



# Psychoneural reduction: a perspective from neural circuits

David Parker<sup>1</sup>

Received: 4 February 2019 / Accepted: 30 July 2019 / Published online: 6 August 2019  
© The Author(s) 2019

## Abstract

Psychoneural reduction has been debated extensively in the philosophy of neuroscience. In this article I will evaluate metascientific approaches that claim direct molecular and cellular explanations of cognitive functions. I will initially consider the issues involved in linking cellular properties to behaviour from the general perspective of neural circuits. These circuits that integrate the molecular and cellular components underlying cognition and behaviour, making consideration of circuit properties relevant to reductionist debates. I will then apply this general perspective to specific systems where psychoneural reduction has been claimed, namely hippocampal long-term potentiation and the *Aplysia* gill-withdrawal reflex.

**Keywords** Psychoneural reduction · Memory · Neural network · Neural circuit · Behaviour

## Introduction

Scientists can seemingly have little time for philosophy (see Lipton 2005). For example, Zeki (1993, p 7) claimed that philosophers engage in “endless and ultimately fruitless discussion.” However, scientists take a philosophical position whenever they discuss knowledge or make claims to understanding. For example, Lichtman and Smith (2008, p 5) wrote in support of neural imaging that “just as most humans learn about football by watching... watching may be the most efficient route to understanding”. But can we understand nervous system (or any) activity simply by observing: could a naïve observer infer the rules of a game without some instruction (e.g. ignore the crowd, celebrations etc.)? In discussing connectomic approaches to neural circuits, Lichtman and Sanes (2008, p 349) wrote that structure “may enable predictions of circuit behaviour”, and Morgan and Lichtman (2013, p 496) wrote that structure “will signify a physiological process without the

---

✉ David Parker  
Djp27@cam.ac.uk

<sup>1</sup> Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EJ, UK

requirement of repeating the physiological analysis...a little physiology may go a long way”. But can we infer function from structure? Evidence from many systems over many years suggests a limited ability, at best, even when much of the physiology is known (a lot of physiology may go a little way).

The need for links between science and philosophy has been made. Bertrand Russell (1915, p 13) suggested that philosophy had “achieved fewer results, than any other branch of learning”, and to be successful required people “with scientific training and philosophical interests, unhampered by the traditions of the past”. Einstein claimed that training in philosophy gives “independence from prejudices from which most scientists are suffering”, and that better students “had a vigorous interest in epistemology” (see Howard 2017). However, linking philosophy and science is not simple. Philosophical debates often use language unfamiliar to most scientists to address abstract issues (e.g. philosophical zombies). Conversely, scientists can struggle to explain their methods and results, which can make it difficult for those outside to evaluate scientific claims: Reichenbach (1957) warned that “those who watch and admire scientific research from the outside frequently have more confidence in its results. The scientist knows about its difficulties (and) will never claim to have found the ultimate truth”.

However, Reichenbach may have committed the error he warned of in claiming that scientists are cautious, as there are many tenuous scientific claims. A current example is the Human Brain Project (HBP), many claims of which are more science fiction than science. Supporters of the HBP (e.g. Frackowiak 2014; Frackowiak and Markram 2015) try to negate critiques by arguing that the Human Genome Project (HGP) was successful despite its critics (the HGP is becoming a version of Godwin’s law in defending big science projects; see Lichtman and Sanes (2008) for its use in another context). However, HGP critiques did not focus on the feasibility of genome sequencing but on the hyperbole of HGP leaders (Roberts 1990; see Lewontin 1993): for example, that the HGP would address “basic philosophical questions about human nature” (James Watson); it would reveal “that we are all born with different talents and tendencies” (David Baltimore); and it would “impact on self-identity and self-understanding” (Walter Gilbert; all quotes from Gannett 2016). Critics highlighted that these aspects cannot be simply reduced to genes, let alone a genome sequence (see Gannett 2016). These critiques were not negated when the sequence was completed, but were embraced in the call for transcriptomics and proteomics. In defending HGP claims, Lichtman and Sanes (2008, p 349) wrote, “In one sense the accusation was justified...(but) faced with the need to convince skeptical funders, they needed to present an optimistic view”. This isn’t a defence: scientists have a responsibility to ensure that their language avoids false claims and hopes. The need for caution this caution is exemplified by the claim of Colin Blakemore (later head of the UK Medical Research Council), who wrote in 2000 that with completion of the sequence “will come, very quickly, new, targeted drugs and preventive measures... (so that) whole families are relieved, forever, of the curse of genetic disease” (Blakemore 2000, p 3).

Claims of this sort reflect the particular views of scientists, views that can direct research and the meaning and significance of the results. Scientists usually react to this claim, but research paradigms or programs direct research in ways that may not

conform to the strictly objective views often outlined in scientific textbooks (e.g. the dropping of falsified hypotheses). This is not to say that choosing one hypothesis over another necessarily reflects a lack of reasoning or logic. But the choice is a decision, a bounded or instrumental rationality that considers the best way to achieve a satisfactory solution to a desired goal, and this can introduce an element of subjectivity as decisions are based on multiple factors rather than formal rules. These include the views or assumptions of individual scientists; the assumptions of a scientific field that forms the accepted background knowledge, the questions asked, and the approaches used; and assumptions about the techniques and instruments used (Meehl 1990). Scientists and non-scientists thus need critique when deciding to make or accept claims to explanation or understanding and to consider the language used (Bennett and Hacker 2003).

In this article, I examine metascientific claims of the psychoneural reduction of memory from the perspective of neural circuits. These circuits are not usually considered in reductionist debates, but as they integrate the molecular and cellular properties underlying sensory, motor and cognitive functions, appreciation of circuit understanding is relevant to these debates. I will initially consider how circuit aspects affect reductionist accounts, before focusing on specific examples of psychoneural reduction in memory systems.

## Reductionism

Reductionism has been debated extensively in the philosophy of science (van Riel and Gulick 2018), and in psychology and neuroscience (Endicott 2001; Gold and Roskies 2008). However, most neuroscientists seem unaware of the philosophical debates in their field. The discussions can be complex, even for philosophers (see Endicott 2001), and includes whether there is a need for bridge laws (Nagel 1961) between the reduced and reducing theories (Ager et al. 1974); whether a reducing theory eliminates the reduced theory (van Riel and Gulick 2018); and how reduction relates to explanation (Craver 2007; Bechtel 2007).

While reductionism has the general aim of reducing all science to a common basis, typically physics, in intertheoretic reduction a higher-level theory is reduced to a lower-level that explains the phenomenon rather than necessarily being reduced to chemistry and physics. Identifying the appropriate explanatory level underlies much of the reductionism debate in psychology and neuroscience. Bechtel (2009) has distinguished mechanistic reductionism which requires continuity of effects between initial and ultimate mechanisms (Craver 2007), from ruthless reductionism that claims direct molecular or cellular links to cognition and behaviour (Bickle 2003). Consideration of these aspects is not “fruitless discussion” as claimed by Zeki (see above), as no analysis, technique, or equation no matter how sophisticated or accurate is of use if the underlying concept is flawed. Consider trying to explain behaviour from the interaction of actin and myosin proteins in muscle that underlie muscle contraction. While these molecular interactions are understood and explain movement to some extent, they don’t explain behaviours, which depend on the central nervous system, skeletal properties, and various environmental aspects. For

example, in *Phaedo*, Plato described how in the last hours of his life Socrates denied that his behaviour could be explained mechanistically, his posture in terms of muscles pulling on bones and his speech to the properties of sound, air and hearing, as the cause of his behaviour was that he had been sentenced to death for impiety and he had refused to escape.

Mechanistic and ruthless reductionism both consider that molecular and cellular properties underlie cognition and behaviour (the neuron doctrine; Gold and Stoljar 1999) a view generally accepted by neuroscientists (e.g. Barlow 1972; Hubel 1974; Kandel 1998; Zeki 1993; Crick 1994; Edelman 1989; Changeux 1997). Thus, Churchland and Sejnowski (1992) wrote that “it is highly improbable that emergent properties cannot be explained by low-level properties” (1992, p 2). Paul Churchland (1996, pp 3–4) claimed the neuron doctrine explains, “how the smell of baking bread, the sound of an oboe, the taste of a peach, and the color of a sunrise are all embodied in a vast chorus of neural activity...how the motor cortex, the cerebellum, and the spinal cord conduct an orchestra of muscles to perform the cheetah’s dash, the falcon’s strike, or the ballerina’s dying swan...how the infant brain slowly develops a framework of concepts with which to comprehend the world.” This hyperbole is contrasted by Hubel (1995, p 222) who, while not negating the neuron doctrine, wrote, “the knowledge we have now is really only the beginning of an effort to understand the physiological basis of perception... We are far from understanding the perception of objects”; and Torsten Wiesel who claimed that ‘we need a century, maybe a millennium’ to comprehend the brain, and that beyond understanding a few simple mechanisms ‘we are at a very early stage of brain science’ (cited in Horgan 1999, p 19).

While non-reductionists can accept that brains consist of neurons and these underlie brain functions, they contend that a focus on neurons offers a “terrible explanation” (Putnam 1975, p 296). Consider again trying to explain a movement from all of the actin and myosin interactions that underlie muscle contraction in all of the relevant muscle groups: even if this description was accurate, complete, and could explain behaviour, it would be cumbersome compared to reference to properties like the type of movement (e.g. walking, hopping) or its speed or force. Reference to the behaviour would also be more general and would avoid getting bogged down by variations (two seemingly identical movements made by two people, or by a single person on different occasions, are unlikely to use exactly the same complement and arrangement of actin and myosin proteins). An explanation of cognition or behaviour solely in cellular and molecular terms would thus be cumbersome and complicated by variability.

## Metascientific approach

To avoid fruitless discussion in philosophical accounts of reductionism, Bickle (2008) suggested that the focus should be on what scientists consider to be reductions, a metascientific approach. This approach offers a link between philosophy and neuroscience as debate is informed by scientific data and methods. Bickle focuses on molecular and cellular cognition, especially memory. The literature is vast, and

Bickle (2008, p 34) says that to select studies for metascientific analyses we should take “examples dictated by the field’s most prominent researchers...publication in the most respected journals...and the like”). However, while it is impossible to consider all the evidence, these measures have limits for assessing science. Consider impact factors, which are typically used to measure a journal’s importance. The San Francisco Declaration on Research Assessment states that the impact factor was “created as a tool to help librarians identify journals to purchase, not as a measure of the scientific quality of research...the Journal Impact Factor has a number of well-documented deficiencies as a tool for research assessment” (<https://sfdora.org/>). The quality or reliability of science also does not increase with increasing prestige of a journal, but may actually decrease (Brembs 2018).

There are also issues with appeals to prominence. Scientific quality is only one factor that determines this: it is not difficult to find prominent neuroscientists whose work is not respected, even by those who cite it. This paradoxical behaviour typically reflects deference to power (Fong and Wilhire 2017). Citations are routinely used to measure prominence, but these can be manipulated to increase citations to some and limit them to others (Fong and Wilhire 2017). Even if only used to filter studies for metascientific analyses, appeals to prominence could generate a Matthew effect as a relatively small number of people will dominate analyses. As prominence does not guarantee immunity from error (e.g. the HBP and HGP; see also Hardcastle and Stewart 2002), this could lead to “the canonization of weak and sometimes false facts” (Fortunato et al. 2018, p 2). Sagan (1987, p 7), writing to support science, wrote that dropping erroneous hypotheses “doesn’t happen as often as it should, because scientists are human and change is sometimes painful”.

The need for objective critique rather than appeals to prominence is illustrated by the replication crisis (Ioannidis 2012), and historical acceptance of erroneous neuroscience paradigms (see Gross 2009). Paradigms allow scientists to work and communicate effectively, but they can also influence analyses and conclusions: Kuhn (1970) claimed that the measure of a scientist was how well they fitted to the prevailing paradigm, and this can reduce flexibility and limit critique (he referred to this as the essential tension; Kuhn 1977). While scientists generally reject these claims, they have found some support. Peter Medawar (1963) asked if the scientific paper was a “fraud”, as it gives “spurious objectivity” and “a totally mistaken conception, even a travesty, of the nature of scientific thought” (to really understand scientific work Medawar said you had to “listen at a keyhole”; Medawar 1967, p 7). Error and dogma can be claimed to be overcome by self-correction in science, but while this may apply in the long-run, Max Planck’s aphorism of scientific change reflecting the death of adherents of outmoded ideas rather than logic generally applies. Feynman (1985) used the example of Robert Millikan’s measurement of the charge on an electron. Millikan’s value was too low because he used the incorrect value for the viscosity of air. Values after Millikan gradually increased to reach the correct value after approximately thirty years (Robinson 1937). Feynman claimed the error was not corrected sooner despite higher (correct) values being obtained because it was assumed that the higher values were wrong and reasons were sought to give values closer to Millikan’s. This wasn’t dishonesty, but a lack of confidence in challenging prominence.

Consideration of prominent papers and the claims of prominent individuals is necessary because their prominence, by definition, means that they will have a major impact on a field. But they have to be considered critically rather than used as a measure of the validity of a theory through appeal to authority or tradition.

Mechanistic links between neurons and behaviour have been studied since the 1950's from the perspective of neural circuits (e.g. cerebellum (Eccles et al. 1967); spinal cord locomotor circuits (Stuart and Hultborn 2008); retinal processing (see Hubel 1995; Barlow 1972), and neuroethological analyses of the cellular mechanisms underlying behaviour in various model systems (Zupanc 2010). These circuits integrate the molecular, cellular and synaptic properties contributing to sensory, motor, and cognitive functions. While circuit function alone cannot explain behaviour, by integrating the cellular aspects that contribute to higher-level functions, neural circuits offer a link between physiology and cognition and behaviour. Requirements for explaining circuit function have been outlined by scientists (e.g. Yuste 2008; Koch 2012; Braganza and Beck 2018) and philosophers (Bechtel and Richardson 1993; Bechtel 2008; Machamer et al. 2000; Silva and Bickle 2009; Piccinini and Craver 2011; Craver and Kaplan 2018). These outlines repeat but don't refer to criteria previously applied to neurophysiological analyses of invertebrate and lower vertebrate model systems (Bullock 1976; Selverston 1980; Getting 1989). These were initially called simple systems, but were soon tellingly re-named *simpler* systems. Conventional neurophysiological criteria for linking cellular properties to circuits function are: (1) identify the component cells involved; (2) identify the synaptic connections between these cells; (3) determine the functional properties of the cells and synapses (excitability and synaptic properties and their underlying mechanisms); and (4) if a model of the system using this information mimics actual system features then some claim to understanding can be made (see Selverston 1980).

In their metascientific approach to reductionism, Silva and Bickle (2009) say that molecular explanations of cognition require:

- A. Observation of the hypothesised molecular mechanism strongly correlated with the behaviour.
- B. Negative alteration—decreasing target activity affects the behaviour
- C. Positive alteration—increasing target activity triggers the behaviour
- D. Integration—gather other supporting evidence from previous data or expectations.

These match previous experimental criteria (see Jeffrey 1997). But would a molecular explanation of memory be like trying to explain behaviour from actin and myosin interactions (see above)? Actin and myosin molecules correlate very strongly with movement (A); negative alteration would reduce or totally occlude movement (B); positive alterations that trigger the same pattern of actin and myosin interactions would evoke the movement and overexpressing these proteins would positively alter the behaviour (Antonio and Gonyea 1993; C); and varied integrated evidence would support actin and myosin involvement (D). This may be considered a trivial argument, but is this because we understand the role of

actin and myosin and how they relate to movement and behaviour better than we understand the role of molecules in memory?

Finding correlations between components and behaviours is often not difficult, but they can lead us astray without proper critique. For example, requiring components to be strongly correlated with the behaviour could introduce an error of omission as weakly correlated aspects can have powerful system effects (Schneidman et al. 2006). Positive and negative manipulations can strengthen correlative evidence. In the 19th century David Ferrier said that motor acts resulting from stimulation of the cortex alone did not indicate that the region was a motor one, and suggested combining stimulation and lesion studies (i.e. positive and negative alterations). Issues were soon recognised with these approaches, including diaschisis (i.e. action at a distance, effects that occur at sites other than the intended target; see Carrera and Tononi 2014), a feature that caused Ferrier to mis-localise various functions in the cortex (Jacobson 1993).

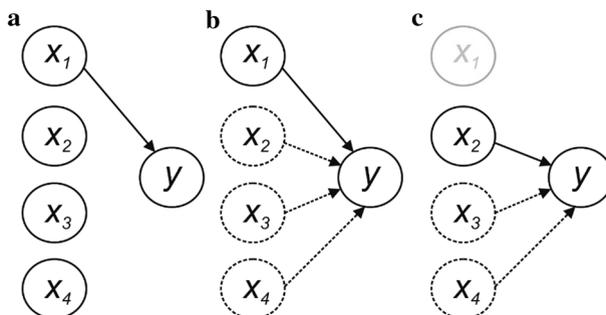
Positive and negative manipulations are used as criteria to identify neurons involved in neural circuits generating behaviours, but there are again issues with this approach (see Selverston 1980). Consider motor neurons. If we didn't know their role we may assume that they are involved in movement as they connect to muscles (point D of Bickle and Silva above). To see if their activity was correlated with the behaviour (point A) we could record from a motor neuron while an animal moved. However, motor neurons have variable sub-types associated with different types of movement: slow motor neurons are used when small muscle forces are required for long periods (standing or walking), and fast motor neurons when movements require brief but large muscle forces (sprinting or jumping). Recording from a fast motor neuron when an animal was making a slow movement (or vice versa) may result in no correlated activity: without awareness of the different motor neuron sub-types we could conclude that motor neurons are not involved in movement, an error of omission. Criterion (A) may also lead to errors of commission: respiratory neuron activity can correlate with locomotor activity (Burrows 1975; Rojas-Libano et al 2018), and we could erroneously assume that respiratory activity generated limb movements. This shows the utility of negative and positive alterations (B and C), as altering respiratory neuron activity would not affect locomotion, thus identifying the error. However, we may see no effect with manipulation of skeletal motor neurons if they were activated in physiologically-irrelevant ways; if fast motor neurons were silenced during slow movements (or vice versa); if the system showed degeneracy (i.e. different components can generate the same output) or redundancy (similar elements can serve the same function); or if there were compensatory adjustments to the manipulation (these can occur in minutes; Frank et al. 2006). All of these effects could thus cause errors of omission. Errors of commission could occur if input neurons to a motor neuron were activated and we assumed that they directly evoked the locomotor activity. Identifying direct vs indirect actions is non-trivial, but can seemingly be treated as such. For example, Olsen and Wilson (2008, p 6) state, "if one can demonstrate that a precisely-timed depolarization of one neuron evokes a short-latency synaptic response in the other neuron, then a direct connection is unequivocal". They refer to this as "a litmus test for synaptic connectivity", but in reality there are several issues with this statement (e.g. should scientists use the term

unequivocal?), but the main one is that the statement is demonstrably wrong (see Berry and Pentreath (1976), who discuss the limitations of latency to measure direct connections).

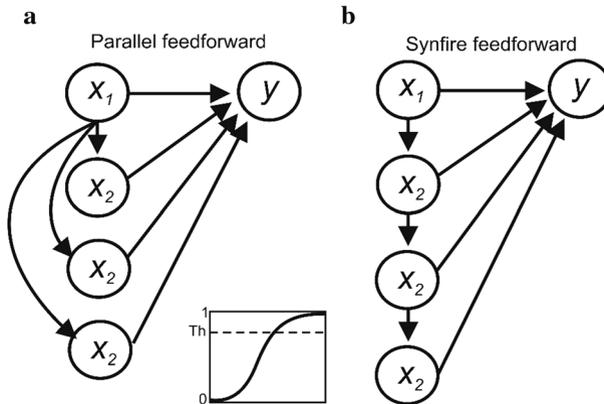
All of these issues can potentially be overcome if factors that could generate false negatives or positives are considered. Negative results are more likely to prompt this consideration to find out why a predicted effect didn't occur, but there is a danger that false positives are accepted because they support a favoured hypothesis. The need to address potential caveats is demonstrated elegantly by Berry and Pentreath (1976): in addition to highlighting approaches that can increase confidence in direct (i.e. single or monosynaptic connections) over indirect (i.e. through multiple or polysynaptic connections) interactions, they illustrate the caution that Reichenbach claimed scientists show by emphasising that even these approaches only provide *putative*, not unequivocal, evidence.

Positive and negative manipulations for identifying components associated with a behaviour relate to criteria of necessity and sufficiency (e.g. Kupferman and Weiss 1978). Negative manipulations that cause loss of function can suggest necessity, while positive manipulations that trigger a function can suggest sufficiency. These interpretations are complicated by the system organisation. Consider the scheme in Fig. 1. In (a),  $x_1$  is the only connection onto the output  $y$ , and  $y$  is evoked/abolished when  $x_1$  is activated/inactivated, suggesting a necessary and sufficient causal link between  $x_1$  and  $y$  (an example of a "command system"; Kupferman and Weiss 1978). However, if the system was redundant or degenerate (Tononi et al. 1999; b) or if there was some compensation for the removal of  $x_1$  (see Davis and Bezprozvanny 2001), the absence of  $x_1$  could allow one of  $x_2$ - $x_4$  to activate  $y$  (c), making  $x_1$  sufficient but not necessary for evoking  $y$ .

Connections between  $x_{1-4}$  add additional issues (Fig. 2). In (a),  $x_1$  sends parallel feedforward projections to  $x_2$ - $x_4$ , which sum to activate  $y$ . Activating  $x_1$  will evoke  $y$ , and blocking  $x_1$  will block  $y$ , suggesting  $x_1$  is necessary and sufficient for  $y$ . But  $x_1$  would only be necessary, not sufficient, if  $y$  needed the summed input from  $x_1$ - $x_4$ . In the case of a synfire-like chain (i.e. where activity is strictly feed-forward), activating  $x_1$  will evoke  $y$  and  $y$  will be blocked when  $x_1$  is knocked out or inhibited, again



**Fig. 1** **a** A unique direct pathway from  $x_1$  to  $y$  makes  $x_1$  necessary and sufficient for activating  $y$ . **b** With other inputs to  $y$ ,  $x_1$  may no longer be necessary as removal of  $x_1$  may evoke compensations that allow one of  $x_2, x_3$  or  $x_4$  to activate  $y$  (**c**)

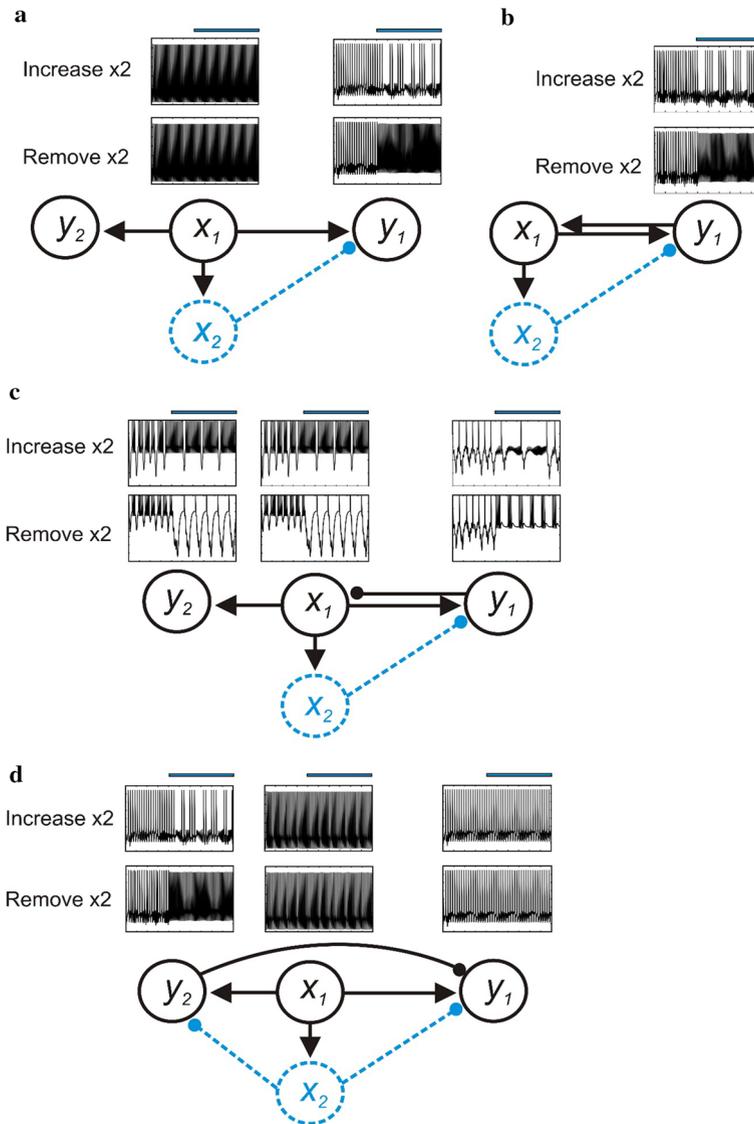


**Fig. 2** **a** A feedforward chain of connections where  $x_1$  activates  $x_2-x_4$  which all feedforward to activate  $y$ .  $x_1$  would be necessary for activating  $y$ , but not sufficient if the summed activity from  $x_1-x_4$  was needed to exceed the threshold (Th) for activating  $y$ . **b** In a synfire chain, activity in  $x_1$  would sequentially activate  $x_2-x_4$ , which all feedforward to activate  $y$ .  $x_1$  would again not be sufficient for  $y$  if the summed activity from  $x_1-x_4$  was needed, and would also not be necessary if  $x_2-x_4$  was sufficient to activate  $y$

suggesting  $x_1$  is necessary and sufficient. But  $x_1$  may not be sufficient for the reason given above, and would not be necessary if stimulation of  $x_2$  could recruit  $x_3-x_4$  to exceed the threshold for evoking  $y$ .

Feedback connections, a common feature in nervous and molecular pathways, add further complications. The circuits in Fig. 3 show effects in a simple computer simulation where  $x_1$  sends parallel excitatory inputs to output neurons  $y_1$  and  $y_2$ , and to an interneuron  $x_2$ . Assume  $y_1$  generates the output underlying a behaviour we are trying to explain, and we positively and negatively manipulate  $x_2$  to test the hypothesis that it inhibits  $y_1$ . Without feedback connections (a), inactivating  $x_2$  removes the inhibition of  $y_1$  and increases its activity, while activating  $x_2$  activity reduces  $y_1$  activity, consistent with the hypothesised inhibitory role of  $x_2$ . However, with feedback excitation from  $y_1$  to  $x_1$  (b), removing  $x_2$  will increase  $y_1$  activity, as hypothesised. However, increasing  $x_2$  activity causes oscillations rather than a reduction in excitability, because: (1) increased inhibition from  $x_2$  reduces  $y_1$  activity; (2) this reduces feedback excitation of  $x_1$ ; (3) this reduces  $x_2$  activity; (4)  $y_1$  activity increases; (5) this increases  $x_1$  and  $x_2$  activity to inhibit  $y_1$ . This cycle will repeat to cause oscillation. With feedback inhibition from  $y_1$  to  $x_1$  (c), increasing  $x_2$  activity reduces  $y_1$  activity as hypothesised, but as this reduces feedback inhibition from  $y_1$ ,  $x_1$  activity will increase to increase  $y_1$  activity to a greater extent than that caused by  $x_2$  manipulation alone. Removing  $x_2$  will increase  $y_1$  activity, but as this will inhibit  $x_1$  the excitatory drive to  $y_1$  will be inhibited, again causing oscillations in  $y_1$  as  $x_1$  activity increases and decreases. Finally, as  $x_1$  connects to  $y_2$ , any changes evoked in  $x_1$  caused by manipulating  $x_2$  will also alter any behaviour influenced by  $y_2$ , even though neither  $x_1$  or  $y_2$  is directly affected by  $x_2$  (i.e. diaschisis).

Changes in  $y_1$  (and other) outputs can thus cause effects that are not predicted from manipulation of  $x_2$ . An added issue is that if  $x_2$  did directly affect  $y_2$ , the  $y_1$  output may be unchanged despite widespread changes in the circuit. In (d),  $x_2$  inhibits



**Fig. 3** Summary of circuit interactions with feedback realised in a simple MatLab computer model (see Jia and Parker 2016 for model details).  $x_1$  drives activity in  $x_2$ ,  $y_1$ , and  $y_2$ , while  $x_2$  provides feedforward inhibition to  $y_1$ . The neurons are modelled using Hodgkin-Huxley kinetics, and inhibitory (filled circle) and excitatory synapses (arrow) are modelled to generate excitatory and inhibitory postsynaptic potentials. The circuit is driven by a constant excitatory input to  $x_1$ . **a** With only feedforward connections, positive and negative manipulations of  $x_2$  decrease or increase  $y_1$  activity, respectively. **b** Feedback excitation from  $y_1$  to  $x_1$  causes oscillations of  $y_1$  activity when  $x_2$  inhibition is increased. **c** With feedback inhibition from  $y_1$  to  $x_1$ , positive and negative manipulations of  $x_2$  evoke varied effects in  $y_1$ ,  $x_1$ , and  $y_2$ . **d** With inhibition from  $x_2$  to  $y_1$  and  $y_2$  and inhibition from  $y_2$  to  $y_1$  positive and negative manipulations of  $x_2$  lead to no change in  $y_1$  output, despite  $x_2$  having its hypothesised effect. See text for explanations

both  $y_1$  and  $y_2$ , and  $y_2$  provides feedforward inhibition to  $y_1$ . As a result, positive or negative manipulations of  $x_2$  will evoke changes in  $y_2$  that ultimately leave  $y_1$  unaffected, erroneously suggesting no influence of  $x_2$  in the circuit.

Linking cellular mechanisms and circuit activity can thus be difficult with the proposed metascientific and neurophysiological criteria. The potential issues could in principle be addressed in reductionist analyses, but only by considering potential caveats. Interventionist approaches using positive and negative manipulations since at least the 19<sup>th</sup> century (e.g. the David Ferrier example above) have been subject to false positives and negatives. The models here are simple, but feedforward and feedback interactions are common features of molecular and neural systems, making it unlikely that these effects will be absent in more complex systems. The various criteria used for explanation of behaviour and psychoneural reduction outlined above can provide useful insight, but we have to remain critical and consider wider circuit interactions that can lead to false positive or negative evidence for a hypothesis. In making a manipulation we cannot assume that everything else is equal, and these effects have been demonstrated to underlie errors of interpretation (see Selverston 1980 for examples). Gregory (1962) gave the example of the vibration, spluttering, and loss of power caused by removing a spark plug from an engine. We could mistakenly claim that the spark plug was an anti-vibration or anti-spluttering device, but our knowledge of the system and spark plugs allows us to correctly interpret the effects (e.g. that vibration and spluttering result from fuel ignition by a hot manifold). This is integrated knowledge (point D), and it can help, but it introduces some circularity as we have to know a certain amount about the system to understand it. Appealing to supporting evidence (point D above) also introduces the danger of affirming the consequent. Avoiding these errors not only requires consideration of potential caveats, but also the ability to specifically address them. The latter is difficult at best (see below; Parker 2010), but even if we can intervene to precisely manipulate known targets, diachisis remains an issue in any feedforward/feedback system.

Not only have we learnt that we can make errors in identifying relevant components and their circuit roles using conventional criteria, but we have also learnt that reducing circuits to their components is not sufficient to explain behaviours (e.g. Selverston 1980; Parker 2006, Parker 2010; Jonas and Kording 2017). The ability to explain an output may depend on the nature of the organisation and interactions between the components of the system, in other words the relations between the components rather than the features of any single component. Simon (1962) suggested that analyses of components and their interactions can explain an output in decomposable systems (i.e. where there is a hierarchy of interrelated subsystems with distinct roles, and interactions within components are stronger than those between them) or in nearly decomposable systems (where components interact but can be studied separately), but not when the system is minimally decomposable (i.e. the organisation of components plays a significant role). Francis Crick's (1994) claim that "your joys and your sorrows, your memories and ambitions, your sense of personal identity and freewill, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules", is trivial, it is the assembly, the organisation and interaction of molecular, cellular and behavioural properties, that is

key to these features. For example, sodium channel activation requires cooperativity between voltage-sensitive S4 regions in sodium channels (Marban et al. 1998), while the cardiac rhythm reflects the interplay of ion channels in cells organised into the 3-D structure of the heart, influenced by the rest of the body and behaviour (Noble 2002). While substantialists see components as discrete (decomposable) entities with intrinsic characteristics, leading to references to memory molecules, inhibitory or excitatory neurons, mood or pain transmitters etc., these properties reflect relationships with other components. For example, excitation can occur from interactions between two (or any even number) of nominally inhibitory neurons; a motor neuron is defined by its relationship to a muscle or gland; neurotransmitters aren't inhibitory or excitatory or for mood, reward, or pain, their functional effects instead depend on the specific receptors that activate on cells in specific circuits. Components form relationships that define components. Charles Sherrington highlighted this for the motor system when he wrote that "a simple reflex is probably a purely abstract conception, because all parts of the nervous system are connected together and no part of it is probably ever capable of reaction without affecting and being affected by various other parts...the simple reflex is a convenient, if not probable, fiction" (quoted in Posner and Raichle 1994, p 5). Relational properties can also generate functional components that do not exist outside of the normal organisation and activity of the system (e.g. ephaptic communication; Weiss and Faber 2010).

This has parallels in genetics: the protein that ultimately results from a DNA sequence depends on the regulatory dynamics of the whole cell, not the DNA sequence alone, and the "gene for" approach is now recognised as simplistic (Dupre 2008). Under experimental conditions, genetically identical strains of animals show variable phenotypes even when all other variables are thought to be controlled (Crabbe et al. 1999), and even simple genetic diseases (e.g. phenylketonuria) differ in severity in individuals with the same genotype, and can be affected by behaviour (e.g. limit foods rich in the amino acid phenylalanine). The Centers for Disease Control's concluded, "variable expressivity is the rule rather than the exception" (see Moore 2001, p 230). This doesn't make molecular genetic analyses irrelevant, but illustrates that on their own they can't provide a full description.

Even if all components and direct interactions are correctly identified, this only describes them, it doesn't explain how they generate outputs. Reductionism, meta-scientific or philosophical, is more than identifying parts. Different analyses are concerned with different properties and use different explanations depending on the focus of the analysis. Molecular and cellular analyses of how component molecules affect action potential generation or synaptic transmission become subsumed in neural circuits into the excitability of neurons and the sign and amplitude of synaptic inputs in structural motifs where function depends on the relationship of one component to another; molecules underlie these effects, but descriptions of excitability and synaptic transmission do not refer directly to these features. Circuits interact with other circuits, the body, and the external environment to generate behaviours: action potentials and synaptic potentials underlie these effects, but mechanisms do not usually refer to action potentials or synaptic potentials but to the regions activated and their inputs/outputs. This progression is mirrored in computational analyses: single cell computations can focus on the properties of structurally detailed ion

channels and associated changes in ions inside and outside the cell, or cable equations describing voltage changes along anatomically detailed 3D reconstructions of neurons. As circuits are simulated these details are dropped and cellular and synaptic effects are described by Hodgkin-Huxley kinetics and statistical or phenomenological models of transmitter release in single compartments: in more complex networks this detail is also lost and we change from the biophysical or biologically realistic to abstract connectionist models (e.g. Hopfield networks) and input/output relationships and information coding. While this progression traditionally reflected limits on computing power, mechanisms are difficult to follow even in powerful simulations when the detail increases beyond very modest levels: extra detail thus does not necessarily improve the explanation (see Greenberg and Manor 2005 for a demonstration of this). The relevant detail and level of analysis thus depends on the question being addressed.

This is again mirrored in genetics. The reductionist focus on DNA is not essential to understand genetics. We understood inheritance decades before the structure and regulation of DNA was elucidated, and more significantly, the molecular detail did not eliminate these earlier descriptions: population and medical (human) genetics and genetic counselling still use conventional hereditary diagrams, Punnett squares, and terms like genotype and phenotype, dominant and recessive etc., rather than refer to DNA sequences and translational and transcriptional mechanisms and regulation. Thus, while not negating their importance, molecular and cellular explanations do not necessarily make higher-level explanations otiose (Bickle 2003). Describing molecular or cellular effects associated with behaviour doesn't explain behaviour, any more than turning a key to start a car makes us mechanics. The limitations of molecule to behaviour leaps are shown by cystic fibrosis and Huntington's disease: while we know a lot about the molecular basis of these diseases, this molecular detail hasn't led to effective interventions.

## Examples of psychoneural reduction—the cellular basis of memory

The issues outlined above relate to the general difficulties of linking component parts to normal and abnormal functions in systems that consist of multiple interacting components. The specific example of memory, a focus of psychoneural reduction and molecular and cellular cognition, will be discussed to illustrate how the issues outlined above affect metascientific claims to direct causal-mechanistic molecular explanations of behaviour.

A major focus of molecular and cellular cognition and psychoneural reduction has been on hippocampal long-term potentiation (LTP), arguably the dominant aspect of neurophysiological analyses over the last 40 years. Cellular and molecular analyses consider memory differently to psychology. In psychology, memory is a multifaceted behaviour that reflects non-declarative and declarative aspects in multiple brain regions (Hardt and Nadel 2018). The hippocampus alone has been implicated in spatial memory, declarative memory, explicit memory, recollection, associative memory, and relational memory (see Konkel and Cohen 2009). Memories are context-dependent and reflect the formation of associations between

new and older memories, are labile on retrieval (an illustration of how behaviour affects lower-level mechanisms), and require reconsolidation using mechanisms that differ to initial consolidation. Consolidation is thus a “never-ending” process (McKenzie and Eichenbaum 2011), negating the view that memories become robust and fixed after initial consolidation (Bechtel 2009; Josselyn et al. 2015). This is reflected in different people remembering the same event differently, or an individual’s memory differing over time (Bartlett 1932). Neuroscience generally sees memory as a representation in some molecular/cellular structure or process that reflects long-term increases in the strength of specific synapses (typically LTP) evoked by the activity associated with the encoding of these memories. Experimentally this has often meant using various tetanisation protocols in isolated regions of the brain (e.g. hippocampal slices; McEachern and Shaw 1999; Gallistel and Matzel 2013). Effects under these conditions are correlated to learning and memory in behavioural studies (e.g. by knocking-out components; see below). However, the behaviours considered are typically quite limited (e.g. to freezing or approach or avoidance behaviour) compared to memory studied in psychological investigations. This probably reflects the focus on the underlying molecular and cellular mechanisms rather than behaviour (Krakauer et al. 2017), and the need to consider simpler behaviours to facilitate lower-level explanations.

Hippocampal LTP has been considered a mechanism for learning and memory (e.g. Bliss and Collingridge 1993, 2013). The 2016 Brain Prize was awarded to three prominent LTP researchers who “have revolutionised our understanding of how memories are formed, retained and lost” ([http://www.thebrainprize.org/files/4/uk\\_press\\_release\\_final.pdf](http://www.thebrainprize.org/files/4/uk_press_release_final.pdf)). Bickle (2003, p 44) suggests that although “Not even its most strident proponents think that LTP is *the* cellular/molecular mechanism for all forms or types of memory (their writing sometimes give this mistaken impression)”. There are many examples of this: Poo et al. (2016, p 1) write, “There is now general consensus that persistent modification of the synaptic strength via LTP ...represents a primary mechanism for the formation of memory”, and Morris (2003, p 646) claimed the link between LTP and memory was at a point where “it is reasonable to set aside a scientist’s natural scepticism about the central principle”. Without discussing whether scepticism is “natural” to scientists, it is surely never reasonable to set it aside (it should be noted that outside of the LTP paradigm Morris retains scepticism; Morris 2014). Even if these statements are not really meant, this can’t be excused: scientists have a responsibility to ensure that the language they use is appropriate; we don’t accept careless methodology, and shouldn’t accept careless language. While many prominent figures in the LTP field claim links between LTP and memory, others present more critical summaries (e.g. Jeffrey 1997; Keith and Rudy 1990; McEachern and Shaw 1999). These are older references, but scepticism hasn’t been set aside. Queenan et al. (2017) recently showed that the strength of certain synapses increased during learning, consistent with the LTP paradigm, but memory persisted when this was abolished. They concluded, “Gross synaptic strengthening can be excluded as a candidate mechanism for memory storage....a paradigm shift may be required to look at the existing data from a more productive angle (Queenan et al. (2017, p 115).” The lower profile of these more critical authors means that these views

may not be considered in metascientific analyses based on prominence, but the central principle, the link between LTP and memory, must remain open.

Bickle (2015) specifically refers to David Marr's three levels of analysis (Marr 1982) when saying that critiques of cellular approaches to cognition no longer apply: "a careful metascientific account...over the last two decades reveals that Marr's criticisms no longer have force, as neuroscience now directly explains cognition, causal-mechanistically (Bickle 2015, p 299)". To support this claim, Bickle (2016) appeals to the development of new experimental tools. He correctly highlights that approaches using drugs or lesions to affect molecular, cellular, or circuit components have marked limitations: pharmacological interventions can lack specificity, while lesions are destructive, affect non-targeted regions, and only allow loss of function. The newer tools for cellular and circuit analyses are gene knock-outs and optogenetics. Bickle (2016) claims that metascientific analyses show that these tools have led to scientific revolutions, claims that are also frequently made by users of these techniques (see Parker 2018 for discussion).

Addressing the LTP/memory link has largely relied on molecular genetic techniques, principally gene knock-outs. The involvement of various molecules in LTP is often persuasive, although the molecular basis of LTP, as with other forms of plasticity, is complex and many molecules are involved (Sanes and Lichtman 1999). At best these approaches offer molecule to behaviour correlations. To say molecular effects are incorporated "into the circuits for particular memory traces", or that the "causal mechanistic story for the cognitive phenomenon in question now resides at the lowest level ...in conjunction with the anatomical circuitry that gets the neuronal activity out to the periphery" (Bickle 2008, p 49), does not explain circuit effects or the behaviour.

Bickle (2016) suggests that work on spatial learning after pharmacologically blocking NMDA-dependent LTP (Morris 1989) using the drug AP5 was the "motivating problem" for the development of gene knock-outs. NMDA receptors are a sub-type of receptor for the neurotransmitter glutamate, the neurotransmitter principally associated with excitation in nervous systems. There were several issues with the pharmacological approach, including disrupted learning with drugs inactive at NMDA receptors (Walker and Gold 1994); slowed rather than no learning when NMDA receptors were blocked (Keith and Rudy 1990); sensorimotor defects with NMDA-receptor blockade (Tricklebank et al. 1989); effects on non-hippocampal sites caused by the spread of AP5 into the forebrain; and the anxiolytic effects of AP5 (benzodiazepines have similar effects on learning). The pharmacological results showed, at best, some correlation between NMDA-dependent effects and memory: Jeffrey (1997, p 101) concluded that "drug-induced interference with LTP has produced mixed results, neither proving nor disproving the LTP-learning hypothesis".

Silva et al. (1992) used knock-outs of CaMKII, a calcium-dependent intracellular pathway (or second messenger) that can alter diverse functional properties of cells and synapses, to study the link between LTP and memory. This tested Lisman's (2003) claim that CAMKII could serve as a memory molecule. The knock-out animals were "jumpy" and more nervous in open field mazes, and tried "frantically to avoid human touch" ("aside" from this they were normal). Bickle (2016, p 5) claims the knock-out "delivered successfully on one key desideratum.... The

targeted mutation did not disrupt synaptic function”. But this wasn’t the case as the knock-out reduced paired pulse facilitation (PPF), an increase in the amplitude of the second of two synaptic inputs evoked within a short time period (Zucker and Regehr 2002). This is an alteration of synaptic function that could make LTP-inducing protocols less effective in knock-out animals (see Brown et al. 1990), and thus indirectly link LTP to memory. Also, LTP could sometimes occur even though CaMKII was completely knocked out, suggesting a lack of necessity for CaMKII under some conditions (Keith and Rudy 1990). Grant et al. (1992) used mutations of non-receptor tyrosine kinases, specifically *fyn*. PPF was again reduced, although Grant et al. say non-significantly. LTP was not blocked in *fyn* knock-out mice, as it could be evoked with increased stimulation or by pairing with postsynaptic depolarisation (conceivably reflecting the reduced PPF). Notably, this knock-out caused marked abnormalities in the hippocampus: overexpression of cells in the dentate gyrus and CA3 regions, and abnormalities of apical dendrites and a change in the density of CA1 cells. Grant et al. begged the question by claiming “mutations in the *fyn* gene result in an impairment of both LTP and spatial learning. This is consistent with the idea that LTP is *causally* important” (my *italic*). Discounting any influence of the marked structural abnormalities to claim a causal link to memory seems strange when viewed from outside the field, but not when seen from the perspective of adherence to the LTP paradigm.

Are the altered structure, changes in synaptic function, and behavioural changes in knock-out studies any less off-target than the caveats associated with the pharmacological approach used by Morris? Jeffrey (1997) wrote, “both positive and negative transgenic results can be accommodated within either the LTP=learning or the LTP≠learning frameworks”. The molecular genetic analyses offer correlations, not causal explanations. A recent example is the effect of knocking-out another second messenger molecule, PKM- $\zeta$ , in LTP. This was claimed as a landmark finding (Takeuchi et al. 2014), but subsequent work showed that the knock-out altered other molecular and cellular effects besides PKM- $\zeta$  (see LeBlancq et al. 2016), and produced unintended compensatory changes, thus weakening the claimed link (see Frankland and Josselyn 2016).

Optogenetics is a more recent technique than molecular genetics. This technique uses genetically-encoded light sensitive proteins that when activated by particular wavelengths of light evoke voltage changes that either activate or inhibit cells depending on the protein expressed. Claims are made to optogenetics high temporal and spatial precision. However, high is a relative term: while only those neurons that have the optogenetic protein in them will be activated, that does not necessarily mean that only the desired cells are labelled and activated, or are activated in physiologically relevant ways (Hauser 2014). Bickle (2016, p 10) rightly says that “nature doesn’t always cooperate to clump target neurons together into discrete cortical columns or microcolumns to make stimulation by microelectrodes feasible”, but nature also doesn’t cooperate by clumping neurons into specific groups that are uniquely labelled by optogenetic probes. Even if a single population of cells could be targeted, neuronal populations have variable properties and manipulating a population essentially averages diverse functional effects (see Soltesz 2006). Even if this can be overcome, diaschisis remains an issue with any optogenetic manipulation

(Otchy et al. 2015), negating the claim that optogenetics offers “direct cellular and molecular mechanisms for higher-level biological phenomena” (Bickle 2016, p 9). Direct links to behaviour are difficult to make, and require more than manipulating a system variable and monitoring the resulting effects (see Krakauer et al. 2017). Reference to memory in physiological analyses does not typically relate to psychology: the latter sees memory as very diverse and there are still questions over the categorisation of memory into different types, while the former typically limits memory to some overt behaviour like freezing or simple approach or avoidance behaviours (Josselyn et al. 2015). Even in these simple cases, optogenetics describes the effect of a manipulation, and in the best cases can make a case for a role of the manipulated components, but this is not an explanation of how the effect is caused.

Woodward (2003) suggests an interventionist approach can distinguish an explanatory model from a descriptive one. If  $x$  and  $y$  are causally related according to  $y=f(x)$ , where  $f$  is some specified function, intervening on  $x$  will cause  $y$  to respond in the way described by  $f$  (but  $y$  should not affect  $x$ ). But to explain a system we need to know how the intervention on molecule or cell  $x$  had the effect it did, and this requires knowing  $f$ , which in a circuit context will require knowing how  $x$  influences and is influenced by the circuit. This applies even if the intervention was “surgical” (Pearl 2000) and specifically affected a single component. But interventions are seldom the surgical molecular scalpels they are claimed to be (Kiehn and Kullander 2004), but lack precision and affect multiple components (for example see Gosgnach et al. (2006), including the supplementary material), making them closer to molecular sledgehammers. It can be argued that targeting a single component is impossible, as interventions in nervous systems are complicated by upstream and downstream diachronic effects caused by altering feedforward and feedback interactions (Carrera and Tononi 2014).

These issues are not restricted to knock-outs and optogenetics, but can occur with any intervention. For example, anisomycin, a drug used in many LTP studies to block protein synthesis (and one I have used in a different context; Parker et al. 1998), can have non-specific cellular effects that complicate conclusions (Sharma et al. 2012). Even temperature due to activity in behavioural studies can introduce artefacts that affected LTP claims (Moser et al. 1993). Jeffrey (1997) wrote, “Dissociating the possibly subtle changes in plasticity related evoked field potential size from the large non-specific changes produced by associated behaviors and affective states presents a formidable task”. This again highlights the need to retain rather than set aside scepticism.

LTP in the hippocampus still seems too complicated with too many caveats to claim a successful psychoneural reduction, despite the prominent claims. Learning of the gill and siphon withdrawal reflex in *Aplysia* offers a simpler model system in which to address the links between molecules, cells and behaviour. Kandel and colleagues have characterised molecular and cellular changes in sensory neurons during learning of the gill withdrawal reflex (see below). Kandel shared the 2000 Nobel Prize for demonstrating “how changes of synaptic function are central for learning and memory” (<https://www.nobelprize.org/prizes/medicine/2000/press-release/>). This work is considered the epitome of a successful reductionist account of behaviour (Gold and Roskies 2008; Bickle 2003; Gold 2009), even by those who take a

critical approach to neuroscience claims (e.g. Hardcastle and Stewart 2002). Bickle (2003) claims that Hawkins and Kandel (1984) provided an alphabet of mechanisms that explain more complex cognitive representations like second-order conditioning (where a learnt stimulus is used as the basis for learning about a new stimulus) and blocking (where a conditioned stimulus is presented with another conditioned stimulus associated with the unconditioned stimulus), to be “completely explained by *theoretical* combinations of these fundamental cellular properties coupled with reasonable anatomical and physiological assumptions”. However, Hawkins and Kandel (1984, p 381) emphasised that “some of the higher order behavioural phenomena discussed have not been tested in *Aplysia*. Our arguments on these points are therefore entirely speculative”.

The Hawkins and Kandel (1984) model consisted of two conditioned stimulus (CS) pathways, CS1 and CS2, represented by sensory neurons SN1 and SN2, and an unconditioned stimulus (US) pathway represented by a facilitatory serotonergic interneuron (FN). Gill motor neurons generate conditioned or unconditioned response (CR and UR). The CS neurons connect to the MN and the FN. In first-order conditioning SN1 and FN are paired, and FN produces associative plasticity at SN1 synapses onto the FN and the motor neuron. In second-order conditioning, since plasticity also occurs at the SN1 to the FN synapse, the CS1 can now become a US and influence the FN. Pairing CS2 and CS1 allows CS1 to activate FN to cause conditioning to CS2. This model predicts that second-order conditioning can be blocked without affecting first-order conditioning. Blocking can also arise if activity in the FN undergoes rapid depression during the second phase of a blocking trial as its output would be too small to support associative conditioning of SN2.

Buonomano et al. (1990) modelled the circuitry proposed by Hawkins and Kandel (1984) and found that it was possible to model either second-order conditioning or blocking depending on the threshold of the facilitator neuron, but both phenomena could not be modelled with a single set of parameters. They proposed an alternative hypothetical model, a lateral inhibitory network, which could simulate second-order conditioning and blocking. Neither model can claim to provide a successful psychoneural reduction without additional direct experimental support. Subsequent analyses in *Aplysia* by Kandel and colleagues reduced the preparation to isolated sensory and motor neurons in culture, with applied serotonin substituting for the facilitatory interneuron (see Kandel 2001). This removed the components needed to address the hypotheses outlined above (but see Trudeau and Castellucci 1993). The molecular basis of simpler forms of learning could thus be examined in detail, but the higher-level phenomena remained hypothetical (Kandel did not mention higher-level effects in his Nobel award speech; Kandel 2001).

While the Nobel award to Kandel suggests ultimate acceptance of the claimed mechanisms for the simple forms of memory, there were issues with even these simpler forms that would not be apparent in a metascientific approach that appeals to prominence (i.e. Kandel). Glanzman used the same *Aplysia* learning model, but unlike Kandel’s focus on the presynaptic sensory neurons, Glanzman (1995) and others (Trudeau and Castellucci 1995) showed learning-induced changes in the post-synaptic motor neurons. Glanzman (2010) said that this work was obscured by “the biases of some investigators” (p R31), and called the attention on the postsynaptic

mechanisms after 2000 the “return of the repressed” (p R34) (Kandel’s Nobel summary did not mention postsynaptic effects; Kandel 2001). Postsynaptic effects have subsequently been highlighted (Kandel et al. 2014), in some cases referring to its “recent” identification (e.g. Hawkins et al. 1981) despite the work being done several years before the Nobel award to Kandel (see Glanzman 1995). Glanzman (2010) summarised the overall state of knowledge of the *Aplysia* field: “At present we have only partial cellular accounts of learning, even for simple forms in relatively simple organisms.

Hickie et al. (1997) also analysed the role of the sensory-motor neuron synapses studied by Kandel et al. and Glanzman et al. in the gill withdrawal reflex. Using voltage-sensitive dyes that report on neuronal activity in the whole region of the CS conorlling gill withdrawal, they claimed their results “casts doubt on the validity of using this synaptic connection as a model for gill-withdrawal behaviour”. Others had considered this: Trudeau and Castellucci (1993) estimated a 25% contribution, and Antonov et al. a 55% contribution. Antonov et al. (1999) attributed the remainder to plasticity of excitatory and inhibitory interneurons in polysynaptic components of the reflex. Hawkins et al. (1981) had very impressively examined these interneurons, concluding that the circuitry was “forbidding in its complexity”, and it was not clear “what functions many of the interneurons serve” (Hawkins et al. 1981, p 311). Far less was subsequently done on this interneuron circuitry (but see, for example, Trudeau and Castellucci 1993), possibly because of the forbidding complexity which made the more tractable sensory to motor neuron synapses more attractive targets. This is common: we know far more about pyramidal neurons in the cortex and hippocampus, Purkinje cells in the cerebellum, and motor neurons in the spinal cord than the associated interneurons. Analyses do not necessarily focus on the most needed aspects but on those that are most tractable. This can reflect various factors: analyses may offer limited chances of success and graduate or postdoctoral researchers may rationally choose aspects that are more likely to generate data. This is an example of Medawar’s requirement to listen at a keyhole: papers do not highlight that an analysis was done because it offered greater success in getting a publication, but are instead required to suggest reasons why the analysis was pressing.

As with the Queenan et al. (2017) study suggesting that LTP is not mediated by persistent synaptic changes, a similar suggestion has been made for some forms of learning in *Aplysia*. Chen et al. (2014) and Bedecarrats et al. (2018) claim that the synaptic changes investigated in detail in *Aplysia* are an epiphenomenon of cellular changes at the cell body, and that learning can be transferred by RNA from trained to untrained animals (this revives the 1960’s ideas of McConnell who suggested learning transfer by RNA in flatworms (see Smalheiser et al. 2001; Rilling 1996).

The actual mechanism underlying the gill withdrawal reflex thus seemingly reflect a combination of presynaptic and postsynaptic cellular and synaptic effects. The successful reductionist account of the simpler forms of learning (classical conditioning) claimed in the Nobel citation to Kandel would require determining the relative contribution of each aspect to the behaviour, or alternatively showing that effects other than the changes in the sensory neurons (e.g. the work of Hickie et al. (1997) and Glanzman (1995)) were erroneous or irrelevant, and thus that the focus on sensory neuron synapses was correct. Neither was done, suggesting that

acceptance of the claimed understanding either reflected a lack of awareness of the competing evidence, or where it was known a subjective decision was presumably made that the presynaptic neurons provided a satisfactory explanation of the effects. This uncertainty is not apparent in metascientific analyses that appeal to prominence (Bickle 2003), but it weakens the frequently claimed reductionist explanation of memory in this system, without requiring the critique that the work was only done in *Aplysia* (Looren de Jong and Schouten 2005), or philosophical debate over whether Kandel's account appeals to psychological aspects and is thus not an actual psychoneural reduction (Gold and Stoljar 2009).

## Conclusion

An animal or person behaves, remembers or forgets, not a molecule or a part of their brain (Bennett and Hacker 2003). Claims to understanding associated with psychoneural reductions often rely on concepts and analyses driven by the need to limit variables to facilitate interpretations. Thus, behaviours may be constrained to simpler forms that can be more easily monitored and quantified, while cellular and circuit analyses traditionally use reduced preparations (e.g. tissue slices) with cellular effects extrapolated to simpler behavioural analyses. Even ignoring the potential for physiological changes under these conditions (e.g. Kuenzi et al. 2000; Hoffman and Parker 2010), these conditions remove effects from their normal functional context. This does not negate the utility of the analyses, but requires that claims to understanding behaviour have to be moderated until effects can be directly related to normal function and behaviour (Krakauer et al. 2017). For the examples of psychoneural reduction discussed here, this requires consideration of the caveats raised by the less prominent researchers in the fields. For example, to understand how the changes in sensory interneurons, motor neurons, and the various interneurons in *Aplysia* act together to generate the behavioural change, or how the cellular or synaptic activity in different regions of the hippocampus and other temporal lobe structures contribute to a memory, and where and how the activity in these centres is transferred, stored, and ultimately retrieved to generate actual behaviours.

Even with this analysis, psychoneural reduction does not make higher levels otiose. Gregory (2005) gave the analogy of a car journey from London to Cambridge: the mechanics of the car are of obvious necessity to the journey, but the journey can be made without knowledge of the engine or combustion (molecular or physiological aspects). Knowledge of these components alone would not allow the journey to be made (behaviour), as we also need a plan of the route (cognition). Lower (mechanical) or higher (route) errors could have the same outcome (i.e. being late or not arriving at all), but the explanation (and solution) will depend on where the error occurs (i.e. fix the engine or consult a map).

Causation is not upwards, behaviours and the context they are performed in also influence molecular and cellular properties. Behaviour is thus not simply an output to be reduced to cellular details to be explained, but part of the explanation. Pain provides an example. While there is significant insight into the molecular mechanisms of nociception (Dubin and Patapoutian 2010), nociceptive sensations

influence pain perception and pain-related behaviours through a distributed network of parallel feedforward and feedback pathways rather than one molecule or brain region (Hardcastle and Stewart 2009). Pain perception is also affected by psychological factors, context and expectation (i.e. the placebo effect), which can affect molecular and cellular properties (e.g. by triggering the release of endogenous opioids; Benedetti 2007; Jepma and Wager 2013). We can correlate pain-related behaviours to nociceptive transduction using positive and negative manipulations, and this could provide useful anti-nociceptive strategies, but this wouldn't explain pain-related behaviours, even if the component was necessary and significant to initiate the cascade of effects that led to it.

Despite its frequent use in neuroscience, the meaning of explanation or understanding is still debated in philosophy (see Craver and Kaplan 2018; Woodward 2017). A detailed account cannot be given here but will need a separate account. Propositional understanding (e.g. understanding “that” sodium is needed for an action potential) differs to explanatory understanding (why does sodium cause an action potential). Whether explanation requires mechanistic detail will probably depend on the type of analysis (Woodward 2017b), but while mechanistic atomism has been replaced in physics it still dominates biology. Midgely (2004) claims this could reflect a cargo cult: by following the example or early physics, we hope to repeat its successes. This is seen in claims that more data will lead to understanding. The Brain Activity Map project claims that understanding will result from mass recording of electrical activity (Alivisatos et al. 2013), and the project aims to “record neuronal activity at a single-cell level from ever more cells over larger brain regions” (Mott et al. 2018, p 3). The precursor of the Human Brain Project, the Blue Brain Project, went further, claiming that a more detailed brain emulation will cause a Copernican revolution (see Lehrer 2008). But while certain detail seems necessary to explain and rationally intervene in a system, understanding, let alone a Copernican revolution, does not “drop out” once a critical amount of detail is obtained. We have a glut of data about various aspects of the nervous system, but understanding has not followed (Midgely (2004) claimed that no amount of knowledge of Einstein's brain would tell a neuroscientist about relativity). This doesn't negate the importance of molecular, cellular or circuit analyses, but mechanistic detail needs the appropriate context and concept (Hardt and Nadel 2018). An example of this is the spinal cord half-centre underlying basic locomotor activity (Brown 1914), a concept with no cellular detail but one that still provides the foundation of all analyses of spinal cord locomotor network function (Stuart and Hultborn 2008).

It is not hard to see the attraction of reductionist analyses given the availability of tools that allow molecular and cellular manipulations (Bickle 2016). Parsimony and the search for simple explanations dominate, but should we expect neuroscience explanations to be simple? Neuroscience approaches remove subjective values and replaces qualities with quantities (e.g. the replacement of subjective feelings with transmitter receptors or levels in psychiatry; see Laing 1985). Claims to simplicity or that reductionist approaches are more objective are subjective. We can explain molecular, cellular and to some extent circuit properties, or why a person or animal behaves in a certain context. These approaches are not more or less objective, useful or otiose, but different ways of addressing the same question. Molecular and cellular

analyses need concepts or models that allows them to be related to behaviour, and behavioural analyses need cellular details to constrain potential mechanisms and provide rational targets for interventions. Understanding for much of neuroscience, reflects “non-ideal” explanations (Railton 1981), reduced uncertainty about a phenomenon but lacking the information needed for an “ideal” explanation; or objectual understanding, which uses a body of information together with logical analyses and inference to the best explanation, to offer the most plausible account of a phenomena (Lipton 2000).

In terms of what we know, there has clearly been great progress in neuroscience that has gone beyond the analysis specifically of components (action potentials and synaptic transmission) where classical neurophysiology has had great success, to examining the cellular and then molecular correlates of behaviours. However, detail (knowledge) and description (e.g. a neuron spikes during a certain behaviour or cognitive phenomenon) is not explanation or understanding. We can claim that the behaviour explains why the neuron spikes or that we understand that the behaviour is associated with spiking. But ideally we want to explain a function like memory in the same way that we can explain biochemical pathways like the urea cycle (the conversion of ammonia from the breakdown of amino acids to urea). In these pathways, identified components interact in a causal sequence; the role of each step in the sequence and how it explains the behaviour of the whole system is known; and this information can be translated clinically to control the system in specific ways. We have not achieved this for psychoneural reduction, the claims of the brain's complexity are far closer to the truth than any translational claims, but this doesn't mean it is necessarily impossible. But explanation will require consideration of the complexities rather than simplistic claims or assumptions, and closing the divide between the “internal” aspects that neuroscience generally considers and the “external” aspects that are the concern of psychology.

Instead of discussing or debating “what-ifs”, Bickle's metascientific approach that grounds debate on scientific data seems sensible. Philosophers can offer critiques of language and logic, something that scientists may not show, but they need to consider Reichenbach's warning and not be seduced by scientific claims (see McCabe and Castel 2008). Appealing to prominence or authority seems prone to this danger, especially in fields where issues and caveats are less obvious as appeal to authority allows erroneous views of dominant individuals to persist. Exceptions and variability abound: we can try to control aspects, but we are either unable (e.g. diaschisis) or unaware of what needs to be controlled, or we constrain aspects to an extent that may remove them from their normal context (Krakauer et al. 2017). Given the limits on explanation and understanding in neuroscience all claims need to be critically evaluated.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Ager TA, Aronson JR, Weingard R (1974) Are bridge laws really necessary? *Noûs* 8(2):119–134
- Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, Greenspan RJ, McEuen PL, Roukes ML, Sejnowski TJ, Weiss PS, Yuste R (2013) The brain activity map. *Science* 339:1284–1285
- Antonio J, Gonyea W (1993) Skeletal muscle fiber hyperplasia. *Med Sci Sports Exerc* 25:1333–1345
- Antonov I, Kandel ER, Hawkins RD (1999) The contribution of facilitation of monosynaptic PSPs to dishabituation and sensitization of the Aplysia Siphon withdrawal reflex. *J Neurosci* 19:10438–10450
- Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1:371–394
- Bartlett F (1932) *Remembering: a study in experimental and social psychology*. Cambridge University Press, New York
- Bechtel W (2007) Reducing psychology while maintaining its autonomy via mechanistic explanation. In: Schouten M, Looren de Jong H (eds) *The matter of the mind: philosophical essays on psychology, neuroscience and reduction*. Basil Blackwell, Oxford
- Bechtel W (2009) Molecules, systems, and behaviour: another view of memory consolidation. In: Bickle J (ed) *The oxford handbook of philosophy and neuroscience*. Oxford University Press, New York, pp 13–40
- Bechtel W, Richardson RC (1993) *Discovering complexity: decomposition and localization as strategies in scientific research*. Princeton University Press, Princeton
- Bedecarrats A, Chen S, Pearce K, Cai D, Glanzman DL (2018) RNA from trained aplysia can induce an epigenetic engram for long-term sensitization in untrained aplysia. *Eneuro* 5:3. <https://doi.org/10.1523/eneuro.0038-18.2018>
- Benedetti F (2007) Placebo and endogenous mechanisms of analgesia. *Handb Exp Pharmacol* 177:393–413
- Bennett M, Hacker P (2003) *Philosophical foundations of neuroscience*. Blackwell Publishing Ltd, Oxford
- Berry M, Pentreath V (1976) Criteria for distinguishing between monosynaptic and polysynaptic transmission. *Brain Res* 105:1–20
- Bickle J (2003) *Philosophy and neuroscience: a ruthlessly reductive account*. Springer, Netherlands
- Bickle J (2008) Real reduction in real neuroscience: metascience, not philosophy of science (and certainly not metaphysics!). In: Hohwy J, Kallestrup J (eds) *Being reduced*. Oxford University Press, Oxford, pp 34–51
- Bickle J (2015) Marr and reductionism. *Top Cognit Sci* 7:299–311
- Bickle J (2016) Revolutions in neuroscience: tool development. *Front Syst Neurosci* 10:24
- Blakemore C (2000) Achievements and challenges of the decade of the Brain. *EuroBrain* 2:1–4
- Bliss T, Collingridge G (1993) A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361:31–39
- Bliss TV, Collingridge GL (2013) Expression of NMDA receptor-dependent LTP in the hippocampus: bridging the divide. *Mol Brain* 6:5
- Braganza O, Beck H (2018) The circuit motif as a conceptual tool for multilevel neuroscience. *Trends Neurosci* 41:128–136
- Brembs B (2018) Prestigious science journals struggle to reach even average reliability. *Front Hum Neurosci* 12:37
- Brown T (1914) On the nature of the fundamental activity of the nervous centres; together with an analysis of the conditioning of rhythmic activity in progression, and a theory of the evolution of function in the nervous system. *J Physiol Lond* 48:18–46
- Brown TH, Kairiss EW, Keenan CL (1990) Hebbian synapses: biophysical mechanisms and algorithms. *Annu Rev Neurosci* 13:475–511
- Bullock TH (1976) In search of principles in neural integration. In: Fentress JD (ed) *Simple networks and behavior*. Sinauer Assoc, Sunderland, pp 52–60
- Buonomano DV, Baxter DA, Byrne JH (1990) Small networks of empirically derived adaptive elements simulate higher-order features of classical conditioning. *Neural Netw* 3:507–523
- Burrows M (1975) Co-ordinating interneurons of the locust which convey two patterns of motor commands: their connexions with ventilatory motoneurons. *J Exp Biol* 63:735–753
- Carrera E, Tononi G (2014) Diaschisis: past, present, future. *Brain* 137:2408–2422
- Changeux J-P (1997) *Neuronal man: the biology of mind*. Princeton University Press, Princeton

- Chen S, Cai D, Pearce K, Sun PY-W, Roberts AC, Glanzman DL (2014) Reinstatement of long-term memory following erasure of its behavioral and synaptic expression in aplysia. *ELife* 3:e03896
- Churchland P (1996) *The engine of reason, the seat of the soul*. MIT Press, Cambridge
- Churchland P, Sejnowski T (1992) *The computational brain*. MIT Press, Cambridge
- Crabbe JC, Wahlsten D, Dudek BC (1999) Genetics of mouse behavior: interactions with laboratory environment. *Science* 284:1670–1672. <https://doi.org/10.1126/science.284.5420.1670>
- Craver C (2007) *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. Clarendon Press, Oxford
- Craver CF, Kaplan DM (2018) Are more details better? On the norms of completeness for mechanistic explanations. *Br J Philos Sci*. <https://doi.org/10.1093/bjps/axy1015>
- Crick F (1994) *The astonishing hypothesis*. Scribners, New York
- Davis G, Bezprozvanny I (2001) Maintaining the stability of neural function: a homeostatic hypothesis. *Ann Rev Neurosci* 63:847–869
- Dubin AE, Patapoutian A (2010) Nociceptors: the sensors of the pain pathway. *J Clin Investig* 120:3760–3772
- Dupre J (2008) What genes are, and why there are no “genes for race”. In: Koenig BA, Lee S, Richardson S (eds) *Revisiting race in a genomic age*. Rutgers University Press, New Jersey
- Eccles JC, Ito M, Szentágothai J (1967) *The cerebellum as a neuronal machine*. Springer, Berlin
- Edelman G (1989) *The remembered present: a biological theory of consciousness*. Basic Books, New York
- Endicott R (2001) Post-structuralist angst—critical notice: John Bickle, *Psychoneural reduction: the new wave*. *Philos Sci* 68:377–393
- Feynman R (1985) *Surely you're joking, Mr. Feynman!*. W. W. Norton & Company, New York
- Fong EA, Wilhite AW (2017) Authorship and citation manipulation in academic research. *PLoS ONE* 12:e0187394
- Fortunato S, Bergstrom CT, Borner K, Evans JA, Helbing D, Sa M, Petersen AM, Radicchi F, Sinatra R, Uzzi B, Vespignani A, Waltman L, Wang D, Barabasi A-L (2018) Science of science. *Science*. <https://doi.org/10.1126/science.aao0185>
- Frackowiak R (2014) Defending the grand vision of the human brain project. *N Sci* 2978
- Frackowiak R, Markram H (2015) The future of human cerebral cartography: a novel approach. *Philos Trans R Soc B Biol Sci* 370:20140171
- Frank C, Kennedy M, Goold C, Marek K, Davis G (2006) Mechanisms underlying rapid induction and sustained expression of synaptic homeostasis. *Neuron* 52:663–677
- Frankland PW, Josselyn SA (2016) In search of the memory molecule. *Nature* 535:41–42
- Gallistel CR, Matzel LD (2013) The neuroscience of learning: beyond the Hebbian synapse. *Annu Rev Psychol* 64:169–200
- Gannett L (2016) The human genome project. In: Edward N. Zalta (ed) *The stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/sum2016/entries/human-genome/>. Accessed Summer 2016
- Gettings P (1989) Emerging principles governing the operation of neural networks. *Ann Rev Neurosci* 12:185–204
- Glanzman D (1995) The cellular basis of classical conditioning in aplysia californica—it's less simple than you think. *TINS* 18:30–36
- Glanzman DL (2010) Common mechanisms of synaptic plasticity in vertebrates and invertebrates. *Curr Biol* 20:R31–R36
- Gold I (2009) Reduction in psychiatry. *Can J Psychiatry* 54:506–512
- Gold I, Roskies A (2008) Philosophy of neuroscience. In: Ruse M (ed) *The Oxford handbook of philosophy of biology*. Oxford University Press, Oxford
- Gold I, Stoljar D (1999) A neuron doctrine in the philosophy of neuroscience. *Behav Brain Sci* 22:809–869
- Gosgnach S, Lanuza GM, Butt SJB, Saueressig H, Zhang Y, Velasquez T, Riethmacher D, Callaway EM, Kiehn O, Goulding M (2006) V1 spinal neurons regulate the speed of vertebrate locomotor outputs. *Nature* 440:215–219
- Grant S, O'Dell T, Karl K, Stein P, Soriano P, Kandel E (1992) Impaired long-term potentiation, spatial learning, and hippocampal development in fyn mutant mice. *Science* 258:1903–1910
- Greenberg I, Manor Y (2005) Synaptic depression in conjunction with a-current channels promote phase constancy in a rhythmic network. *J Neurophysiol* 93:656–677

- Gregory R (1962) The logic of the localization of function in the central nervous system. In: Bernard E, Kare M (eds) *Biological prototypes and synthetic systems*. Plenum Press, New York, pp 51–53
- Gregory RL (2005) The Medawar Lecture 2001 knowledge for vision: vision for knowledge. *Philos Trans R Soc Lond B Biol Sci* 360:1231–1251
- Gross C (2009) Three before their time: neuroscientists whose ideas were ignored by their contemporaries. *Exp Brain Res* 192:321–334
- Hardcastle V, Stewart C (2002) What do brain data really show? *Philos Sci* 69:72–82
- Hardcastle VG, Stewart CM (2009) fMRI: a modern cerebrascope? The case of pain. In: Bickle J (ed) *The Oxford handbook of philosophy and neuroscience*. Oxford University Press, Oxford
- Hardt O, Nadel L (2018) Systems consolidation revisited, but not revised: the promise and limits of optogenetics in the study of memory. *Neurosci Lett* 680:54–59
- Hawkins R, Kandel E (1984) Is there a cell-biological alphabet for simple forms of learning? *Psychol Rev* 91:375–391
- Hawkins RD, Castellucci VF, Kandel ER (1981) Interneurons involved in mediation and modulation of gill-withdrawal reflex in *Aplysia*. I. Identification. *J Neurophysiol* 45:304–314
- Hickie C, Cohen L, Balaban P (1997) The synapse between LE sensory neurons and Gill motoneurons makes only a small contribution to the *Aplysia* Gill-withdrawal reflex. *Eur J Neurosci* 9:627–636. <https://doi.org/10.1111/j.1460-9568.1997.tb01411.x>
- Hoffman N, Parker D (2010) Lesioning alters functional properties in isolated spinal cord hemisegmental networks. *Neuroscience* 168:732–743
- Horgan J (1999) *The undiscovered mind*. Widenfeld and Nicholson, London
- Howard DA (2017) Einstein's philosophy of science. In: Edward N. Zalta (ed) *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/fall2017/entries/einstein-philsience>. Accessed Autumn 2017
- Hubel D (1974) Neurobiology: a science in need of a Copernicus. In: Neyman J (ed) *The heritage of Copernicus: theories "pleasing to the mind"*. MIT Press, Cambridge
- Hubel D (1995) *Eye, brain, and vision*. Henry Holt and Company, New York
- Ioannidis JPA (2012) Why science is not necessarily self-correcting. *Perspect Psychol Sci* 7:645–654. <https://doi.org/10.1177/1745691612464056>
- Jacobson M (1993) *Foundations of neuroscience*. Springer, New York
- Jeffrey K (1997) LTP and spatial learning—where to next? *Hippocampus* 7:95–110
- Jepma M, Wager TD (2013) Multiple potential mechanisms for context effects on pain. *Pain* 154:629–631
- Jia Y, Parker D (2016) Short-term synaptic plasticity at interneuronal synapses could sculpt rhythmic motor patterns. *Front Neural Circuits*. <https://doi.org/10.3389/fncir.2016.00004>
- Josselyn SA, Kohler S, Frankland PW (2015) Finding the engram. *Nat Rev Neurosci* 16:521–534
- Kandel E (1998) A new intellectual framework for psychiatry. *Am J Psychiatry* 155:457–469
- Kandel E (2001) The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 294:1030–1038
- Kandel ER, Dudai Y, Mayford MR (2014) The molecular and systems biology of memory. *Cell* 157:163–186
- Keith J, Rudy J (1990) Why NMDA-receptor-dependent long-term potentiation may not be a mechanism of learning and memory: reappraisal of the NMDA-receptor blockade strategy. *Psychobiology* 18:251–257
- Kiehn O, Kullander K (2004) Central pattern generators deciphered by molecular genetics. *Neuron* 41:317–321
- Koch C (2012) Modular biological complexity. *Science* 337:531–532
- Konkel A, Cohen NJ (2009) Relational memory and the hippocampus: representations and methods. *Front Neurosci* 3:166–174
- Krakauer JW, Ghazanfar AA, Gomez-Marín A, MacIver MA, Poeppel D (2017) Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93:480–490
- Kuenzi F, Fitzjohn S, Morton R, Collingridge G, Seabrook G (2000) Reduced long-term potentiation in hippocampal slices prepared using sucrose-based artificial cerebrospinal fluid. *J Neurosci Method* 100:117–122
- Kuhn T (1970) *The structure of scientific revolutions*. University of Chicago Press, Chicago
- Kuhn T (1977) The essential tension: tradition and innovation in scientific research? The essential tension. University of Chicago Press, Chicago, pp 225–239
- Kupferman I, Weiss W (1978) The command neuron concept. *Behav Brain Sci* 1:3–39
- Laing R (1985) *Wisdom, madness and folly: the making of a psychiatrist*. MacMillan, London

- LeBlancq MJ, McKinney TL, Dickson CT (2016) ZIP It: neural silencing is an additional effect of the PKM-zeta inhibitor zeta-inhibitory peptide. *J Neurosci* 36:6193–6198
- Lehrer J (2008) Can a thinking, remembering, decision-making, biologically accurate brain be built from a supercomputer? *Seed Mag* [http://seedmagazine.com/content/print/out\\_of\\_the\\_blue](http://seedmagazine.com/content/print/out_of_the_blue). Accessed Summer 2015
- Lewontin R (1993) *Biology as ideology: the doctrine of DNA*. Harper Collins, New York
- Lichtman JW, Sanes JR (2008) Ome sweet ome: What can the genome tell us about the connectome? *Curr Opin Neurobiol* 18:346–353
- Lichtman J, Smith S (2008) Seeing circuits assemble. *Neuron* 60:441–448
- Lipton P (2000) Inference to the best explanation. In: Newton-Smith W (ed) *A companion to the philosophy of science*. Blackwell, Oxford, pp 184–193
- Lipton P (2005) The truth about science. *Philos Trans R Soc B* 360:1259–1269
- Lisman J, Lichtman J, Sanes J (2003) LTP: perils and progress. *Nat Rev Neurosci* 4:926–929
- Looren de Jong H, Schouten MKD (2005) Ruthless reductionism: a review essay of John Bickle's *Philosophy and neuroscience: a ruthlessly reductive account*. *Philos Psychol* 18:473–486
- Machamer P, Darden L, Craver CF (2000) Thinking about mechanisms. *Philos Sci* 67:1–25
- Marban E, Yamagishi T, Tomaselli GF (1998) Structure and function of voltage-gated sodium channels. *J Physiol* 508:647–657
- Marr D (1982) *Vision*. Walter Freeman and Co, San Francisco
- McCabe DP, Castel AD (2008) Seeing is believing: the effect of brain images on judgments of scientific reasoning. *Cognition* 107:343–352
- McEachern J, Shaw C (1999) The plasticity-pathology continuum: defining a role for the LTP phenomenon. *J Neurosci Res* 58:42–61
- McKenzie S, Eichenbaum H (2011) Consolidation and reconsolidation: two lives of memories? *Neuron* 71:224–233
- Medawar P (1963) Is the scientific paper a fraud? *Listener* 70:377–378
- Medawar (1967) *The part of the soluble*. Methuen and Co., Ltd., London
- Meehl PE (1990) Appraising and amending theories: the strategy of Lakatosian defense and two principles that warrant it. *Psychol Inq* 1:108–141
- Midgely M (2004) *What action is. The myths we live by*. Abingdon, Routledge Classics, pp 68–80
- Moore DS (2001) The dependent gene: The fallacy of “nature vs. nurture”. Freeman, New York
- Morgan JL, Lichtman JW (2013) Why not connectomics? *Nat Methods* 10:494–500
- Morris R (1989) Synaptic plasticity and learning: selective impairment of learning rats and blockade of long-term potentiation in vivo by the N-methyl-D- aspartate receptor antagonist AP5. *J Neurosci* 9:3040–3057
- Morris R (2003) Long-term potentiation and memory. *Philos Trans R Soc Lond B Biol Sci* 358:643–647
- Morris R (2014) Brain project leaders need an open mind. *Nature* 511:292
- Moser E, Mathiesen I, Andersen P (1993) Association between brain temperature and dentate field potentials in exploring and swimming rats. *Science* 259:1324–1326
- Mott MC, Gordon JA, Koroshetz WJ (2018) The NIH brain initiative: advancing neurotechnologies, integrating disciplines. *PLoS Biol* 16:e3000066
- Nagel E (1961) *The structure of science: problems in the logic of scientific explanation*. Harcourt, Brace & World, San Diego
- Noble D (2002) *Modelling the heart: insights, failures and progress*. *BioEssays* 24:1155–1163
- Olsen S, Wilson R (2008) Cracking neural circuits in a tiny brain: new approaches for understanding the neural circuitry of *Drosophila*. *Trends Neurosci* 31:512–520
- Otchy T, Wolff S, Rhee J, Pehlevan C, Kawai R, Kempf A, Gobes S, Ölveczky B (2015) Acute off-target effects of neural circuit manipulations. *Nature* 528:358–363
- Parker D (2006) Complexities and uncertainties of neuronal network function. *Philos Trans R Soc B Biol Sci* 361:81–99
- Parker D (2010) Neuronal network analyses: premises, promises and uncertainties. *Phil Trans R Soc Lond B* 365:2315–2328
- Parker D (2018) *Kuhnian revolutions in neuroscience: the role of tool development*. *Biol Philos* 33:17
- Parker D, Zhang W, Grillner S (1998) Substance P modulates NMDA responses and causes long-term protein synthesis-dependent modulation of the lamprey locomotor network. *J Neurosci* 18:4800–4813
- Pearl J (2000) *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge
- Piccinini G, Craver C (2011) Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese* 183:283–311

- Poo M-M, Pignatelli M, TsJ R, Tonegawa S, Bonhoeffer T, Martin KC, Rudenko A, Tsai L-H, Tsien RW, Fishell G, Mullins C, Gonzalves JT, Shtrahman M, Johnston ST, Gage FH, Dan Y, Long J, Gr B, Stevens C (2016) What is memory? The present state of the engram. *BMC Biol* 14:40
- Posner M, Raichle M (1994) *Images of mind*. Scientific American Books, New York
- Putnam H (1975) *Philosophy and our mental life*. In: *Mind, language and reality*. Philosophical papers, vol 2. Cambridge University Press
- Queenan BN, TsJ R, Gazzaniga M, Gallistel CR (2017) On the research of time past: the hunt for the substrate of memory. *Ann N Y Acad Sci* 1396:108–125
- Railton P (1981) Probability, explanation, and information. *Synthese* 48:233–256
- Reichenbach H (1957) *The rise of scientific philosophy* 305
- Rilling M (1996) The mystery of the vanished citations: James McConnell's forgotten 1960s quest for planarian learning, a biochemical engram, and celebrity. *Am Psychol* 51:589–598
- Roberts L (1990) Genome backlash going full force. *Science* 248:804
- Robinson HR (1937) The charge of the electron. *Rep Prog Phys* 4:212
- Rojas-Libano D, Solar JWd, Aguilar-Rivera M, Montefusco-Siegmund R, Maldonado PE (2018) Local cortical activity of distant brain areas can phaselock to the olfactory bulb's respiratory rhythm in the freely behaving rat. *J Neurophysiol* 120:960–972
- Russell B (1915) *Our knowledge of the external world as a field for scientific method in philosophy*. Routledge
- Sagan C (1987) The burden of skepticism. *Skept Inq* 12:28–46
- Sanes J, Lichtman J (1999) Can molecules explain long-term potentiation? *Nat Neurosci* 2:597–604
- Schneidman E, Berry M, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012
- Selverston A (1980) Are central pattern generators understandable. *Behav Brain Sci* 3:535–571
- Sharma AV, Nargang FE, Dickson CT (2012) Neurosilence: profound suppression of neural activity following intracerebral administration of the protein synthesis inhibitor anisomycin. *J Neurosci* 32:2377–2387
- Silva A, Bickel J (2009) Science of research and the search for the molecular mechanisms of cognitive functions. In: Bickel J (ed) *Oxford handbook of philosophy and neuroscience*. Oxford University Press, New York, pp 91–126
- Silva A, Stevens C, Tonegawa S, Wang Y (1992) Deficient hippocampal long-term potentiation in alpha-calcium-calmodulin kinase II mutant mice. *Science* 257:201–206. <https://doi.org/10.1126/science.1378648>
- Simon H (1962) The architecture of complexity. *Proc Am Philos Soc* 106:467–482
- Smalheiser NR, Manev H, Costa E (2001) RNAi and brain function: Was McConnell on the right track? *Trends Neurosci* 24:216–218
- Soltesz I (2006) *Diversity in the neuronal machine*. Oxford University Press, Oxford
- Stuart D, Hultborn H (2008) Thomas Graham Brown (1882–1965), Anders Lundberg (1920–), and the neural control of stepping. *Brain Res Rev* 59:74–95
- Takeuchi T, Duszkiwicz AJ, Morris RGM (2014) The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos Trans R Soc B Biol Sci* 369:20130288
- Tononi G, Sporns O, Edelman G (1999) Measures of degeneracy and redundancy in biological networks. *Proc Natl Acad Sci* 96:3257–3262
- Tricklebank D, Singh L, Oles RJ, Preston C, Iversen SD (1989) The behavioural effects of MK-801: a comparison with antagonists acting competitively at the NMDA receptor. *Eur J Pharmacol* 167:127–135
- Trudeau L-E, Castellucci V (1993) Sensitisation of the gill and siphon withdrawal reflex of *Aplysia*: multiple sites of change in the neuronal network. *J Neurophysiol* 70:1210–1220
- Trudeau L-E, Castellucci V (1995) Postsynaptic modifications in long-term facilitation in *Aplysia*: upregulation of excitatory amino acid receptors. *J Neurosci* 15:1275–1284
- van Riel R, Van Gulick R (2018) Scientific reduction. In: Edward N Zalta (ed) *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/sum2018/entries/scientific-reduction>. Accessed Summer 2018
- Walker D, Gold P (1994) Intrahippocampal administration of both the D- and L-isomers of AP5 disrupt spontaneous alternation behavior and evoked potentials. *Behav Neural Biol* 62:151–162
- Weiss S, Faber D (2010) Field effects in the CNS play a functional role. *Front Neural Circuits* 4:1–10
- Woodward J (2003) *Making things happen*. Oxford University Press, New York

- Woodward J (2017a) Explanation in neurobiology: an interventionist perspective, James Woodward. In: Kaplan D (ed) Explanation and integration in mind and brain science. Oxford University Press, Oxford, pp 70–100
- Woodward J (2017b) Scientific explanation. In: Edward N. Zalta (ed) The Stanford encyclopedia of philosophy. <https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation>. Accessed Autumn 2017
- Yuste R (2008) Circuit neuroscience: the road ahead. *Front Neurosci* 2:6–9
- Zeki S (1993) A vision of the brain. Blackwell Scientific Publications, Oxford
- Zucker R, Regehr W (2002) Short-term synaptic plasticity. *Annu Rev Physiol* 64:355–405
- Zupanc GKH (2010) Behavioral neurobiology: an integrative approach, 2nd edn. Oxford University Press, Oxford

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.