Synthese (2009) 167:511–539
DOI 10.1007/s11229-008-9389-4

The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience

Jacqueline A. Sullivan

Received: 1 July 2008 / Accepted: 1 July 2008 / Published online: 29 August 2008
© Springer Science+Business Media B.V. 2008

Abstract Descriptive accounts of the nature of explanation in neuroscience and the global goals of such explanation have recently proliferated in the philosophy of neuroscience (e.g., Bechtel, *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Lawrence Erlbaum, 2007; Bickle, *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer Academic Publishing, 2003; Bickle, *Synthese, 151*, 411–434, 2006; Craver, *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press, 2007) and with them new understandings of the experimental practices of neuroscientists have emerged. In this paper, I consider two models of such practices; one that takes them to be *reductive*; another that takes them to be *integrative*. I investigate those areas of the neuroscience of learning and memory from which the examples used to substantiate these models are culled, and argue that the multiplicity of experimental protocols used in these research areas presents specific challenges for both models. In my view, these challenges have been overlooked largely because philosophers have hitherto failed to pay sufficient attention to fundamental features of experimental practice. I demonstrate that when we do pay attention to such features, evidence for reduction and integrative unity in neuroscience is simply not borne out. I end by suggesting some new directions for the philosophy of neuroscience that pertain to taking a closer look at the nature of neuroscientific experiments.

Keywords Experiment · Experimentation · Explanation · Learning · Long-term potentiation (LTP) · Mechanism · Protocol · Reduction · Reliability · Unity · Validity

J. A. Sullivan (✉)
Department of Philosophy, University of Alabama at Birmingham,
HB 414A, 900 13th Street South, Birmingham, AL 35294-1260, USA
e-mail: jas1@uab.edu

1 Introduction

A primary aim of recent philosophy of neuroscience has been to characterize the nature of neuroscientific explanation. Part of the project has been to describe the nature of the *experimental practices* that yield neuroscientific explanations. To date, a predominant number of the case studies used as a basis for such descriptive accounts are derived from the neurosciences of learning and memory. Common examples include research on (1) long-term potentiation (LTP), the purported neural correlate of learning and memory, (e.g., Bechtel 2007; Bickle 2003, 2006; Craver 2001, 2002, 2003, 2007) (2) spatial memory in the rodent (e.g., Craver 2007; Craver and Darden 2001) and (3) cross-species memory consolidation processes (Bickle 2003, 2006). For example, John Bickle has used neurobiological experiments on *LTP*, *classical conditioning* and *social recognition memory* to defend the view that in low-level neuroscience, experimental practices are '*ruthlessly reductive*'—namely, they reduce or have the potential to reduce higher-level sciences to lower-level sciences and mind to molecular activity. Carl Craver, in contrast, has used neurobiological research on LTP and spatial memory to argue that experimental practices in neuroscience are directed at the discovery of *mechanistic explanations* of phenomena. On Craver's model, instead of one area of science being reduced to another, fields of research in neuroscience become *unified* or *integrated* as investigators across fields contribute, either directly or indirectly, to the goal of providing a mechanistic explanation for a phenomenon. Discoveries in various fields purportedly provide constraints on the components that may be included in the explanation until it becomes adequate or complete.

A principal feature of both reductive and mechanistic accounts of neuroscientific practice is that they originate in the context of the *philosophy of explanation*. Furthermore, and crucially, these accounts treat almost exclusively of those features of research on LTP and memory that can be used to validate the particular model of scientific explanation being argued for while overlooking or failing to emphasize *differences* in the experimental protocols that different laboratories use when investigating LTP and learning and memory. Consequently, they have failed to recognize the implications of such differences for global issues concerning reduction and unity of science.

In Sect. 2 of this paper, I briefly define what an experimental design and experimental protocol are and describe a very basic set of features that experimental protocols in the field of cellular and molecular cognition have. This is meant to lay the groundwork for the analysis I undertake in Sects. 3 and 4, where I demonstrate that a consideration of the multiplicity of experimental protocols in cellular and molecular cognition alone suggests that evidence for reductionism and methodological or explanatory unity in neuroscience is simply not borne out. What we encounter instead is a plurality of different experimental strategies, which in turn raises interesting challenges to understanding how the results that such diverse strategies yield fit together. Such challenges must be met by proponents of either the reductionist or unity pictures. In Sect. 5, I use a conceptual framework to explain why such differences in experimental protocols across laboratories emerge, given specific constraints on the experimental process that may be differentially realized across laboratories. I use this framework to better articulate the kinds of challenges faced by the reductionist and unity of science accounts.

I end, in light of such challenges, with a consideration of some new investigative directions for the philosophy of neuroscience.

2 The structure of experimental protocols in cellular and molecular cognition: a first pass

What follows is a preliminary framework for thinking about some basic features of the types of experimental designs and protocols that we encounter in the field of molecular and cellular cognition. Here, I am only interested in behavioral and electrophysiological experiments, so I am setting aside those features of, for example, the biochemical and immunocytochemical experiments that often accompany such studies. The framework I offer is derived primarily from a general investigation of the methods sections of journal articles from the field of cellular and molecular cognition and from my own experience working in this area (Sullivan 2003). Although some of the details captured in this framework emerge in Bickle's (2003, 2006) and Craver's (2001, 2002, 2007) descriptions of the features of particular neurobiological experiments on learning and synaptic plasticity, neither author treats of these features in the abstract as I do here. Nor, as I shall argue, do they recognize their consequences.

The basic goals of experimentation in cellular and molecular cognition are to identify the cellular and molecular processes productive of learning and memory. There are two primary types of experiments in cellular and molecular cognition: (1) *learning or "memory consolidation" experiments*, in which animals are trained in a learning paradigm, and which are run in conjunction with intervention techniques that inhibit or exacerbate the activity of a molecule prior to, during or after training and (2) *synaptic plasticity experiments*, done either in vivo, at some synaptic location in the brain (e.g., Schaffer commissural projections from CA3 to CA1 of the hippocampus) or in vitro, at some synaptic location in a slice of brain tissue (e.g., hippocampal slices). In plasticity experiments, an investigator applies electrophysiological stimulation to one or more pre-synaptic neurons and records from one or more post-synaptic neurons before and after the application of a pattern of electrophysiological stimulation designed to produce a change in the strength of the synapse. These experiments are also accompanied by intervention techniques that inhibit or exacerbate molecular activity before, during and/or after the application of the plasticity-inducing electrophysiological stimulation.

All learning and synaptic plasticity experiments have to be *designed*. The outputs of what I refer to as the "design stage" of the experimental process (See Sect. 5) are an *experimental design* and an *experimental protocol*. I treat of both items in the abstract. An experimental design includes the overall set-up of the experiment, in so far as it specifies such things as the experimental context (e.g., how and where objects are to be arranged) and the materials and methods to be used (e.g., See Hacking 1992). The experimental protocol is the set of step-by-step instructions that an investigator follows each time he or she runs an experiment. An experimental protocol essentially specifies how each individual learning or plasticity experiment is to be run from start to finish. When an investigator is in the middle of an experiment and confused about what to do next—he or she will refer to the experimental protocol (not the experimental design).

When I say that experimental designs and their associated protocols can differ across laboratories, I mean that features of the entire overarching experimental design and protocol may differ with respect to such things as what kinds of drugs will be used for pharmacological manipulations and their concentrations, how the animals are to be handled prior to training them in a learning paradigm, how many times a stimulus may be applied at a synapse, which steps will precede or follow other steps. In those instances in this paper where I identify differences across laboratories in terms of experimental protocols (sp., in Sects. 3 and 4), my focus is primarily on differences in those *protocols* associated with learning and electrophysiological stimulation paradigms (described below). However, clearly other differences can arise across laboratories in terms of the features of entire experimental designs and protocols. For the sake of brevity, I can only tell part of the story and its implications here.

Learning and plasticity experiments require the identification of a form of learning or synaptic plasticity of interest and the selection of (a) a model organism (e.g., the sea mollusk, the rat, the mouse), (b) an area of the brain or brain correlate within that organism in which the learning is purported to take place or in which a plasticity inducing electrophysiological stimulation will be applied (e.g., the hippocampus) and (c) a set of synapses or a synaptic location in that area thought to be involved in some form of learning or where a plasticity-inducing electrophysiological stimulation is to be applied (e.g., sensory-motor neuron pathway in *Aplysia*, Schaffer-collateral pathway in the rat).

The form of learning or plasticity of interest is always *operationally defined*. An operational definition is built directly into the design of an *experimental paradigm*. An experimental paradigm is a standard method or procedure for producing an effect of a specific type (i.e., a form of learning or a form of synaptic plasticity). The following features are typically included in the design of an experimental paradigm for learning or synaptic plasticity: (1) *production procedures*, namely, a specification of the stimuli (independent or input variables) to be presented to the organism or synapses, how those stimuli are to be arranged (e.g., spatially, temporally) and how many times they are to be presented during phases of (a) pre-training/pre-tetanus, (b) training/tetanic stimulation and (c) post-training/post-tetanus; (2) *measurement procedures* that specify the response variables to be measured in the (a) pre-training/pre-tetanus (b) post-training/post-tetanus phases of the experiment and how to measure them using apparati designed for such measurement; (3) *detection procedures* that specify what the comparative measurements of the response variables from the different phases of the experiment must equal in order to be able to ascribe learning to the organism or a plastic change to a synapse or set of synapses. This detection procedure is simply an operational definition that specifies the measurable "change" in response variables from pre- to post-training as a result of the stimuli presented that must be observed in order to say that the form of learning or type of plastic change has occurred. It is a comparison of the data points produced pre- and post-training/or pre- and post-stimulation that is used to make such a determination.

What I will herein refer to as a "*sub-protocol*" is a production procedure, written up step by step, which corresponds to a learning or electrophysiological stimulation paradigm. It will, for example, specify (1) the duration of time of the presentation of each stimulus to be used in an experiment, (2) the duration of time that is to elapse between

presentation of the stimuli used in an experiment, or the inter-stimulus interval (ISI), (3) the amount of time that is to elapse between individual trials, or the inter-trial interval (ITI) and (4) the amount of time that is to elapse after the last trial before the brain is removed and the relevant brain areas or synaptic locations are removed for biochemical analysis (that is, if brain removal is part of the experiment).

In molecular and cellular cognition, these features of sub-protocols are often dictated by such considerations as the number of trials, stimulus intensity and stimulus duration taken to be required to produce detectable and measurable molecular effects and knowledge of the time points at which specific molecules will purportedly be operative during a learning or plasticity event. Yet, as I will later show (Sect. 5), the precise features of sub-protocols often differ at least subtly across different laboratories that study learning and synaptic plasticity, mostly for idiosyncratic reasons. It perhaps goes without saying, then, that the overarching experimental designs and protocols may differ, also for idiosyncratic reasons.

By using the framework I have provided here, we can readily approach the methods section of any published research paper in the area of molecular and cellular cognition and pick out the primary features of the overall experimental design and experimental protocol used. Then we can compare designs and protocols across research papers to identify similarities and differences between them, and raise the question of what the implications of such similarities and differences are for our philosophical projects. What I aim to show in my analysis of Bickle's and Craver's work is that when we do this, we encounter a multiplicity of different approaches to the study of learning and synaptic plasticity in cellular and molecular cognition and this has specific implications for reductionist and mechanistic pictures of unity of science and the kind of evidence that cellular and molecular cognition alone can provide with respect to shedding light on the relationship between the mind and the brain.

3 Multiplicity and "ruthless reductionism"

John Bickle has perhaps been the most ardent proponent of reductionism in recent years. In *Philosophy and Neuroscience: A Ruthlessly Reductive Account* (2003) he uses data from experiments on synaptic plasticity and learning and memory to argue that the reduction of cognitive psychology to cellular and molecular neurobiology and of mind to molecules is forthcoming in contemporary neuroscience. More recently (Bickle 2006, 2007), he has claimed that experiments in cellular and molecular neuroscience are by their very nature 'reductive-in-practice'. This term of art, coined by Bickle, is meant to capture the idea that the reduction we find in contemporary neuroscience is not classical intertheoretic reduction (Nagel 1961). Rather, low-level areas of neuroscience redefine the psychological entities and processes studied by cognitive neuroscience in terms of observable changes in behavior. They then use molecular intervention techniques, which, when successful, show that these behavioral changes may be accounted for by changes in molecular activity.

In the kinds of experiments Bickle has in mind, an investigator begins with a "well-accepted" behavioral protocol associated with a given learning paradigm, which specifies how to produce a psychological function and how to detect when it has

occurred—the latter by reference to observable changes in behavior. Animals in which certain pharmacological or genetic interventions have been undertaken are trained using this protocol. Statistically significant differences in the behavior of these animals compared to that of wild-type normal controls are taken as indicative of the necessity or irrelevance of the activity of the manipulated variable (e.g., the activity of a molecule) in the production of the observed differences in behavior.[1]

To get clear on the nature and import of this strategy, it is worthwhile to consider a case study that Bickle (2006) puts forward to illustrate it: experiments undertaken to determine if cyclic-AMP response element binding protein (CREB)—a transcriptional regulator poised to target immediate early genes and impact structural changes at brain synapses—is involved in social recognition memory in mice. On a most basic definition, social recognition memory consists in the ability to recognize another individual at a later point in time after an initial episode of interaction with that individual. It is thought to be dependent in part on the hippocampus, a brain structure that has been implicated in a variety of different forms of learning and memory. Because CREB has been shown to be involved in other forms of hippocampus-dependent long-term memory, the authors whom Bickle cites, Kogan et al. (2000), were primarily interested in determining if CREB activation was also involved in the formation of long-term social recognition memory in mice.

To achieve this goal, Kogan et al. (2000) obtained knock-out mice lacking expression of the α and δ isoforms of CREB from the lab of Gunther Schutz (Hummler et al. 1994), and then crossed them with wild-type mice to produce $\text{CREB}^{\alpha\Delta^-}$ mutant mice–mice deficient in two isoforms of CREB, and in which the amount of activated CREB in the brain is significantly attenuated (i.e., levels are lower than normal levels).[2] They then trained a group of these mutants ($n = 7$) and a group of normal wild-type mice ($n = 7$) using a modified version of a behavioral protocol associated with a previously developed learning paradigm (Thor and Holloway 1982, p. 1001). In these experiments:

> a male juvenile mouse [sp., a novel juvenile wild-type male less than 5 weeks of age] was placed into a cage with an adult [a CREB mutant or wild-type mouse who had been habituated to the cage for 15 min] for an initial interaction trial of 2 min. […] Following the intertrial delay [24 h], the same juvenile was placed back into the adult's cage [for a 2-min test trial]. The social investigation of the juvenile by the adult mouse was observed continuously by a trained observer who timed the duration of investigation behavior with a hand-held stopwatch. Behaviors that were scored as social investigation […] included the following: direct contact with the juvenile while inspecting any part of the body surface (including grooming, licking, and pawing), sniffing of the mouth, ears, tail, anogenital area, and close following (within 1 cm) of the juvenile. (Kogan et al. 2000, p. 48)

[1] This description corresponds roughly to those details about learning experiments and their associated designs and protocols that I provided in Sect. 2.

[2] As Bickle (2006) provides extensive detail about the processes by which these mutants are produced, and as such details are not immediately relevant to the criticisms I wish to make, I will refrain from providing them here. I simply refer the reader to Bickle (2006) and/or Kogan et al. (1997, 2000).

Using this experimental design and protocol, Kogan et al. (2000) found that, in comparison to "normal" wild-type controls, $CREB^{\alpha\Delta-}$ mutant mice engaged in stereotypic examination of the juvenile mouse to the same extent after 24 h as they did upon the initial presentation of the juvenile. They took this as indicative of a failure to encode, in the long-term, the information requisite to recognize the juvenile on the subsequent presentation, which they attributed directly to the CREB mutation. So, experiments on long-term social recognition memory in $CREB^{\alpha\Delta-}$ mutant mice can be shown, as Bickle claims, to instantiate the basic features of the so-called "reductive" strategy of intervening at the level of molecular activity and tracking observable changes in behavior.

Now, Bickle explicitly asserts that when research strategies like the one employed in the Kogan et al. (2000) study are successful, the molecular interventions undertaken *directly explain the behavioral data*:

> the cellular and molecular events in specific neurons into which experimenters have intervened, in conjunction with the neuronal circuits in which the affected neurons are embedded, leading directly to the neuromuscular junctions bridging nervous and muscle tissue, *directly explain* the behavioral data. (Bickle 2006, p. 426)

However, there is an important question what exactly this claim amounts to. On a plausible as well as charitable interpretation, we might take Bickle here simply to mean that the molecular interventions undertaken by an investigator (or multiple investigators) directly explain the behavioral data observed by *that* investigator in *that* laboratory. Certainly, this seems to be what he *ought* to mean, because he clearly cannot mean that these interventions directly explain behavioral data observed by another investigator in another laboratory studying long-term social recognition memory in the mouse (unless the behavioral protocol and intervention techniques used were identical). Nor can he mean that these interventions can be used to directly explain long-term social recognition memory broadly construed as some phenomenon we are able to detect in the world, which organisms exhibit. Yet, if all that Bickle *does* mean is that the interventions explain the behavioral data in this one laboratory, an interesting problem arises for his attempt to mount a case for the reduction of psychology to cellular and molecular neuroscience and of mind to molecules. Getting clear on the precise nature and dimensions of this problem will be my aim in the rest of this section.

Let me begin by teasing apart four types of reductive claims to which Bickle appears to be committed—two of which are "local" and two "global". The first 'local' reductive claim captures an uncontroversial sense in which experiments in cellular and molecular neuroscience are 'reductive-in-practice'. Reduction-in-practice is something that occurs in an individual experimental laboratory when an investigator operationalizes a psychological function (e.g., "long-term social recognition memory consolidation") by developing a protocol that specifies how to produce that function (e.g., 2 min presentation of novel conspecific, 24 h ITI) and detect when it occurs, by reference to observable changes in behavior (e.g., increases or decreases in "investigatory behavior" such as sniffing). When the investigator intervenes at the molecular level (e.g., "knocks-out" a gene), observes a statistically significant difference in the

behavior of these animal subjects in comparison to normal ("wild-type") controls, and has good reason to believe (i.e., adequate controls were in place) that this difference can only be attributed to the intervention, he/she can interpret the data as indicating that the intervention resulted in the differences in behavior observed between the two groups. Therefore, those differences at the molecular level between the two groups can be used to *directly explain* the observed behavioral data. Note, that what we have here is at best a "local reduction", because it is in this local context that the investigator has "set aside" the need to refer to the original psychological function of interest—for he/she has operationalized it and linked it directly to molecular activity (e.g., disruption in normal CREB function). So, it is in *this* laboratory that those areas of science that make reference to "intervening explanatory levels, including the psychological, cognitive/information processing, even the cognitive neuroscientific" (Bickle 2006, p. 426) may be claimed to have no place. This, in sum, is how, in molecular and cellular cognition, investigators effect local reductions of psychology and cognitive science to molecular and cellular neurobiology. Similarly, it must be exclusively in this local experimental context that a traditional problem in philosophy of mind may be claimed to have been solved. A psychological or mental kind has been operationalized and linked to molecular activity, and Bickle's second local reductive claim, which pertains to the philosophy of mind, is that the investigator, as a consequence, need make no reference to a psychological or mental kind.[3]

Such local reductions—if we want to call them reductions at all—are at best extremely modest, and in no way warrant the designation of "ruthless". Indeed, it is an open question if these are reductions at all for any number of reasons. For one thing, the strategy of investigators in molecular and cellular cognition is to *assume* that the operationalizations that they provide are actually indicative of the function of interest—this is, of course, an open question. Other reasons include the fact that, sometimes what has been learned is not exhibited in overt behavior (e.g., Taylor 1964); sometimes one's itemized list of possible behaviors associated with the function is incomplete, or the way in which one has selected to produce a behavior only captures one way when there are others. There are a number of interesting issues that arise for the methodology we encounter in molecular and cellular cognition. As these issues are not the direct focus of this paper, I will set them aside here. Now, given his commitment to 'ruthlessness' on a grand scale, Bickle would not be satisfied with such 'local reductions' as investigators in this area may be capable of achieving. This brings me to what I take to be the two types of *global* reduction that Bickle propounds.

The first type of global reduction may be formulated as follows: Given the current success of the strategy of bringing about local within-experimental-protocol-reductions in molecular and cellular cognition, such as the successful strategy we encounter in the social recognition memory case, we can anticipate the gradual elimination of all of the psychological/functional kinds posited by high-level cognitive science/psychology. This will occur as investigators working in different laboratories in molecular and

[3] Admittedly, this is an odd 'local' claim, because investigators in molecular and cellular cognition *assume* the priority of their approach for understanding psychological functions, so issues about the mind never arise for them. For the sake of addressing scientific problems that interest them, they necessarily have to assume such priority. But then, this looks more like eliminativism than reductionism.

cellular cognition, select psychological kind after psychological kind for experimental study, operationalize each kind in terms of a protocol that will produce changes in behavior indicative of the kind, and link those changes in behavior to molecular activity. This is what we encountered in the case of long-term social recognition memory consolidation and other types of long-term memory consolidation (Bickle 2003), and we will witness the gradual 'reduction' of all psychological functions including attention, motivation, etc. to molecules. Once these psychological kinds studied by the higher-level sciences are eliminated, the areas of science that study them will also be eliminated. Investigators in cellular and molecular cognition, or so the claim goes, have already shown that once these areas of science have localized functions to specific brain areas, they are of no further use.

The second type of global reduction is similar to the first, but applies exclusively to the philosophy of mind. The idea here is that, gradually, local reductions-in-practice will converge so as to yield the reduction of *all* mental functions to cellular and molecular processes in neuroscience. Reductive physicalism in philosophy of mind will be realized; and all physicalists will be required to do is appeal to the vast body of data obtained in molecular and cellular cognition in order to defend their arguments.

With these four types of reduction identified, I am now in a position to characterize what I take to be two fundamental problems for Bickle's (2006, 2007) position, given how he understands local reductions and given his ultimate commitment to global reduction.[4] The first problem concerns how to get 'ruthless' reductionism out of very weak local claims. At best what Bickle (2003, 2006) has achieved by appeal to experiments in molecular and cellular cognition is a case for many local "within-lab" reductions. What is entirely lacking, however, is a convincing, or even initially plausible account of how these local reductions will ultimately converge so as to effect global reduction. The second problem is how—even granting that these local claims are reductionist—one could be entitled to *generalize* them beyond the context in which they were produced. As the extent of these problems may not be immediately clear, I will appeal to the case of long-term social recognition memory in order to make them more salient.

The first feature to note about forms of learning and memory like social recognition memory is that a large number of neurobiological laboratories other than the Silva lab study its cellular and molecular bases in the mouse (See references below). It is thought to depend not only on the hippocampus, but also on the amygdala (e.g., Ferguson et al. 2001). CREB is not the only molecule of interest; some labs have used gene knock-outs to determine the role of oxytocin (e.g., Ferguson et al. 2001) and

[4] Bickle's reductionist claims have been the target of much recent philosophical criticism and these criticisms may be regarded as directed at Bickle's global reductionist claims. Aizawa (2007), for example, has argued convincingly that Bickle fails to dodge Putnam's traditional argument from *multiple realizability*, in so far as the molecules that realize memory consolidation are multiply realizable across species at the level of amino acid sequences. Therefore, he takes Bickle to lack adequate evidence for the reduction of mind to brain. Schouten and Looren de Jong (2005) and Looren de Jong independently (2006) have accused Bickle of taking the reductive claims of cellular and molecular neurobiologists even farther than they themselves are willing to take them. Also assuming that Bickle (2003) has the global reduction in mind, I argued (Sullivan in press), that the type of reduction he is entitled to lay claim to is modest at best. However, none of these criticisms of Bickle have been informed by a consideration of the nature of experimentation in molecular and cellular cognition.

vasopression (e.g., Bielsky et al. 2005) in social recognition memory in the mouse.[5] Other labs (e.g., Richter et al. 2005) have sought to determine the role of protein synthesis in long-term social recognition memory. From one laboratory to another, the experimental protocols with respect to those production procedures used to produce social recognition memory in the mouse differ. Furthermore, these protocols differ with respect to other features, which I will not elaborate on here.

I want to consider one example from the literature on social recognition memory in the mouse to make my case. Richter et al. (2005) were interested in investigating the role of *protein synthesis* in long-term social recognition memory consolidation. They used a behavioral protocol to produce and detect social recognition memory in mice in which an adult male was exposed to a novel juvenile conspecific for a 4 min initial exposure trial (compared to Kogan and colleagues' 2 min initial exposure trial). This was followed by a 24 h intertrial interval (identical to Kogan and colleagues' 24 h (ITI)) and then a 4-min test trial in which the juvenile mouse presented during the initial trial was presented to the adult mouse together with a novel juvenile mouse of the same strain (compared to Kogan and colleagues' 2-min test trial in which only the juvenile mouse presented on the 1st trial was re-presented). Although Richter et al. (2005) employed what they refer to as a "social discrimination" task, they identify the effect that they observe in the laboratory as "social recognition memory".[6] Richter et al. (2005) used an inhibitor of protein synthesis, anisomyicin, which Kogan et al. (2000) also used in one set of experiments.[7] However, the behavioral protocols still differed between the two laboratories in the anisomyicin experiments. There were other differences between the entire experimental designs and protocols used in these two studies. For example, the adult mice used in the Kogan et al. (2000) study were 4–8 months old, whereas the adult mice used in the Richter, Wolf and Engelmann study were of a different strain, and were 9–16 weeks old. And, it should be noted, these are only some of the differences that clearly need to be taken into account.

If we were to follow Bickle's line of reasoning, what we should say about the Richter et al. (2005) study is that they intervened at the level of protein synthesis and tracked an observable change in behavior and therefore could appeal to the intervention to explain the behavioral data. However, given the differences in the features of the protocols between these two studies, in what interesting and plausible way can we say that the intervention that Richter et al. (2005) employed can be used to explain the behavioral data obtained in the Kogan et al. (2000) study, when the features of the behavioral protocols were so dissimilar? At best what we can say is that the intervention undertaken

[5] Oxytocin and Vasopressin are neuropeptides that have been implicated in a number of different social behaviors. In the studies that I cite, their role as neuromodulators in the formation of social recognition memory was under investigation. However, each has been implicated in a number of both social and non-social functions (See Bielsky et al. 2005; Ferguson et al. 2001).

[6] Are the two labs studying the same phenomenon? I think this is an open question, but one for which Bickle should have an answer.

[7] For the sake of ease of presentation of Bickle's position, I did not mention the anisomycin experiments in the context of talking about CREB knock-outs. However, the study (Kogan et al. 2000) also involved a set of experiments in which rats were trained using the same sub-protocol as was used in the experiments with CREB knock-outs. The only difference was that wild-type animals were used and injected with either anisomycin or saline (control group) 30 min prior to the initial interaction trial with the novel juvenile.

by Kogan, Frankland and Silva explains their data and the intervention that Richter, Wolf and Engelmann took explains their data. It is an open question how these two local claims would converge towards a more global claim. The very presence of this open question *undermines* Bickle's entitlement to move from such local claims to global reductive claims.

To emphasize just how open-ended this question is, I want to put forward some further evidential support. Experimentalists appear to be willing to attribute differences in data obtained in their lab and data obtained in another lab studying the "same effect" to differences between the experimental protocols used in the two labs. For example, Richter et al. (2005) point out differences between their experimental protocol and a protocol used by Ferguson et al. (2001), in order to account for differences in FOS-immunoreactivity in the brain following training in a social recognition task.[8] In their own words, differences in the data result from such "procedural" differences (Richter et al. 2005, p. 410). Another finding, in a study conducted by Dudek and Fields (2001), is that in at least one brain area, the hippocampus, "the intensity and frequency of synaptic input to CA1 hippocampal neurons are critically involved in determining the path by which second-messenger cascades are activated to activate" the molecule "mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK)" (RC122). Activation of MAPK/ERK, has been shown to be necessary for phosphorylation of CREB in some forms of hippocampal synaptic plasticity (for review see Impey et al. 1999). This is suggestive of the possibility that in learning experiments, differences in features of the sub-protocols used to produce learning may result in differences in the molecular pathways that become subsequently activated and/or which may be necessary for learning to occur. Interestingly, Richter and colleagues do not mention differences between their experimental protocol and the one used by Kogan et al. (2000) to study social recognition memory. From what I can gather, this is because of the similarity of the data arising out of the two laboratories. It would seem that just so long as the data are similar, differences in experimental protocols do not seem to be acknowledged. However, it is *simply incorrect* to assume that just because the data are similar between two sets of experiments, the internal cellular and molecular processes are also similar or identical. Given differences in the experimental protocols, this is an entirely open question.[9]

[8] C-Fos is one of a family of immediate early genes and increases in FOS expression in cells in a relevant brain area following a learning event are taken as indicative of the kinds of long-term changes at the molecular level that are requisite for the persistence of the learned information.

[9] The original Thor and Holloway (1982) study from which the behavioral protocol used by Kogan et al. (2000) was taken was modified. The initial interaction period with the juvenile conspecific was 5 min, followed by an "interexposure interval" (i.e., intertrial interval) of anywhere from 10 to 80 min. I want to mention here that the most likely reason for Kogan et al.' (2000) modification of the original experimental protocol has to do with then-current understandings of the timeframe of the molecular activity requisite to initiate memory consolidation in the hippocampus. To this end, Kogan and colleagues most likely picked the shortest time period possible for exposing animals to the novel conspecific on the first trial, and to keep the protocol across trials in which different pharmacological or genetic manipulations were used consistent. This corresponds to a claim I make in Sect. 5, that what seems to drive the structure of experimental protocols in cellular and molecular cognition is the goal of detecting molecular effects in the brain—clearly essential if one's goal is to detect such effects. However, oftentimes the choices made appear to be idiosyncratic rather than purely systematic.

If I am correct, Bickle has to provide answers to exactly these kinds of open questions if he is to make good on his inadequately supported thesis that local reductions are going to converge on global reductions. To date, he has been silent on these issues, but as I hope I have made clear, he has to take them seriously if he is going to have any chance of making a convincing case for his belief that the two forms of global reductions that he is advocating are anywhere on the horizon. Whatever philosophical position he is able to achieve by appeal to local claims made in the context of individual laboratories in cellular and molecular cognition, it is consistent with both non-reductive and locally-reductive forms of physicalism.

This brings me to a second problem for Bickle, with which I will end this section. I have made the case that even if we grant local reductive claims within the context of individual laboratories that study the molecular and cellular bases of social recognition memory in the mouse, we are not entitled to generalize those claims to data obtained in another laboratory. So, we can anticipate the nature of the difficulties that we will encounter in trying to extend such local reductive claims to instances of social recognition memory in humans.[10] To appreciate the extent of such difficulties, let's again consider the case.

Social recognition memory is commonly defined as the ability of an organism to distinguish familiar from unfamiliar individuals. Ferguson et al. (2002), the authors who Bickle cites, define it as:

> the ability to encode and recall very specific individual information [i.e., details of individual social status or kinship], [which] is required of almost all organisms living in complex social systems. In humans and other primates, individual recognition relies mostly on visual and auditory cues. Indeed, in the human brain, a specific visual association area, the right fusiform gyrus, appears to be critical for face recognition […]. In most other mammals, social information is encoded via olfactory or pheromonal signals, although auditory and visual signals may have important influences. (Ferguson et al. 2002, pp. 200–201)

Several interesting features of social recognition memory emerge on this definition. First, even if social recognition is a form of memory shared across species, there are *species-specific* differences with respect to the type of sensory information attended to and used for subsequent recognition of an individual. For example, *face recognition* appears to be a human-specific form of social recognition, relying on different structures than, say, *kinship recognition* in rodents, which predominantly involves olfactory structures. So, the systems-level processes (i.e., the areas of the brain that are involved) that underlie what are taken to be different instances of social recognition memory differ across species. Given such species-specific differences in the information attended

[10] I want to emphasize, however, that generalizing one local claim from the mouse to the human case is not merely problematic because of differences between the mouse brain and the human brain. It will be problematic given differences between how learning was produced in the local context, given the features of the experimental protocol used to produce it, and how learning is produced in laboratory compared to non-laboratory environments. For example, we might identify an event in the world that looks a lot like social recognition memory in the mouse, and yet the conditions under which this type of learning was formed may be wholly dissimilar from those we encounter in the laboratory. It is then an open question whether we are entitled to generalize from the one context to the other.

to, one may also anticipate differences in the temporal and spatial profiles of the underlying cellular and molecular processes that accompany different forms of social recognition memory across species.[11]

Note that the protocol that Kogan and colleagues use to produce social recognition memory in their mouse subjects has specific features pertaining to, for example, the duration of time of the presentation of the novel juvenile conspecific used in the experiment (=2 min), and the duration of time that elapses between the first and second presentations of this conspecific (=24 h). This is merely *one* experimental protocol that specifies a set of procedures for producing one form of long-term social recognition memory in the mouse. Yet, what is the likelihood given differences in those brain areas involved in the formation of different types of social recognition memory across species that the causal processes operative in the mouse with respect to this single experimental protocol, or even multiple experimental protocols will be identical to those occurring in non-human primates and human beings? It is very unlikely that there will be a simple story to tell. For, even if we could link up all the local reductive claims about social recognition memory in the mouse and achieve global reduction in the case of the mouse, it is not clear kind of light this sheds on the dynamics of long-term memory formation in non-human primates and human beings.

This is another issue that Bickle simply does not address. Yet, the potential lack of generalizability of results obtained in the Kogan et al. study to long-term social recognition in non-human primates and human beings suggests that the areas of science that do treat of differences at the systems-level across species cannot simply be set aside. They will continue to play a fundamental role in shedding light on the nature of the spatial and temporal dynamics of social recognition processes in higher-order species, and insights as to how to relate the findings of low-level neuroscience to such cases. And, with respect to goal of reducing mind to molecular pathways, philosophers of mind of a physicalist bent ultimately desire to understand the mind–brain relationship in the human case. While few would deny the relevance of studying non-human animals to shed light on the nature of this relationship, linking one operationalization of long-term memory formation in the mouse to molecular activity will not constitute a reduction of mind to molecular activity. So, reduction of the mind to brain is not achieved by Bickle's appeal to the social recognition memory case study, either.

I now want to turn to an alternative to Bickle's reductionist account that claims that the unity of neuroscience is achieved as different fields *integrate* their research by adding constraints on multilevel mechanistic explanations (Craver 2007). I aim to show that the challenges that arise for Bickle's account can also be shown to arise for this alternative.

4 Multiplicity and "mosaic unity"

A central aim of Carl Craver's recent book: *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience* (Oxford 2007) is to provide an alternative to

[11] Clearly much more could be said about the implications of these differences. For the sake of brevity, I am saying very little here.

reductionist accounts (e.g., Churchland 1982; Bickle 1998) of how unity is achieved in neuroscience. In my view, the alternative he offers is no better grounded than its reductionist counterparts. Indeed, I shall argue that it falls far short of descriptive adequacy. Craver's account is quite complicated, and how the pieces of the story fit together is not entirely clear. As a consequence, my assessment of it will proceed in stages. I will begin by providing an outline of his model of the unity of neuroscience and of my argument against it. Once I have done that, I shall look more closely at the elements of Craver's account of the nature of explanation in neuroscience that are crucial when assessing the plausibility of his model of the unity of neuroscience. I will then go on to demonstrate the implausibility of the model by appealing to the nature of experimentation in neuroscience as it is exhibited in neurobiological research on LTP.

Craver argues for a specific kind of non-reductive unity of science in which no single area of neuroscience is more fundamental than any other; rather, all areas jointly contribute to the goal of providing mechanistic explanations of phenomena. On this picture, neuroscience is comprised of "fields" of researchers sharing "common problems, techniques and vocabularies" (Craver 2007, 228fn). Some fields in neuroscience, such as biochemistry, anatomy and electrophysiology are specialized with respect to their objects of study and the techniques they use. Other fields, like cellular and molecular cognition, are "integrative", employing diverse techniques to study multiple kinds of objects. Every field in neuroscience yields important results relevant to providing mechanistic explanations of phenomena. The unity of neuroscience is achieved as each field contributes *part* of an explanation for a phenomenon (e.g., spatial memory), part of the story about what the nature of the mechanism productive of a phenomenon is. In Craver's words, observations and findings of individual fields "are used, like the tiles in a mosaic, to elaborate the mechanism sketch, showing piecemeal which components and properties are relevant to the mechanism and how those components are organized spatially, temporally, and actively within the mechanism" (Craver 2007, p. 267). By filling out part of the picture, the "findings from different fields" act as constraints to "shape the space of possible mechanisms" (Craver 2007, p. 228). Thus, they serve to lay down *constraints* on how the rest of the mechanistic explanation is going to turn out.

Now, the model that Craver puts forward—according to which all fields of neuroscience are ultimately headed towards the same explanatory goals—is interesting and highly optimistic. However, I aim to show that it is simply not borne out by the facts about experimental practice in neuroscience. My argument runs as follows. At first sight, Craver's key claim—that findings of an individual field serve to lay down constraints on what the (relevant) mechanism is—may seem plausible. However, it is in fact highly problematic. To begin with, we should note that it is only in a *derivative* sense that a 'field' can be thought of as a source of constraints on mechanistic explanation. For, the findings Craver refers to are findings of *individual researchers and laboratories*. This may seem a trivial point, but it is not. For it is only if we pay attention to *how* individual researchers arrive at their findings that we can make well-grounded claims about the manner in which experimental practice in neuroscience serves to place constraints on mechanistic explanations. And once we take a closer look—or so I shall argue—the initial plausibility of Craver's story evaporates.

Ultimately, then, it is not fields that place constraints on mechanistic explanations but individual researchers working within a field. Furthermore, they place such constraints on both an explanandum phenomenon and the mechanism that produces it—and they do so specifically in virtue of the *experimental protocols* that they select to study a phenomenon and its mechanism. Given this situation, the fundamental point is that the corresponding mechanistic explanation (however partial) that any one investigator can provide will be very specific to that local experimental context—i.e., the lab—in which the explanandum phenomenon was produced.

If I am correct about where and how constraints on mechanistic explanation actually originate, then Craver faces a serious problem. For, within any one "field" in neuroscience a multiplicity of experimental protocols are used to study what is taken to be (or at least labeled as) the "same phenomenon" (e.g., "social recognition memory"). Thus, Craver must provide an account of how results arrived at in different laboratories using different experimental protocols come to be integrated *within a field*, before he is in any position to suggest that the combined results of *that field* can be used to place constraints on a mechanistic explanation of a phenomenon. Barring this, his integrative model of the mosaic unity of neuroscience breaks down. And, what I will show is that it does just that: given the multiplicity of experimental protocols used to study the 'same' phenomenon that we encounter in fields like molecular and cellular cognition, the prospect of a 'mosaic unity' of neuroscience is a distant one indeed—little more, I should think, than a twinkle in the eye of an optimistic theorist.

With the general line of my argument sketched, I want to lay out some of the key aspects of Craver's account that are fundamental for assessing the plausibility of his "mosaic unity" conception of neuroscience as well as my own argument against it. First, I will consider his overall view of the nature of neuroscience and the kinds of explanations it seeks to give. Having done so, I shall provide a detailed assessment of the plausibility of his idea that the basic fields of neuroscience can be said to jointly contribute to an explanation for how a given phenomenon is produced. This I will do by looking at scientific experimentation as it is *actually carried out* in neuroscience. Specifically, I demonstrate that an investigation of the nature of the activity of neuroscientists and the kinds of goals they set for themselves, even in a relatively simple research area like LTP, provides the evidence requisite to demonstrate the failure of Craver's account. First, it is not clear that neuroscientists working within the same field are even talking about the same phenomenon. Second, even if we leave this problem aside, given the way that experimental science actually operates, there is no way the results of different laboratories can be conceived as fitting together in the manner Craver suggests.

Now, there are two primary components of relevance for understanding Craver's (2007) account of the nature of explanation in neuroscience: (1) what a *mechanism* is, and (2) what it means to provide a *mechanistic explanation* for a phenomenon. Craver defines a *mechanism* as "a set of entities and activities organized such that they exhibit the phenomenon to be explained" (Craver 2007, p. 5). On this definition, *entities* are objects that can be discriminated from other objects. Examples of entities include such things as organisms, nervous systems, neurons, and molecules. These entities engage in *activities*, which may be understood as identifiable behaviors that entities can exhibit. Examples of *activities* include: phosphorylating, firing and

learning. Entities and activities must be *organized* in a specific way so as to "exhibit" the *explanandum phenomenon*. Craver identifies three types of organization: (a) *active*, (b) *spatial* and (c) *temporal* (Craver 2007, pp. 136–138). *"Active* organization" is meant to capture the idea that the entities and activities in a mechanism must act and interact in such a way that the phenomenon is *actively produced* by these actions and interactions. According to Craver, such active organization depends upon the entities and activities of a mechanism having specific spatial and temporal organizations. *Spatial organization* in a mechanism includes such things as the locations and orientations of entities and activities. As Craver points out, if the entities and activities occur in the wrong location or if the entities are in the wrong orientation, the explanandum phenomenon will not be produced. *Temporal organization*, on the other hand, pertains to the fact that "the order, rate and duration of [the] successive component activities" (Craver 2007, p. 138) of the mechanism are crucial for the explanandum phenomenon to come about. The explanandum phenomenon, the last component in the definition of a mechanism, is understood generally as a detectable *property* (e.g., excitement, blueness, tiredness) or *behavior* (e.g., firing, shivering, yawning) exhibited by an entity (Craver 2007, p. 6) that is the target of explanation.

Given this definition of a mechanism, a *mechanistic explanation*, for Craver, is one that identifies that set of entities and activities and their organization that are productive of the explanandum phenomenon. According to Craver, mechanistic explanations of an explanandum phenomenon can vary with respect to the degree of detail that they include about the *actual* mechanism that produces the phenomenon. Whereas a *sketch* of a mechanism "characterizes some parts, activities or features of the mechanism's organization, but it leaves gaps", which are often filled in by "filler terms" (Craver 2007, p. 113), a complete mechanistic explanation, characterizes *all* of the relevant entities, activities and features of the mechanism's organization that result in the production of the explanandum phenomenon. And, in the time between a sketch of a mechanism for the explanandum phenomenon is introduced and a complete mechanistic explanation of it is provided, a number of different *mechanism sketches* may be offered as a means to capture *how possibly* the explanandum phenomenon is produced—each sketch specifying "what possibly" are the entities, activities and their organization that comprise the mechanism.

On Craver's model, then, the unity of neuroscience is achieved as the details of a mechanism sketch for an explanandum phenomenon (e.g., spatial memory) are filled in. He claims that the various fields of neuroscience, working together, fill in the details. They do this by means of two kinds of experiments (a) *intralevel* and (b) *interlevel* experiments. Intralevel experiments are carried out primarily in more specialized fields, such as biochemistry, anatomy and pharmacology, where investigators are interested in the nature of only some of the components or stages of a mechanism (e.g., cells, molecules and their adjoining activities). Other neuroscientists interested in a wider expanse of the mechanism that spans multiple components and stages, conduct *interlevel* experiments in which they intervene in the operation of one of the components in the mechanism and look at the effect of that intervention on the phenomenon of interest. In such experiments, investigators appeal to findings from both intralevel experiments and findings made by other investigators conducting similar kinds of interlevel experiments. Interlevel experiments, then "establish

which parts [those entities and activities identified and described by other sciences] are components in a mechanism [productive of that phenomenon that is of interest to them] and which are not" (Craver 2007, p. 144).

Craver has a neat story to tell, then, about how fields in neuroscience come to be integrated. For any given phenomenon, according to Craver, there is a space of possible mechanisms by virtue of which an explanandum phenomenon may be produced. The observations and findings obtained in different fields of neuroscience serve to narrow down the space of possible components that the mechanism productive of a phenomenon may contain, and hence, place constraints on what will qualify as an adequate mechanistic explanation of a phenomenon. The mosaic unity of neuroscience is achieved as the "findings from different fields" act as constraints to "shape the space of possible mechanisms" (Craver 2007, p. 228) productive of a phenomenon of interest. As such constraints are laid down, the findings from different fields are integrated. Such integration continues, according to Craver, until a time at which all the relevant details of the sketch have been filled in and the mechanistic explanation is complete. Then, perhaps the various fields of neuroscience go on to explain other phenomena in a similar fashion. The problem with this story, however, lies in its lack of plausibility and descriptive accuracy—a matter to which I now turn.

I want to concede, for the sake of argument, the aforementioned features of Craver's general account of mechanisms and mechanistic explanations. Furthermore, I also take his claim that experimental practice in neuroscience aims at providing mechanistic explanations of phenomena to be accurate. However, with these background assumptions in place, I find his model of how unity is achieved in neuroscience difficult to reconcile with the actual experimental practices of neuroscientists. My difficulty arises in the following way. Fields are comprised of individual investigators working within laboratories (e.g., "the Kandel lab", "the Silva lab", "the Morris lab") running experiments. It is the investigators within the laboratories that pick the experimental protocols that are used to both produce and study an explanandum phenomenon and identify the components of the mechanism by virtue of which *it* is produced. The first thing we can say, then, is that the spatial and temporal components of the mechanism of an explanandum phenomenon are likely to correspond to specific features of the experimental protocol used to produce that phenomenon and, experimental protocols can vary from one lab to the next (which I have already pointed out with respect to social recognition memory in Sect. 3). The second claim we can make, is that given Craver's own commitment to the idea that mechanisms have specific components and temporal and spatial organizations, and given that such features of mechanisms are sensitive to what set that mechanism into motion, i.e., the events that preceded it, then it is entirely possible, indeed quite likely, that different experimental protocols will yield different mechanisms for potentially different phenomena. In sum, the problem that Craver faces is how—and indeed whether—he can make good on the claim that the unity of neuroscience can be achieved in the manner he suggests, given the wide range of differences across experimental protocols that we encounter in neuroscience.

At this point, I will turn to LTP research as a way of demonstrating just how serious the problem is for Craver's account. So serious, in fact, that as long as we

can assume that LTP research is representative of ongoing neuroscientific research, Craver's 'mosaic unity' account is little more than an optimistic promissory note. In the next section, I shall, among other things, go on to explain what it is that *gives rise* to the multiplicity of experimental protocols that we find in the neurosciences, and thereby explain why it is highly unlikely that the situation will change. Here, I want to focus on only a small subset of the experiments in the learning and memory research field, specifically, those interlevel experiments in which LTP is produced in area CA1 of the hippocampus in vitro (i.e., in slices). To restrict the class of experiments in this research area even further, I want to focus on LTP experiments geared to study the role of one entity (one molecule), and its activity, which has been studied in the context of LTP in area CA1 of the hippocampus in vitro: the mitogen-activated protein kinase (MAPK1/2)/extracellular regulated signaling kinase (ERK1/2).[12] The only aspects of such LTP experiments with which I will be concerned are differences in the features of the stimulation protocols used across these experiments to produce LTP.[13]

Research studies on the role of MAPK/ERK in LTP are representative of the kinds of differences in experimental protocols that we find in different studies on LTP throughout its ~35 year history. The first two groundbreaking studies identifying a role for the ERK cascade in area CA1 of the hippocampus in vitro were undertaken by English and Sweatt (1996, 1997). For their in vitro electrophysiology experiments, they used a high-frequency stimulation paradigm (HFS) consisting of "three sets of tetani, each set spaced 10 min apart", with "each set consist[ing] of two trains of 100 Hz stimulation, each lasting 1 s, with an intertrain interval of 20 s, at a stimulus intensity that generated 75% of the maximal pEPSP"[14] (English and Sweatt 1996, p. 24329). In another set of experiments conducted to determine the role of ERK in LTP, Tom O'Dell and colleagues, induced LTP by pairing "100 EPSPs evoked at 2 Hz with a constant injection of current through the recording electrode to depolarize the postsynaptic membrane potential to between 0 and +20 mV" in a whole cell recording (Watabe et al. 2000, p. 5925). In another set of experiments, Dudek and Fields (2001), also conducting in vitro LTP experiments in hippocampal slices, used *theta-burst stimulation* (TBS) to induce LTP in area CA1. This stimulation "consisted of 10 bursts of four pulses at 100 Hz, delivered at 5 Hz, given a total of three times, with a

[12] I want to point out that LTP research is one of two cases that Craver uses in order to make his argument for mosaic unity. My focus on a restricted area of LTP research in no way weakens my argument against his account. Rather, it strengthens it. For, if it can be shown within this very restricted area of LTP research that there is a multiplicity of experimental protocols and a lack of clarity about how the results arising from these different protocols are to be fit together in the 'mosaic unity' way, then broadening the scope of our analysis will likely make the overall task more difficult.

I also want to mention that I select ERK primarily because I spent several years studying its role in one form of LTP, high-frequency stimulation induced LTP, at the CA3–CA1 commissural fiber pathway in the rat hippocampus in vivo (Sullivan 2003). The stimulation protocol that I employed consisted of 4 trains of 100 pulses delivered over 420 s, at a stimulus intensity the generated 60% of the maximal excitatory post-synaptic potential (pEPSP).

[13] In other words, there are other features with respect to which the experimental protocols contained in these different studies differ.

[14] The acronym "pEPSP" stands for "excitatory post-synaptic potential" (IEPSP) of the population (p) of cells from which an electrophysiological recording was taken.

15–30 s interval" (Dudek and Fields 2001, p. RC122). In the lab of Emmanuel Landau, in a 2001 in vitro research study, "LTP was induced in area CA1 by applying 1 μM isoproterenol (ISO)[15] in the bath for 10 min, followed by theta-pulse stimulation (TPS; 150 pulses at 10 Hz) of the Schaffer collaterals (*stratum radiatum* [i.e., the dendritic layer] in area CA3)" (Giovannini et al. 2001, p. 7054). Finally, in the lab of David Sweatt, LTP was induced by means of three different stimulation protocols including, HFS, which "consisted of two trains of 1-s, 100-Hz stimulation with an intertrain interval of 20 s = 200 pulses total", theta-frequency stimulation (TFS), which "consisted of 30 s of stimulation delivered at 5 Hz (200 ms between pulses) = 150 total pulses and theta" and "TBS", which "consisted of three trains of stimuli delivered at 20-s intervals, each train composed of 10 stimulus bursts delivered at 5 Hz, with each burst consisting of four pulses at 100 Hz = 120 total pulses" (Selcher et al. 2003, p. 37). While there are other similar kinds of studies that I could consider here, I think I have made the point that I wish to make, namely, that there is a wide variety of different stimulation protocols that may be used to induce LTP and to tease out the role of a given molecule, such as ERK, in LTP. In some cases, for example, one protocol will be used by multiple labs. TBS-induced LTP is one such example. In other cases, however, there will be differences in the stimulation protocols employed—some subtle, some much less so.

If we consider early as well as current LTP research it is apparent that at some point after the original discovery of LTP (Bliss and Lømo 1973), the field *exploded* in a number of different investigative directions. It became swamped with new investigators, and for each individual lab that began to work on the mechanisms of LTP, there were differing opinions about what was the best stimulation protocol to induce it (e.g., HFS or TBS or paired-pulse stimulation (PPS), or spike-timing) and what features a stimulation protocol ought to have (e.g., inter-stimulus and inter-train intervals, pulse number, duration and frequency, train number). There was even controversy as to how long the potentiated effect had to last in order for it to qualify as a viable instance of LTP.

On Craver's account it would seem that with respect to the case of LTP, we should be able to see how different protocols can each come to contribute a component or somehow constrain the multi-level mechanistic explanation of this single phenomenon. However, nothing of the sort appears to happen in this case. Rather, an investigation of the multitude of experimental protocols in the LTP field alone suggests that it has predominantly been and actually still remains an unconstrained free-for-all. The interesting question then, is, if this is true, how will research, even in so restricted an area as that of LTP ultimately integrate into a mosaic unity? Because clearly, given the differences in the experimental protocols used to produce it across different laboratories, there will be different temporal and spatial constraints on the activities and entities involved in each case, and these will not readily translate from the context of one experimental protocol to another. Which experimental protocol and which constraints ought we to prefer? If the answer is that we will know which to prefer given results obtained from those protocols used to investigate (some form of) learning at

[15] Isoproteronal is a synthetic catecholamine that can be used to stimulate activation of beta adrenergic receptors, as it was in these experiments.

the level of the organism, well, I have already shown in my refutation of Bickle, that the multiplicity of experimental protocols problem also arises there. So, Craver must first provide an account of how results obtained from different laboratories studying phenomena grouped under a general heading (e.g., "spatial learning") fit together, before he can say how these results in combination can be used to place constraints on components operative at lower levels of the mechanism. Yet, Craver's 'mosaic unity' model contains no suggestion about how to tackle these complex issues—and yet the plausibility of his account depends precisely on there being some way of doing just that. Clearly, the story will not be a simple one—and it is unlikely that such an account will be forthcoming anytime in the near future.

To recapitulate, given differences in plasticity-producing protocols across labs, it is not likely that integration in neuroscience, or "mosaic unity" can be achieved in the way that Craver claims, because spatial, temporal, and componency constraints on a mechanism for LTP arising out of one lab will not necessarily be informative with respect to spatial, temporal and componency constraints on a mechanism for LTP arising out of another lab. The only thing that the 'mosaic unity' model seems to capture is the fact that research undertaken on one form of LTP in one lab may yield testable *hypotheses* as to what entities and activities will be operative in another form of LTP in another lab. However, this does not mean that the ways in which the entities are operative will be identical—rather, it is likely that there will be differences. For example, parametric studies undertaken to determine the role of ERK in LTP in the hippocampus (e.g., Dudek and Fields 2001; Selcher et al. 2003) have revealed differential activation of the ERK cascade in response to differences in stimulation protocols/stimulus parameters. Here it is worth noting that the fact that neurobiologists are privy to such differences, suggests that they themselves do *not* take LTP to be a unitary phenomenon, *nor* do they take there to be only one mechanistic pathway to LTP. To achieve the kind of mosaic unity that Craver desires or thinks is in the offing, either neurobiologists would have to be aiming explicitly for consistency across experimental protocols with respect to how LTP is produced or there would have to be something in the nature of experimental practice that resulted in such consistency. But this is not a current goal in LTP research, nor is there any sign of its coming about by some other means. Rather, in the case of LTP research, methodological *plurality* rather than unity is the norm. Whatever LTP is—defined generally, as a change in synaptic strength—there are many different instances of it and each different instance has potentially different spatial and temporal constraints operating over it. Given Craver's 'mosaic unity' conception, we would have to be able to see the field of LTP research as comprised of a group of researchers and laboratories collectively—albeit perhaps not *self-consciously*—engaged in the project of providing an explanation of *the* mechanism underlying *the* phenomenon of LTP, or something near enough. This is clearly *not* what is going on in LTP research—even if we were to allow ourselves some liberties in the name of providing a 'rational reconstruction' of the activities in the field. The 'mosaic unity' conception, in sum, falls far short of descriptive adequacy.

I want to conclude this section by drawing the reader's attention to the fact that Craver's mosaic unity conception also runs into trouble when we consider different labs studying phenomena labeled under the same name (e.g., "social recognition

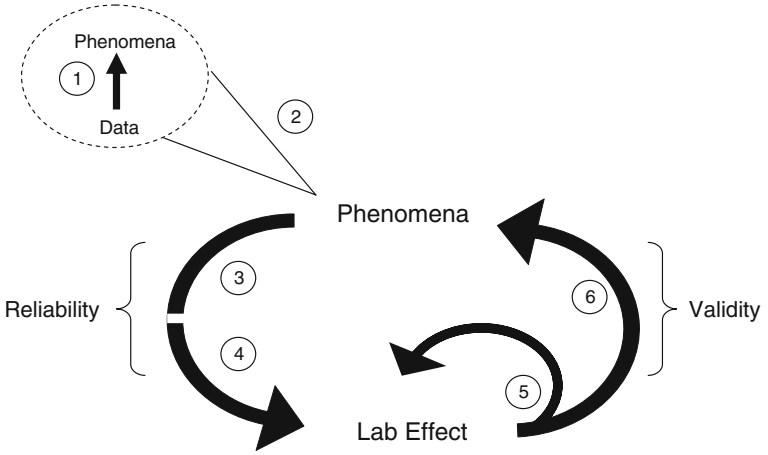

Fig. 1 The Experimental Process. (1) An investigator begins with an empirical question about a phenomenon of interest. This question is then redirected at an effect to be produced in the laboratory. Thus initiates the (3) design and (4) implementation stages of data production. If the data production process is reliable, it results in the discrimination of one hypotheses from a set of competing hypotheses about the effect produced in the laboratory. This initiates the stage of data interpretation, in which the discriminated hypothesis is treated as a claim and is taken as true of (5) the effect produced in the laboratory and (6) the original phenomenon of interest in the world. If the claim was produced by a reliable data production process and it is true of the effect produced in the lab it is valid (internal validity). If it was produced by a reliable data production process and it is true of the effect in the world, it is valid (external validity)

memory") in *different* species (e.g., mouse, human). For, even if an investigator working within a given laboratory arrives at some conclusion about a component in the mechanism of social recognition memory in *the rat*, there is no guarantee, given differences across species, that this conclusion will shed light on the mechanism of social recognition memory in *humans*. So, here, we have a special case of the kind of problem that the multiplicity of experimental protocols raises for the mosaic unity view, just as it did for Bickle's reductive position. Specifically, we can anticipate that the temporal and spatial constraints that are operative within the confines of one experimental protocol to study social recognition in *one species* will not be identical to those operative in another laboratory employing a different protocol to study it in *another*, *different species*. In some sense, we might say, that the multiplicity of experimental protocols arises with respect to a phenomenon like "social recognition memory" in part because it manifests itself differently in different species and, so, will involve different overall mechanisms in different species. Yet, this necessarily leaves open the possibility that the constraints (e.g., spatial and temporal) that are discovered to be operative in one instance of social recognition memory in one species in one lab will not be the same as those operative with respect to a different species in another lab. Thus, it is extremely hard to see how integration of two fields working on the 'same' phenomenon but in different species could be achieved in the way that Craver suggests.

Having demonstrated that the multiplicity of experimental protocols in neurobiology poses a serious problem for the positions of both Bickle and Craver, I now want to turn to the issue of the origin of this multiplicity in contemporary neuroscience.

5 The experimental process and the origin of multiplicity

In arguing for their positions, Bickle and Craver crucially appeal to case studies in the contemporary neurobiology of learning and memory. My claim has been that their arguments fail because both overlook the fact that there is a multiplicity of experimental protocols in neurobiology. I shall now turn to the question of *why* we encounter such a multiplicity of protocols. Doing so serves to strengthen my criticism of the two—because once we have seen why the multiplicity arises, it will also become clear that it is unlikely to go away anytime soon, if ever. In order to answer the above question, one needs to have an understanding of the nature of experimentation in neurobiology (or, at least, of some of its fundamental features). For that reason, I will begin by providing such an account, before using it to provide an explanation of how multiplicity arises.[16] Before doing so, however, it seems appropriate to issue a caveat. To date, scant philosophical attention has been paid to the nature of experimentation in neuroscience.[17] My own thinking on experimentation in neurobiology is informed by four different sources. First, I rely heavily on my own experience working at the bench in a neurobiological laboratory for insight into the structure of neurobiological experiments. In addition, I rely on insights about knowledge-producing processes arising from (a) general epistemology (sp., Goldman 1988), (b) the epistemology of experiment (e.g. Bogen and Woodward 1988; Cartwright 1991, 1999; Franklin 1986, 1999; Hacking 1983, 1992; Mayo 1996, 2000; Wimsatt 1981; Woodward 1989, 2000), and (c) other areas of science (e.g., Cook and Campbell 1979; Messick 1989). I want to stress that I do not seek to offer the final word on the nature of experimentation in neurobiology, but only what seems to me to be a worthwhile initial contribution.[18] I take the multiplicity of experimental protocols to be a fact—a fact that has important philosophical consequences, and my goal is to get clear on what it is about experimentation in neurobiology that has led to this fact. By focusing on it, an interesting area of philosophical research opens up that has hitherto been unexplored in philosophy of neuroscience, and I aim merely to set the ball rolling.

With respect to neurobiological experiments of the kind considered in this paper, what I will refer to as the "experimental process" is set in motion by an empirical question about a phenomenon of interest.[19] I use the phrase "phenomenon of interest" only to very loosely capture the idea that *something* prompts an investigator to conduct an experiment—*something* of interest to him/her. How he/she comes to identify

[16] The foundations of this account of experiment and the claims I make about reliability and validity are contained in my dissertation (Sullivan 2007).

[17] However, some attention has been paid to specific epistemological problems that arise with respect to neuropsychological and cognitive neuroscientific experiments. See, for example, Bechtel and Stufflebeam (2001), Bogen (2002), Feest (2003).

[18] Bickle and Craver have also made significant contributions to the development of our understanding of some fundamental features of neurobiological experiments. However, as I mentioned at the outset of this paper, because their discussions of experimentation have been primarily in the service of their accounts of explanation, they have overlooked crucial features of experimentation and experimental practice in neuroscience.

[19] I take experimentation to be *a process* having certain discrete stages, as depicted in Fig. 1. Here, I am taking a cue from Woodward (2000).

or detect that something is an open question, but one way the story might go, if we consider the case of a learning phenomenon, is that an investigator notices before and after changes in the behavior of an organism that serve as data points for its detection (Bogen and Woodward 1988) (Step 1 in Fig. 1).[20] For example, I know that a student has learned the distinction between normative and descriptive ethics if I can point to a time at which she did not appear to have such knowledge, and the time of the exam, when she was able to answer the relevant question correctly. Experiments on learning in neurobiology may have their starting point in such changes in behaviors as exhibited by organisms "in the world", or they may begin with a phenomenon that has been detected by means of data points in the controlled environment of the laboratory (Step 2 in Fig. 1).[21] There is most likely a complicated story that could be told as to how an investigator arrived at a particular empirical question. Some of the details of Bickle's and Craver's accounts will certainly be relevant here. However, I am interested in what happens after such questions are posed. Essentially, using Woodward's (2000) terminology, this is when the process of *data production* begins (Step 3 in Fig. 1).

On my account, data production may be divided into two discrete stages: (1) *design* and (2) *implementation*. The design stage, in basic terms, involves the development of an experimental design and protocol, the general features of which, I described in Sect. 2. It typically proceeds in discrete stages: questions are posed and suggestions about how to address them are provided; projections are then made about potential problems that might be encountered in the course of implementing the design and tentative solutions to these problems are offered; finally, the combined considerations are worked into the design and protocol. Essentially, at this stage the empirical question of interest is directed at some effect to be produced in the lab.

The implementation stage of data production (Step 4 in Fig. 1) begins at some point after an experimental design and protocol has been completed. It involves individual instantiations of the experimental design by means of the systematic following of the experimental protocol using the equipment, materials, and techniques assembled during the design stage. At this point, an investigator takes an individual subject or a group of subjects, and runs it through the steps of the protocol, following those steps

[20] I am here extending a claim made by Bogen and Woodward (1988). They introduce the distinction between "data" and "phenomena" and commit themselves to the idea that phenomena are not observable, but only detectable by means of reference to data (See Bogen and Woodward 1988, p. 305), which are observable. I am here trying to reconcile this claim that they make with the fact that investigators begin experiments with questions about something that *is* detectable. But I concede their point, that whatever this thing is, it is only detectable derivatively, by means of reference to "data points". Sometimes data points are obtained in the controlled environment of the laboratory and sometimes they are obtained in less controlled environments. So, while I think that Bogen and Woodward (1988) primarily had in mind data obtained in laboratories, in order to detect a phenomenon like learning, out in the world, we can similarly make reference to "before and after" data points.

[21] By drawing this distinction between "within lab" effects and worldly phenomena, I am just trying to capture two potential sources of the origin of a phenomenon of interest. It is a rational reconstruction on my part to claim that neurobiological investigators actually begin with clear concepts of the phenomena in which they are interested. Oftentimes, I think they simply begin with standard operationalizations of a given form of learning that have been used in other laboratories (as Bickle described), but then they tweak those operationalizations so as to best suit their empirical interests. Feest (2003) provides a nice treatment of the complex issues that arise with respect to concept formation in the context of experimentation in experimental and neuropsychology.

as precisely as possible. The immediate output of each individual implementation of the design is an individual data point or set of data points.

Once enough data points for each type of experimental manipulation have been collected, the data points are combined and each complete data set is analyzed statistically. The statistically analyzed data is then used to discriminate one hypotheses from the set of competing hypotheses about the role of the molecule in the effect of interest produced in the laboratory. This initiates what I refer to as the first phase of *data interpretation* (Step 5 in Fig. 1). In this phase, the hypothesis discriminated by the data is taken as true with respect to the effect produced in the laboratory. That same claim may then be extended back to the original effect of interest in the world that prompted the empirical question about the phenomenon of interest in the first place and taken as true of it (Step 6 in Fig. 1).

This is one basic framework for thinking about experimentation in neurobiology. One of its advantages is that it can be used to illuminate normative constraints on the experimental process that contribute to the development of a multiplicity of experimental protocols. Two fundamental constraints are (1) *reliability* and (2) *external validity*.

Reliability has traditionally been characterized by philosophers of science as a feature that an experiment has when it can be used to produce evidence for knowledge claims (e.g., Bogen and Woodward 1988; Cartwright 1991; Franklin 1986, 1999; Mayo 1991, 1996, 2000; Woodward 1989, 2000). Yet, in both philosophical and scientific usage, the term "reliable" is often ascribed to different aspects of experiments including data, instruments, techniques, methods (and even) claims about data. What differentiates my understanding of reliability from alternative understandings is that I take it to be a feature that can be ascribed to a data production process as a whole. Briefly, on my account, a complete data production process is *reliable* if and only if it results in statistically analyzed data that can be used to discriminate one hypothesis from a set of competing hypotheses about an effect produced in the laboratory. A definition of what reliability is and how it functions in neurobiological experiments should accompany any account of the nature of experimentation. I am offering only one such possibility here.

A second desirable feature of the experimental process, which differs from reliability, is *validity*. Traditional scientific accounts frequently make use of a general notion of validity, which is taken to be a feature that is ascribed to experiments or tests (e.g., Campbell and Stanley 1963; Messick 1989). On such accounts, an experiment (or test) is taken to be *valid* if it supports the intended conclusion that is drawn from its results (e.g. Campbell and Stanley 1963; Messick 1989). However, if this is regarded as a *sufficient* condition, then validity appears to be tantamount to reliability, traditionally construed. So, in order to develop a notion of validity that appreciates how it can operate as a constraint on the experimental process in a way different than reliability, I want to appeal to a distinction drawn in the scientific (e.g., Cook and Campbell 1979) and philosophical (e.g., Latour 1988; Guala 2003, 2005) literature between *external* and *internal* validity.[22]

[22] Validity is a complex notion and many different types of validity have been identified in the scientific literature including: *construct validity* (e.g., Cronbach and Meehl 1955; Shadish et al. 2002), *predictive*

An interest in *external* validity emerges in both the philosophical and scientific literature (e.g. Latour 1988; Cook and Campbell 1979). Recently, Francesco Guala (2003, 2005) has sought to understand the problems that arise when an investigator desires to have the conclusion of an experimental result apply not only to the experimental situation in the laboratory but also to "circumstances of interest" (Guala 2003, p. 1198) in the outside world. Guala (2003, 2005) uses the distinction between internal and external validity (Cook and Campbell 1979) to shed light on these problems. On Guala's account, the *internal validity* of an experimental result is established when that result captures a causal relationship that is operative in the context of the laboratory. That experimental result is *externally* valid, when it captures a causal relationship that is operative in "a set of circumstances of interest", outside the laboratory.

On my understanding of the internal/external validity distinction with respect to data interpretation, I take validity itself to be a feature of interpretive claims rather than of experimental results. Experimental results, on my account, are statistically analyzed sets of data and interpretive claims are what arise when a hypothesis that has been discriminated from a set of competing hypotheses by a set of data is applied to/taken as true of (a) an effect produced in the laboratory, (b) the original phenomenon of interest outside the laboratory. The primary difference between my account and Guala's is that I restrict the application of "validity" to interpretive claims. An interpretive claim, on my account, may be "internally" or "externally" valid. A claim about an effect produced in a laboratory is *internally* valid, if and only if that claim is true about the effect produced in the laboratory. A claim about a phenomenon of interest outside the laboratory is externally valid, if and only if that claim is true about that phenomenon.

On my account, reliability and validity both operate as normative constraints on the experimental process. Yet, they give rise to conflicting prescriptions. Reliability prescribes simplifying measures in the context of the laboratory in order to narrow down a set of competing hypotheses with respect to the effect produced in the laboratory. As it operates to constrain the process of data production, it thereby inevitably restricts the extension of interpretive claims to the laboratory. Validity, on the other hand, pulls in the opposite direction in so far as it prescribes that an investigator build into an experimental design those dimensions of complexity that accompany the phenomenon of interest in the world about which an investigator would like to say something. Yet, to follow the normative prescriptions of validity would inevitably lead to a decrease in the simplicity of the effect produced in the laboratory and an expansion of the set of competing hypotheses that pertain to that effect—in essence, a decrease in reliability. And yet, without reliability, nothing is gained—for if control is lost in the laboratory, nothing true can even be said about the effect produced there—internal validity is lost as well.

Footnote 22 continued
validity (e.g., Cook and Campbell 1979) and *ecological validity* (e.g., Brunswick 1943; Bronfenbrenner 1979). Getting clear on the various types of validity that there are and how they are related is a project in itself. However, in order to make the point I wish to make in this section, it is sufficient to focus exclusively on external validity.

So, how do these normative constraints elicit a multiplicity of experimental protocols with respect to learning/memory consolidation and plasticity experiments in cellular and molecular cognition? They do so when taken in conjunction with the investigative aims and interests of the individual researcher or laboratory. Sometimes an investigator studying memory consolidation will select the experimental protocol that is most likely to produce a robust molecular effect, even if the spatial and temporal patterns of stimulus presentation do not resemble patterns that occur in the world outside the laboratory, or patterns which are used by investigators in other laboratories. Here, idiosyncratic views about the reliability of an experimental design and protocol will be relevant. Another investigator may wish to pick an experimental protocol that contains temporal and spatial patterns of stimulus presentation that *do* resemble the patterns that an organism encounters in the external world in some common learning situation. Some investigators may expose their animals to enriched environments as part of an experimental protocol to make the animals more closely resemble animals in the external world or to make them more closely resemble the human organism. In these latter cases, concerns about validity are clearly playing a decisive role.

With respect to plasticity experiments, the story will be similar. It may be that an investigator wants to use an experimental protocol that he/she believes will produce a robust and detectable form of LTP that will be accompanied by a robust and detectable molecular effect. Concerns about reliability would thus be operative here. Alternatively, an investigator may select to use an experimental protocol that includes a stimulation protocol having parameters that are physiologically plausible. Concerns about validity would in turn be operative here.

So, on my account of the experimental process, the diversity that we encounter with respect to experimental protocols in neurobiology is, at least in part, a byproduct of the differential activity of the constraints of reliability and validity—as determined by investigative aims and interests of the individual researcher or laboratory. Furthermore, such diversity is in no way discouraged, nor do I know of any reason why it ought to be. Yet, as long as such diversity persists the evidence requisite to ground global pictures of the kind propounded by Bickle and Craver will not be forthcoming.

6 The next frontier and the space of possible positions

I have shown that the multiplicity of experimental protocols in contemporary neuroscience, at least with respect to cellular and molecular cognition, poses a serious problem for two global positions that are at opposite ends of the philosophical spectrum: reductive physicalism and what might best be referred to as 'integrative unitarianism'. If this problem proves to be insurmountable, we would thereby have a compelling argument for the only remaining kind of position on the philosophical spectrum: non-*reductive pluralism* (See for example Mitchell 2003; Mitchell and Dietrich 2006). Although I am tempted towards the general position of non-reductive pluralism, it has not been my aim to provide an argument for it here. Indeed, I am skeptical of the value of advancing global accounts at the present point in time precisely for the reasons I have specified: there are a number of fundamental and thorny issues that to date have not received adequate philosophical attention; issues that philosophers of science must get

clear on before there is going to be any real prospect of providing adequate evidence for a global position in the philosophy of neuroscience. Thus, I think this is where current efforts in philosophy of neuroscience are best directed. Not only for the sake of seeking support for global positions, but also because the issues in question are of profound importance in their own right.

Acknowledgements The author owes a debt of gratitude to Carsten Hansen for extensive discussion of the issues dealt with in this paper and for invaluable comments on earlier versions. The author would also like to thank Ken Aizawa for very helpful comments. Thanks also to the attendees of the *Neuroscience and the Mind-Brain Relationship Workshop* including Carl Gillett, Tom Polger, Bob Richardson, Larry Shapiro and Ken, for helpful comments during the discussion period.

References

Aizawa, K. (2007). The biochemistry of memory consolidation: A model system for the philosophy of mind. *Synthese, 155*, 65–98.

Bechtel, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Lawrence Erlbaum.

Bechtel, W., & Stufflebeam, R. S. (2001). Epistemic issues in procuring evidence about the brain: The importance of research instruments and techniques. In W. Bechtel, P. Mandik, J. Mundale, & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 55–81). Oxford: Blackwell.

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

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

Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular and molecular neuroscience. *Synthese, 151*, 411–434.

Bickle, J. (2007). Ruthless reductionism and social cognition. *Journal of Physiology-Paris, 101*, 230–235.

Bielsky, I., Hu, S., Ren, X., Terwilliger, E., & Young, L. (2005). The V1a vasopressin receptor is necessary and sufficient for normal social recognition: A gene replacement study. *Neuron, 47*, 503–513.

Bliss, T., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology, 232*(2), 331–356.

Bogen, J. (2002). Epistemological custard pies from functional brain imaging. *Philosophy of Science, 69*, S59–S71.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review, 97*, 303–352.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge: Harvard University Press.

Brunswick, E. (1943). Organismic achievement and environmental probability. *The Psychological Review, 50*, 255–272.

Campbell, D. D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Cartwright, N. (1991). Replicability, reproducibility and robustness: Comments on collins. *History of Political Economy, 23*(1), 143–155.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge: Cambridge University Press.

Churchland, P. S. (1982). Mind brain reduction—new light from the philosophy of science. *Neuroscience, 7*(5), 1041–1047.

Cook, T. D., & Campbell, D. D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.

Craver, C. (2001). Role functions, mechanisms and hierarchy. *Philosophy of Science, 68*, 31–55.

Craver, C. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science Supplement, 69*, S83–S97.

Craver, C. (2003). The making of a memory mechanism. *Journal of the History of Biology, 36*, 153–195.

Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Craver, C., & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. K. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences*. Pittsburgh: University of Pittsburgh Press.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Dudek, S., & Fields, R. D. (2001). Mitogen-activated protein kinase/Extracellular signal-regulated kinase activation in somatodendritic compartments: Roles of action potentials, frequency, and mode of calcium entry. *The Journal of Neuroscience, 21*, RC122.

English, J. D., & Sweatt, J. D. (1996). Activation of p42 mitogen-activated protein kinase in hippocampal long-term potentiation. *Journal of Biological Chemistry, 271*(40), 24329–24332.

English, J. D., & Sweatt, J. D. (1997). A requirement for the mitogen-activated protein kinase cascade in hippocampal long term potentiation. *Journal of Biological Chemistry, 272*(31), 19103–19106.

Feest, U. (2003). *Operationism, experimentation, and concept formation*. Dissertation, University of Pittsburgh.

Ferguson, J., Aldag, M., Insel, T., & Young, L. (2001). Oxytocin in the medial amygdala is essential for social recognition in the mouse. *Journal of Neuroscience, 21*(20), 8278–8285.

Ferguson, J. N., Young, L. J., & Insel, T. R. (2002). The neuroendocrine basis of social recognition. *Frontiers in Neuroendocrinology, 23*, 200–224.

Franklin, A. (1986). *The neglect of experiment*. New York: Cambridge University Press.

Franklin, A. (1999). *Can that be right?: Essays on experiment, evidence, and science*. Boston: Kluwer.

Giovannini, M., Blitzer, R., Wong, T., Asoma, K., Tsokas, P., Morrison, J., Iyengar, R., & Landau, E. (2001). Mitogen-activated protein kinase regulates early phosphorylation and delayed expression of Ca^{2+} Calmodulin-dependent protein kinase II in long-term potentiation. *Journal of Neuroscience, 21*(18), 7053–7062.

Goldman, A. (1988). *Epistemology and cognition*. Cambridge: Harvard University Press.

Guala, F. (2003). Experimental localism and external validity. *Philosophy of Science Supplement, 70*, 1195–1205.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge: Cambridge University Press.

Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.

Hacking, I. (1992). The self-vindication of the laboratory sciences. In A. Pickering (Ed.), *Science as practice and culture*. Chicago: University of Chicago Press.

Hummler, E., Cole, T., Blendy, J., Ganss, R., Aguzzi, A., Schmid, W., Beerman, F., & Schutz, G. (1994). Targeted mutation of the CREB gene: Compensation within the CREB/ATF family of transcription factors. In *Proceedings of the National Academy of Sciences USA* (Vol. 91, pp. 5647–5651).

Impey, S., Obrietan, K., & Storm, D. R. (1999). Making new connections: Role of ERK/MAP kinase signaling in neuronal plasticity. *Neuron, 23*, 11–14.

Kogan, J. H., Frankland P. W., Blendy, J. A., Coblentz, J., Marowitz, Z., Shüz, G., & Silva, A. J. (1997). Spaced training induces normal long-term memory in CREB mutant mice. *Current Biology, 7*(1), 1–11.

Kogan, J. H., Frankland, P. W., & Silva, A. J. (2000). Long-term memory underlying hippocampus-dependent social recognition in mice. *Hippocampus, 10*(1), 47–56.

Latour, B. (1988). *The pateurization of France*. Cambridge: Harvard University Press.

Looren de Jong, H. (2006). Explicating pluralism: Where the mind to molecular pathway gets off-track. *Synthese, 151*(3), 435–443.

Mayo, D. (1991). Novel evidence and severe tests. *Philosophy of Science, 58*, 523–552.

Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.

Mayo, D. (2000). Experimental practice and an error statistical account of evidence. *Philosophy of Science, 67*(3), S193–S207.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Mitchell, S. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Mitchell, S., & Dietrich, M. (2006). Integration without unification: An argument for pluralism in the biological sciences. *American Naturalist, 168*, S73–S79.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace and World.

Richter, K., Wolf, G., & Engelmann, M. (2005). Social recognition memory requires two stages of protein synthesis in mice. *Learning and Memory, 12*, 407–413.

Schouten, M., & Looren de Jong, H. (2005). Ruthless reductionism. *Philosophical Psychology, 18*(4), 473–486.

Selcher, J. C., Weeber, E. J., Christian, J., Nekrasova, T., Landreth, G., & Sweatt, J. D. (2003). A Role for ERK MAP Kinase in Physiological Temporal Integration in Hippocampal Area CA1. *Learning and Memory, 10*, 26–39.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

Sullivan, J., Memory consolidation, multiple realization and modest reductions. *Philosophy of Science Supplement* (in press).

Sullivan, J. (2003). *Regulation of extracellular signal-regulated kinase during long-term potentiation in area CA1 of the rat hippocampus in vivo*. Master's Thesis, University of Pittsburgh.

Sullivan, J. (2007). *Reliability and validity of experiment in the neurobiology of learning and memory*. Dissertation, University of Pittsburgh.

Taylor, C. (1964). *Explanation of behavior*. Amherst, NY: Prometheus.

Thor, D. H., & Holloway, W. R. (1982). Social memory of the male laboratory rat. *Journal of Comparative and Physiological Psychology, 96*(6), 1000–1006.

Watabe, A. M., Zaki, P. A., & O'Dell, T. J. (2000). Coactivation of β-adrenergic and cholinergic receptors enhances the induction of long-term potentiation and synergistically activates mitogen-activated protein kinase in the hippocampal CA1 region. *The Journal of Neuroscience, 20*(16), 5924–5931.

Wimsatt, W. (1981). Robustness, reliability, and overdetermination. In M. B. Brewer & B. E. Collins (Eds.), *Scientific inquiry and the social sciences*. San Francisco: Jossey-Bass Inc.

Woodward, J. (1989). Data and phenomena. *Synthese, 79*, 393–472.

Woodward, J. (2000). Data, phenomena and reliability. *Philosophy of Science, 67*(3), S163–S179.