse does not require a strong functional identity between extended and purely internal cognitive processes. Nor is it necessary that it must be possible for extended cognitive processes to also be done "in the head". Cognitive processes that extend into the body and environment do so because, often, these types of cognitive processes cannot efficiently, or easily, be performed in the head. In some cases, they cannot be performed in the head at all and, therefore, cognitive extension allows embodied organisms to acquire cognitive functions that might not otherwise be available to their (imagined) disembodied counterparts. In this way, certain cognitive processes rely as much on the body and the ability to manipulate the environment as they do on the brain. Remove any part and the mechanism collapses.

Much of the empirical research into the extended mind has drawn a distinction between what Kirsh and Maglio (1994) call "epistemic actions" and "pragmatic actions". Pragmatic actions are actions aimed at achieving some particular goal, such as walking towards an object that is desired. Epistemic actions, on the other hand, are physical actions that aid mental computation. This might be achieved by reducing memory requirements for, increasing the speed of, or increasing the reliability of, a particular computation. To illustrate the role of epistemic actions we turn to an example from the experimental literature.

Human working memory has a limited capacity (Cowan, 2001), which imposes constraints on any computational process that includes working memory. A computational advantage is to be had then if a cognitive system does not utilize working memory to its full capacity, but instead offloads information onto the environment or, alternatively, simply does not represent information in working memory at all, but rather extracts information from the environment when necessary. Evidence from eye-tracking studies suggest that people routinely do leave information out in the world accessing it as it is needed rather than representing it all internally. These studies suggest that eye movements can act as epistemic actions, aiding computation. An account of the role of eye movements (fixations and saccades) has been developed by Ballard et al. (1997) with their notion of deictic pointers.

One pragmatic reason for fixation with the fovea is to increase the spatial resolution of the object being looked at. Ballard et al. (1997), however, suggest that fixation plays a role beyond merely increasing spatial resolution; they suggest that it plays a direct role in information processing. In a number of experiments using a block-copying paradigm, Ballard et al. asked subjects to duplicate a particular spatial arrangement of colored blocks while their eye and hand movements were tracked. The results of these experiments showed that subjects made far more eye movements than would be expected if they were using working memory to its maximum capacity. Often, subjects would perform a saccade from the model to the block resource area, select a block, perform a saccade back to the model, and then to the assembly area before dropping the block in the correct location. Ballard et al. suggest that the saccade back to the model after the block has already been selected suggests that this saccade serves the purpose of retrieving the block's spatial location from the model, with the block's color already retrieved (and possibly now discarded from working memory) in the first fixation of the model. They suggest that this minimal memory strategy is preferred because of the high cost of maintaining items in working memory. Saccades, however, also carry a cost in terms of time, with saccades and online environmental retrieval being slower than internal retrieval from working memory.

The cost of online retrieval can be increased by increasing the spatial separation between the model block pattern and the block resource area, thereby introducing the need for head movements. In this case, the two competing costs (of body movements for online retrieval and working memory usage) are weighed against each other when selecting the strategy for performing the block-copying task. If the increase in temporal or metabolic costs is too great, then eye movements dramatically decrease and subjects utilize more working memory for this task, employing a maximal memory strategy (Ballard et al., 1997).

Ballard et al. (1997) suggest that these two strategies amount to two distinct computational approaches to solving the single problem of block copying. In the minimal memory strategy, eye movements serve as deictic pointers. A deictic pointer, like pointers in computer programming terminology, is a variable that stores the location of a piece of stored data. In this way, they differ from regular variables that store the data itself. In the block-copying example, the piece of stored data might be the block's color or spatial location. In the minimal memory strategy, this data is stored out in the environment; in the maximal memory strategy, this data is stored directly in working memory. Thus, eye movements in the minimal memory strategy serve to retrieve the data from the data store (the environment) based on the information stored in the (internal) pointer. In the maximal memory strategy, a functionally similar process retrieves the same data from the internal data store (working memory). In this way, each computational strategy can be functionally decomposed into parts (data store, central buffer) and operations (retrieval) yielding two, functionally similar, but not identical, cognitive/information-processing mechanisms. In one case, the parts of the mechanism are

located solely inside the brain and the operations are performed solely by the brain. In the other case, however, the parts are located in both the brain and the environment, and operations are performed by the brain and the body. According to the parity principle, this causal spread requires a spread of causal credit (Clark & Chalmers, 1998), and the result is causal equivalency between the neural and the extra-neural.

Bickle (2008) has criticized the idea of causal equivalency between the neural and the extra-neural. He argues that the neural events have a "causal priority over all other causes of behavior, beyond the fact… that brain events are the most proximal causes." (p. 472). In particular, Bickle takes issue with Clark and Chalmers (1998) when they state: "[I]f we retain internal structure but change the external features, behavior may change completely. The external features here are just as causally relevant as typical internal features of the brain." (p. 9). According to Bickle there are two ways to interpret this statement. The first, he claims, is trivial and amounts to nothing more than the claim that the external world produces inputs for the brain with the brain maintaining causal priority over behavior. The second, however, is "spooky" and "simply empirically implausible" (p. 472). This reading is "spooky" because, as Bickle quite rightly points out, changes in behavior are not possible unless there is also a change in the particulars of neurotransmitter release at the neuromuscular junction. This, however, is not something that anybody would deny. Furthermore, accepting that changes in neurotransmitter release at the neuromuscular junction is required for changing behavior amounts to nothing more than accepting that neural events are the most proximal causes of behavior (save possibly for chemical changes in muscle fibers). However, for Bickle, neural events are the locus of intelligent behavior over and above the fact that they are the most proximal causes of intelligent behavior. Thus, it is clear that the "spooky" reading is not the one to be endorsed by extended mind theorists and Bickle's refutation of the "spooky" reading is accepted, although it does nothing to advance his argument. Therefore, it is clear that the issue is with the first reading. In particular, the issue is whether the environment should be merely characterized as input, as Bickle's first reading suggests, or whether it is something more. The studies by Ballard et al. (1997) might help to illustrate how the environment is more than just input.

The key finding from Ballard et al. (1997) is that changing the environment by moving the position of the model, the block resource area, and the assembly area has the effect of changing the computational algorithm or mechanism that is implemented: closer together and the larger mechanism includes neural operations that function as deictic pointers; further apart and the larger mechanism includes neural operations that "represent" objects and features of the outside world. Changes in the environment result in changes to the mechanism employed to complete the task, by altering the division of labor between neural and extra-neural processes. Furthermore, if, when examining the

environment-spanning mechanism, we conclude that those aspects of the physical environment that are causally relevant to behavior are merely inputs then what is to stop one deciding that in the purely internal case the physical aspects of particular neural structures that are causally relevant to behavior are also merely inputs: are the neural operations in V1 merely inputs to "higher" cortical areas or are they to be considered part of the mechanism? If the former is true, then where exactly in the causal chain does one draw the line between what is input and what is not? We agree with Chemero (2007) that psychology, and cognitive science, as we have presented it here, is the study of the *brain*-body-environment system. By "refus[ing] to grant the psychological any genuine causal explanatory role" (2003, p. 115), Bickle is suggesting that we ignore the way in which environmental differences change the mechanisms that are implemented and he thereby fails to appreciate the true causal role of the environment. In brushing over these differences, Bickle's reductionism results in a science of mind that focuses on only part of a much larger mechanism, and while it is true that the effects of environmental changes on behavior are mediated/modulated by neural processes, it is certainly not the case that one can explain or predict behavior based solely on a description of the neural parts and operations of this larger mechanism. Indeed, such a restriction fails to even recognize the extra-neural structures and processes as part of the mechanism! Bickle's commitment to a brain-only account of behavior amounts to nothing more than an *a priori* assertion that those facts about the world that are important for understanding intelligent behavior *are* those facts about the world that are studied by neuroscience.

Conclusion

For much of its history cognitive science has operated in a biological vacuum. This is a direct result of the view that the mind is the software of the brain. We have argued, however, that a clear-cut division between software and hardware does not exist. Rather, facts about how a system is implemented can (and does) have consequences for the functional (behavioral) properties of that system. Furthermore, we have tried to highlight the role that neuroscience has to play in theoretical progress in the field of cognitive psychology by examining how neuroscience, and neuroscientific reasoning about cognition, can focus attention on what might otherwise be neglected data. Neuroimaging, with its ability to uncover facts about brain function during certain tasks, informs cognitive psychology by enabling the generation of novel cognitive theories and behavioral predictions. However, although we urge a greater role for neuroscience within cognitive science, we suggest that reductionist tendencies should be resisted. If to understand intelligent behavior we need to understand the dynamic interactions between brain, body and environment, as we have claimed, then brain reductionism with its singular focus only results in a decrease of explanatory power. Understanding the brain is vitally important to

understanding intelligent behavior, but it is only one part of the greater puzzle. Along with the downward trend of reductionism there needs to be an outward expansion into the body and environment of biological organisms. It is only through this interdisciplinary endeavor that it will be possible to develop a better understanding of how intelligent behavior is produced.

Acknowledgments

References

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *The Behavioral and Brain Sciences*, *20*, 723–742.

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer Academic Publishers.

Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge Mass.: MIT Press.

Bickle, J. (2008). The molecules of social recognition memory: Implications for social cognition, extended mind, and neuroethics. *Consciousness and Cognition*, *17*, 468–474.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, *13*(2), 400–404.

Calvo-Merino, B., Glaser, D., Grezes, J., Passingham, R., & Haggard, P. (2004). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cereb. Cortex*, *15*, 1243–1249.

Chemero, A. (2007). Asking what's inside the head: Neurophilosophy meets the extended mind. *Minds and Machines*, *17*, 345–351.

Clark, A. (1989). *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. Cambridge Mass.: MIT Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.

Colling, L. J., Thompson, W. F., & Sutton, J. (2009). Action synchronisation with biological motion. Presented at the 9th Conference of the Australasian Society for Cognitive Science, Sydney, Australia.

Coltheart, M. (2004). Brain imaging, connectionism, and cognitive neuropsychology. *Cognitive Neuropsychology*, *21*, 21–25.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, *35*, 125–129.

Flach, R., Knoblich, G., & Prinz, W. (2003). Off-line authorship effects in action perception. *Brain and Cognition*, *53*, 503–513.

Gibson, J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

James, W. (1905). *The principles of psychology* (Vol. 2). New York: Henry Holt and Company.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science: A Multidisciplinary Journal*, *18*, 513–549.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965* (pp. 91–196). Cambridge UK: Cambridge University Press.

McClelland, J. L. (1985). Putting knowledge in its place: A scheme for programming parallel processing structures on the fly. *Cognitive Science: A Multidisciplinary Journal*, *9*, 113–146.

Mozer, M. C. (1983). Letter migration in word perception. *Journal of Experimental Psychology. Human Perception and Performance*, *9*, 531–546.

Newell, A., & Simon, H. A. (1997). Computer science as empirical inquiry: Symbols are search. In J. Haugeland (Ed.), *Mind design II: philosophy, psychology, artificial intelligence* (Revised and enlarged edition., pp. 81–110). Cambridge Mass.: MIT Press.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*, 419–421.

Shapiro, L. A. (2000). Multiple realizations. *The Journal of Philosophy*, *97*, 635–654.

Stevens, J., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: selective neural encoding of apparent human movements. *NeuroReport*, *11*(1), 109–115.

Thagard, P. (1986). Parallel computation and the mind-body problem. *Cognitive Science: A Multidisciplinary Journal*, *10*, 301–318.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*, 460–473.

Citation details for this article: