TOPiCS
TOPICS IN COGNITIVE SCIENCE

Marr and Reductionism

John Bickle[a,b]

[a]*Department of Philosophy and Religion, Mississippi State University*
[b]*University of Mississippi Medical Center*

Abstract

David Marr's three-level method for completely understanding a cognitive system and the importance he attaches to the computational level are so familiar as to scarcely need repeating. Fewer seem to recognize that Marr defends his famous method by criticizing the "reductionistic approach." This sets up a more interesting relationship between Marr and reductionism than is usually acknowledged. I argue that Marr was correct in his criticism of the reductionists of his time—they were only describing (cellular activity), not explaining (cognitive functions). But a careful metascientific account of reductionistic neuroscience over the past two decades reveals that Marr's criticisms no longer have force. Contemporary neuroscience now explains cognition directly, although in a fashion—causal-mechanistically—quite different than Marr recommended. So while Marr was correct to reject the reductionism of his day and offer an alternative method for genuinely explaining cognition, contemporary cognitive scientists now owe us a new defense of Marr's famous method and the advantages of its explanations over the type now pursued successfully in current reductionist neuroscience. There are familiar reasons for thinking that this debt will not be paid easily.

Keywords: Reductionistic approach; Electrophysiology; Molecular and cellular cognition; Convergent 3 analysis; Non-intervention experiments; Causal-mechanistic explanation

1. Introduction

David C. Marr's three-level methodology for cognitive science is so well known as to scarcely need restating. Fig. 1 (Marr, 1982, figure 1–4) succinctly illustrates "the different levels at which an information-processing device must be understood before one can be said to have understood it completely" (Marr, 1982, p. 24). For Marr, the

Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be physically realized?

Fig. 1. Marr's (1982) figure 1.4. Marr's caption reads: "The three levels at which any machine carrying out any information-processing task must be understood" (p. 25).

level of computational theory holds special importance because the computations the system computes depend more on the computational problems it must solve than on the hardware which implements those computations (pp. 27–29). So far, so familiarly Marr.

Marr prefaces this discussion, however, with more than 15 pages aimed at an alternative, competing methodology, the one that informed his own earlier work, namely, "the reductionistic approach." This alternative approach seeks explanations of cognition in the "electrophysiology ... of the brain" (Marr, 1982, p. 11). The importance of this prior discussion is missed, or neglected, by most proponents and opponents of Marr-style cognitive science (but see Eliasmith and Kolbeck, this volume). After his brief recounting of the history of reductionism, ending with Barlow's (1972) "neuron" doctrine of perception from just a few years prior, Marr admits that he was "fully caught up in this excitement" (pp. 13–14). He also tells us explicitly that it was "deeper reflections" (his own and others) on what exactly neurophysiology was providing that led him to realize that "something important was missing" (p. 15). A close reading of this typically neglected discussion makes clear that for Marr, the three-level methodology and the "importance of computational theory" rest upon the failure of reductionism to explain cognition. (Hardcastle and Hardcastle [this volume] likewise see the importance of this argument for Marr's levels.) His methodology is his specific suggestion about what else was needed.

Interestingly, Marr's reductionist critics have also missed the role of Marr's criticism of reductionism for his defense of his methodology. Mostly reductionists have simply challenged the computer metaphor of mind that clearly motivates Marr's specific program for advancing beyond reductionism; Patricia Churchland is typical. After reciting a brief history of neuroscientific research on learning and memory, spanning behavior, neural systems, structures, neurons, and macromolecules, Churchland notes: "It is simply not rewarding to sort out this research in terms of the tri-level computer analogy, nor is there any useful purpose to be served in trying to force a fit" (Churchland, 1986, p. 360). That Marr anticipated and sought to address this criticism himself (Marr, 1982, p. 27) typically goes unmentioned. Once we see the place of Marr's challenge to reductionism as a key step in his defense of his method, a much more interesting dialectic opens between the two camps. Perhaps Marr was correct in his challenge to the reductionism of his

time—perhaps the evidence its experiments uncovered did not and could not explain. Yet perhaps reductionistic neuroscience over the past 30 years has acquired both the experimental tools and designs that make genuine explanations of cognitive functions now available. And most interestingly, perhaps these new reductionist explanations are of a different variety than the ones Marr recommended for cognitive science. This metascientific outcome would both undercut Marr's own explicit defense of his method and leave Marr-inspired cognitive scientists with the task of defending their method and the kind of explanations it seeks against a genuinely explanatory reductionism. This is the argument I will develop in this short paper.

2. Marr on describing versus explaining

My argument requires that we get clear on Marr's exact challenge to reductionism and its importance for his defense of his methodology. I have already mentioned his explicit admiration for early neurophysiological work on vision. He was especially impressed with technological developments in signal amplification and electrode size that led to the concept of a sensory neuron's receptive field—the region of modality-specific space and the features of the stimulus that maximally drive the neuron's physiological response. He quotes extensively from Barlow's (1972) "neuron" doctrine for perception built on this experimental work. Barlow insists that neurophysiology had revealed that a given single visual neuron can perform much more complex and subtle tasks than had previously been thought, including detecting patterned elements, discriminating object depth, ignoring irrelevant stimuli variations, giving prominence to what is informationally important in stimuli, responding reliably, and having this selectivity of response modified by previous visual experiences. For Barlow, these neurophysiological discoveries prompt "a revolution in our thinking," requiring us "to regard single neurons as the prime movers of . . . mechanisms . . . of mental operations" (Marr, 1982, p. 13). Single-cell electrophysiology, along with the detailed microanatomy that reveals each neuron's patterns of connectivity with other neurons, was poised to provide "a complete enough description for a functional understanding of the nervous system."

Then something unexpected happened—namely, nothing of much scientific significance! Marr puts this point bluntly: "The initial discoveries of the 1950s and 1960s were not being followed by equally dramatic discoveries in the 1970s" (Marr, 1982, p. 14). According to Marr, this change in experimental fortunes was not just happenstance. There lurked a deep and important gap in what reductionistic neuroscience could provide. Something important about perception—and about cognition in general—could not be found in neurophysiology alone: explanations over and above descriptions. Focusing on vision, Marr asks: "What are the visual areas of the cerebral cortex actually doing? What are the problems in doing it that need explaining, and at what level of description should such explanations be sought?" (p. 15) Electrophysiology alone could not address these requests for explanations.

To drive home this point Marr invites us to speculate: Suppose electrophysiology (and microanatomy) succeeded in meeting Barlow's and others' wildest dreams—suppose electrophysiologists "actually found the apocryphal grandmother cell," the neuron which fires "only when one's grandmother comes into view" (Marr, 1982, p. 15)—Marr asks (rhetorically): "would that really tells us much of anything at all?" Marr answers:

It would tell us that it existed—Gross's hand-detectors tell us almost that—but not why or even how such a thing can be constructed from the outputs of previously discovered cells. Do the single-unit recordings—the simple and complex cells—tell us much about how to detect edges or why one would want to, except in a rather general way through arguments based on economy and redundancy? If we really knew the answers, for example, we should be able to program them on a computer. But finding a hand-detector certainly did not allow us to program one. (p. 15)

Marr's point is that electrophysiologists—the "reductionists"—were only describing the behaviors of the cells they isolated. But describing is not explaining. And vision, of course, was just one of many cognitive functions to which his worry applied.

To really grasp Marr's distinction and point, it is worth considering in some detail a single example of an early groundbreaking electrophysiological result. So consider Charlie Gross and colleagues' discovery of "hand-detector" neurons in primate inferotemporal (TE) cortex. In the paper Marr himself cites (Gross, Rocha-Miranda, & Bender, 1972), and credits in the long quote cited above with "almost telling" us about the existence of grandmother cells, Gross and collaborators recorded activity to a variety of visual stimuli in 263 neurons in inferotemporal cortex and surrounding regions in 17 paralyzed and anesthetized rhesus monkeys. The receptive fields of these neurons were comparatively large: Most overlapped the foveal region (of greatest visual acuity) and the fields of some crossed the ipsilateral-contralateral midline (relative to the neuron's location, left hemisphere or right hemisphere). Activity in most TE neurons was driven by a variety of familiar visual aspects, including wavelength, size, shape, orientation, and direction of movement.

What about those fabled hand-detector TE neurons? These were cells, according to Gross and collaborators, that displayed "highly specific and unique triggering features" (Gross et al., 1972, p. 110). The "rank order of adequate stimuli" that drove activity in these unusual neurons "did correlate with similarity (for us) to the shadow of a monkey hand" (pp. 103–104) (see Fig. 2 below). Gross and colleagues report that fingers pointed downward elicited little response in these neurons, as compared to fingers pointed upward or to the sides, and noted that the latter are "the usual orientations in which the animal would see its own hand" (p. 104).

These hand detectors were not numerous. Gross and collaborators note that of the 262 TE neurons investigated for their entire study, 128 responded to dark stimuli, with 50 of those firing best to moving small rectangles. For the remainder—78/262—"particular complex dark stimuli were the best" (Gross et al., 1972, p. 104). One presumes that "hand detectors" comprised a much smaller proper subset of these remaining neurons.
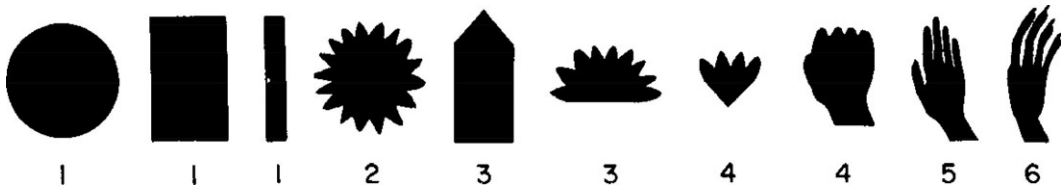
Fig. 2. Gross et al. (1972) figure 6. The original caption reads: "Examples of shapes used to stimulate a group TE unit apparently having very complex trigger features. The stimuli are arranged from left to right in order of increasing ability to drive the neuron from none (1) or little (2 or 3) to maximum (6)" (p. 104).

Gross and collaborators call their investigations into stimulus shape "limited and nonsystematic" (p. 104). In the paper's Discussion, they mention "a number" of inferotemporal neurons with "striking specific and complex trigger features," but they also admit that estimating "the incidence of such cells" could not be done (p. 107).

These few details from this landmark study in early visual neurophysiology nicely illustrate Marr's "deeper worry" about reductionism. His account of what these and other results provide, namely, descriptions of cellular activities rather than explanations of cognitive functions, is right on target. Even discovering the stimulus field properties of neurons throughout the visual hierarchies, and all the connectivity patterns obtaining between them, would not explain visual functioning, in any genuine sense of "explain." Not even the most careful and complete recordings of cellular activities and mappings of interneuronal connections could test hypothesized explanations of cognitive functions. Marr saw exactly the limits that "electrophysiology," the reductionistic standard of his day, was bound to. His three-level methodology and the "importance of computational theory" was his specific response to these limitations, his proposals for how to make cognitive science genuinely explanatory.

3. Marr and multiple realizability: A quick aside

One of my principal claims in this paper is that Marr's "descriptive, not explanatory" challenge to reductionistic neuroscience is his most important argument for his methodology. Some fans of Marr will demur. What about Marr's plump for multiple realizability—cognitivism's most popular challenge to reductionism (Fodor, 1974; Putnam, 1967), and still advocated to the present day? Without question Marr advocated multiple realizability. (Cooper and Peebles [this volume] also discuss this argument.) In fact, he did so two-fold: the multiple realizability of computational theory at the algorithmic level (Marr, 1982, p. 23) and of algorithmic theory at the physical implementation level (p. 24). But unlike functionalist philosophers, he does not make explicit argumentative use of these relations. He especially does not emphasize either multiple realizability's purported inconsistency with reductionism or its most controversial claim, the strict identity of the realized kind in distinct realizing kinds. Without strict identity of the realized kind across distinct realizers, the reductionist easily accounts for the differences among the realizing

kinds. Thirty years ago the popular reductionist response to the multiple realizability challenge was to acknowledge its existence but deny its relevance for scientific reduction. More recently, critics of the multiple realizability argument have challenged the multiple realizability premise itself. Proponents of multiple realizability are replying to these challenges, but some of these replies concede reductionists' principal points (Bickle, 2013) This vexed history of the multiple realizability argument, the recent swings in its fortunes that seem to favor its critics, and Marr's own minimal argumentative use of it in Chapter 1 of *Vision*, inclines me to set it aside in articulating Marr's argument for his methodology. Marr spends roughly 13 printed pages developing his "only describes, doesn't explain" argument against reductionistic neuroscience, as opposed to less than 1 printed page simply mentioning multiple realizability (without even citing Putnam, 1967, or Fodor, 1974).

4. Genuine explanations in reductionistic neuroscience post-Marr: Causal-mechanistic explanations

To this point I have defended three steps in this paper's argument:

1. Marr argues explicitly for his methodology for cognitive science, and his most important argument is the failure of reductionistic neuroscience of his time, namely electrophysiology, to provide explanations of cognitive functions (like vision).
2. Marr's methodology (the three levels and the importance of the computational level) were the specific additions he hypothesized to produce genuine explanations of cognitive functions.
3. Marr was correct in his assessment of state-of-the-art reductionistic neuroscience in the 1970s.

But important new experimental tools, and more importantly experimental designs that yielded evidence for a genuine kind of explanation, were about to enter reductionistic neuroscience's mainstream. As I will argue, these new tools enabled neuroscientists to conduct new types of experiments, yielding data that support a new kind of explanation, challenging Marr's specific critique of reductionism.

This new kind of explanation is distinctly different from what Marr urges cognitive scientists to pursue. Neuroscience began offering causal-mechanistic explanations: explanations of the target phenomena as the outcome of its component parts, the dynamics of these parts, and their organization. The new tools allowed experimenters to intervene into the workings of the hypothesized components—increasing or decreasing their activities, slowing them down, or speeding them up—and to track the effects of these interventions on the target phenomena. That is no longer merely describing, as electrophysiological recordings are limited to; in the right combinations of experimental interventions and results, that is gathering evidence that supports or challenges genuine causal-mechanistic explanations. This addition to the reductionistic neuroscience of the 1970s was not what Marr recommended for cognitive science. But it is difficult to argue that the resulting

accounts of the behavior indicating a specific cognitive function was not thereby explained.[1]

Two experimental tools were especially prominent in generating new data for causal-mechanistic explanations in neuroscience. One was stimulating electrodes, now spun small enough to be embedded, in living, behaving animals, into cortical microcolumns composed of neurons with similarly tuned receptive fields. This tool gained quick application across cognitive neuroscience; in vision research William Newsome's early work was especially prominent (e.g., Salzman, Britten, & Newsome, 1990). Over the past two decades, cortical microstimulation has been applied to other sensory modalities, motor systems, and even memory. Using electrodes now capable of directly stimulating only an estimated 80 to 100 neurons, experimenters have reported profound effects on behavior routinely assumed to be "cognitive" (typically in non-human primates, no less). Integrating these new interventional experimental results with the background descriptive knowledge that electrophysiology and microanatomy had already provided generates causal-mechanistic explanations of the target behavior at the cellular level, and hence of the cognitive functions those behaviors are taken to indicate.

Recent reductionstic neuroscience does not stop its interventions with small populations of neurons. Developments in the field of molecular and cellular cognition (MCC) over the past 20 years stemmed from the application of transgenic techniques from molecular genetics into neuroscience. These features allow experimenters to mutate any cloned gene in living, behaving mammals, and thereby manipulate key proteins in intracellular signaling pathways. A key early result in this field was Silva's and collaborators (Silva, Paylor, Whener, & Tonegawa, 1992; Silva, Stevens, Tonegawa, & Wang, 1992) manipulation of the gene for the alpha isoform of calmodulin kinase II (α-CaM-KII), a protein that had already been implicated in synaptic long-term potentiation (LTP) mediated by *N*-methyl-D-aspartate (NMDA) receptor activation.

Silva, Stevens, et al. (1992) engineered the α-CaMKII mutant mice using the gene-targeting techniques that eventually won Capecchi, Evans, and Smithies the 2007 Nobel Prize for Physiology or Medicine. The targeted gene is first mutated in embryonic stem cells, and these cells in turn are used to colonize host embryos and eventually generate genetic lines of mice with the mutation in every cell. Because α-CaMKII is expressed almost exclusively in post-natal forebrain excitatory neurons, this mutation did not compromise the survival, development, or the general health of the mutant mice. The α-CaMKII mutants were used in electrophysiological slice studies to reveal that the α-CaMKII deletion caused clear deficits in LTP measured in neurons in the CA1 region of the hippocampus (Silva, Stevens, et al., 1992). Since that region had already been established as critical for some learning tasks, they next did behavioral studies to show that the mutants also had severe hippocampal learning and memory deficits (Silva, Paylor, et al., 1992). For example, the mutants took many more trials than their wild-type (non-mutated) siblings to learn the location of the hidden platform in the Morris water maze. They also failed to exhibit freezing, a stereotypic rodent fear response, in contextual fear conditioning.

Soon after the Silva and collaborators α-CaMKII papers appeared, similar papers were published focusing on different kinases and other synaptic molecules. For example, Seth Grant, Tom O'Dell, and colleagues in Eric Kandel's laboratory showed that mice engineered with a knockout mutation of the fyn tyrosine kinase had deficient CA1 LTP and learning impairments in the Morris water maze and contextual fear conditioning (Grant et al., 1992). The convergence among the α-CaMKII and Fyn genetic findings was striking. In both cases, an experimental manipulation of a single intracellular signaling kinase that blocked hippocampal LTP also blocked hippocampal-dependent learning and memory.

These early transgenic studies integrating molecular biology, physiology, and behavior attracted considerable attention because of the potential promise the experimental approach offered for investigating directly hypothesized causal mechanisms of cognitive behaviors. Of course, these were not the first neuroscience experimental tools that manipulated hypothesized causal factors and then tracked their behavioral effects. Pharmacological tools, experimental surgical, electrical, and chemical ablations, and stimulating micro-electrodes had been used for decades to do exactly this. What MCC added to the experimental mix was the capacity to increase manipulation specificity and control. By manipulating the gene that codes for a protein known to be involved in intracellular function, first in embryonic stem cells but later (now) directly in the developed adult animal through viral vector insertions, experimenters can be much more precise about which neurons they manipulate. This contrast is striking with even the most precise surgical lesions or pharmacological manipulations (or even increasingly smaller stimulating electrodes), where questions about the extent of damage and drug leakage beyond the targeted neurons routinely confound interpretations of results. The myriad experimental controls perfected by decades of prior research in molecular biology generally enable a clear picture of not only which neurons have been manipulated but also the specific intracellular signaling pathways affected in those neurons. With this increased experimental power, causal mechanical hypotheses literally linking intracellular and intercellular signaling molecules and pathways to cognitive behaviors are directly testable on the neuroscientists' benches. State-of-the-art mainstream contemporary neuroscience is no longer just measurement and description. It now tests, directly and experimentally, causal-mechanistic hypotheses that purport to explain cognition. Not surprisingly, neuroscientists have flocked to these new tools, experimental designs, and causal-mechanistic hypotheses of behavior.[2]

5. From cellular and molecular experimental interventions to causal-mechanistic explanations of cognitive behavior

In addition to the three claims stated at the beginning of the previous section, I am now defending a fourth:
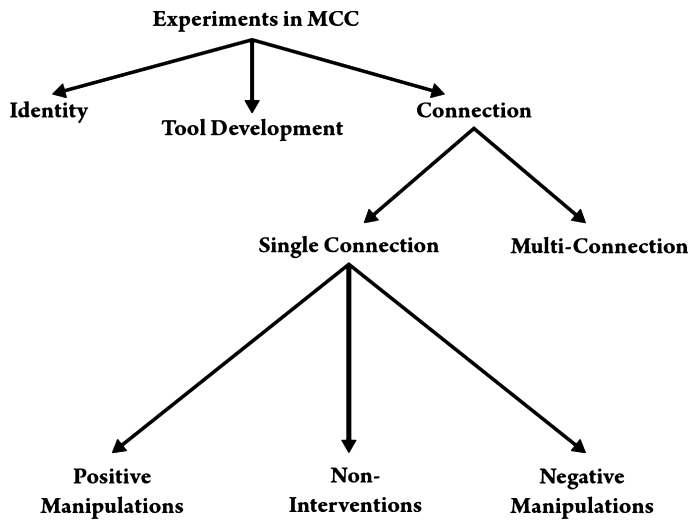
Fig. 3. Components of a framework for classifying experiments in molecular and cellular cognition (MCC). See text for descriptions. (From Silva et al., 2013, figure 2.1, p. 19.)

4. Current reductionistic neuroscience now provides causal-mechanistic explanations of behaviors routinely taken to indicate cognitive functions; a kind of explanations quite different from the type Marr recommended, but genuinely explanatory nonetheless.

Yet before I take (4) as defended, I need to say more about how current neuroscience gets from the kinds of experimental results sketched in the previous section to genuine causal-mechanistic explanations. The key resources are the variety of experiments typically pursued and the way these results are integrated with others.

Working with some landmark experimental publications from the 20-year history of MCC, Silva et al. (2013) derive a framework for experiments and the integration of their results into explanations. Concerning connection experiments, which test causal hypotheses,[3] we distinguish three types: positive manipulations, which increase the probability or intensity of a hypothesized causal factor (mechanism) and measure the hypothesized effect; negative manipulations, which decrease the probability or intensity of a hypothesized causal factor and measure the hypothesized effect; and non-intervention experiments, which measure activities of the hypothesized causal mechanism and the effect without directly manipulating either experimentally. (See Fig. 3.)

One important type of integration of experimental results is convergent three analysis, which investigates whether all three of these types of experiments have actually been performed for a specific causal hypothesis, and whether their results are consistent. Each type of connection experiment, by itself, has serious limitations for establishing a causal hypothesis. The old scientific adage "correlation is not causation" is sufficient to demonstrate the limitations of non-intervention experiments. Their results are limited exclusively to correlating activities of the hypothesized causal factor and effect. Well-controlled negative manipulations can show that a hypothesized mechanism is necessary

for the effect to occur. If we reduce the probability or intensity of the hypothesized causal agent and measure a decrease in the effect, then we have shown the necessity of the hypothesized cause for the effect, and that is an important step toward establishing a causal hypothesis directly. But negative manipulations alone cannot yield fully sufficient evidence for a causal hypothesis. Negative manipulations are typically experimental manipulations of hypothesized causal factors beyond the limits of biologically normal functioning (e.g., knocking out an entire gene). Hence, they must be combined with consistent converging results from non-manipulation experiments that correlate activities of the hypothesized mechanism and the effect under biologically realistic conditions. And negative manipulations typically cannot distinguish between background factors necessary for the effect and a genuinely triggering cause. That is the converging evidence that positive manipulations provide. By increasing the probability or intensity of actual triggering causal factors, we typically raise the probability or intensity of the effect. Typically this is not true for background conditions necessary for the effect to occur, but not actually part of the effect's triggering causal mechanism.[4]

The important point to emphasize here is that contemporary MCC recognizes the importance of converging evidence from all three types of connection experiments to establish even the simplest causal hypothesis with any strong degree of confidence. And a successful convergent three analysis alone is not sufficient to establish accepted causal-mechanical explanations in actual neuroscience practice. Establishing that a molecule like α-CaMKII is part of the causal mechanism for both synaptic LTP and spatial memory does not tell us how it is so causally linked—about what other mechanisms mediate α-CaMKII activity in hippocampal neurons and trained behavior in the Morris water maze, for example. That requires more connection experiments investigating related causal hypotheses, Convergent 3 analyses for each, and other forms of integration of results.[5]

Armed with even this brief discussion of our framework, we can now even more clearly see the force or Marr's criticism of the reductionistic neuroscience of his time and the failure of that criticism for current reductionism. Consider again Gross et al.'s (1972) discovery of hand-detector neurons in primate inferotemporal cortex. What kind of experiments were Gross and collaborators pursuing? Answer: non-intervention experiments, exclusively. They mapped activities in individual neurons through single-cell recording electrodes while a variety of visual stimuli was presented to a monkey's visual field. That is mere correlation, insufficient by itself to establish cause, and it is not even correlation of the cellular activity with any behavioral measure indicating the content of visual experience (because the monkeys were anesthetized). Furthermore, the experimental conditions required to conduct these single-cell electrophysiological recordings in living monkeys stretched biological normalcy considerably: heads secured via bolts surgically implanted in skulls, microdrive bases for electrode insertions surgically implanted over a portion of the excised temporal bone, visual stimuli presented, and single-cell recordings made under paralysis and anesthesia. So even considered as individual non-intervention experiments, these were less than ideal for providing even the limited evidence for causal explanations that non-intervention experiments can at most provide.[6] Similar remarks

hold for the other early experimental landmarks of "the reductionistic approach" that Marr discusses, correctly, as non-explanatory.

With the help of our framework and integration procedures for relating experimental interventions to causal-mechanical explanations, I now take claim (4) above as defended (at least to the extent possible in a short paper). The electrophysiologists who constituted "the reductionistic approach" in Marr's time were simply describing (correlating, actually), not explaining, and more of those kinds of (non-intervention) experiments, no matter how careful and thoroughly, could not alter that limitation. But the new experimental tools that were soon to take over mainstream neuroscience, and more importantly the new kinds of experiments they permitted and the forms of integration of results they made possible, made subsequent neuroscience genuinely explanatory. The causal-mechanistic explanations of behavior are not the kind of explanations Marr recommended that cognitive scientists pursue. But they are genuine explanations, and this fact undercuts Marr's own principal argument for his famous three-level methodology.

6. Conclusion, Circa 2015

Claims (1)–(4) argued for in this paper support the following conclusion:

5. While Marr had a sound argument against the explanatory potential of the reductionistic neuroscience of his time, that argument is no longer sound in light of developments in neuroscience post-Marr, and so Marr's methodology is striped of the principal argument he offered for it.

Contemporary cognitive scientists can no longer be so quick to reject "the reductionistic approach" and pursue Marr-style cognitive science. What we now have are two approaches for explaining cognitive functions, which yield different kinds of explanations. They compete because both purport to fully explain the behavior taken to indicate a specific cognitive function. Contemporary Marr-inspired cognitive scientists at least owe us a new argument for why "reductionism" fails; Marr's "describes, doesn't explain" gambit no longer holds. (Notice that some of the defenders/elaborators of Marr in this volume—Bechtel and Shagrir, Eliasmith and Kolbeck, Copper and Peebles—do not offer us this.) The modern-day Marrians' task here grows increasingly difficult by the avalanche of research these new neuroscientific methods are generating (check out the latest issues of *Science* or *Nature*), pesky questions about the nature of levels and the relations between them, and philosophical worries about "downward" causation and explanatory exclusion. Current reductionism ascendant?

Notes

1. The importance of causal-mechanistic explanations in biology, and the importance of experimental interventions to discover evidence for them, has not been lost on

recent philosophy of science. Machamer, Darden, and Craver (2000) are widely acknowledged for introducing "the new mechanism" more broadly into philosophy of science, with Craver (2007) acknowledged for its detailed application to neuroscience. These works trace back to Bechtel and Richardson (1993), and more distantly to scattered essays by Wimsatt. Woodward (2003) is the classic recent statement of the role of experimental interventions in testing causal-mechanistic explanations. Silva, Landreth, and Bickle (2013) reach a similar picture, working exclusively with experimental landmarks in "molecular and cellular cognition."

2. Another new tool that has garnered widespread attention is optogenetics. Experimenters insert the gene for a light-sensitive protein from algae (e.g., channelrhodopsin), which forms an ion channel that opens in response to blue light to pass sodium ions into the neuron, depolarizing the membrane. Those specific neurons receiving the transgene can then be caused to fire simply by flashing them with blue light. Susumu Tonegawa's laboratory recently exposed mice mutated with this transgene to a novel environment and allowed them to explore it. This activity produced new expression of the *channelrhodopsin-2* transgene, and hence the synthesis of new light-sensitive ion channels, specifically in the neurons encoding the memory for the new context. They then transferred the mutant mice to a different environment, shined blue light onto the region of the brain containing the neurons with the newly synthesized light-sensitive ion channels, and delivered a conditioning aversive foot shock. Mutated mice "froze" when placed back into the first, "safe" context, displaying the classic rodent fear response. One interpretation is that this optogenetic intervention technique thus engineered a false memory, a context-shock association to a context in which no shock had ever been delivered (Ramirez et al., 2013).

3. Connection experiments and techniques of integrating their results are the primary focus of Silva et al. (2013).

4. For a detailed presentation of this argument for convergent three integration, illustrated by landmark experimental studies from the history of MCC, see Silva et al. (2013, Chapter 4).

5. Detailed comparisons and contrasts between Woodward's (2003) and Silva et al.'s (2013) accounts of causal-mechanistic hypothesis testing remain to be determined. One difference is clear, however: MCC's focus on cellular and molecular mechanisms of behavior makes Silva, Landreth, and Bickle's account much more reductionistic than Woodward's purports to be.

6. This discussion illustrates an important point about our use of the term, "non-intervention." Of course, "interventions" were performed in Gross et al.'s (1972) experiments! But these interventions were not designed to manipulate variables of causal interest, but rather simply to set up conditions necessary for the required observed correlations.

References

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, *1*, 371–394.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity*. Princeton, N.J.: Princeton University Press.

Bickle, J. (2013). Multiple realizability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Available at http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/.

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Craver, C. (2007). *Explaining the brain*. New York: Oxford University Press.

Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese 28*: 97–115.

Grant, S. G., O'Dell, T. J., Karl, K. A., Stein, P. L., Soriano, P., & Kandel, E. R. (1992). Impaired long-term potentiation, spatial learning, and hippocampal development in fyn mutant mice. *Science*, *258*, 1903–1910.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in infrotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*, 96–111.

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.

Marr, D. (1982). *Vision*. San Francisco: Walter Freeman and Co.

Putnam, H. (1967). Psychological predicates". In W. H. Capitan, & D. D. Merrill (Eds.), *Art, mind and religion* (pp. 37–48). Pittsburgh: University of Pittsburgh Press.

Ramirez, S., Liu, X., Ren, P.-A., Suh, J., Pignatelli, M., Redondo, R., Ryan, T. J., & Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science*, *341*, 387–391.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgments of motion direction. *Nature*, *346*, 174–177.

Silva, A. J., Landreth, A., & Bickle, J. (2013). *Engineering the next revolution in neuroscience*. New York: Oxford University Press.

Silva, A. J., Paylor, R., Whener, J. M., & Tonegawa, S. (1992). Impaired spatial learning in alpha-calcium-calmodulin kinase II mutant mice. *Science*, *257*, 206–211.

Silva, A. J., Stevens, C. F., Tonegawa, S., & Wang, Y. (1992). Deficient hippocampal long-term potentiation in alpha-calcium-calmodulin kinase mutant mice. *Science*, *257*, 201–206.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.