Metadata of the chapter that will be visualized online

Chapter Title	Explaining Capacities: Assessing the Explanatory Power of Models in the Cognitive Sciences	
Copyright Year	2014	
Copyright Holder	Springer Science+Business Media Dordrecht	
Corresponding Author	Family Name	**Gervais**
	Particle	
	Given Name	**Raoul**
	Suffix	
	Division	Centre for Logic and Philosophy of Science
	Organization	Ghent University
	Address	Blandijnberg 2, 9000, Ghent, Belgium
	Email	Raoul.Gervais@UGent.be
Abstract	It has been argued that only those models that describe the actual mechanisms responsible for a given cognitive capacity are genuinely explanatory. On this account, descriptive accuracy is necessary for explanatory power. This means that mechanistic models, which include reference to the components of the actual mechanism responsible for a given capacity, are explanatorily superior to functional models, which decompose a capacity into a number of sub-capacities without specifying the actual realizers. I argue against this view by considering models in engineering contexts. Here, other considerations besides descriptive accuracy play a role. Often, the goal of performance trumps that of accuracy, and researchers are interested in how cognitive capacities *as such* can be realized, rather than how it is realized in a given system.	
Keywords (separated by "-")	Explaining capacities - Cognitive sciences - Mechanistic models - Functional models	

Chapter 3
Explaining Capacities: Assessing the Explanatory Power of Models in the Cognitive Sciences

Raoul Gervais

3.1 Introduction

As Robert Cummins notes, capacities are an important type of explanandum addressed by psychologists (Cummins 2000). In fact, this does not only hold with respect to psychology, but seems to apply in equal measure to the other disciplines that fall under the label 'cognitive sciences'. All kind of cognitive capacities are in need of explanation, from face recognition to the ability to play chess; from motor skills to language acquisition. Now whereas most other types of explanandum (events, occurrences, states of affairs etc.) are, at least intuitively, explained by identifying their *causes*, capacities are typically explained in terms of a *model*.[1,2] To put the difference between these two explanations in pragmatic or erotetic terms, the former are answers to why-questions (Van Fraassen 1980), the latter to how-questions.[3]

[1]Throughout this paper, the term 'model' is used in a loose sense, to encompass any schema that mimics a certain pattern of behaviour that constitutes the explanandum. Of course, not all such models are scientifically or even philosophically interesting. However, in what follows, some specific *types* of models that *are* of interest will be considered in more detail.

[2]Of course, this is not to say that models cannot be causal in themselves, or that we cannot model causes. Rather, the difference is that the explanation of an event, occurrence or state of affairs typically refers to the cause of that event, occurrence or state of affairs, while the explanation of a capacity refers to a model, which may include *descriptions or simulations* of causes, but not the actual cause responsible for the capacity. In the former case, the explanans is located in reality, in the latter, it is a description or simulation of the cause, not the cause itself that does the explaining.

[3]This is not to say that one cannot ask how-questions about events, or why-questions about capacities (evolutionary explanations of biological traits provide examples of the latter strategy).

R. Gervais (✉)
Centre for Logic and Philosophy of Science, Ghent University,
Blandijnberg 2, 9000 Ghent, Belgium
e-mail: Raoul.Gervais@UGent.be

E. Weber et al. (eds.), *Logic, Reasoning, and Rationality*, Logic, Argumentation & Reasoning 5, DOI 10.1007/978-94-017-9011-6_3,

In the cognitive sciences, two types of model are used to explain capacities: functional and mechanistic models. Functional models explain capacities by decomposing them into ever smaller sub-capacities or -routines, and then attempt to show how the overall capacity arises as a result of the way these sub-routines are organized (a useful metaphor here is that of the assembly line, where a complex task is divided into several simpler ones). These functional models can be highly abstract, putting more emphasis on the function to be performed than what actually performs it. Mechanistic models on the other hand, are less abstract. They too involve decomposing a capacity into a hierarchy of sub-functions or -capacities, but also include data on what type of entity is actually responsible for this or that (sub-)function (I will explain these two types of models in more detail in Sect. 3.2).

According to some authors, mechanistic models are superior to functional models precisely because they incorporate this additional information. While the latter are merely loose conjectures, the former are, at least in the ideal case, complete descriptions of the mechanism responsible for the explanandum. Indeed, Craver goes so far as to say that only to the degree it describes the actual entities by means of which a mechanism performs a capacity, can the model be said to *explain* that capacity (Craver 2006). Functional models can be useful for the purposes of prediction and control (they can successfully map the input-output patterns of the target system) but explanation requires something further. In the case of cognitive capacities, the model should at least be somewhat accurate ('plausible') from a neurophysiological point of view, if it is to explain those capacities. In short, it seems that on this view, *accuracy with regard to a mechanism's components is necessary for a model to have explanatory power*.

In this paper, I will argue against this view. Of course Craver is right in stating that in cases where we try to explain a capacity as it is realized in some particular system (which, of course, is what Craver and the mechanists in general are interested in), mere phenomenal models are not explanatory. However, this conclusion does not carry over to models in general: it is not correct to claim that descriptive accuracy is necessary in every context. The argument I present takes the form of a reductio: if it were necessary, this would exclude a whole range of models that are not only useful in the phenomenal sense (for the purposes of control or prediction), but intuitively also have explanatory power. These models are found in the context of *engineering*. A particularly promising way to account for these models is to employ the pragmatic perspective on explanation I hinted at above. We should realize that models need not be answers to how-questions relative to some set of systems S, but can also answer how-questions about capacities *as such*. The picture that emerges suggests that explaining capacities is a much more dynamic affair—consequently, a simple insistence on descriptive accuracy is too simplistic and does no justice to scientific practice.

The point is simply that in the cognitive sciences, explaining how a capacity comes about by constructing a model is simply a very prominent research strategy, as we shall see, which makes it philosophically interesting.

3.2 Functional Versus Mechanistic Explanations

Traditional functional explanations work by decomposition. They explain a capacity by breaking it down into sub-capacities or -functions, and then show how the overall capacity is a result of the organization of these sub-functions. Returning to the metaphor of the assembly line, let us consider a factory churning out radios. This factory effectively performs the function of taking parts as input and producing radios as output. This function can be explained by dividing the assembly process into several sub-routines carried out by workers standing alongside a conveyor belt, where each subsequent worker adds a specific component to the radio, until the finished product appears at the end of the belt, ready for transport. Once we know all the sub-routines that make up the assembly process, and understand the way they are organized (the order in which the parts are added) we can explain how the factory performs its function by means of a flow-chart or box diagram.

This explanatory strategy was widely used in the cognitive sciences, especially in the 1980s and 1990s. Cognitive capacities like memory storage, face recognition and numerical cognition were explained by construing models of how these capacities might be divided up into sub-functions. In psycholinguistics for example, a particularly influential functional model for the capacity of speech production was offered by Levelt (1989). Roughly, the process was divided into three steps: first, the person conceptualizes what he wants to say, second, he formulates this into language (this step is in turn divided into two sub-tasks, one of lexicalization, which produces the words needed, and one of syntactic planning, which provides order and grammatical structuring to these words) and finally, he engages in articulation (see Fig. 3.1).

Of course, this is a rough sketch of how the capacity might be realized, but it need not be wholly speculative. For example, the distinction between lexicalization and syntactic planning may be grounded in experimental evidence: some test subjects might be able to produce the right words, but fail to put them in the correct order. In general then, functional models need not be merely phenomenal (input-output mapping devices): with respect to the partitioning of a capacity into sub-routines,

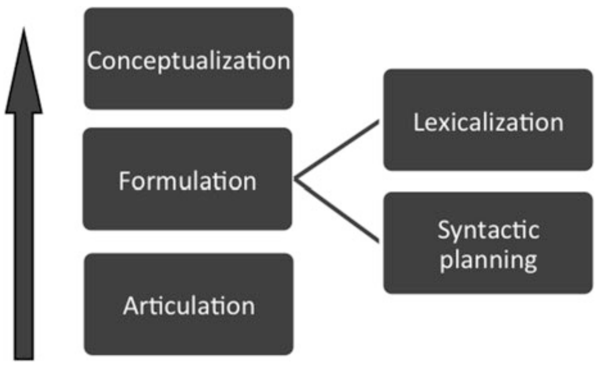

Fig. 3.1 An adaptation of Levelt's functional model of speech production. The sequence of operations is depicted vertically, with one step (formulation) being divided into two sub-routines (lexicalization and syntactic planning)

one can be detailed or abstract, and this partitioning might be supported by experimental evidence to a greater or lesser degree.

Yet however much informed a functional model like this might be, there is one issue with respect to which it remains silent: it has nothing to say about what actually performs all these sub-tasks. To put the point differently, it specifies functions, but not the realizers of these functions. In the example of the assembly line, imagine that in another factory, the different assembly tasks are realized by robots instead of workers. From a certain level of abstraction, the two factories are functionally equivalent, as they both perform the function of taking in parts as input and producing radios as output. More formally, if we want to explain a capacity C of a system S, we have to construct a functional model M which performs C, such that for each input, output and input-output relation in S there is a corresponding input, output and input-output relation in M.

In philosophy, this abstraction from what performs a function is often paired with the thesis of multiple realizability, and has been a key motivator to argue in favour of the autonomy of the special sciences (Fodor 1981). However, what was once hailed as an advantage is now increasingly criticised as a weakness. To be sure, functional models may succeed in correctly mapping the input-output relation of the target system, and for the purposes of control or prediction this may suffice, but does that make the model explanatory? Even though a particular partitioning of a function into subroutines is supported by evidence, if we want to understand how we, as humans, perform some kind of cognitive capacity, it seems imperative that we know something of the brain regions involved. Too often, the critics say, researchers are at a loss about what is really behind the boxes in their diagrams. For heuristic purposes, e.g. when we are just mapping out a certain capacity, this may be fine,[4] but if the original status of these boxes as mere placeholders is forgotten, they only serve to mask gaps in our understanding (hence the derogatory term 'boxology' that is sometimes applied to pure functional analysis).

In any case, a growing body of literature is devoted to an alternative approach to explaining cognitive capacities: mechanistic explanations. Like functional explanations, mechanistic explanations decompose the target capacity into several sub-capacities. Unlike functional explanations however, mechanistic explanations also incorporate information about *what* performs a certain (sub-)function. They explain a capacity of a system by modelling the mechanism responsible for it: its operations, its entities or parts and the way the operation and parts are organized come into play.[5] Of course, this model need not be a complete description of the mechanism.

[4]See for example Machamer et al., who write that a mechanistic explanation typically starts by providing a mechanism sketch, which is "…an abstraction for which bottom out entities and activities cannot (yet) be supplied or which contains gaps in its stages. The productive continuity from one stage to the next has missing pieces, black boxes, which we do not yet know how to fill in" (Machamer et al. 2000, p. 18).

[5]Another way to put the difference is that mechanistic explanations, besides decomposition, also involve localization, where the latter notion is understood as the identification of activities with parts (Bechtel and Richardson 1993).

Ideally complete descriptions only serve as a regulative ideal: the degree of completeness required depends on our purposes at the time.

So far so good. But some authors do not stop at that. They believe that if our purposes are *explanatory*, then the model cannot afford to remain silent about the parts or entities of a mechanism:

> In order to explain a phenomenon, it is insufficient merely to characterize the phenomenon and to describe the behavior of some underlying mechanism. It is required, in addition, that the components described in the model should correspond to components in the mechanism. . . . (Craver 2006, p. 361)

Note that in this quote, Craver no longer talks about capacities as they are realized by humans, or indeed by any specific system: the claim he makes is about explaining 'a phenomenon', that is, about the explanatory power of models in general, not as they apply to any particular system. Thus Craver seems to endorse the following thesis:

(T) *For a model to have explanatory power, it is necessary that it corresponds to the target system, both with respect to its operations and the parts carrying out these operations.*

Now I agree that if we want to explain a capacity *as it is performed by some system or set of systems*, we must say something about the parts or components involved and, what is more, what we say should be correct. That is, the accuracy of the model should extend beyond the input-output relations to the actual mechanism itself. However, if from this concession **T** follows, we are in trouble, for not only do the traditional functional models described above not give accurate descriptions of a system's components, they typically remain silent about them altogether! According to **T** then, purely functional models are not explanatory. Nevertheless, from the 1970s onward, they have been used in cognitive psychology to explain all kind of capacities. With this discrepancy in mind, in Sect. 3.3, I will try to account for explanatory, yet purely functional models by considering some pragmatic aspects of explanation, while in Sect. 3.4, I will give an example of an explanatory context in which these aspects typically play a role.

3.3 Pragmatic Aspects of Explanation Considered

Although traditional functional models like the one sketched above are more abstract than mechanistic explanations in that they remain silent about a system's components, it would be wrong to infer from this that they have no explanatory power at all. To make this point, I will turn to a pragmatic account of explanation. The account I shall develop is pragmatic in the sense that it elaborates on van Fraassen's erotetic model of explanation.

According to van Fraassen, explanations are answers to why-questions (Van Fraassen 1980). However, as I have mentioned in the introduction, when dealing with *capacities*, it is often more appropriate to say that explanations are answers to how-questions. Fortunately, it has been argued persuasively that how-questions are valid explanation-seeking questions in their own right (Scriven 1962; Salmon 1989). Again, while answers to the former typically consist of identifying or referring to causes, the answers to the latter take the form of models. Recall how functional models work: if we want to explain a capacity C of a system S, we have to construct a functional model M which performs C, such that for each input, output and input-output relation in S there is a corresponding input, output and input-output relation in M. That is, if we want to answer a question like:

(1) How is C realized in S?

we should construct a model M that maps the input-output relations that make up C. Having done that, we can answer (1) by saying:

(2) C is realized in S the same way that C is realized in M.

Note that although it looks like (2) just restates the mystery, it does not, for we must remember that M is not a mechanism or system in nature, but a model that we have constructed ourselves, so that we know in detail how it realizes C. However, and this is where I agree with Craver, the question seems to ask something beyond input-output mapping. For a simple example, consider:

(3) How is the capacity to recognize faces realized in the human brain?

Now some face-recognition systems have been developed that perform this capacity very well, in that they are able, in experimental setups, to map the input-output relations of the brain (they are presented with examples of faces and non-faces and are able to tell the difference with more or less the same degree of accuracy as humans), but do so in a fundamentally different way. Up until recently for example, they could only use two-dimensional geometrical data. Of course we do not want to count:

(4) The capacity to recognize faces is realized in the human brain by applying algorithms to exclusively 2-D geometrical data.

as an answer to (3). As we know ourselves to see, e.g., chins and noses as protrusions, (4) is clearly inaccurate. Beyond this appeal to 'first person knowledge' however, there is also some 'harder' evidence. For example, 2-D face systems notoriously suffer from what is known as the 'lighting problem': their ability to recognize faces deteriorates significantly when the strength of the light coming from the image they are presented with is varied, while humans tend to retain their abilities in such circumstances. No matter how perfectly such systems may mimic our performance in this task, we have to concede that, being 2-D, they are not explanatory models for face recognition as it is performed by humans.

Granted then, a model may to a certain extent map the human input-output relation for a capacity, without being explanatory with respect to the human realization of that capacity. However, **T** makes a stronger claim than that. Craver went beyond models for capacities as they are performed by humans or systems, to claim that *any* model that does not offer an adequate description of a system's components has no explanatory power. But do models always have to be models of a capacity as it is performed in a specific (set of) system(s)? The erotetic approach we have explored so far says that if a capacity is the explanandum, the explanans can be viewed as an answer to a how-question. There is nothing to restrict this type of question to include only capacities *as they are realized in some system*, we can also ask how-questions about capacities *as such*, that is, without any particular descriptive or correspondence constraints. Instead of (1), we might ask:

(5) How is C (as such) possible?[6]

The point here is not that researchers will actually be interested in how capacities could be realized without *any* constraints: capacities are of course always realized in some system. Rather, the point is that one can have legitimate motives in placing *as little constraints on the system as possible*. In Sect. 3.4, I will consider one context in which this strategy is commonplace, namely the context of engineering. For now, note that at least in psychology and the cognitive sciences, asking explanatory questions about capacities as such forms an important part of scientific practice, if only as a preliminary strategy (that is, preliminary to the business of answering the question how the capacity is realized in some particular system). In fact, this was AQ1 already noted by Dennett back in 1978:

> Faced with the practical impossibility of answering the empirical questions of psychology by brute inspection (how *in fact* does the nervous system accomplish X or Y or Z), psychologists ask themselves an easier preliminary question: How could any system (ldots) possibly accomplish X? This question is easier because it is 'less empirical'; it is an engineering question, a quest for a solution (*any* solution) rather than a discovery. (…) Seeking an answer to such a question can sometimes lead to the discovery of general constraints on all solutions (…), and therein lies the value of this style of aprioristic theorizing. (…). For instance, one can ask how any neuronal network with such-and-such physical features could possibly accomplish human color discriminations (…). Or, one can ask, with Kant, how anything at all could possibly experience or know anything at all. Pure epistemology, thus viewed (…) is simply the limiting case of the psychologist's quest. (Dennett 1978, pp. 110–111)

Thus viewed, the 'Kantian' question (How is X possible at all?) can be interpreted as constituting the extreme end of a continuum, while enquiries about how a particular AQ2 system performs that function occupies the opposite end (Fig. 3.2).

As Dennett notes, it is possible to begin with more general questions, discovering constraints having to do more with C itself, and work your way to a particular realization of C in S. However, explanation can also work in the opposite direction.

[6] Note that this question does not fall into the category of Craver's how-possibly questions (Craver 2006). For Craver, how-possibly questions are loose inquiries that are made in the early stages of an investigation, in which a lot of data is still missing: they are attempts to put some initial constraints on the explanandum, prior to constructing a more informed (how-plausibly), and ultimately ideally complete description (how-actually). Nevertheless, how-possibly questions in Craver's sense are still asked with respect to a capacity as it is performed by some system. The question under consideration differs because it is asked about a capacity *as such*, regardless of any particular realization.

How does S perform C?   How is C performed in S and $S^{1\ldots}$?   How is C possible at all?

Fig. 3.2 Different levels of abstraction at which one might seek to explain a capacity

As one moves to the right of the spectrum, the number of constraints will decrease. This means that somewhere along the line you get to the point where C is described in such a general way that it applies to more than one system. In other words, the scope increases. Examples of this can be found in medicine. If an impaired capacity in a brain damaged patient has somehow been restored by the brain, we might be interested to know just exactly how that capacity is carried out in this damaged brain. In circumstances like these, we are actually looking to move toward the right end of the spectrum. Of course, detail matters: as soon as we reach the point where all the relevant systems fall under the scope of that capacity, we stop. In the example, as soon as we have described the capacity in such general terms that it applies both to healthy patients and the brain damaged patient, we stop jettisoning constraints. This stopping has to do with our methodological interests: it is simply the act of eliminating variables.[7]

In Sect. 3.4, I will give a more detailed example of this explanatory strategy. For now, the point to note here is that abstraction is a matter of degree. How many constraints one places on the system responsible for a certain capacity will be decided by pragmatic issues. This however, seems at odds with **T**, which endorses descriptive accuracy about implementational details as necessary for a model to have explanatory power. Of course, this is particularly striking for questions located near the right end of the spectrum: surely, one cannot expect a model answering (5) to excel in descriptive accuracy, for there is no mechanism specified to describe. In fact, *any* model of *any* system that realizes C is a valid answer. Again, scientists are rarely (if ever) interested in capacities under no constraint whatsoever. Nevertheless, the continuum sketched above suggests a more dynamic and more tolerant picture of model-explanation; a picture which **T**, with its simple assertion that descriptive accuracy about entities and parts is necessary for a model to have explanatory power, is too rigid to encompass.

3.4 Explaining Capacities in Engineering Contexts

Explanation-seeking how-questions about capacities as such are often asked in cases where the research is driven by engineering interests. In the case of the cognitive sciences for example, type (5) questions might arise in artificial intelligence. Let us consider one specific example of a cognitive capacity: exact calculation.

[7] Also, think of animal testing: here we continue to drop constraints until the capacity is described in such a way as to apply across species. Again, S can be any system, natural or artificial.

Humans are endowed with the capacity to perform exact calculations accurately, up to a certain level of complexity. If we ask how we perform this capacity, the model that answers this question indeed derives its explanatory power from (among other things) its neurophysiological accuracy. That is, if we want to answer:

(6) How is the capacity to perform exact calculations realized in humans?

the model that we use to answer (6) has to reproduce the capacity under a number of constraints. For example, some artificial computing devices might make poor models, as they are disanalogous to human brains in important respects: they might be neurophysiologically implausible, or they might fail to reproduce the capacity to perform exact calculations (e.g., they might be less exact, or they might take far longer to solve arithmetic problems).

However, although these respects are important to contexts like the one referred to in question (6), there are other contexts in which they are less important, or even irrelevant, and these other context might still have to do with explaining the capacity. In other words, descriptive accuracy or correspondence is not the only explanatory context in which we could be interested in the capacity: there are other reasons we might want to explain the capacity to perform exact calculations. Suppose an engineer wants to construct a desk calculator. Now of course, his goal is not to construct a model of how humans perform complex calculations: after all, he is designing a tool that, hopefully, surpasses our own ability. In fact, he seeks to *duplicate* the capacity. Motivated by this interest of duplication, he might ask:

(7) How is exact calculation as such possible?

However, this is somewhat artificial. In fact, when constructing a desk calculator, there are all kinds of constraints he needs to take into account.[8] The point is that these constraints are different from the ones applying to exact calculations as it is performed by humans. Thus, a sensible strategy would be to put fewer constraints on the capacity, until the scope is broad enough to apply to both humans and certain artificial devices. In terms of the continuum sketched above, we stop somewhere in the middle, at the point where the scope is just broad enough to encompass both the human realization of the capacity and an artificial one. To put it in other terms, we stop where the forces pulling in opposite directions, namely level of detail (to the left) and duplication (to the right), balance out for the task at hand.

But that is not all. In engineering contexts, it is not uncommon to jettison the requirement of descriptive accuracy completely. To appreciate this, let us continue to pursue the example of the engineer trying to construct his desk calculator. Now there are a number of models that can perform exact calculations. For reasons of clarity, let us consider classic computationalism and connectionism. The symbolic architecture of classic computationalism, where symbols are manipulated according

[8]Examples of such constraints are: the materials available, convenience of use and time considerations (we want the calculator to perform calculations rapidly—within a timeframe that is of use to us, that is).

to a pre-programmed set of rules, is very good at performing very complex calculations with great accuracy, far surpassing that of any human. On the other hand, as a model of the mind, computationalism is outdated. The serial nature of its operations and its consequent brittleness does not compare to the robustness of our brains. Connectionism on the other hand, resembles our brains more closely. In fact, in the original debate between computationalism and connectionism as candidate models for the mind, the latter's neural plausibility (in the form of distribution of activity over a network of nodes, graceful degradation, its ability to recognize patterns etc.) counted as an important point in its favour (McClelland and Rumelhart 1986).[9] However, despite all these advantages, they perform poorly when it comes to exact calculations. In fact, connectionist networks have been ridiculed for answering a question like "What is two plus two?", after much crunching, with "About four".

Clearly, exactness is a virtue when it comes to desk calculators. In fact, when engineering interests drive model construction, *performance trumps accuracy*. Duplication therefore, is only a subsidiary goal: it is really the desire to make a system that outperforms humans that motivates the engineer, and the model he finally constructs will reflect this. Of this model, that is of the flow chart representing how the calculator performs the exact calculations, we can say three things. First, with regard to how humans perform exact calculations, it is an inaccurate model and fails to explain it. Second, with regard to how the calculator performs it, it is an ideally complete description and explains it, but that is hardly surprising, since it is the very blueprint the engineer used to make the calculator in the first place. Third, with regard to the capacity to perform *exact calculations* as such, it explains how that capacity *can be* performed. When the engineer asked (7) and started decomposing exact calculation down into sub-routines, he was looking for an explanation, only not with neurophysiological accuracy on his mind, but performance.

Yet there are other interests besides duplication or performance that might prompt the search for an explanation of such capacities. Another interest is *unification*. Once an artificial system has been designed and constructed, then to anyone besides the engineers involved in this process of designing and construction, the explanatory question might arise as to what these artificial systems have in common with, e.g., natural systems. Again, the term 'system' has been chosen to reflect the fact that we might not only be interested in a capacity as performed by humans (or natural systems in general), but also by artificial ones. Thus, one might ask the following question:

(8) How is the capacity to perform exact calculations performed in this desk calculator and in humans?

This question is situated somewhere in the middle of the continuum presented in Sect. 3.3. In effect, what we are asking for here is what two realizations of the capacity of exact calculations have in common with each other. These comparative

[9] As the debate currently stands though, connectionist networks are considered to be highly idealized models too—but still more plausible than classic computationalist architectures.

question-types are often motivated by unification: in revealing features that are common to the operations of both types of systems, an answer to (8) brings together information from multiple and diverse sources. And of course, an answer to comparative question-types like (8) will typically take the form of a model—precisely the kind of functional model introduced in Sect. 3.2. In the case of question (8), this is especially clear, since any similarity between humans performing complex calculations and desk calculators exercising the same capacity will not be found in the entities, but will be confined solely to the domain of the operations. Yet, despite its abstract nature, and *pace* **T**, such a model would clearly be of explanatory value to those who are interested in the similarities between human and artificial performances of exact calculation.

Again, all this does not tarnish the explanatory importance of mechanistic models when it comes to explaining capacities as they are realized in particular systems. Of course we need the models of, e.g., biological functions to be accurate, and not only phenomenally adequate. It might even follow that for particular systems, this accuracy is necessary for a model to have any explanatory power regarding that capacity. What does not follow however, is that phenomenal and functional models have no explanatory power in *any* context. Reiterating Dennett's point, asking about capacities under fewer constraints can be a valuable research strategy. Ultimately, how many constraints one takes into account is decided by one's interests: in the case of performance, an interest typical of engineering contexts, these constraints will surely be determined by practical considerations, but not empirical adequacy. Nevertheless, this does not undermine the explanatory power of answers to such questions. Hence, it seems that Craver's thesis **T** is false as it stands. However, although strictly speaking correct, this conclusion should not be the main point to take away from this discussion, if only for the fact that Craver and the mechanists have a very different context in mind from some of the ones considered in this paper. Of greater importance is the observation, borne out by the continuum sketched in Sect. 3.3 and illustrated in this section, that the business of explaining capacities by constructing models is far more diverse and dynamic than Craver suggests. This more constructive conclusion might serve as a starting point to reformulate **T** in a way that either restricts its scope, so that it applies only to those contexts which Craver had in mind, or to drop the requirement of descriptive accuracy, so that it does justice to the practice of explaining capacities by constructing models.

3.5 Some Concluding Remarks

Two final remarks are in order. First, although distinct, engineering and accuracy interests are often present at the same time and can even be complementary. This is especially the case when a model has to be constructed of a capacity at which, unlike exact calculations, humans are particularly good. Face recognition for example, is a capacity in which we excel, and many of the early artificial systems badly underperformed compared to us, being sensitive to all kind of distortions (we already encountered the lighting problem, faces presented at angles is another one) that human test persons just see right through. In such cases of course, an engineer wanting to design such an artificial system has everything to gain by first asking how the capacity is realized in us. The point is though, that even here, accuracy is only a sub-goal. As soon as artificial systems are starting to equal or outperform us, engineers will drop accuracy as a goal, as it no longer serves the greater goal of performance.[10] Finally, one may wonder whether the capacities targeted by functional explanations in engineering contexts, such as the one described in Sect. 3.4, are still properly called *cognitive* capacities. Can we still talk of subtraction as a cognitive capacity when it is performed by a humble desk calculator instead of a person? Here, one might point out that the engineering sciences (artificial intelligence in particular) have a history of fruitful interaction with the cognitive sciences. Artificial systems can help us understand our own capacities, while knowledge of these may in turn lead engineers to improve the performance of these systems. After all, the point made in this article is that accuracy and explanatory power can, and in some cases do, operate separately from each other, not that they always do so.

Acknowledgements The research for this paper was supported by the Research Fund Flanders (FWO) through project nr. G.0031.09.

References

Bechtel, W., & Richardson, R. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Craver, C. (2006). When mechanistic models explain. *Synthese, 153*, 355–376.

Cummins, R. (2000). "How does it work?" versus "What are the laws?" Two conceptions of psychological explanations. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117–145). Cambridge: MIT.

Dennett, D. (1978). Artificial intelligence as philosophy and as psychology. In D. Dennett (Ed.), *Brainstorms* (Philosophical essays on mind and psychology, pp. 109–126). Montgomery: Bradford Books.

Fodor, J. (1981). Special sciences. In *Representations: Philosophical essays on the foundations of cognitive science* (pp. 127–145). Harvester: Hassocks.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: MIT.

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

McClelland, J., & Rumelhart, D. (1986). *Parallel distributed processing: Explorations in the micro-structure of cognition* (Vol. 2). Cambridge: MIT.

[10] And in fact, with the example of face recognition systems we considered earlier, this is beginning to happen right now; see the results from the 2006 Face Recognition Vendor Test (available for download at: http://www.frvt.org/).

Salmon, W. (1989). Four decades of scientific explanation. In P. Kitcher & W. Salmon (Eds.), *Minnesota studies in philosophy of science* (Vol. VIII, pp. 3–10). Minneapolis: University of Minnesota Press.

AQ3 Scriven, M. (1962). Explanation, predictions and laws. In H. Feigl & G. Maxwell (Eds.), *Scientific explanation, space and time* (Minnesota studies in the philosophy of science, Vol. III, pp. 170–229). Minneapolis: University of Minnesota Press.

Van Fraassen, B. (1980). *The scientific image*. Oxford: Clarendon Press.

AUTHOR QUERIES

AQ1. Please advise shall we change "(ldots)" as "(. . .)" in the sentence starting "Faced with the practical. . . ".

AQ2. Please check if inserted citation for "Fig. 3.2" is okay.

AQ3. Please check if updated publisher location for "Scriven (1962)" is okay.