



Philosophia Scientiæ

Travaux d'histoire et de philosophie des sciences

17-2 | 2013

The Mind–Brain Problem in Cognitive Neuroscience

What's Old Is New Again: Kemeny-Oppenheim Reduction at Work in Current Molecular Neuroscience

Kari Theurer and John Bickle



Electronic version

URL: <http://journals.openedition.org/philosophiascientiae/856>

DOI: 10.4000/philosophiascientiae.856

ISSN: 1775-4283

Publisher

Éditions Kimé

Printed version

Date of publication: 1 May 2013

Number of pages: 89-113

ISBN: 978-2-84174-631-6

ISSN: 1281-2463

Electronic reference

Kari Theurer and John Bickle, « What's Old Is New Again: Kemeny-Oppenheim Reduction at Work in Current Molecular Neuroscience », *Philosophia Scientiæ* [Online], 17-2 | 2013, Online since 27 June 2013, connection on 03 November 2020. URL : <http://journals.openedition.org/philosophiascientiae/856> ; DOI : <https://doi.org/10.4000/philosophiascientiae.856>

Tous droits réservés

What's Old Is New Again: Kemeny-Oppenheim Reduction at Work in Current Molecular Neuroscience

Kari Theurer

Trinity College (USA)

John Bickle

Mississippi State University (USA)

Résumé : Nous introduisons un nouveau modèle de réduction inspiré par celui de Kemeny et Oppenheim [Kemeny & Oppenheim 1956] et soutenons que ce modèle est opérationnel dans une conception « imputoyablement réductionniste » des neurosciences actuelles. Le modèle de Kemeny et Oppenheim a été rapidement rejeté dans la philosophie des sciences du milieu du xx^e siècle et remplacé par des modèles développés par Ernest Nagel et Kenneth Schaffner [Nagel 1961], [Schaffner 1967]. Nous pensons que le modèle de Kemeny et Oppenheim a été rejeté à juste titre, compte tenu de ce dont une « théorie du réductionnisme » devait rendre compte à l'époque. Mais leurs idées directrices à propos de ce qui constitue une réduction scientifique — l'accroissement de la portée explicative et la systématisation — reflètent les véritables pratiques des neurosciences réductionnistes actuelles. L'étape clé d'une réhabilitation visant à accorder leurs idées avec les détails de la science actuelle consiste à les reformuler à partir de travaux récents sur l'explication causale-mécanique.

Nous commençons avec une étude de cas scientifique, tirée du domaine relativement nouveau de la « cognition moléculaire et cellulaire ». Cette étude fournit une explication de l'effet d'espacement d'Ebbinghaus sur l'apprentissage et la mémoire, en terme d'interactions entre une protéine activatrice de la transcription et sa phosphatase inhibitrice de neurones recrutés dans la trace mnésique. Ensuite, nous décrivons brièvement certains modèles réductionnistes populaires de la philosophie des sciences du milieu du xx^e siècle. Nous remarquons comment ces modèles échouent à éclairer des caractéristiques principales de notre étude de cas scientifique. Finalement, nous présentons notre modèle causal-mécaniste qui est une mise à jour du modèle de Kemeny et Oppenheim et soutenons qu'il décrit d'une manière convaincante les éléments de notre

étude de cas scientifique. Nous terminons avec une remarque qui, avec un peu de chance, sapera la surprise que beaucoup peuvent ressentir d'apprendre qu'une explication philosophique longtemps rejetée du réductionnisme est en réalité à l'œuvre dans l'une des plus importantes entreprises réductionnistes de la science actuelle.

Abstract: We introduce a new model of reduction inspired by Kemeny and Oppenheim's model [Kemeny & Oppenheim 1956] and argue that this model is operative in a “ruthlessly reductive” part of current neuroscience. Kemeny and Oppenheim's model was quickly rejected in mid-20th-century philosophy of science and replaced by models developed by Ernest Nagel and Kenneth Schaffner [Nagel 1961], [Schaffner 1967]. We think that Kemeny and Oppenheim's model was correctly rejected, given what a “theory of reduction” was supposed to account for at that time. But their guiding insights about what constitutes scientific reduction—increases in explanatory scope and systematization—reflect actual practices of current reductionistic neuroscience. The key rehabilitative step to make their insights fit current scientific details is to restate them using resources from recent work on causal-mechanistic explanation.

We begin with a scientific case study, drawn from the relatively new field of “molecular and cellular cognition”. It provides an explanation of the well-known Ebbinghaus spacing effect on learning and memory in terms of interactions between a transcriptional enhancer protein and its inhibiting phosphatase in neurons recruited into the memory trace. Next we briefly describe some popular models of reduction from mid-20th-century philosophy of science. We point out how these models fail to illuminate key features of our scientific case study. Finally we present our causal-mechanistically updated Kemeny and Oppenheim-inspired model and argue that it nicely accounts for the details of our scientific case study. We close with a remark that will hopefully undercut the surprise many may feel to learn that a long-rejected philosophical account of reduction actually is at work in one of the most prominent reductionistic endeavors in current science.

1 A case study from recent “ruthlessly reductionistic” neuroscience

A recent exchange between William Bechtel [Bechtel 2009] and John Bickle [Bickle 2012] suggests that neuroscience may not speak with one voice about what reduction(ism) is. Different fields comprising the discipline may work with fundamentally different understandings of the concept. But it's difficult to miss the “ruthlessly” reductionistic tone of a mainstream part of the discipline for the past two decades. In their monumental textbook, *Principles of Neural Science* (4th ed.), Eric Kandel, James Schwartz, and Thomas Jessell remark:

This book [...] describes how neural science is attempting to link molecules to mind—how proteins responsible for the activities of individual nerve cells are related to the complexity of neural processes. Today it is possible to link the molecular dynamics of individual nerve cells to representations of perceptual and motor acts in the brain and to relate these internal mechanisms to observable behavior. [Kandel, Schwartz & Jessell 2001, 3–4]

The “today” they speak of was more than one decade ago, and the field of molecular and cellular cognition continues to do exactly what these authors describe: manipulate specific genes and proteins in central neurons (in animal models), and track the effects directly on behavior.

The most direct way into the field of molecular and cellular cognition is through detailed descriptions of its landmark experimental results. We describe one such result in this section, but at least 100 other case studies would illustrate equally well the reductionism that characterizes this field. Like most first-rate scientific research, our case study was built upon previous work: on the role of the intraneuronal cyclic adenosine monophosphate (cAMP)-protein kinase A (PKA)-cyclic AMP-responsive element-binding protein (CREB) pathway in inducing both late-phase long-term potentiation (L-LTP) in mammalian hippocampus and cortical neurons, and long-term memory consolidation.¹

Cyclic AMP is a classic second messenger of molecular biology, relaying signals from the cell surface to other locations in the cell. In one prominent form of L-LTP (in mammals), a form in which synaptic potentiation can endure for days up to weeks, high frequency activity in pre-synaptic neurons activate dopaminergic interneurons that synapse on G-protein coupled receptors on the post-synaptic target cell.² Dopamine binding to these receptors prime adenylyl and adenylate cyclase molecules, which convert adenosine triphosphate (ATP) molecules into cellular energy and cAMP. Increased cAMP molecules in the post-synaptic neuron bind to regulatory subunits of PKA molecules, freeing up

1. Consolidation is the process by which labile, easily disrupted short-term memories are converted into more stable long-term form.

2. We are about to inundate readers with a large dose of molecular neuroscience. This background is important for understanding the specific manipulations performed in the landmark experiment we'll describe below, and for understanding the rationale behind the experimental manipulations performed and the significance of the results. One of us (Bickle) often describes work at this level of neurobiological detail to develop philosophical points, and tends to aim at the excessive descriptive extreme rather than the deficient. Other philosophers of science are less patient with this amount of scientific detail. (We thank Colin Allen for calling our attention to this particular worry.) The key point to emphasize here is that the scientific background details matter to the argument to come. Besides, what's wrong with learning some details about state-of-the-art work in current neuroscience? For those who find the background details presented here too sparse to be comprehensible, see Bickle [Bickle 2003, chap. 2] for a more complete description and some additional helpful diagrams reprinted from the original publications.

catalytic PKA subunits to translocate to the post-synaptic neuron's nucleus. There PKA catalytic subunits phosphorylate CREB molecules, with phosphorylated (pCREB) α and δ isoforms serving as gene transcriptional enhancers. pCREB enhances gene transcription and subsequent protein synthesis for regulatory proteins (like ubiquitin carboxyl-terminal hydrolase), which keep catalytic PKA subunits unbound from regulatory subunits and active even after the transient rise in cAMP returns to baseline levels. It also enhances transcription for other transcriptional enhancers (like CCAAT enhancer binding protein), which activate gene transcription for structural proteins that reshape the post-synaptic density, anchoring once-hidden AMPA receptors into active sites. The result is a potentiated synapse with increased numbers of active AMPA receptors, poised now to deliver stronger excitatory post-synaptic potentials (EPSPs) upon subsequent pre-synaptic glutamate release, and ultimately higher probabilities of action potential generation in response to the inducing pattern of excitatory pre-synaptic neurotransmission.

Work from Alcino Silva's lab in the mid-1990s established long-term memory deficits in CREB "knock-out" mice [Bourtchouladze, Frenguelli, Blendy *et al.* 1994], i.e., genetically engineered mutant mice lacking the gene for CREB α and δ isoforms. These mice were deficient in remembering the location of the hidden platform in the Morris water maze task. More impressively, the mutants were intact on a short-term context-shock association task (compared to wild-type [non-mutated] littermate control mice), but were markedly deficient on the long-term version of the same task.³ This second experiment

3. The Morris water maze is a spatial memory task in which rodents are plunged into a pool of opaque liquid. Being hydrophobic, they try to escape the pool, to no avail, but eventually discover a platform underneath the surface of the opaque liquid they can climb on and take refuge. The pool is located in a room with variously shaped visually accessible displays at specific locations on the walls. On one standard protocol, the platform remains at the same location through numerous training trials, and mice are released into the pool at random locations on each trial. With repeated training trials, mice eventually learn the location of the platform relative to the visual displays, and quickly head for it regardless of where in the pool they are released. Reduction in time required to locate and stand on the platform is one standard measure of spatial memory for the platform's location vis-à-vis the visual clues. Another common measure is percentage of time spent in the quadrant where the platform had been located during training on "probe" trials in which the platform is removed.

Context-shock association involves exposing the rodent for a short exploratory period (e.g., 2 minutes) to a novel training context (e.g., a new cage environment). At the end of the exploratory period a significant aversive stimulus is delivered (typically a strong foot shock through a floor grid). The animal is removed from the context for a specified delay period (30 minutes to 1 hour for short-term memory testing, 24-48 hours for long-term memory testing), then returned to the environment. The common measure for context-shock fear memory is the amount of time the rodent spends "freezing" when returned to the context. Freezing is a stereotypic rodent fear response in which the animal hunches over, tucks its front paws under its chest, and ceases all movements except breathing. Nice features of context-fear conditioning are

implicated the cAMP-PKA-CREB pathway as a key molecular mechanism specifically for memory consolidation. Since that publication, experimenters have used a variety of more sophisticated techniques for intervening into CREB gene transcription, and a wide variety of behavioral memory tasks. A common pattern of results consistently has been found: intact short-term memory but impaired long-term memory on the same task in mice with reduced hippocampal and cortical CREB activity; and enhanced long-term memory in mice with increased CREB activity. Recently Han *et al.* have presented impressive evidence that pCREB functions directly in interneuronal competition for inclusion in memory traces for specific stimuli [Han, Kushner, Yi *et al.* 2007]. Neurons with more pCREB activity at the time of training are recruited with significantly higher probability into the memory trace for that stimulus than are those with less pCREB activity at training time.

pCREB activity itself is modulated in neurons by protein phosphatase-1 (PP1). PP1 when activated removes the phosphate group from CREB molecules, rendering them inactive as gene transcriptional enhancers. Both CREB and PP1 activity is neuron-activity dependent. The higher frequency of action potentials a neuron generates (up to a ceiling frequency), the higher will be its concentration of pCREB molecules available as transcriptional enhancers. But the same is true for CREB's inhibiting phosphatase, PP1. Furthermore, PP1 activity is modulated by another protein, Inhibitor 1 (I1). This pathway of intraneuronal interactions suggests the following molecular-genetic intervention strategy: increasing the level of active I1 molecules should decrease the activity of PP1, which in turn should increase the transcriptional activity of pCREB, which in turn should increase L-LTP at active synapses and generate better performance on the memory tasks known to involve pCREB. Furthermore, by blocking the normal interplay of pCREB and PP1 by activating more I1, experimenters could investigate the role that pCREB-PP1 interaction plays in normal neuronal function, and learning and memory.

David Genoux and his collaborators in Isabelle Mansuy's lab succeeded in constructing viable I1* mutant mice [Genoux, Haditsch, Knobloch *et al.* 2002]. They used a reversible tetracycline-controlled transactivator system so that the I1* transgene (inserted into embryonic stem cells) was only expressed when the engineered mice were dosed with doxycycline. Standard biochemical and molecular biological assays showed that the transgene was properly placed, and functional in both hippocampal and cortical neurons. The experimenters also linked the I1* transactivator system to a α -calmodulin kinase II (α CamKII) promoter, which limited transgene expression under doxycycline dosage principally to forebrain regions. Transgenic I1* mice were viable and showed no apparent phenotypic abnormalities.

that its long-term version is a hippocampus-dependent task that can be learned in a single trial, and it can be coupled with an amygdala-dependent fear conditioning task (e.g., Pavlovian tone-shock conditioning) to control for motivational and behavioral confounds of the genetic mutations [Abel, Nguyen, Barad *et al.* 1997].

For their behavioral protocol, Genoux and colleagues chose a much-studied rodent non-spatial object recognition task. I1* transgenic and wild-type littermate control mice were exposed to a variety of differently shaped mouse-sized objects presented alone (small toys, refuse objects, and the like) and were permitted to explore that object for a specified training time (more on the exact times utilized below). After a delay period, mice were exposed to a pair of objects, one from the training set previously encountered and another novel object not presented before. (Locations of previously presented and novel objects in the test apparatus were varied randomly to insure the non-spatiality of this memory test.) Mice are naturally curious, and when offered a choice will explore novel objects for greater durations than objects they remember encountering. So the exploratory discrimination time ratio during the testing phase, time exploring the novel object/summed total time exploring both objects, gives a useful measure of strength of memory for the previously presented object. Exploratory ratios $>50\%$ indicate some memory for the previously encountered object, since they indicate that the mouse spent more time exploring the novel object during the testing phase. Higher exploratory ratios indicate stronger memories for the previously presented object, since the mouse virtually ignored it during the testing phase.

Genoux and colleagues chose their time parameters to reflect the well-known Ebbinghaus spacing effect for learning and memory. This effect is so named because it was first explored systematically by German psychologist Hermann Ebbinghaus in the late 19th century using human memory for nonsense syllables. This effect is one thing we warn our undergraduate students to avoid: don't cram for the exam starting the night before! Learning and memory improves when training is divided into multiple sessions, with longer times (up to a limit) separating the learning sessions. In more technical terms, longer-interval distributed training produces better learning and stronger memories for training materials than does either brief-interval distributed training—shorter intervals separating the multiple learning sessions—or massed training—where all learning is crammed into a single session. One cognitive-level explanation for this effect is that, with frequent repetitive training (massed or brief-interval distributed), information acquired on the previous trial is still being processed when new input arrives, and that longer delays are required to complete input integration [Spear & Riccio 1994]. This explanation is not only limited in scope: why does learning and memory require “complete integration”? Why can't later stages of the memory process function on previously acquired information while earlier stages handle the new inputs? The cognitive-level story doesn't answer those crucial questions. This cognitive-level explanation also seems *ad hoc*—motivated solely to capture Ebbinghaus spacing data, but with no obvious generalization capacities or explanatory motivations beyond these data. As we've already stressed above, this is not true of the molecular-level explanation. It connects quite naturally with already-established molecular-mechanism work. So as Genoux *et al.* correctly assess, the Ebbinghaus spacing effect and

other “time-dependent constraints on learning and memory are poorly understood” [Genoux, Haditsch, Knobloch *et al.* 2002, 971].

For their I1* transgenic and wild-type control mice, Genoux and colleagues gave all groups 25 minutes training time with each object presented individually in the training set [Genoux, Haditsch, Knobloch *et al.* 2002]. The massed training group had all 25 minutes crammed into a single training session with each object. The briefer-interval distributed training group had their 25 minutes training with each object broken into 5×5 -minute training episodes, each separated from the next by a 5-minute non-training interval. The longer-interval distributed training group also had their 25 minutes training with each object broken into 5×5 -minute training sessions, but each training session was separated from the next by a 15 minute non-training interval. Mice were further divided into three control groups (wild-types, I1* transgenics not dosed with doxycycline [off dox], and I1* transgenics dosed previously with doxycycline but now off doxycycline due to normal drug metabolism [on/off dox]), and one experimental group (I1* transgenics on doxycycline [on dox]). Tests with pairs of a previously presented object and a novel object occurred 5 minutes after the last training session, 3 hours after, and 24 hours after. (See Figure 1a.)

Behaviorally, with all control groups, Genoux and collaborators got the standard Ebbinghaus spacing effect (see Figure 1b) [Genoux, Haditsch, Knobloch *et al.* 2002]. Exploratory ratios for control mice were statistically significantly higher only with longer-interval distributed training. I1* transgenics on dox, however, showed statistically significantly higher exploratory ratios with both briefer-interval distributed training and longer-interval distributed training, with no statistically significant differences between these two training regimens. In other words, increasing I1 to block PP1 to keep pCREB active during training episodes produces a kind of Ebbinghaus spacing effect super-learner: mice who remember encountering objects previously under briefer-interval distributed training with a strength wild-type mice can only achieve with longer-interval distributed training. Equally importantly, discrimination ratios returned to wild-type levels in I1* transgenics on/off dox (see Figure 1c), indicating that increased I1 activity prior to training did not improve memory performance. Exploration times during training sessions also did not differ between I1* transgenics on dox and all control groups (see Figure 1d). All groups spent statistically identical amounts of time during training exploring the individually presented objects.

This landmark MCC study presents evidence for an intriguing causal-mechanistic molecular explanation of the Ebbinghaus spacing effect. It suggests that the effect results from an intraneuronal interaction between activity-dependent pCREB and activity-dependent PP1. In massed training (and briefer-interval distributed training), action potential frequency in neurons recruited into the memory trace for the stimulus activates both, and the PP1 activity inhibits the pCREB activity, limiting L-LTP and memory (see Figure 2a). In longer-interval distributed training, action potential frequency

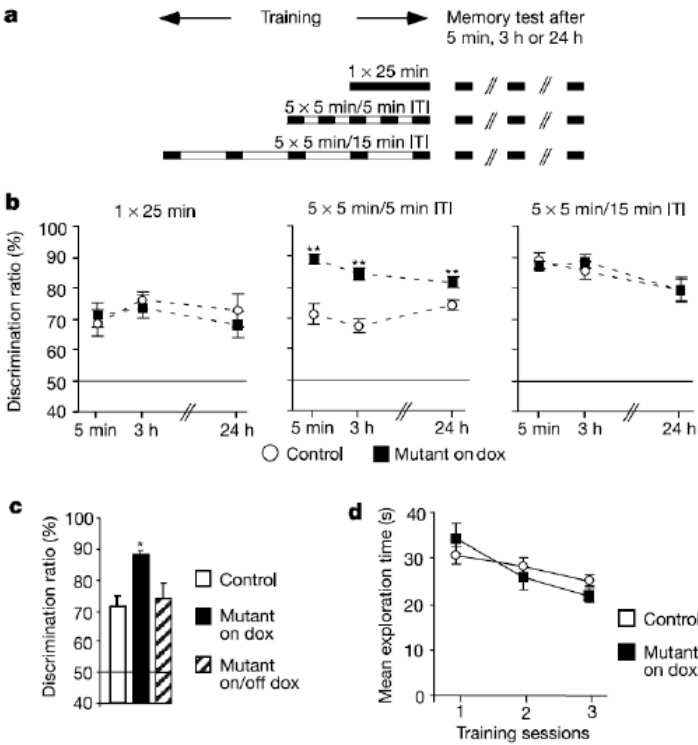


FIGURE 1: Results from [Genoux, Haditsch, Knobloch *et al.* 2002] experiment. (a) Training and testing regimens for massed (top row), briefer-interval distributed (middle row), and longer-interval distributed training (bottom row) groups. (b) Control mice showed significantly improved object recognition memory only during longer-interval distributed training, displaying the characteristic Ebbinghaus spacing effect. I1* mutant mice on dox showed significantly improved object recognition memory on both briefer-interval distributed training and longer-interval distributed training, displaying discrimination ratios on briefer-interval distributed training statistically identical to ratios displayed by control mice only on longer-interval distributed training. (c) This improvement was not present in I1* mutants dosed with doxycycline previously but off dox at training time due to normal drug metabolism. (d) This effect was not due to different exploratory times with individual objects during training sessions. See text for detailed explanation. Reprinted by permission from Macmillan Publishers Ltd: *Nature*, [Genoux, Haditsch, Knobloch *et al.* 2002, 971, Figure 1], ©2002.

in recruited neurons is diminished enough (by the longer intervals separating training sessions) to generate sufficient pCREB activity but dampened PP1 activity, generating strong L-LTP and stronger memory performance for the training stimulus (see Figure 2b). The I1* mutants on dox have engineered decreases in PP1 due to the additional activity of the I1* transgene and its protein products, enabling pCREB to have similar effects during briefer-interval distributed training that it only has during longer-interval distributed training in experimental control mice (see Figure 2c). But could Genoux *et al.* also show that pCREB activity in hippocampal and cortical neurons was similar in I1* mutants on dox during briefer-interval distributed training, compared to all groups during longer-interval distributed training [Genoux, Haditsch, Knobloch *et al.* 2002]? That result, against the backdrop of what was already known about pCREB as a molecular mechanism of certain forms of memory consolidation, would be strong additional evidence for the pCREB-PP1 interaction hypothesis as a molecular mechanism of the Ebbinghaus spacing effect.

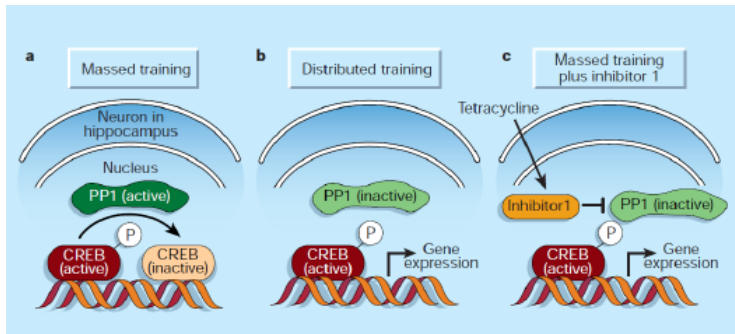


FIGURE 2: Learning and forgetting as a tale of molecular antagonism in neurons recruited into the memory trace. (a) During massed training (and briefer-interval distributed training), increased action potential frequency in neurons recruited into the memory trace for the stimulus activates both pCREB and its inhibiting phosphatase PP1, rendering CREB inactive, and producing little synaptic potentiation or learning. (b) During (longer-interval) distributed training, lessened action potential frequency in recruited neurons is sufficient to activate pCREB, but insufficient to activate PP1, producing long-lasting synaptic potentiation and enhanced learning and memory. (c) In I1* mutant mice on dox, increased I1 activity inactivates increased PP1 activity produced by massed or briefer-interval distributed training, producing long-lasting synaptic plasticity, and learning and memory, statistically identical to wild-type mice on longer-interval distributed training. Reprinted by permission from Macmillan Publishers Ltd: *Nature*, [Silva & Josselyn 2002, 930, Figure 1], ©2002.

Genoux *et al.* obtained this last result [Genoux, Haditsch, Knobloch *et al.* 2002]. They inserted a β -galactosidase gene whose expression and subsequent protein synthesis were tied to a sudden high concentration of pCREB in the neuron. In other words, neurons in these mice with high pCREB activity dur-

ing object recognition training will express a gene whose synthesized protein makes the neuron glow blue under standard light microscope observation. Mice from each of the groups were sacrificed after object recognition training, slices were cut through cortex and hippocampus, and highly pCREB-positive neurons (those that glowed blue) literally were counted. If increased pCREB activity due to lowered PP1 inhibition in longer-interval distributed training is an intraneuronal molecular mechanism of the Ebbinghaus spacing effect, then in [Genoux, Haditsch, Knobloch *et al.* 2002] various control groups we should see a statistically significant increase in high pCREB-neurons only in the longer-interval distributed training groups. In I1* transgenics on dox, however, we should see similar numbers of high pCREB-neurons in both briefer-interval distributed and longer-interval distributed training groups. That's exactly what Genoux *et al.* found (see Figure 3) [Genoux, Haditsch, Knobloch *et al.* 2002]. The number of neurons with high pCREB concentrations per mm³ more than doubled in control animals for only the longer-interval distributed training groups. The significant increases were found in I1* mutants on dox in both briefer-interval distributed and longer-interval distributed training groups.

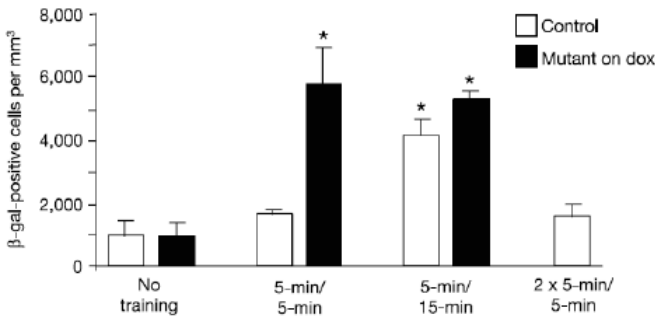


FIGURE 3: Control mice showed statistically significant increased numbers of neurons with high pCREB levels only on longer-interval distributed training. I1* mutant mice showed statistically identical increases on both briefer-interval distributed and longer-interval distributed training. See text for detailed discussion. Reprinted by permission from Macmillan Publishers Ltd: *Nature*, [Genoux, Haditsch, Knobloch *et al.* 2002, 972, Figure 3a], ©2002.

These data are impressive experimental evidence for an intriguing “ruthlessly reductive” hypothesis: The Ebbinghaus spacing effect, a paradigmatic psychological kind, reduces to the intraneuronal interactions between pCREB, a transcriptional enhancer already established to be a molecular mechanism of a variety of forms of memory consolidation, and PP1, its inhibiting phosphatase, in hippocampal and cortical neurons recruited into the memory trace. Genoux *et al.* established this hypothesis for the paradigmatic case of non-spatial hippocampus-dependent object recognition memory in mice [Genoux,

Haditsch, Knobloch *et al.* 2002]. The broader background of pCREB's already-established role in a variety of forms of memory consolidation, the ubiquity of CREB-PP1 interactions in all types of neurons (and in biological tissue types generally), and the comparative molecular genetics and proteomics of CREB and PP1 across species⁴ suggest a far broader application of this reductive hypothesis. And the principal experimental evidence offered in its support involves tracking the behaviors that experimentally operationalize the psychological kind, in light of precise experimental interventions that positively or negatively manipulate components of intraneuronal molecular signaling pathways.⁵ For it is the outcomes these mechanisms generate in central neurons whose effects are transmitted through the nervous system to the behavioral periphery, to generate the observed patterns of behavioral data.

Clearly the science here is impressive. But in our interpretation of what the science accomplished, do we have an instance of philosophers suggesting lessons that the scientists would resist? We do not. Genoux *et al.* clearly advocate having found a molecular mechanism for the Ebbinghaus spacing effect [Genoux, Haditsch, Knobloch *et al.* 2002]. They note that their results “show that protein phosphatase 1 (PP1) determines the efficiency of learning and memory by limiting acquisition and favoring memory decline” [Genoux, Haditsch, Knobloch *et al.* 2002, 970]. PP1 “regulates both the acquisition and retention of information” [Genoux, Haditsch, Knobloch *et al.* 2002, 974]. These are claims of a discovered molecular mechanism directly for a cognitive function, made by the scientists who performed the experiments. In a “News and Views” commentary published in the same issue, neurobiologists Alcino Silva and Sheena Josselyn drew equally bold conclusions. Studies of “the molecular and cellular foundations of cognitive processes” had now “come of age” with “the birth of a field that is unraveling the basis of learning, remembering, and now [...] forgetting” [Genoux, Haditsch, Knobloch *et al.* 2002,

4. The nucleotide sequence of a cloned *Aplysia CREB1* gene and the amino acid sequence of one of its predicted protein products, the CREB1a polypeptide isoform, displayed 95% amino acid sequence homology to mammalian CREB proteins—meaning that, across a gastropod and mammals, 19 out of every 20 amino acids residues in the protein sequences were identical [Bartsch, Casadio, Karl *et al.* 1998]. The key phosphorylation consensus site in the proteins' phosphorylation (P) boxes, the site where freed PKA catalytic subunits attach a phosphate group to activate the transcription enhancer, was completely conserved across *Aplysia CREB1a* and mammalian CREB, meaning that every amino acid residue in the P box sequences across species was identical. Based on these and other cross-species comparisons, insect biologists Josh Dubnau and Tim Tully, in a much-cited *Annual Review of Neuroscience* publication, wrote “elucidation of the role of CREB-mediated transcription in long-term memory in flies, LTP and long-term memory in vertebrates, and long-term facilitation in *A. californica* suggest that CREB may constitute a universally conserved molecular switch for long-term memory” [Bartsch, Casadio, Karl *et al.* 1998, 438; our emphasis].

5. For a detailed description of the full experimental methodology at work in MCC, see [Silva & Bickle 2009] for a chapter-length treatment, and [Silva, Landreth & Bickle in press] for the full book-length treatment.

929]. Genoux and colleagues “report that an enzyme known as protein phosphatase 1 (PP1) actively suppresses memories in mice” [Genoux, Haditsch, Knobloch *et al.* 2002, 929].

Like other biological processes, memory is regulated by the yin-and-yang-like interactions between molecules with opposing functions—in this case, protein phosphatases and kinases, which respectively remove and add phosphate groups on target proteins, thereby altering their properties. [Genoux, Haditsch, Knobloch *et al.* 2002, 929]

Notice the phrasing: memory is regulated by these molecules. CREB and PP1 are claimed to be memory’s mechanisms. These are the published interpretations of the scientists themselves, not of some reckless philosophical reductionists.

With a detailed example of current “ruthlessly reductive” neuroscience before us, let’s next assess how well some prominent models of reduction from the philosophy of science illuminate its key aspects.

2 Classic philosophical models of reduction

The still-dominant theory of reduction from the philosophy of science is Ernest Nagel’s model [Nagel 1961], according to which reduction is logical deduction (derivation) of the reduced theory (\mathbf{T}_R) from the reducing (\mathbf{T}_B). In most interesting scientific reductions, where the \mathbf{T}_B is more than just a simple extension of \mathbf{T}_R , \mathbf{T}_R will contain theoretical terms not contained in \mathbf{T}_B . ([Nagel 1961] called such cases “heterogeneous” reductions.) To affect anything other than a trivial derivation of \mathbf{T}_R in these cases, various statements must be added to \mathbf{T}_B to “connect” the disparate vocabularies. The logical status and ontological import of these ‘bridge principles’, ‘correspondence rules’, ‘coordinating definitions’, or ‘connectability conditions’ has been a central concern of the [Nagel 1961] model since its inception.

Paul Feyerabend quickly and famously took Nagel’s model to task, arguing that it did not accurately represent reduction in actual scientific practice [Feyerabend 1962]. This is because actual reductions do not proceed via logical deduction and it is unreasonable to reconstruct them as doing so. According to Feyerabend, Nagel’s condition of derivability is not met even in the supposedly paradigmatic reduction of Galilean to Newtonian mechanics. What is derivable from Newtonian mechanics is a set of laws that is empirically indistinguishable from those of Galilean mechanics—indistinguishable based on the observable evidence that forms the inductive basis for the latter. However, this derivable set is also logically inconsistent with actual Galilean mechanics. And the Galilean-to-Newtonian mechanics case is not an anomaly. It is precisely what we should expect from attempts to meet Nagel’s condition

of derivability. So the condition of derivability is an unreasonable constraint on explanation and reduction. Instead of its laws being logically derived from Newtonian mechanics (plus the appropriate correspondence rules), Feyerabend argued that the ontology of Galilean mechanics is instead eliminated in favor of that of Newtonian mechanics. Reduction is thus not a logical relationship between theories. It is the replacement of the reduced theory's ontology by the reducing theory's incommensurate alternative ontology.

Kenneth Schaffner [Schaffner 1967] incorporated key elements of Nagel's [Nagel 1961] model while acknowledging some of Feyerabend's lessons about corrective reductions in science. According to Schaffner's "general" model, it isn't the actual reduced theory \mathbf{T}_R that gets deduced in a reduction. Rather, what gets deduced is a corrected version of \mathbf{T}_R , \mathbf{T}_R^* , strongly analogous to \mathbf{T}_R . And Nagel's logically and ontologically problematic rules of correspondence for Schaffner became "reduction functions" from terms of \mathbf{T}_B onto the terms from \mathbf{T}_R^* missing in \mathbf{T}_B 's laws or generalizations. Reduction functions had the ontological import of synthetic identities between the referents of the terms from the reducing and corrected reduced theories. In this fashion, interesting scientific reductions show where and why the actual reduced theory \mathbf{T}_R worked as well as it did when it did, and where and how it was mistaken.

Does the Nagel (1961)-Schaffner (1967) model of reduction illuminate the type of reduction at work in MCC, documented in the detailed case study in section 1 above? It does not. Yet contra Feyerabend, Nagel's derivability requirement, even modulo Schaffner's \mathbf{T}_R^* supplement, is only a symptom of the real problem. The real problem is that in MCC reductions there are no theories to be derived from one another. The psychological "theory", the Ebbinghaus spacing effect, is simply a pattern of behavior in the training regimen and subsequent memory tests. The neurobiological "theory" is simply a description of a causal mechanism, the interaction between an activity-dependent transcriptional enhancer protein and its activity-dependent inhibiting phosphatase, the subsequent effects on gene transcription and protein synthesis in hippocampal and cortical neurons, and neural transmission effects out to the motor peripheries that generate the measured behavior. However it turns out to be that we best characterize what is going on in cases like this, reconstructing it as a logical derivation of one theory from another will do this science a disservice in two ways: derivation as the appropriate reduction relationship, and theories as the appropriate relata.

Fortunately there is a model of reduction from the philosophy of science, from the decade prior to Nagel [Nagel 1961] and Schaffner [Schaffner 1967], built on insights that do illuminate our neuroscience example. Paradoxically, it is a model that has been by and large forgotten in the philosophy of science except for its historical role: John Kemeny and Paul Oppenheim's model [Kemeny & Oppenheim 1956]. Unlike Nagel's (1961) model and Schaffner's (1967) revision, Kemeny and Oppenheim's (KO) model does not require any cross-theoretical correspondence rules or reduction functions. Early in the essay they insist: "It is not the case that the vocabulary of [reduced theory

$\mathbf{T_R}$] is in any simple way connected with the vocabulary of [reducing theory $\mathbf{T_B}$]” [Kemeny & Oppenheim 1956, 13].⁶ Their account only requires that the reducing theory explain or predict all of the observational data explained or predicted by the reduced theory.

Since KO reduction has been virtually forgotten in current philosophy of science, we elaborate on some details. A KO reduction has two basic requirements. First, it requires that $\mathbf{T_B}$ “fulfill the role” played by $\mathbf{T_R}$, meaning its explanatory role. $\mathbf{T_B}$ must explain all the observational data that $\mathbf{T_R}$ explained. Typically a $\mathbf{T_B}$ has greater explanatory scope—it explains more observational data than $\mathbf{T_R}$ explains. Second, $\mathbf{T_B}$ must be at least as well systematized as $\mathbf{T_R}$. For Kemeny and Oppenheim, reduction is a kind of scientific progress that increases both of these measures [Kemeny & Oppenheim 1956].

Systematization is a virtue of theories that balances simplicity with explanatory power. It is a measure of how well the increased complexity of $\mathbf{T_B}$ is compensated for by the increase in the amount of observational data it explains. While simplicity is an important virtue of theories, for Kemeny and Oppenheim typically it is outweighed by increased explanatory power. A simpler theory can be reduced to a more complex one if the more complex theory is significantly stronger in explanatory power. In such a case, it “would seem reasonable to allow it some additional complexity” [Kemeny & Oppenheim 1956, 11]. Kemeny and Oppenheim continue:

What our intuition tells us is that we must be satisfied that any loss in simplicity is compensated for by a sufficient gain in the strength of the body of the theories. We need some measure that combines strength and simplicity, in which additional complexity is balanced by additional strength. Let us express this combined concept by talking about how well a theory is systematized[...] We will then require that the reducing theory be at least as well systematized as the theory reduced. [Kemeny & Oppenheim 1956, 11]

6. Interestingly, Kemeny and Oppenheim [Kemeny & Oppenheim 1956] began their account with a criticism of Ernest Nagel’s account—[Nagel 1951], that is. According to Kemeny and Oppenheim, Nagel’s [Nagel 1951] account was logically too strong: it excluded actual cases of reduction that any account should include. They argued that this failure was unavoidable on any model of reduction that asserts direct connections between theoretical terms across $\mathbf{T_R}$ and $\mathbf{T_B}$ via connecting principles of any sort.

One further note: Kemeny and Oppenheim’s account is explicitly one of theory reduction. We’ve emphasized above that molecular and cellular cognition doesn’t contain theories in the standard philosophy of science senses. So obviously we’ll later be offering a Kemeny-Oppenheim-inspired account of reduction, modified to relate something other than theories in the sense that their original account proposes. That will be the topic of section 3 below. Thanks to Colin Allen for pointing out this potential worry about our goal.

In other words, \mathbf{T}_B 's increased complexity must be balanced by a similar increase in explanatory power, as compared to the ratio we find in \mathbf{T}_R . However, Kemeny and Oppenheim never develop this notion in formal detail. As Schaffner noted a decade later, "the notion is apparently an intuitive one" [Schaffner 1967, 140].

We stress two features of KO reduction that are especially salient for our argument. First, any "connection" between \mathbf{T}_B and \mathbf{T}_R is indirect. Speaking explicitly of Nagel (1951), Kemeny and Oppenheim acknowledge that any attempt to establish a "direct connection" between two theories will require translation of the vocabulary of one into the other. But, they argue, "it is entirely possible that a theory should be able to explain all facts that another can, without there being any method of translation" [Kemeny & Oppenheim 1956, our emphasis]. Notice that in an important sense their criticism anticipated Feyerabend's [Feyerabend 1962] better-known attack on Nagel's [Nagel 1961] conditions of derivability and connectivity. A second feature of KO reduction is that, although it was formulated within a broadly logical empiricist conception of science, it is less dependent on some central—and no longer widely accepted—details of the logical empiricist program than is the Nagel/Schaffner model.

The Nagel/Schaffner model relies heavily on the now-widely dismissed Deductive-Nomological (D-N) model of explanation.⁷ The D-N model is not only at work in their presentations of stock examples of reductionist explanations (e.g., the free-fall of bodies near the earth's surface, mercury's expansion in a thermometer). Nagel and Schaffner both assume the D-N model in a more fundamental way. A Nagel/Schaffner reduction is itself a species of D-N explanation, where the \mathbf{T}_B is part of the explanans and the \mathbf{T}_R or \mathbf{T}_R^* is the explanandum. A successful Nagel/Schaffner reduction—a valid derivation of \mathbf{T}_R or \mathbf{T}_R^* from \mathbf{T}_B and correspondence rules or reduction functions—demonstrates that the reduced theory is a logical consequence of the reducing theory. In this way, and directly in keeping with the D-N model, \mathbf{T}_B explains \mathbf{T}_R or \mathbf{T}_R^* . Separating Nagel's and Schaffner's guiding intuitions about what a reduction is from the broader philosophy of science it was expressed within is thereby not an easy task. Without the D-N model's backing, why else think that reduction involves logical deduction of the reduced (corrected or not) from the reducing theory?

Kemeny and Oppenheim's motivating insight—that reduction is a kind of scientific progress that involves increased explanatory scope and systematization—does not rely on the D-N account for its principal motivation.

7. The D-N model goes back to Carl Hempel and Paul Oppenheim [Hempel & Oppenheim 1948], reprinted in [Hempel 1965]. According to this view, explanations are arguments, with the statement to be explained (the 'explanandum') derived validly (hence 'deductive') from a set of premises (the 'explanans'), at least one of which is a scientific law (hence 'nomological'). Hempel and others added numerous additional conditions on the basic D-N model to try to account for a variety of different kinds of scientific explanations.

True, they did express their motivating insights using concepts from the logical empiricist program, including D-N explanation. But one can hold these insights about what reduction is and express them using a variety of different models of scientific explanation. The independence of their insights from the concepts they happened to use to express them makes their model of reduction more easily adaptable to fit within, for example, the recently revived framework of causal-mechanistic explanation.

Before we develop that account, we hope the reader feels the intuitive pull of where we're headed. Psychology's contribution to the case study in section 1 above, the Ebbinghaus spacing effect, is a pattern of behavior in different training regimens and subsequent memory tests. Our case study described direct experimental evidence for the hypothesis that the Ebbinghaus spacing effect reduces to the interactions between a transcriptional enhancer molecule (pCREB) and its inhibiting phosphatase (PP1) in central (hippocampal and cortical) neurons recruited into the memory trace during training, and suitably connected to the motor periphery to affect the measured behavioral patterns. The evidence for the reductive hypothesis is that manipulating the hypothesized molecular mechanism, suitably embedded in the larger nervous and motor systems, generates—and thereby causal-mechanistically explains—the behavioral data. This molecular mechanistic explanation both increases explanatory scope and systematization: exactly the features that KO reduction emphasized. Let's now turn to some details of the account.

3 KO reduction and causal-mechanistic explanation

Kemeny and Oppenheim's account of reduction was dismissed quickly [Kemeny & Oppenheim 1956]; consensus in mid-20th-century philosophy of science quickly settled on the stronger conditions on reduction offered by Nagel [Nagel 1961] and Schaffner [Schaffner 1967]. But in light of the nature of reductions being developed in current mainstream neuroscience, illustrated by our case study and not illuminated by the Nagel/Schaffner model, we think it is time to reconsider the principal insights of KO reduction. However, doing this will require reconceiving those insights in more contemporary fashion. Our strategy will be to employ some resources developed in current work on causal-mechanistic explanation.⁸ We'll then argue that the resulting KO-inspired model captures the kind of reductionism illustrated by our case study, better than any other reduction model in the philosophy of science.

A mechanistic KO reduction will involve reducing one level of mechanism to another by demonstrating that the lower-level mechanism has an explanatory

8. Kari Theurer developed this KO-inspired account of reduction in her recently defended doctoral dissertation [Theurer 2012].

scope at least as great as that of the higher-level mechanism, and that the resulting trade-off in complexity of the lower-level mechanism is balanced by this increase in explanatory scope. That is, the lower-level mechanism must explain all the data that the higher-level mechanism explains, and typically more, and the increased complexity of these lower-level explanations must not exceed this increased explanatory scope. In keeping with recent work in “new mechanistic” philosophy of science, we conceive the levels-ordering relation in terms of containment, component, or parthood: the parts or components of a mechanism, performing their activities in organized fashion to generate the mechanism’s behavior, are at the next level down from the mechanism itself. And those parts or components may themselves be mechanisms which have active, organized parts. Those latter parts or components are at the next mechanistic level down.⁹

Mechanistic explanation has seen an important revival over the past two decades. Anti-reductionism is prominent among some influential new mechanists (e.g., [Craver & Kaplan 2011]; but see [Bechtel 2009] for a different view). But connections between mechanism and reduction were prominent throughout 20th-century philosophy of science. Some of the strongest defenders of law-based explanation recognized the importance of reduction via elucidation of lower-level mechanisms, and they were not alone. Nearly two decades before the start of mechanism’s current revival, William Wimsatt noticed these attempts as a genuine kind of reduction:

At least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms, rather than explaining theories by deriving them or reducing them to other theories, and this is seen as reduction, or as integrally tied to it. [Wimsatt 1976, 671; our emphasis]

Craver himself has made a related point:

Wimsatt [Wimsatt 1976] and Sarkar [Sarkar 1992] recommend that the term “reduction” be used to describe the explanation of a phenomenon by its mechanism. Peter Smith [Smith 1992] suggests that the reductive ideal can be maintained by appeal to “modest reductions”. Many scientists use the term “reduction” in this way. [Craver 2007, 18, fn 19]

Craver even suggests that a mechanistic explanation of a phenomenon is “weakly reductionistic” [Craver 2007]. However, the KO-inspired model we’re developing here is stronger than what is indicated by most accounts of mechanistic explanation. The reason for our model’s increased strength has to do

9. See [Craver 2007, chap. 5]. Craver, for the now-standard account of levels of mechanisms, with useful diagrams and applied directly to neuroscience examples. Craver usefully points out, against models of *intertheoretic* reduction, that levels of mechanisms don’t always match up well with levels of theory. We agree with him on this point.

with one of the central KO insights: increased explanatory power of lower-level mechanisms in reductive causal-mechanistic explanations.

In the cases that interest us, as we descend downward through nested mechanisms, opening black boxes at progressively lower levels along the way, we are uncovering mechanisms that are progressively more explanatory. That is, we are elucidating mechanisms that allow us to account for increasingly more observational data. Crucially, in these cases, mechanisms at lower levels provide explanatory information that is unavailable at higher levels. Lower-level explanations of cognitive functions of the sort exemplified in molecular and cellular cognition are paradigmatic examples of this kind of reductive causal-mechanistic explanation. As we saw in our discussion in section 1 above, psychological and cognitive level mechanistic explanations of the Ebbinghaus spacing effect (such that frequent repetitive inputs interfere with complete integration of information acquired on previous trials) are severely limited in explanatory scope. They seem to be *ad hoc* add-ons designed simply to capture the specific behavioral data characterizing the effect.¹⁰ Discovering that the causal mechanisms of this effect involve the intraneuronal interactions between an activity-dependent transcriptional enhancer and its inhibiting phosphatase in the neurons recruited into the memory trace embeds a causal-mechanistic explanation for the Ebbinghaus effect into the full panoply of explanations from molecular biology involving these molecules.

This constitutes a huge gain in explanatory scope. CREB and PP1 aren't specific to cortical and hippocampal neurons, nor even to neurons. They interact to a variety of effects in a wide range of biological tissues. The discovery that their interaction causally-mechanistically explains the Ebbinghaus spacing effect now enables us to compare that effect to other biological processes mechanistically explained by CREB-PP1 interactions. The explanatory scopes of the best causal-mechanistic cognitive, even cognitive-neuroscientific explanations are, by contrast, greatly limited in scope. They apply only to specific cognitive systems, or to specific brain regions, or to neurofunctional constructs. Typically they explain only the behavioral data they are designed to accommodate. Finding increasingly lower-level mechanisms for higher-level cognitive effects brings all the cellular physiology or molecular biology known about the mechanism into the explanatory fold, along with the behavioral data for the specific cognitive function. In our example, the CREB-PP1 mechanism for the Ebbinghaus spacing effect now also accommodates the addition and removal of

10. This is not always true, as Colin Allen reminded us. More sophisticated kinds of cognitive, including cognitive-neuroscientific modeling, often make connections with other results, independently of the specific phenomena directly at explanatory issue. There is also the familiar point that the greater detail offered by the lower-level mechanism can obscure explanatory patterns that are more visible at higher levels. But a mechanized Kemeny-Oppenheim-inspired account of reduction isn't blind to these concerns. They're the very concerns that go into considerations of systematicity: the trade-off between increased explanatory scope and increased explanatory complexity. We elaborate on this point a bit more toward the end of this section.

phosphate groups from the mechanism's components, configurational changes in the proteins due to these additions and removals, and connections that the resulting form of late-phase synaptic plasticity has with CREB-PP1 interactions in other kinds of cells.¹¹ The Genoux results immediately suggest looking into the role of I1 as a natural inhibitor of cognitive forgetting and of PP1 as a potential mechanism of cognitive decline with aging. That's massively increased explanatory scope, actual and potential, over what even a developed psychological or cognitive-neuroscientific mechanism for the Ebbinghaus spacing effect could provide.

What about the second requirement of a KO reduction, systematization? Does expressing KO reduction insights using resources from causal-mechanistic explanation give us any way to assess whether our hypothesized reductive explanations are growing increasingly systematized? Two resources might be helpful. The first is the distinction among a complete description of a mechanism, a mechanism sketch, and a mechanism schema. A sketch is an incomplete model of a mechanism that leaves gaps in characterizing its entities, activities, or their organization. These gaps are marked by black boxes and filler terms. Peter Machamer, Lindley Darden, and Carl Craver (MDC) tell us that

a sketch [...] serves to indicate what further work needs to be done in order to have a mechanism schema [...] [and can serve] as an abstraction that can be instantiated as needed for the tasks mentioned above, e.g., explanation, prediction, and experimental design. [Machamer, Darden & Craver 2000, 18]

Mechanism schemata lie on a spectrum between a complete description of a mechanism and a sketch. A schema contains more explanatory information than does a sketch, and schemata closer to complete descriptions contain more than those further away.

Psychological and cognitive-neuroscientific mechanistic explanations of a cognitive function are inevitably sketches. Black boxes and filler terms abound. Cellular and molecular mechanistic explanations, when available, are at least steps toward schema. Some of those black boxes and filler terms get opened and filled. This is one kind of increase in explanatory scope that can be compared, at least intuitively, with the added complexity of molecular mechanistic explanations.

Craver provides the second potentially useful dimension for our purpose [Craver 2007]. He distinguishes between "how-possibly" and "how-actually"

11. Based on some of the earliest results in the search for cellular and molecular mechanisms of learning and memory, Eric Kandel recommended that we "conceive of learning as [...] a late stage of neuronal differentiation" [Kandel 1979, 76]: in other words, as one of the developmental processes that eventually individuate neurons from one another! Learning theory joins up with developmental neuroscience to increase the explanatory scope of both. Who knew?! Certainly nobody who rested content with psychological or even cognitive-neuroscientific mechanistic explanations of learning and memory!

mechanistic models. How-possibly models “are only loosely constrained conjectures about the sort of mechanism that might suffice to produce the explanandum phenomenon. They describe how a set of parts and activities might be organized together such that they exhibit the explanandum phenomenon” [Craver 2007, 112]. How-possibly models “are often heuristically useful in constructing and exploring the space of possible mechanisms, but they are not adequate explanations” [Craver 2007, 112]. How-actually models lie at the other end of a spectrum, and “describe real components, activities, and organizational features of the mechanism that in fact produces the phenomenon” [Craver 2007, 112]. How-plausibly models lie between these extremes. An experimentally justified case can be offered that some of the hypothesized components, activities, and organization specified by the how-plausibly mechanism exist, and could function to generate the behavior to be explained.

For Craver, a mechanistic explanation’s location on both of these spectra carries normative implications:

Progress in building mechanistic explanations involves movement along both the possibly-plausibly-actually axis and along the sketch-schema-mechanism axis. [Craver 2007, 114]

Progress in neuroscience involves elucidating mechanisms. This process requires showing that possible mechanisms are plausible, even actual, and filling in black boxes and filler terms in mechanism sketches to produce schemata, and ideally complete descriptions. These activities usually require elucidating other mechanisms. In “ruthlessly reductive” scientific fields like MCC, these other mechanisms are inevitably at a lower level than those of psychology, cognitive science, even cognitive neuroscience. By both of Craver’s measures, successful results of this process will yield increased explanatory information: a schema closer to a complete description of a mechanism, a how-plausibly, maybe even a how-actually mechanism. This is what we saw demonstrated in our detailed MCC case study in section 1, and it is exactly the process of effecting a mechanized KO reduction. Those same intuitions about what reduction is seem clearly at work in “ruthlessly reductionistic” MCC.

The idea of an explanation sketch is, in fact, another old reductionist idea. Hempel, for example, describes an explanation sketch as such:

What the explanatory analyses of historical events offer, is, then, in most cases not an explanation in one of the senses indicated above, but something that might be called an *explanation sketch*. Such a sketch consists of a more or less vague indication of the laws and initial conditions considered as relevant, and it needs “filling out” in order to turn into a full-fledged explanation. This filling-out requires further empirical research, for which the sketch suggests the direction. [Hempel 1965, 238]

Craver does not locate the roots of his appeal to ‘mechanism sketch’ in Hempel. But the similarities apparent in the quotes above, including the normative assumptions about better explanations and progress in science, are further representative of the shared historical roots of reductionism and mechanism. “New” mechanism, by the nature of its key resources, may be far more reductionistic than some of its proponents notice (or admit).

So, do MCC mechanistic explanations, reductive in the fashion described by our mechanized KO model, increase explanatory systematization? They certainly increase explanatory scope, both intuitively and along the two dimensions emphasized by the new mechanists. But do these increases override their added “complexity”? Any answer to this question is difficult to assess. Explanatory complexity is an elusive concept, difficult to measure and even more difficult to compare with an equally elusive measure like increased explanatory scope. Explanatory systematization remains very much in the eye of the beholder. But we shouldn’t confuse complexity with unfamiliarity. MCC experiments and causal-mechanistic explanations are unfamiliar to most philosophers and cognitive scientists. The molecular biology and genetics is typically daunting, especially when expressed in compressed scientific-journal form. But did you follow the experiments as described in section 1 above? Did you understand the results and the mechanistic explanation of the Ebbinghaus spacing effect built on them? When expressed in less compressed form, is the explanatory complexity of the reductive explanation really so daunting? When enough of the molecular biology is revealed, is the increase in explanatory complexity really so great? So much that it overcomes the obvious increases in explanatory scope?

One final point: A KO-inspired account of causal-mechanistic reduction allows cognitive functions to reduce directly to molecular mechanisms. So long as the hypothesized mechanism explains all the behavioral data, the psychological kind reduces directly to the molecular components, activities, and organization. Intervening levels of mechanisms aren’t required, although of course they’re often developed in the search for increasingly lower-level mechanisms. Our KO-inspired model thus differs from Bechtel’s [Bechtel 2009] “mechanistic reduction” model, which advocates “nested mechanisms within mechanisms” and limits reductions of a phenomena only to its mechanisms at the next level down. We don’t find these limitations in actual MCC research (as illustrated by our case study in section 1). So while Bechtel’s “mechanistic reduction” might correctly characterize what reduction is in cognitive/systems/behavioral-level neuroscience, we think our KO-inspired account better captures reduction on the hoof in really reductionistic neuroscientific fields.¹²

12. Bickle has argued for the direct reduction of cognitive functions to cellular-physiological and molecular-biological mechanisms since his [Bickle 2003, 2006]. Theurer argues that, due to the very notions that define new mechanism, new mechanists are likewise committed to “direct” reductions of psychological kinds to their cellular and molecular mechanisms [Theurer 2012].

4 Conclusion: Should we be surprised?

KO reduction quickly gave way in mid-20th-century philosophy of science to “direct” reduction paradigms proposed by Nagel [Nagel 1961] and Schaffner [Schaffner 1967]. Schaffner’s dismissal of KO reduction was a popular complaint. After giving his revised Nagel derivability condition involving “corrected” \mathbf{T}_R^* and his “strong analogy” requirement (see section 2 above), Schaffner insisted that when these conditions fail to obtain,

then one might have to fall back on something like the Kemeny-Oppenheim paradigm [...] There is no theory relation in this case—only adequate explanation of the observable predictions of the previous theory. [Schaffner 1967, 145; our emphases]

This criticism had force when intertheoretic reduction was paradigmatic of scientific reduction—and not coincidentally, when physics still ruled as the paradigmatic science for philosophical reflection. As the “new” mechanists have rightly emphasized, biology, including neuroscience, is less concerned with theory construction and more concerned with discovering mechanisms. Despite the initial surprise that an old, widely discredited account of reduction is actually at work in a “ruthlessly reductionistic” field central to mainstream current neuroscience, on reflection perhaps our thesis isn’t so surprising. Theory reduction is certainly less operative in neuroscience than in physics. Neuroscience doesn’t trade in theories the way physics does. Shouldn’t we expect, then, that something other than an implicit model of intertheoretic reduction guides its reductionistic proclivities? Kemeny and Oppenheim’s old “indirect” reduction paradigm, correctly criticized by mid-20th-century philosophers of science for failing to account adequately for theory reduction, seems nicely suited to illuminate neuroscience’s continued reductionistic search for lower-level mechanisms of cognition.

Bibliography

- ABEL, T., NGUYEN, P., BARAD, M., DEUEL, T., KANDEL, E. R. & BOURTCHOULADZE, R.
1997 Genetic demonstration of a role for PKA in the late phase of LTP and in hippocampus-based long-term memory, *Cell*, 88, 615–626.
- BARTSCH, D., CASADIO, A., KARL, K. A., SERODIO, P. & KANDEL, E. R.
1998 CREB encodes a nuclear activator, a repressor, and a cytoplasmic modulator that form a regulatory unit critical for long-term facilitation, *Cell*, 95, 211–223.

BECHTEL, W.

- 2009 Molecules, systems, and behavior: Another view of memory consolidation, in *Oxford Handbook of Philosophy and Neuroscience*, edited by BICKLE, J., New York: Oxford University Press, 13–39.

BICKLE, J.

- 2003 *Philosophy and Neuroscience: A Ruthlessly Reductive Account*, Dordrecht: Springer.
- 2006 Reducing mind to molecular pathways: Explicating the reductionism implicit in current mainstream neuroscience, *Synthese*, 152, 411–434.
- 2012 A brief history of neuroscience's actual influence on mind–brain reductionism, in *New Perspectives on Type Identity*, edited by GOZZANO, S. & HILL, C., Cambridge: Cambridge University Press, 88–110.

BOURTOULADZE, R., FRENGUELLI, B., BLENDY, J., CIOFFI, D., SCHUTZ, G. & SILVA, A. J.

- 1994 Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element binding protein, *Cell*, 79, 59–68.

CRAVER, C.

- 2007 *Explaining the Brain*, New York: Oxford University Press.

CRAVER, C. & KAPLAN, D.

- 2011 Toward a mechanistic philosophy of neuroscience: A mechanistic approach, in *Continuum Companion to the Philosophy of Science*, edited by FRENCH, S. & SAATSI, J., New York: Continuum Press, 268–292.

FEYERABEND, P. K.

- 1962 Explanation, reduction, and empiricism, in *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, vol. 3, 28–97.

GENOUX, D., HADITSCH, U., KNOBLOCH, M., MICHALON, A., STORM, D. & MANSUY, I. M.

- 2002 Protein phosphatase 1 is a molecular constraint on learning and memory, *Nature*, 418, 970–975.

HAN, J.-H., KUSHNER, S. A., YI, A. P., COLE, C. A., MATYNIA, A., BROWN, R. A., NEYE, R., GUZOWSKI, J. F., SILVA, A. J. & JOSSELYN, S. A.

- 2007 Neural competition and selection during memory formation, *Science*, 316, 457–460.

HEMPEL, C. G.

- 1965 *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.

- HEMPEL, C. G. & OPPENHEIM, P.
1948 Studies in the logic of explanation, *Philosophy of Science*, 15, 135–175.
- KANDEL, E. R.
1979 *Cellular Insights into Behavior and Learning*, New York: Academic Press.
- KANDEL, E. R., SCHWARTZ, J. R. & JESSELL, T. (EDS.)
2001 *Principles of Neural Science*, New York: McGraw-Hill, 4th ed.
- KEMENY, J. & OPPENHEIM, P.
1956 On reduction, *Philosophical Studies*, 7, 6–19.
- MACHAMER, P., DARDEN, L. & CRAVER, C.
2000 Thinking about mechanisms, *Philosophy of Science*, 67, 1–25.
- NAGEL, E.
1951 Mechanistic explanation and organismic biology, *Philosophy and Phenomenological Research*, 11(3), 327–338.
1961 *The Structure of Science*, New York: Harcourt, Brace and World.
- SARKAR, S.
1992 Models of reduction and categories of reductionism, *Synthese*, 91, 167–194.
- SCHAFFNER, K. F.
1967 Approaches to reduction, *Philosophy of Science*, 34, 137–147.
- SILVA, A. J. & BICKLE, J.
2009 The science of research and the search for molecular mechanisms of cognitive functions, in *Oxford Handbook of Philosophy and Neuroscience*, edited by BICKLE, J., New York: Oxford University Press, 91–126.
- SILVA, A. J. & JOSSELYN, S. A.
2002 Cognitive neuroscience: The molecules of forgetfulness, *Nature*, 418, 929–930.
- SILVA, A. J., LANDRETH, A. W. & BICKLE, J.
in press *Engineering Future Revolutions in Neuroscience*, New York: Oxford University Press.
- SMITH, P.
1992 Modest reductions and the unity of science, in *Reduction, Explanation and Realism*, edited by CHARLES, D. & LENNON, K., Oxford: Clarendon Press, 19–43.
- SPEAR, N. E. & RICCIO, D. C.
1994 *Memory: Phenomena and Principles*, Needham Heights, MA: Allyn and Bacon.

THEURER, K.

2012 *Rethinking Reductionism: From 17th-century Mechanism to Contemporary Molecular Neuroscience*, Ph.D. thesis, Indiana University.

WIMSATT, W.

1976 Reductive explanation: A functional account, in *Boston Studies in the Philosophy of Science*, edited by MICHALOS, A. C., HOOKER, C. A., PEARCE, G. & COHEN, R. S., Dordrecht: Reidel, vol. 30, 671–710.