A Mechanistic Account of Computational Explanation in Cognitive Science

Marcin Miłkowski (mmilkows@ifispan.waw.pl)
Institute of Philosophy and Sociology, Polish Academy of Sciences
ul. Nowy Świat 72, 00-330 Warsaw, Poland

Abstract

Explanations in cognitive science rely predominantly on computational modeling. Though the scientific practice is systematic, and there is little doubt about the empirical value of numerous models, the methodological account of computational explanation is not up-to-date. The current paper offers a systematic account of computational explanation in cognitive science in a largely mechanistic framework. The account is illustrated with a short case study of modeling of the mirror neuron system in terms of predictive coding.

Keywords: computation; computational modeling; explanation; mechanism; levels; information-processing.

Importance of Computational Modeling

Computational modeling plays a special role in contemporary cognitive science; over 80 percent of articles in theoretical journals focus on computational[1] models (Busemeyer & Diederich, 2010). The now dominating methodology forcefully defended by (Marr, 1982) has turned out to be fruitful. At the same time, the three-level account of Marr is not without problems. In particular, the relationship among the levels is interpreted in various ways, wherein the change of level is both the shift of grain and the shift of the boundary of the system under explanation (McClamrock, 1991); it is not at all clear what is the proper relation between competence and its realization or whether bottom-up modeling is entirely mistaken; and, last but least, whether one model should answer how, what and why questions related to the explanandum.

My goal in this paper is to offer a descriptive account, which is close in spirit to the recent developments in the theory of mechanistic explanation (Bechtel, 2008; Craver, 2007; Glennan, 2002; Machamer, Darden, & Craver, 2000). According to mechanism, to explain a phenomenon is to explain the underlying mechanism. Mechanistic explanation is a species of causal explanation, and explaining a mechanism involves the discovery of its causal structure. While mechanisms are defined variously, the core idea is that they are organized systems, comprising causally relevant component parts and operations (or activities) thereof. Parts of the mechanism interact and their orchestrated operation contributes to the capacity of the mechanism. Mechanistic explanations abound in special sciences and it is hoped that the adequate description of the principles implied in explanations generally accepted as sound will furnish researchers also with normative guidance.

[1] I am *not* using the word 'computational' here in the sense used by Marr to define one of the levels in his account.

The claim that computational explanation is best understood as mechanistic gains popularity (Piccinini, 2007), and I have defended it at length against skeptical doubt elsewhere (Miłkowski, 2013). Here, I wish to succinctly summarize the account and, more importantly, add some crucial detail to the overall mechanistic framework proposed earlier. I cannot discuss Marr's theory in detail here (but see (Miłkowski, 2013, pp. 114–121)) and it is used only for illustration purposes. My remarks below are not meant to imply a wholesale rejection of his largely successful methodology.

Marr's account did not involve any theory of how computation is physically realized, and it is compatible with a number of different accounts. I will assume a structural account of computational realization here, defended also by Piccinini (2008) and Chalmers (2011). For an extended argument, see also (Miłkowski, 2011, 2013).

One particular claim that is usually connected with the computational theory of mind is that the psychologically relevant computation is over mental representation, which leads to the language of thought hypothesis (Fodor, 1975). Here, no theory of mental representation is presupposed in the account of computation, one of the reasons being that representation is one of the most contentious issues in contemporary cognitive science. As the present account is intended to be descriptively adequate, assuming one particular theory of representation as implied by computation would make other accounts immediately non-computational, which is absurd. Another reason is that mechanistic accounts of computation do not need to presuppose representation (Fresco, 2010; Piccinini, 2006), though they do not exclude the representational character of some of the information being processed. In other words, it is claimed that only the notion of information (in the information-theoretic sense, not in the semantic sense, which is controversial) is implied by the notion of computation (or information-processing).

Explanandum phenomenon

Marr stressed the importance of specifying exactly what the model was supposed to explain. Specifying the explanandum phenomenon is critical also for the mechanistic framework, as several general norms of mechanistic explanation are related to the specification of the capacity of the mechanism. All mechanisms posited in explanations have an explanatory purpose, and for this reason, their specification is related to our epistemic interest. For the same reason, the boundaries of the mechanism, though not entirely arbitrary, can be carved in different ways depending on what one wishes to explain.

The explanandum phenomenon has to be described precisely in a mechanistic model; otherwise, the model's use and value will be unclear. The specification of the model is not to be confused with raw, unrefined observation or common-sense intuition about the capacity under consideration. The specification of the capacity may be (and usually is) improved during the modeling process, wherein the model allows to understand the capacity better. What the mechanistic model explains is the real mechanism, but how the explanandum phenomenon is delineated is decided in what was called "the model of data" in philosophy of science (Suppes, 1962). For example, models of language production usually presuppose that user's productivity is the phenomenon to be explained, even though it is impossible to empirically observe a language user producing an infinite set of sentences. If there are theoretical reasons to believe that language users have this capacity, it will be described in a model of data. In this respect, mechanistic explanation is in accord with Marr's plea for explicit specification of *what* is computed.

To some degree, the specification of the explanandum phenomenon corresponds to description of the cognitive competence (understood generically as the capacity of the mechanism). However, in contrast to traditional competence accounts, descriptions of the explanandum need not be idealized. Also, the competence is explained with realization, and its realization by underlying levels of the mechanism is explanatorily relevant. This stands in contrast to traditional symbolic cognitive science.

Explanatory focus and target

In the context of computational modeling, which nowadays uses different computer simulations and embodied robots, it becomes clear that properties of a model are not limited to the ones related directly to the explanandum phenomenon. For example, a robotic model of cricket phonotaxis (Webb, 1995) has to include, for technical reasons, a circuit board even if there is nothing that corresponds to the board in the cricket. Such boards are ignored when evaluating the adequacy of the robotic explanation. I propose to distinguish the *explanatory focus* of the model from its *target*, which is the real robot. In particular, all embodied mechanistic models are *complete* with respect to the capacities of the target, while their explanatory focus may still include gaps: we may still not know how certain properties of the insect give rise to the explanandum phenomenon even if we have a robotic replica. The same goes for purely computational models that contain numerous ad hoc additions (Frijda, 1967; Lewandowsky, 1993). These additions are not parts of the explanatory focus.

Whenever the causal model of the explanatory focus of the mechanism is complete with respect to the explanandum phenomenon (note: not complete in an absolute sense), the model is a mechanistic how-actual explanation; if the model includes some black boxes, whose function is more or less well-defined, it is a mechanism schema; otherwise, it remains a mechanism sketch.[2] Note that even a grounded, embodied, robotic model of visual perception may still be a mechanism sketch with respect to human vision. Also, a model in which the explanatory focus is just a minor part of the mechanism, while the parts included in the target are predominant, violates the principle of parsimony.

Three levels of constitutive explanation

Constitutive mechanistic explanation is the dominant form of computational explanations in cognitive science, and I will focus on it in what follows. This kind of explanation includes at least three levels of the mechanism: a constitutive (-1) level, which is the lowest level in the given analysis; an isolated (0) level, at which the parts of the mechanism are specified along with their interactions (activities or operations); and the contextual (+1) level, at which the function of the mechanism is seen in a broader context (e.g., the context for cricket phonotaxis includes the dispersion of sound in the air). In contrast to how Marr (1982) or Dennett (1987) understand them, levels here are not just levels of abstraction; they are levels of *composition*. Hence, they are tightly integrated but not entirely reducible to the lowest level.

Computational models explain how the computational capacity of a mechanism is generated by the orchestrated operation of its component parts. To say that a mechanism implements a computation is to claim that the causal organization of the mechanism is such that the input and output information streams are causally linked and that this link, along with the specific structure of information processing, is completely described (for more on various mathematical notions of information, see Miłkowski (2013, chap. 2); note that I do not presuppose Church/Turing thesis). Importantly, the link might be cyclical and as complex as one could wish.

There are two ways in which computational models may correspond to mechanisms; first, they may be *weakly equivalent* to the explanandum phenomenon, in that they only describe the input and output information; or *strongly equivalent*, when they also correspond to the process that generates the output information. Note that these notions have been used in methodology of computer simulation since 1960s (Fodor, 1968, chap. 4). Only strongly equivalent models are explanatory according to the mechanistic framework.

Mechanistically adequate model of computation

The description of a mechanistically adequate model of computation at the 0 level usually comprises two parts: (1) an abstract specification of a computation, which should include all the causally relevant variables; (2) a complete blueprint of the mechanism at this level of its organization. I will call the first part *formal model of the mechanism* and

[2] These distinctions were used by Craver (2007), but were unrelated to the distinction between the target and the explanatory focus.

the second *instantiation blueprint* of the mechanism, for lack of a better term. While it should be clear that a formal model needs to be included, it is probably less evident why the instantiation blueprint is also part of the mechanistically adequate model. The causal model must include all causally relevant parts and operations without gaps or placeholder terms (think of generic and unspecific terms such as "representation" or "activation"). Yet formal models cannot function as complete causal models of computers. For example, to repair a broken old laptop, it is not enough to know that it was (idealizing somewhat) formally equivalent to a universal Turing machine. Similarly, how mental deficits will manifest themselves is not obvious based on a description of ideal cognitive capacity. One needs to know its implementation.

Hence, the mechanistic model of a computational phenomenon cannot be limited to its formal properties. Accordingly, merely formal models of, say, linguistic competence, which abstract away from its realization, are assessed as essentially incomplete. They are either mere specifications of the explanandum phenomenon, but not explanatory in themselves, or, when accompanied with a rough theory of how they are related to experimental data, mechanism sketches (Piccinini & Craver, 2011). This means that computational explanations of psychological capacities need to be integrated, for completeness, with models of their realization. Otherwise, they may posit epiphenomenal entities without any causal relevance. Contrary to the functionalist theory of psychological computational explanation (Cummins, 1983), mechanism requires it to be causal. It follows that some symbolic models in psychology, even if they are weakly equivalent to the model of input/output data, are not considered to be fully explanatory because of the inherent danger of positing entities that are causally irrelevant.

Just because the usual description of the computational mechanism usually involves two different models, the formal one and the instantiation blueprint, and these may be idealized, computational modeling requires complex integration, similar to one described as multiple-models idealization (Weisberg, 2007).

Note that my mechanistic account of computation does not stipulate that there be a single formal model of computation that would fit all purposes. Rather, it adheres to transparent computationalism (Chrisley, 2000): any formal model that can be specified in terms of information-processing is fine here, be it digital, analog, or hybrid, as in contemporary computational neuroscience (Piccinini & Bahar, 2012).

The empirical adequacy of the mechanistically adequate model of computation can be tested. As such models are strongly equivalent to processes being modeled, usual process-testing methods apply, including chronometry (Posner, 2005), various kinds of experimental and natural interventions (Craver, 2007), brain imaging – though with usual caveats (Trout, 2008), and task decomposition (Newell & Simon, 1972). All in all, the more independent observables are tested, the more robust the model. Note that the phenomenological validation modeled after the Turing test (Turing, 1950) is not taken to be evidence of the model's empirical adequacy.

Marr's cash register

The account may be illustrated with the example used by Marr (1982, pp. 22–24): a cash register in a supermarket. The explanandum phenomenon is the capacity to add prices of individual items and determine the overall sum to be paid. At the contextual level, one describes the cash register as playing a certain role in the supermarket, by allowing easy calculation of the sum to be paid, and making the work of the cashier clerk easier. This includes a bar-code scanner, a conveyor belt, etc. At the isolated level, a dedicated computer using special software is described. The constraints mentioned by Marr, such as commutativity or associativity of addition, are included in the description of the software. Yet without describing the machine that can run the software, this level of description is incomplete. Various failures of the cash register (e.g., dimming of the display), can be explained not only in terms of the software bugs but also as hardware failures. Also, the particular display configuration, which can be related to user preferences at the contextual level, is usually not described fully in the software specification. It is the isolated level where one describes the physical machine that can display the product name for the cashier clerk and, more fundamentally, can run code by reading it from external memory (not all computers do so; a mechanical cash register, even if it performs computations, cannot run different software). The formal description, usually in terms of the programming language or diagrams, is put into correspondence with the machine. At the constitutive level, the operations of the electronic parts of the machine are explained by reference to their properties, relationships, and organization. Just because vast differences between different types of registers are possible (witness the differences between the self-checkout register and the ones used during the American Civil War), the exact explanations will differ. Also, self-checkout machines will have the capacity to collect cash automatically, which needs to be explained as well (the explanandum will be different), and so forth.

The purpose of this toy example is to show that the mechanistic explanation differs a bit from Marr's account by explicitly tightly integrating the levels. Also, at all levels one can ask the why-question: why is the design appropriate for the user? Why does the cash register appropriately display figures on the screen? Why does it save energy? The how-answer is specified at a lower level, and the lowest level depends on our epistemic interest. The what-question also concerns operation of all levels.

Case study: Predictive coding in mirror neurons

To demonstrate what methodological guidance is offered by the mechanistic account of computational explanation, let

me briefly describe a recently proposed model of action-understanding in terms of predictive coding (Kilner, Friston, & Frith, 2007). Predictive coding is one of the Bayesian frameworks and is gaining now considerable recognition (Clark, 2013). In the model, it is presupposed that this capacity is realized by the mirror-neuron system (MNS henceforth).[3] The explanandum phenomenon, or action understanding, is described at four levels of hierarchy: (1) the intention-level, which includes long-term goals of actions; (2) the goal-level, which includes short-term goals necessary to realize (1); (3) the kinematic level, which is the shape of the movement of limbs in space and time; and (4) the muscle level, which is the pattern of muscle activity underlying the action (Hamilton & Grafton, 2006). People have visual access only to (3) of other agents. Moreover, the same kinematic level information is correlated to different intentions: Mr. Hyde might hurt someone with a scalpel by making the same movements as Dr. Jekyll (Jacob & Jeannerod, 2005). What needs to be explained, therefore, is how one understands actions, given ambiguous visual information; the constraint of the model is that such understanding is to be realized by MNS. Naturally, given relatively scarce evidence about the details of MNS, the model might be currently only biologically plausible. In mechanistic terms, it cannot be a how-actually model, as we lack observables that could confirm that causal factors in the model are actual. We may have only a how-plausible model (for more on this distinction, see (Craver, 2007)), which should ascribe a precise computational role for MNS.

Kilner, Friston & Frith note that other similar explanations of action in terms of MNS posit forward or generative models. Yet these explanations cannot deal with the fact that people easily distinguish between the action of Dr. Jekyll and Mr. Hyde. In other words, they do not explain one important part of the phenomenon.

The contextual level of the proposed predictive coding mechanism includes the context in which the action is observed (e.g., the operation theatre vs. dark streets of London). The context of action, which is not coded by MNS, is hypothesized to be represented by other parts of the larger hierarchy, where intentions are encoded (Kilner et al., 2007, p. 164). Note that such hierarchy can be naturally accounted for in the mechanistic framework, while in the Marrian methodology, nested hierarchies of mechanisms are still analyzed merely on three levels, which are not levels of composition, as in Kilner et al.'s paper (this makes the analysis of the model in Marrian terms all the more difficult).

The 0 level of the mechanism is then described as performing predictive coding of action, i.e., the mechanism predicts the sensory consequences of movements, and the prediction error is minimized through recurrent or reciprocal interactions among levels of a cortical hierarchy. This means that the mechanism posited by authors comprises more than just three levels, which is the minimal number for constitutive explanations. Here, the upper level mechanism employs a generative model to predict representations in the level below. Backward connections are used by the upper level to convey the prediction to the lower level, which is used to produce information about prediction error. The instantiation blueprint of the mechanism includes this hierarchy whose architecture allows adjusting the neural representations of actions in terms of sensory representation of causes of action if prediction error is found. The architecture is self-organizing, and the reciprocal exchange of signals continues until the error is finally minimized.

The formal model of the neural architecture is described here in terms of empirical Bayesian inference (Friston, 2002, 2003, 2005): the prior expectations are generated by the self-organizing information-processing architecture. In other words, this model includes, as usual, two complementary parts: the instantiation blueprint, characterized in terms what is known about MNS, and its formal computational specification. Quite obviously, contrary to the programmable cash register, no stored-program computer is posited.

The constitutive level is merely touched upon; there is no extensive discussion of the precise realization of predictive coding by elementary entities of the neural system. Thus, this model is, at best, a mechanism schema, because it does not explain how MNS comes to operate as it does. The authors stress that to test the model, one would need to characterize the nodes of the cortical hierarchy anatomically and functionally, and such characterization is not available.

The neural plausibility of the predictive coding and its relation to empirical Bayesian modeling is the focus of much current discussion (Blokpoel, Kwisthout, & Van Rooij, 2012). In particular, the question whether the biologically plausible implementation of the predictive coding is equivalent to empirical Bayes or not (it may somewhat approximate it). The mechanistic explanation requires that the mechanisms be not idealized in such a way that would require to ignore tractability questions (Van Rooij, 2008). The data in the original paper makes it impossible to answer critical questions about the mechanism in this context, such as the number of inputs in the Bayesian network, which is essential in assessing the parametrized complexity of the algorithm.

Were the model implemented on the computer, the results of the simulation could be compared to those observed in humans or in macaque monkeys. Alas, no such results are reported by Kilner et al., and since without implemented models detailed testing of hypotheses is impossible, the empirical adequacy of the explanation is not entirely clear. To assess the adequacy properly, one should rather implement several comparable models of the same explanandum phenomenon, which can also help in avoiding the confirmation bias to which researchers are prone (Farrell & Lewandowsky, 2010; Miłkowski, 2013, p. 86).

[3] For my purposes, it is quite irrelevant whether this account of MNS is correct or not (but see (Lingnau, Gesierich, & Caramazza, 2009)). I am merely interested in how the model is vindicated by its authors and how it should be evaluated from the mechanistic standpoint.

Some Bayesian theories in psychology were recently criticized as fundamentalist, i.e., dogmatically trying to model behavior as rational and without mechanistic constraints (Jones & Love, 2011). Note that this is not true of the model under consideration; Bayesian modeling in neuroscience is obviously related to functioning of the brain. Instead of stressing the contrast between the mechanistic account of computational explanation and Bayesian modeling, my intention is to show that the mechanistic framework can be used to evaluate the contribution of the given model to progress in understanding of the explanandum phenomenon.

Summing up this part of the discussion, the mechanistic framework makes it easy to assess the maturity of the model in terms of its completeness and empirical adequacy. Because the computer implementation is lacking, it is impossible to say whether the model contains a lot of empirically undecided decisions that are needed to make it run (hence focus/target evaluation is impossible). At the same time, there is no information about the constitutive level. On the contextual level, placeholder terms such as "intention encoding" are used and they need further explanations in other models. Thus, the model does not include a complete specification of the mechanism.

Also, it is not at all clear how long-term goals might be understood in terms of mere sensory input prediction. Dr. Jekyll's intention to heal a patient (long-term goal) does not seem, prima facie, to be represented just in sensory terms. If it is actually so represented, the model does not explain how. This makes it a mechanism sketch, so its explanatory value is, qualitatively speaking, on the par with traditional symbolic models of competence. (Quantitative evaluation is impossible here, as no results of experiments on computer implementation were reported.)

Conclusion

The mechanistic account of computational explanation preserves the insights of Marr but is more flexible when applied to complex hierarchical systems. It may help integrate various different models in a single explanation. Mechanistic methodological principles are inferred from research practice in life sciences, neurosciences, and cognitive science. Also, by subsuming computational explanation under causal explanation, the mechanistic account is naturally complemented by methodology of causal explanation (Pearl, 2000; Spirtes, Glymour, & Scheines, 2000; Woodward, 2003).

By allowing multiple nested hierarchies, the standard three-level constitutive explanation is naturally expanded when needed. There is also no danger in preferring only the contextual level in the explanation, as it does not furnish us with the constitutive causal factors. The constitutive level will also not obviate the need for the contextual level as it does not contain some of the entities which are found at the contextual level. For example, the encoding of intention is not realized by MNS only, so its explanation cannot be 'reduced' to the description of the lower levels.

The present theory is not intended to settle debates over matters in which modelers explicitly disagree; the only goal is to make as much sense of various modeling approaches as possible, and make cross-approach comparisons possible by showing the common ground between them.

It is also not presupposed that computational explanation is the only proper way to explain cognition (Miłkowski, 2012). On the contrary, only some part of the mechanism model is strictly computational (i.e., uses vocabulary of the theory of computation). The constitutive level of the mechanism has to be framed in non-computational terms; otherwise, the computational operations of the isolated level are not explained, and may turn out to be spurious (Miłkowski, 2013, pp. 82–3). At the same time, the present account leads naturally to explanatory pluralism, as the only requirement for the theoretical frameworks used to describe various levels of composition of mechanisms is that they include causally relevant factors.

Acknowledgments

The author gratefully acknowledges the support of the Polish National Science Center Grant in the OPUS program no. 2011/03/B/HS1/04563.

References

Bechtel, W. (2008). *Mental Mechanisms*. New York: Routledge (Taylor & Francis Group).

Blokpoel, M., Kwisthout, J., & Van Rooij, I. (2012). When Can Predictive Brains be Truly Bayesian? *Frontiers in Psychology*, *3*(November), 1–3. doi:10.3389/fpsyg.2012.00406

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Los Angeles: Sage.

Chalmers, D. J. (2011). A Computational Foundation for the Study of Cognition. *Journal of Cognitive Science*, (12), 325–359.

Chrisley, R. (2000). Transparent computationalism. In M. Scheutz (Ed.), *New Computationalism: Conceptus-Studien 14* (pp. 105–121). Sankt Augustin: Academia Verlag.

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences*, *(in print)*.

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

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, Mass.: MIT Press.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, Mass.: MIT Press.

Farrell, S., & Lewandowsky, S. (2010). Computational Models as Aids to Better Reasoning in Psychology. *Current Directions in Psychological Science*, *19*(5), 329–335. doi:10.1177/0963721410386677

Fodor, J. A. (1968). *Psychological explanation: an introduction to the philosophy of psychology*. New York: Random House.

Fodor, J. A. (1975). *The Language of Thought* (1st ed.). New York: Thomas Y. Crowell Company.
Fresco, N. (2010). Explaining Computation Without Semantics: Keeping it Simple. *Minds and Machines*, *20*(2), 165–181. doi:10.1007/s11023-010-9199-6
Frijda, N. H. (1967). Problems of computer simulation. *Behavioral Science*, *12*(1), 59–67. doi:10.1002/bs.3830120109
Friston, K. (2002). Functional integration and inference in the brain. *Progress in neurobiology*, *68*(2), 113–43.
Friston, K. (2003). Learning and inference in the brain. *Neural networks : the official journal of the International Neural Network Society*, *16*(9), 1325–52. doi:10.1016/j.neunet.2003.06.005
Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *360*(1456), 815–36. doi:10.1098/rstb.2005.1622
Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science*, *69*(S3), S342–S353. doi:10.1086/341857
Hamilton, A. F. de C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(4), 1133–7. doi:10.1523/JNEUROSCI.4551-05.2006
Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences*, *9*(1).
Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(04), 169–188. doi:10.1017/S0140525X10003134
Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159–66. doi:10.1007/s10339-007-0170-2
Lewandowsky, S. (1993). THE REWARDS AND HAZARDS OF COMPUTER SIMULATIONS. *Psychological Science*, *4*(4), 236–243. doi:10.1111/j.1467-9280.1993.tb00267.x
Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(24), 9925–30. doi:10.1073/pnas.0902262106
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25.
Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman and Company.
McClamrock, R. (1991). Marr's three levels: A re-evaluation. *Minds and Machines*, *1*(2), 185–196. doi:10.1007/BF00361036
Miłkowski, M. (2011). Beyond Formal Structure: A Mechanistic Perspective on Computation and Implementation. *Journal of Cognitive Science*, *12*(4), 359–379.
Miłkowski, M. (2012). Limits of Computational Explanation of Cognition. In V. C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence* (pp. 69–84). Berlin - Heidelberg: Springer. doi:10.1007/978-3-642-31674-6_6
Miłkowski, M. (2013). *Explaning the Computational Mind*. Cambridge, Mass.: MIT Press.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.
Piccinini, G. (2006). Computation without Representation. *Philosophical Studies*, *137*(2), 205–241. doi:10.1007/s11098-005-5385-4
Piccinini, G. (2007). Computing Mechanisms. *Philosophy of Science*, *74*(4), 501–526. doi:10.1086/522851
Piccinini, G. (2008). Computers. *Pacific Philosophical Quarterly*, *89*(1), 32–73. doi:10.1111/j.1468-0114.2008.00309.x
Piccinini, G., & Bahar, S. (2012). Neural Computation and the Computational Theory of Cognition. *Cognitive Science*, n/a–n/a. doi:10.1111/cogs.12012
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese*, *183*(3), 283–311. doi:10.1007/s11229-011-9898-4
Posner, M. I. (2005). Timing the brain: mental chronometry as a tool in neuroscience. *PLoS biology*, *3*(2), e51. doi:10.1371/journal.pbio.0030051
Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, Mass.: The MIT Press.
Suppes, P. (1962). Models of Data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress* (pp. 252–261). Stanford: Stanford University Press.
Trout, J. D. (2008). Seduction without cause: uncovering explanatory neurophilia. *Trends in cognitive sciences*, *12*(8), 281–2. doi:10.1016/j.tics.2008.05.004
Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460. doi:10.1093/mind/LIX.236.433
Van Rooij, I. (2008). The tractable cognition thesis. *Cognitive science*, *32*(6), 939–84. doi:10.1080/03640210801897856
Webb, B. (1995). Using robots to model animals: a cricket test. *Robotics and Autonomous Systems*, *16*(2-4), 117–134. doi:10.1016/0921-8890(95)00044-5
Weisberg, M. (2007). Three kinds of idealization. *Journal of Philosophy*, *104*(12), 639–659.
Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.