Applications of an Implementation Story for Non-sentential Models

Jonathan Waskan

Abstract. The viability of the proposal that human cognition involves the utilization of non-sentential models is seriously undercut by the fact that no one has yet given a satisfactory account of how neurophysiological circuitry might realize representations of the right sort. Such an account is offered up here, the general idea behind which is that high-level models can be realized by lower-level computations and, in turn, by neural machinations. It is shown that this account can be usefully applied to deal with problems in fields ranging from artificial intelligence to the philosophy of science.

1 Introduction

Here I offer an elaboration and defense of the cognitive models hypothesis (CMH), which is the proposal that human cognition is sometimes constituted by the utilization of non-sentential representations that are like scale models in crucial respects. Scale models are representations that can be used to effect truth-preserving inferences regarding modeled systems in virtue of their instantiation of the very same properties as those systems. They are also sometimes termed *physically isomorphic* representations [1]. On this broad construal, even simple spatial matrix representations of spatial properties and relations count as scale models.

There is, of course, ample historical precedent for the CMH, having been advanced by Aristotle, Berkeley, Locke, and many others. Many contemporary cognitive scientists also favor the CMH, an hypothesis to which they often refer (or at least allude) with terms like "mental image", "depictive" or "non-propositional representation", "mental model", or "analog representation". The CHM has in recent decades been fruitfully applied in the context

Jonathan Waskan
Department of Philosophy, University of Illinois at Urbana-Champaign
e-mail: waskan@illinois.edu

L. Magnani et al. (Eds.): Model-Based Reasoning in Science & Technology, SCI 314, pp. 463–476.
springerlink.com

of theories of perception, comprehension, grammar, classification, deductive reasoning, planning, abduction, naîve physics, and mind reading. Despite its many useful applications, it has been clear for some time that the CMH faces a major realization crisis. Here I will describe the nature of this crisis and chart out what, at present, appears to be the only way past it. I will then show that this way of resolving the crisis can be used to address a set of issues that has recurred in various guises across a number of fields, including artificial intelligence, logic, psychology, and the philosophy of science.

2 The Realization Crisis

By way of introducing the realization crisis facing the CMH, consider first another hypothesis in whose favor the realization crisis has been resolved – namely, the computational theory of cognitition (CTC). The CTC is just the idea that cognition is constituted (some think entirely) by representation-transforming processes that involve the application of syntax-sensitive rules to sentential representations. McCulloch and Pitts provided one of the earliest indications of how a bridge might be built from neural machinations to strict computations with their description of how collections of neuron-like processing units might implement a set of logic gates and, ultimately, a universal Turing machine [2]. More recently, it has been shown that recurrent neural networks are (infinite memory notwithstanding) universal-Turing equivalent [3]. This sort of research has left few doubts that the brain is the sort of system that can, in principle, realize the sorts of strict (i.e., syntax-crunching) computations posited by proponents of the CTC.

In contrast, a satisfactory demonstration that neural machinations might realize the sorts of representations posited by the CMH has proven far more elusive. The CMH is, recall, the proposal that humans utilize non-sentential representations that are like scale models in crucial respects. Unfortunately, when attempts have been made to specify in precisely in what respects these putative cognitive models are like scale models, the resulting proposals have come out looking either too weak to adequately distinguish cognitive models from other sorts of representations or too strong to be compatible with basic brain facts.

A proposal of the latter sort is that the cognitive representations in question are like scale models in that they too are physically isomorphic with what they represent. Kosslyn, for example, has claimed, based upon the fact that certain areas of visual cortex exhibit retinotopy, that "[t]hese areas represent depictively in the most literal sense [. . .]" [4]. Unfortunately, any physical isomorphisms exhibited by these areas are highly distorted due, for instance, to the disproportionate amount of area devoted to the fovea. Moreover, such areas exhibit at best 2-D isomorphisms, but in order for cognitive representations to play the role for which they are slated by most theories that invoke the CMH they generally need to exhibit isomorphisms with regard to

3-D spatial properties, not to mention kinematic and dynamic ones. In other words, in many cases there would, *per impossibile*, need to be literal buckets, balls, doors, and so forth, in head.

Other ways of fleshing out the notion of a cognitive model have proven too weak to adequately distinguish cognitive models from other sorts of representations. Some say, for instance, that cognitive models are representations that are *merely* isomorphic (i.e., isomorphic without any further restrictions) with what they represent. This seems to be what Craik had in mind when he claimed, "By a model we thus mean any physical or chemical system which has a similar relation-structure to that of the process it imitates" [5]. However, a very wide range of representations will count as models on this view, including those created using the notations of formal logic and, relatedly, those harbored by production systems. Craik's own example of a representational system with a similar relation structure to what it represents – namely, Kelvin's Tide Predictor[1] – does little to narrow down the relevant field. If anything it helps prove the point that in order to support truth-preserving inferences regarding a represented domain a representational system *of any sort* must exhibit isomorphisms with that domain.

Also widely regarded as too weak to distinguish cognitive models from other sorts of representations is the proposal that cognitive models are *functionally* isomorphic with what they represent – that is, that they function in the ways that physically isomorphic representations such as scale models function. Unfortunately, it was recognized some time ago that one may always constrain a syntax-crunching system so that it functions like a scale model [1, 6, 7, 8]. Case in point, computational matrix representations have been constructed that function in the ways that 2-D and 3-D spatial matrix representations function (e.g., in terms of how changes in relative location are updated). However, accepted wisdom has it that these are computational representations in the strict sense described above and, accordingly, that they should be considered sentential representations in good standing. As Block puts it, "[o]nce we see what the computer does, we realize that the representation of the line is descriptional" [9]. In other words, the received view, which has gone nearly unchallenged, is that if a representation of spatial, kinematic, or dynamic properties (let us call these *corporeal* properties) is implemented using a high-level computer program, then it must be sentential in character [10, 11, 12].

In sum, proponents of the CMH have, as yet, failed to adequately articulate what makes cognitive models distinct from other sorts of representations in a way that is compatible with basic brain facts. This concern, which has been forcefully articulated by Pylyshyn and other proponents of the CTC, by no means entails that research in the CMH tradition should grind to a halt. However, it does detract from the relative plausibility of theories that invoke the CMH (e.g., as compared to those that invoke the CTC). As a proponent

[1] See `http://www.math.sunysb.edu/~{}tony/tides/machines.html#kp1` (last accessed January 9, 2010).

of the CMH, I found it important to search for a favorable way of resolving this realization crisis facing the CMH.

3 The Realization Story

What soon became apparent was that, with their insistence that computational implementation entails sentential representation, proponents of the CTC were turning their backs on the principle of property independence (POPI), a principle that more than any other helped pave the way for a favorable resolution of the realization crisis facing their own position.

POPI: Properties found to characterize a system when it is studied at a relatively low level of abstraction are often absent when it is studied at a higher level, and vice versa.

This principle is in large part what justifies one in saying that at level n a certain system contains a set of electronic switches, relays, and so forth, that at level $n+1$ it is best described in terms of the storing of bits of information in numerically addressable memory registers, and that at level $n+2$ it is best understood in terms of the application of syntax-sensitive rules to syntactically structured representations. However, nothing about POPI entails that at the *highest* level these systems must be characterized in terms of sentences and inference rules. Indeed, POPI opens up at least logical space for systems that engage in syntax crunching at one level and that harbor and manipulate non-sentential models at a higher level. As it turns out, in this logical space reside actual systems, such as those that implement finite element models (FEMs). These computationally realized representations of actual and possible physical conditions were first developed in the physical (e.g., civil and mechanical) engineering disciplines, but now they are used in a variety of fields for purposes of testing designs, exploring the ramifications of theories, generating novel predictions, and facilitating understanding. What is important about FEMs for our purposes is that they constitute an existence proof that computational processes can realize non-sentential representations that are like scale models and unlike sentential representations in crucial respects.

To see why, notice first that there are at least two levels of abstraction at which a given FEM may be understood. Just as with scale models, there is, to start with, the relatively low level of the modeling medium. In the case of FEMs, at this level what one finds are sentential specifications of coordinates (e.g., polygon vertices) along with rules constraining how they may change (e.g., due to collisions and loads) (see Figure 1). This, clearly, is the level upon which enemies of the idea of computationally realized non-sentential models are fixating when they suggest that computational realization entails sentential representation. It is, however, not obvious that at this level what we are even dealing with are representations (i.e., of worldly objects and properties), any more than we are when, for instance, we fixate on the constraints

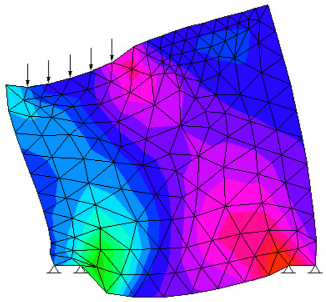

Fig. 1 Polymesh representation of a blunt impact to a semi-rigid sheet of material.

governing the behavior of individual Lego blocks. Rather, the representations are to be found when we turn our attention to the higher level of the models that are realized, *and multiply realizable*, by such modeling media. Moreover, when we take a closer look at the properties of the high-level FEMs realized through massive amounts of low-level number crunching, we find that they share several characteristics that are – most notably, by those who suggest that computational implementation entails sentential representation – taken to distinguish sentential representations from scale models.

To start with, scale models are often said to be distinct from sentential representations in that, *taken by themselves*, the former are all but incapable of representing either non-concrete properties (e.g., the property of being a war criminal), genera (e.g., triangularity), or the singling out of specific properties of specific objects (e.g., the mere fact that Joe's house is blue)[2]. Taken by themselves FEMs suffer from these exact same limitations. Like a scale model, an FEM is always a representation of a specific, concrete object, though it can be used as a proxy for many different objects of the same type. In addition, it always represents many properties of a given object rather than singling out any specific one. Thus, by these reasonable standards, FEMs ought to be considered computationally-*realized* non-sentential models that are at least the close kin of scale models.

What is far more important for present purposes is that representational genera are also often distinguished in terms of whether they constitute extrinsic or intrinsic representations. This distinction, which was introduced by Palmer [1], has been picked up on by opponents of the idea that computers realize non-sentential models. On Palmer's view, extrinsic representations

[2] In Waskan [17] I argue, without undermining the present point, that these representational limitations might be overcome in the human case through the use of extra-representational cognitive resources, such as those implicated in emotional, attentional, and analogical processes. On this view, the sentences with which we give voice to mental states are, somewhat as Dennett and Churchland would have it, compressed representations of a far more complex underlying cognitive reality.

are those that need to be arbitrarily constrained in order to respect the non-arbitrary constraints governing their represented systems, whereas intrinsic representations do not need to be so constrained. Sentence-and-rule-based representations (e.g., formal-logic and production-system representations) are thought to best exemplify the former [11, 13]. Scale models are thought to best exemplify the latter. While this distinction does get at an important difference between the two sorts of representation, it has unfortunately been drawn in a way that leans far too heavily upon the unclear notion of an arbitrary constraint. However, it is possible to preserve the key intuitions behind it in a way that does away with the questionable appeal to arbitrary vs. non-arbitrary constraints.

A better way to draw the distinction is to take as extrinsic those representations that only support predictions concerning particular types of alterations to a represented systems on the basis of distinct data structures. Extrinsic representations are those wherein the consequences of different types of alteration must be spelt out and built in – that is, by hand, learning, or evolution – antecedently and explicitly. For instance, in order to predict the consequences of alterations to even a simple system, such as one containing a doorway, a bucket, and a ball, a production system must incorporate distinct statements or operators that represent the consequences of those alterations. With intrinsic representations, on the other hand, the consequences of different types of alterations can instead be determined on demand and as needed simply by manipulating the representation in the relevant ways and reading off the consequences. For instance, a scale model of the system containing a door, bucket, and a ball can be manipulated in countless ways in order to predict how the consequences of many distinct alterations might play out. In order to predict what happens when the bucket is placed over the ball and slid through the door, one simply carries out the corresponding alteration to the model. Thus, one need not incorporate information about the consequences of this, and countless other alterations antecedently and explicitly.

Now the received view is that FEMs and their brethren are extrinsic representations because the constraints governing how the coordinates of primitive modeling elements may change must be explicitly, antecedently, and arguably even arbitrarily imposed [13]. Indeed, at the level of coordinates and transformation rules nothing is gotten for free in the case of FEMs, for both the coordinate system and the constraints governing changes to vertex coordinates must be antecedently and explicitly imposed. However, once a modeling medium has been used to construct a suitable FEM of a collection of objects, it can be altered in any of countless ways in order to determine the (at least possible) consequences of the corresponding alterations to the represented objects. One can, for instance, use an FEM of the door, bucket, ball system to infer, among countless other things, what would happen were we to place the bucket over the ball and slide the bucket through the door, what would happen were the bucket used to throw the ball at the open doorway, what would happen were the air pressure dramatically decreased, and so on indefinitely [14]. The consequences of these alterations need not be anticipated and explicitly incorporated into the system. Indeed, the very point of constructing FEMs is to *find out* how a system will behave in light of whichever alterations an engineer or scientist can dream up.

Those who would contend that FEMs are, *qua* computational, necessarily extrinsic representations once again overlook the fact that there are multiple levels of abstraction at which a given FEM model can be understood. There is, to be sure, the relatively low level of the modeling medium, and, insofar as there are representations at this level at all, the representations in question are unquestionably extrinsic. It is clear that the goings-on at this level inspire the above-mentioned contention, but what is once again being overlooked is that there is also a higher level, a level at which one finds models of collections of objects. These models are, every bit as much as the scale models they were created to replace, unquestionably intrinsic representations. Thus, once again, by the very standards employed by critics of the idea that some computers realize non-sentential models, FEMs are like scale models and unlike paradigmatic sentential representations.

All of this is bears directly on the longstanding concern that there is no way to bridge the gap between neural machinations and the non-sentential models hypothesized by proponents of the CMH. What the foregoing makes clear is that computational systems can realize non-sentential models that share with scale models the main characteristics that have long been used, even by opponents of the CMH, to distinguish scale models from sentential representations. Insofar as one already thinks that the brain is capable, at least in principle, of realizing computational processes, then one must also agree that brains can realize non-sentential models. This, I submit, is not just the most promising, but also (as yet) the only satisfactory account whatsoever of how a set of electrochemical circuits might realize non-sentential models of the sort posited by proponents of the CMH[3]. These considerations, in turn, give a real boost to the credibility of the CMH and, by extension, to its many specific applications. Indeed, FEMs are generally intrinsic representations, not just of spatial properties, but of kinematic and dynamic properties as well, and so their hypothesized cognitive counterparts ought to be fully capable of playing the roles for which they are slated by the proponents of the CMH.

4 Applications

The foregoing realization story turns out to have ramifications for work in a number of different fields, ranging from artificial intelligence (A.I.) to the philosophy of science. Of particular concern here is the notorious, albeit somewhat ephemeral, frame problem.

[3] Elsewhere I claim that the underlying recipe that is typically followed when constructing FEMs suggests that there may also be a kind of Northwest Passage, one that takes us directly from neural goings-on to non-sentential models [17].

4.1 Artificial Intelligence

Though it first came to light as a consequence of early work in logic-inspired, sentence-and-rule based A.I., we shall see that there are good reasons for understanding the frame problem in a more generic way, as that of determining how, through finite means, a creature or device can come to have human-like knowledge of the consequences of alterations to world. This sort of knowledge often enables us to choose beneficial and avoid harmful courses of action, and it also often enables us to formulate creative solutions to the many challenges that we face.

The frame problem can be broken up into at least two component problems. One, the prediction problem, has to do with fact that we humans have ability to predict the (at least possible) consequences of countless alterations to the world [15]. For instance, with regard to the ball, bucket, door scenario discussed earlier, we all know (to some admittedly fallible degree) that were the bucket placed over the ball and moved through the doorway the ball would also move through the doorway. We also know what would happen were the bucket used to throw the ball through the doorway, and so on indefinitely. While our knowledge of such alterations is immense, we are here still only dealing with a quite limited, 'toy' world.

Another component of the frame problem – namely, the qualification problem [16] – has to do with the fact that we humans are also able to envision countless possible defeaters of specific predictions. For instance, what we actually know about placing the bucket over the ball and moving it through the doorway is actually far more complex than was described above, for what we really know is something like this: If the bucket is placed over the ball and moved through the doorway *and it is not the case that either* there is a hole in the floor, or there is a hole in the side of the bucket, or the ball is affixed to the floor, or what have you, then the ball will move through the doorway.

The prediction problem, as concerns sentence-and-rule-based A.I., is that when we try to endow a system with knowledge of the consequences of countless alterations to a given situation using, at the highest level, sentence-and-rule-based representations of objects, we find that that we must incorporate countless, separate data structures (statements or rules) for each alteration-consequence pair. This problem is compounded by the qualification problem, for in order to truly match what we know, each such statement or rule would also have to incorporate countless distinct qualifications. Put formally, a sentence-and-rule-based system would have to contain countless distinct, endlessly qualified statements or rules of the following form (S's represent starting conditions, A's alteration conditions, Q's qualifiers, and C's consequences)[4]:

[4] There are also possible problems having to do with the sets of S's, A's, and C's.

$$[(S_1 \,\&\, S_2 \,\&\, \ldots \,\&\, S_n) \,\&\, (A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n) \,\&\, (\sim Q_1 \,\&\, \sim Q_2 \,\&\, \ldots \,\&\, \sim Q_n)] \rightarrow (C_1 \,\&\, C_2 \,\&\, \ldots \,\&\, C_n)$$

Restricting ourselves to systems that employ, at the highest level, sentential representations of the world, these problems look to be insoluble. However, they also look insoluble for any approach on which high-level extrinsic representations carry the inferential load, whether it be sentences and rules or activation and weight vectors. There is simply too much knowledge for it to be explicitly encoded in any form.

These problems do, nevertheless, admit of a determinate computational solution [17]. The solution is to constrain syntax-crunching operations so that they realize modeling media from which can be built *intrinsic*, non-sentential models (e.g., FEMs) of mechanisms. A device that can construct such models of its environment and wield them as its core inference engine will be endowed with what might be termed *inferential productivity*, the capacity for boundless inferences through finite means. Such models are, we saw, like scale models in that they can be manipulated in any of countless ways in order to make inferences about how alterations to the world might play out and, by the same token, about the ways in which those consequences might be defeated. Admittedly, this solution does engender problems all its own, but they are far more tractable by comparison [18].

4.2 Psychology and Logic

The frame problem is actually even more generic than a mere problem facing A.I., for it must also be dealt with by any theory of how humans (and perhaps other creatures) are able engage in this sort of reasoning. Even in the human case, there is just too much knowledge for it all to be encoded explicitly. Indeed, centuries ago, European rationalists were so impressed by this ‘universal’ reasoning ability as to conclude that no mere machine (biological or otherwise) could possibly account for it. The human mind, they thought, had to be of non-corporeal origin.

Perhaps, however, the vast bulk of what we know about the consequences of worldly alterations is only *tacit*, which is to say that it is not stored explicitly anywhere in memory but is rather produced on demand and as needed. Take, for instance, the knowledge that virtually all of us possess about how an airship can (i.e., unless it is transparent, there is an elaborate set of mirrors, a cloaking device, etc.) prevent a flagpole from casting a shadow. We all possess this knowledge, though few of us have ever had occasion to encode it explicitly. One way to account for how we come to possess knowledge about this and countless other scenarios through finite biological means is to say that what we have is an ability to construct non-sentential, intrinsic models of objects and to manipulate them in relevant ways on demand and as needed. Indeed, this may be the first full-blown mechanical explanation for our 'universal' reasoning ability.

This suggests, in turn, that we humans possess another mode of monotonic reasoning apart from deduction. Deduction, it is well-known, is monotonic in that if we validly deduce some conclusion which turns out to be incorrect, some of the information from which the conclusion was derived must also be incorrect. But deduction is also formal in that representations of abstract logical particles and principles are what bear the inferential load; the specific contents consistently quantified over and connected drop out as largely irrelevant. Deduction is, of course, sometimes effected externally through the use of truth tables and formal logical notations. Many psychologists believe that deduction is also effected internally through cognitive counterparts to these external methods – namely, through so-called *mental models* [19] or a *mental logic* [20].

Consider, however, the sorts of spatial inferences we are able to make using external, intrinsic representations. For instance, suppose we know that Linus is about a 1/4th taller than Prior and Prior is about 1/4th taller than Mabel [7]. Using an intrinsic representation of their relative heights (e.g., broken matchsticks), we can effect some simple monotonic inferences, such as that Linus is taller than Mabel or that arranging them side-by-side with Linus in the middle would form a kind of pyramidal shape. We also use more sophisticated intrinsic models to make monotonic inferences about kinematic and dynamic happenings. Importantly, in all of these cases, insofar as our representations are accurate, our conclusions must be as well. Conversely, insofar as the conclusions reached on the basis of these models are inaccurate, so too must be the representations from which they were derived. This form of monotonic reasoning is, however, clearly not deductive in nature. It is not effected by abstracting away from specific contents and allowing representations of logical particles and principles to bear the inferential load. Instead, it is the representations of specific contents that bear this load. As yet, however, no name has been assigned to this non-formal mode of monotonic reasoning. To assign it one, let us call it *exduction* (*ex-* out + *duce-* lead). Exduction is obviously effected externally using scale models and, more recently, FEMs and the like. The above proposal is just that we also sometimes engage in exduction internally through the use of non-sentential, intrinsic cognitive models. Indeed, we have seen that there are good reasons for thinking that we do engage in this non-formal mode of monotonic reasoning internally.

If all of this is correct, then exduction must be added to our taxonomy of reasoning processes alongside deduction, both of which are to be classified as monotonic. Inductive generalization, analogical reasoning, and abduction, on the other hand, count as non-monotonic. It also bears mentioning that abduction (by which I mean inference to the best explanation) is, though non-monotonic, also unique in that it may be partly constituted by *any* of the other forms of reasoning. Indeed, matters are complicated further by the fact that explanations lie at the core of all abductive reasoning, and explanations may themselves involve reasoning of a certain sort.

4.3 Explanation

The idea that monotonic reasoning lies at the core of all explanations is not new, as it formed the basis for what for a long time was, and in some quarters still is, the dominant model of explanation – namely, the deductive-nomological (D-N) model[5]. This model fell out of favor in mainstream philosophy of science as problems with it began to accrue. Two of the best known were its seeming inability to account for statistical explanations and its failure to distinguish explanations from non-explanatory deductions. Even more germane to the present discussion, however, are the surplus meaning problem and the problem of provisos.

The first of these has to do with the fact that explanations have countless implications beyond the happenings that they explain. To take a non-scientific example, consider that a mechanic may explain why an automobile engine exhibits a loss of power in terms of its possessing faulty rings. On the D-N model, this explanation involves a deduction of the happening to be explained from information about laws and boundary conditions in something like the following manner:

- If an engine's cylinder has faulty rings, then the engine will exhibit a loss of power.
- One of the engine's cylinder has faulty rings.
- Therefore, the engine exhibits a loss of power.

However, even where one is able to provide a plausible-sounding D-N reconstruction of an explanation such as this one, such reconstructions seldom do justice to the full complexity of the explanations they represent. Consider, for instance, that what the mechanic knows is not only that the faulty rings will result in a loss of power, but the many other implications of his explanation being correct, such as that oil will leak into the combustion chamber, the exhaust will look smoky, the end of the tailpipe will become oily, the spark-plugs will turn dark, replacing the rings will restore power, replacing the filter will not restore power, and so on indefinitely. Any suitable reconstruction of the explanation must thus imply not only the *explanandum*, but countless other things as well. The problem with the D-N model is that it relies on an extrinsic representational scheme, and so no D-N reconstruction can embody all of an explanation's *surplus meaning* [21]. The problem here looms especially large given that these additional explanatory implications are not idle justificatory bystanders. They are what we largely rely upon when assessing the adequacy of explanations.

---

[5] Admittedly, the D-N model was not meant to be in any way psychological, though elsewhere this view has been contested [17].

Making matters worse, one who possesses an explanation such as this one also knows of the countless ways in which each of its countless implications is qualified. The mechanic, for instance, knows that bad rings will lead to a loss of power, but only if the engine is not augmented with an NO2 supply, the other cylinders are not bored out to a higher displacement, and so on. The D-N model is, however, no more able to account for this kind of knowledge than is any other theory that relies upon extrinsic representational apparatus. What makes this problem of *provisos* [22] especially troubling is that, as Quine famously noted, the knowledge at issue here is what enables us to hang on to our explanations in the face of unruly evidence.

Though no alternative theory of explanation has yet proven capable of filling the substantial void left by the D-N model's demise, the mechanistic approach to explanation is increasingly viewed as a promising contender [23, 24]. This approach was pioneered in large part by Salmon, who claimed that explanations are to be identified with the objective mechanisms at work in the world. On his view, an explanatory mechanism is roughly just an arrangement of parts that act and interact so as to collectively yield the happening in question. One limitation of this *ontic* version of the mechanistic approach is that it fails to allow for the possibility of explanations that are either right or wrong, good or bad. Nor, therefore, does it leave room for a process of inference to the best explanation. This limitation is overcome by adopting a psychologistic version of the approach [25]. Broadly speaking, according to the psycho-mechanistic approach, to have an explanation is to have the belief that a certain mechanism is, or may be, responsible for producing some happening, where such beliefs are constituted by mental representations of those mechanisms. It is largely in virtue of our awareness of the information conveyed by these representations that events and physical regularities are rendered intelligible.

The specific variant of the psycho-mechanistic approach suggested by the foregoing is that the mental representations in question are intrinsic cognitive models. It should now be clear that this *model model* of explanation can overcome such limitations of its deductive counterpart as the surplus meaning problem and the problem of provisos. Though these problems were discovered on quite independent grounds by philosophers of science, they are just variants on the prediction and qualification problems uncovered through work in deductive-logic-inspired A.I. Accordingly, the same solution seems to apply – namely, to eschew the appeal to formal-deductive reasoning processes in favor of an appeal to exductive reasoning effected through the manipulation of non-sentential, intrinsic cognitive models. On this view, our exductive inferences give us explicit knowledge of the mechanisms by which a happening may have been produced, but, constituted as they are by intrinsic models, they also endow us with boundless tacit knowledge of an explanatory mechanism's further implications and of the countless ways in which those implications are qualified. This knowledge is, once again, what enables us to determine the testable implications of our explanations and to hang onto those explanations come what may, all of which is essential to forward progress in science [26].

5 Conclusion

The CMH has long been plagued by the concerns about the in-principle neurological plausibility of appeals to non-sentential cognitive models. I have proposed here a particular way of addressing these concerns according to which models that share central and important characteristics with scale models piggy-back atop a thick layer of strict computational processing and, in turn, upon a neurophysiological bedrock. Further research is needed in order to show that, and precisely how, the human brain implements cognitive models, and it is to be expected this research will reveal other ways of bridging the gap between brain and model, and to refine our views about the differences between cognitive models and scale models. I would caution, however, that in order to be considered true extensions of the CMH the central characteristics of non-sentential models discussed here will need to be preserved.

References

1. Pamer, S.: Fundamental aspects of cognitive representation. In: Rosch, E., Lloyd, B. (eds.) Cognition and Categorization, pp. 259–303. Lawrence Erlbaum Associates, Hillsdale (1978)
2. McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 113–115 (1943)
3. Franklin, S., Garzon, M.: Computation by discrete neural nets. In: Smolensky, P., Mozer, M., Rumelhart, D. (eds.) Mathematical Perspectives on Neural Networks, pp. 41–84. Lawrence Earlbaum Associates, Mahwah (1996)
4. Kosslyn, S.M.: Image and Brain: The Resolution of the Imagery Debate. The MIT Press, Cambridge (1994)
5. Craik, K.J.W.: The Nature of Explanation. Cambridge University Press, Cambridge (1952)
6. Shepard, R.N., Chipman, S.: Second-order isomorphism of internal representations: Shapes of states. Cognitive Psychology 1, 1–17 (1970)
7. Huttenlocher, J., Higgins, E.T., Clark, H.: Adjectives, comparatives, and syllogisms. Psychological Review 78, 487–514 (1971)
8. Anderson, J.R.: Arguments concerning representations for mental imagery. Psychological Review 85, 249–277 (1978)
9. Block, N.: Mental pictures and cognitive science. In: Lycan, W.G. (ed.) Mind and Cognition, pp. 577–606. Basil Blackwell, Cambridge (1990)
10. Pylyshyn, Z.W.: Computation and Cognition: Toward a Foundation for Cognitive Science. The MIT Press, Cambridge (1984)
11. Sterelny, K.: The imagery debate. In: Lycan, W.G. (ed.) Mind and Cognition, pp. 607–626. Basil Blackwell, Cambridge (1990)
12. Fodor, J.A.: The Mind Doesn't Work That Way. The MIT Press, Cambridge (2000)

13. Pylyshyn, Z.W.: Mental imagery: In search of a theory. Behavioral and Brain Sciences 25, 157–182 (2002)
14. Waskan, J.A.: Intrinsic cognitive models. Cognitive Science 27, 259–283 (2003)
15. Janlert, L.: The frame problem: Freedom or stability? With pictures we can have both. In: Ford, K.M., Pylyshyn, Z.W. (eds.) The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence, pp. 35–48. Ablex Publishing, Norwood (1996)
16. McCarthy, J.: Applications of circumscription to formalizing common-sense knowledge. Artificial Intelligence 28, 86–116 (1986)
17. Waskan, J.: Models and Cognition. The MIT Press, Cambridge (2006)
18. Waskan, J.: A virtual solution to the frame problem. In: Proceedings of the First IEEE-RAS International Conference on Humanoid Robots. Electronic only (2000), `https://netfiles.uiuc.edu/waskan/www/77.pdf`
19. Johnson-Laird, P.N., Byrne, R.M.J.: Deduction. Lawrence Erlbaum Associates, Hillsdale (1991)
20. Rips, L.J.: Cognitive processes in propositional reasoning. Psychological Review 90, 38–71 (1983)
21. MacCorquodale, K., Meehl, P.E.: On a distinction between hypothetical constructs and intervening variables. Psychological Review 55, 95–107 (1948)
22. Hempel, C.G.: Provisoes: A problem concerning the inferential function of scientific theories. Erkenninis 28, 147–164 (1988)
23. Salmon, W.: Scientific Explanation and the Causal Structure of the World. Princeton University Press, Princeton (1984)
24. Bechtel, W., Richardson, R.C.: Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research. Princeton University Press, Princeton (1993)
25. Waskan, J.: Knowledge of counterfactual interventions through cognitive models of mechanisms. International Studies in Philosophy of Science 22, 259–275 (2008)
26. Lakatos, I.: Falsification and the methodology scientific research programmes. In: Lakatos, I., Musgrave, A. (eds.) Criticism and the Growth of Knowledge, pp. 91–195. Cambridge University Press, Cambridge (1970)