Euro Jnl Phil Sci (2015) 5:349–375
DOI 10.1007/s13194-015-0111-3

ORIGINAL PAPER IN PHILOSOPHY OF SCIENCE

Mechanistic explanation in engineering science

Dingmar van Eck

Received: 7 October 2013 / Accepted: 28 February 2015 / Published online: 19 March 2015

Abstract In this paper I apply the mechanistic account of explanation to engineering science. I discuss two ways in which this extension offers further development of the mechanistic view. First, functional individuation of mechanisms in engineering science proceeds by means of two distinct sub types of role function, *behavior* function and *effect* function, rather than role function simpliciter. Second, it offers refined assessment of the explanatory power of mechanistic explanations. It is argued that in the context of malfunction explanations of technical systems, two key desiderata for mechanistic explanations, 'completeness and specificity' and 'abstraction', pull in opposite directions. I elaborate a novel explanatory desideratum to accommodate this explanatory context, dubbed 'local specificity and global abstraction', and further argue that it also holds for mechanistic explanations of malfunctions in the biological domain. The overall result is empirically-informed understanding of mechanistic explanation in engineering science, thus contributing to the ongoing project of understanding mechanistic explanation in novel or relatively unexplored domains. I illustrate these claims in terms of reverse engineering and malfunction explanations in engineering science.

Keywords Mechanistic explanation · Engineering science · Mechanistic role function · Behavior function · Effect function · Explanatory power

1 Introduction

The philosophy of mechanistic explanations is increasingly expanding its theatre of operations. Analyses have extended from the more 'traditional' domains—like (cell) biology, neuroscience and cognitive science—, and now also include fields like natural selection (Barros 2009; Mckay Illari and Williamson 2010), astrophysics (Mckay Illari and Williamson 2012), and technology (de Ridder 2006).[1] Moreover, explanations

[1]Natural selection is a debated area. Skipper and Milstein (2005), for instance, argue against mechanistic understanding of natural selection. Barros (2009) and McKay Illari and Williamson (2010) argue that natural selection can be explained mechanistically.

D. van Eck (✉)
Centre for Logic and Philosophy of Science, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium
e-mail: Dingmar.vanEck@Ugent.be

often draw on resources from multiple fields, cognitive neuroscience (Bechtel 2008a, b; Piccinini and Craver 2011) being a case in point. This expansion and, often, multi-field character of mechanistic explanations led to separate analyses on what mechanistic explanations have in common across fields (Mckay Illari and Williamson 2010, 2012).

One common key factor that is stressed in the construction of mechanistic explanations concerns the functional individuation of mechanisms in terms of role function ascriptions to activities of entities (Craver 2001, 2012a; Craver and Darden 2001; Mckay Illari and Williamson 2010; see Machamer et al. 2000; Craver 2007; Sustar 2007). In this paper I apply and further develop this view in the context of mechanistic explanation in engineering science. It is shown that in engineering science two distinct sub types of role function are invoked for the functional individuation of mechanisms, rather than a single role concept of function.[2]

Engineering scientists do no use the notion of role function *simpliciter*. Rather they use different sub types of role function in design and explanatory practice, and the level of detail of the explanations they construct hinges on the specific sub type of function employed. Engineers simplify or increase the details of explanations depending on the explanatory purpose at hand, and these adjustments are made using specific sub concepts of function (van Eck 2014). Capturing these explanatory dynamics calls for regimenting the mechanistic concept of role function into domain-specific interpretations of engineering function, to wit: *behavior* function and *effect* function. In this paper I advance this regimentation, focusing on the explanatory contexts of reverse engineering and malfunction explanation.

This analysis not only offers insight into the structure of mechanistic explanation in engineering science. It also provides means to refine current thinking about the explanatory power of mechanistic explanations in important ways. According to one influential perspective, the power of mechanistic models is (almost) always increased when these refer to both functional and structural features of mechanisms (Machamer et al. 2000; Craver 2007). On the counterview, mechanistic models have in certain contexts more explanatory traction when reference to structural aspects of mechanisms is suppressed. Models that solely describe functional characteristics, i.e., causal relations between components, explain better how organization impacts the behavior of mechanisms (Levy and Bechtel 2013). The engineering cases presented here allow for a more fine-grained understanding of the relationship between these views. Based on the reverse engineering case, I argue that these perspectives are not in competition but emphasize different explanatory virtues that hold in different explanation-seeking contexts. I then show that the virtues of 'completeness and specificity' and 'abstraction' pull in opposite directions in the context of malfunction explanation, and that a novel desideratum is required to accommodate this explanatory context. I elaborate a novel desideratum for malfunction explanation, dubbed 'local specificity and global abstraction', and argue that it applies to both malfunction explanation in the engineering sciences and the biological domain. Finally, the cases show, against widespread

[2] I take engineering science to be a 'science of making'. The view of engineering as a 'science' can be defended in terms of key objectives that it has in common with other 'traditional' sciences; explanation, prediction, and understanding of complex systems, in casu technical systems. Explanation, for instance, is a key element in a variety of engineering designs methods, such as diagnostic reasoning methods, reverse engineering and redesign methods, and in knowledge base-assisted designing (see sections 3 and 4).

assumption, that behavioral explanations have explanatory leverage as well.[3] Rather than merely describing phenomena to be explained, they enable explaining (at a course-grained system level) the *contrast* between normal functioning and malfunctioning technical systems/mechanisms.

I continue the paper in the next section with a brief description of the relevant mechanistic concepts: mechanistic explanation, functional individuation of mechanisms, and mechanistic role function ascription. In section 3, I first discuss engineering notions of function and their relation with explanations and explanatory objectives, and then regiment the mechanistic concept of role function (and functional individuation) in terms of these engineering notions of function, arriving at a fine-gained description of mechanistic explanation in engineering science. In section 4, I engage and extend current thinking on the explanatory power of mechanistic explanations. I end with conclusions in section 5.

2 Mechanistic explanation

2.1 Mechanistic explanation: functional individuation and mechanistic role functions

Mechanistic explanation starts by identifying a phenomenon to be explained; then the activities and entities relevant for this phenomenon are identified; finally, the temporal and spatial organization holding between the activities and entities by which they produce the phenomenon is specified (Machamer et al. 2000; Bechtel and Abrahamson 2005; Craver 2007). Mechanistic explanations thus explain how mechanisms, i.e., organized collections of entities and activities, produce phenomena.

This general structure of mechanistic explanation finds widespread support in the literature (e.g. Machamer et al. 2000; Craver 2001, 2007; Bechtel and Abrahamson 2005; McKay Illari and Williamson 2010). Craver (2001, 2012a) and McKay Illari and Williamson (2010), in addition, explicitly argued that the ascription of *role functions* to mechanisms and their component activities and entities is crucial for the construction of mechanistic explanations.[4] The mechanistic view on role function is an offshoot of Cummins' (1975) concept of function, extended to mechanisms and mechanistic explanation (Craver 2001; see Sustar 2007; McKay Illari and Williamson 2010).

In Cummins' account, function ascriptions are conceptually dependent on an "analytical explanation" (1975, 762) of a capacity of a containing system. The

[3] With behavioral explanation, I mean an input–output phenomenon description (see Glennan 2005; Matthewson and Calcott 2011).

[4] Several concepts of function are on offer in the philosophical literature, with *selected effect* and *role* concepts of function being the prominent ones (see Walsh and Ariew 1996). The selected effect concept of function is invoked to explain the presence or characteristics of an item (function bearer) in terms of whatever it does (selected effect) that contributed to the fitness of possessors of these items in the evolutionary past (Millikan 1989). The role concept of function is detached from evolutionary selective history. It, rather, is used to explain how an item contributes to some overall capacity of a system of which the item is a part (Cummins 1975; Walsh and Ariew 1996). The role concept of function is considered far more relevant than the selected effect concept of function for explaining how mechanisms produce explananda phenomena (Craver 2001, 2007; McKay Illari and Williamson 2010): selective histories do not explain how items are situated in mechanisms and how they contribute to the mechanisms of which they are a part (Craver 2007; McKay Illari and Williamson 2010). The role concept of function is invoked specifically for this explanatory job.

manifestation of a system capacity, coined analyzed capacity, is explained in terms of a number of other capacities, coined analyzing capacities, of the system's component parts and/or processes that jointly realize the manifestation of the system capacity (1975, 760). Functions are ascribed to those capacities of the component parts/processes that figure in an analytic explanation of a system capacity. In Cummins' account, more formally, the ascription of a function to an item X is specified as follows (1975, 762):

> X functions as a ϕ in S (or the function of X in S is to ϕ) relative to an analytic account A of S's capacity to ψ just in case X is capable of ϕ-ing in S and A appropriately and adequately accounts for S's capacity to ψ by, in part, appealing to the capacity of X to ϕ in S.[5]

Taking the heart as example, Cummins asserts that the heart (X) functions as a pump (ϕ) in the circulatory system (S) relative to an analytical account (A) of the circulatory system's (S's) capacity to transport food, oxygen, and waste products (ψ) just in case the heart (X) is capable of pumping (ϕ-ing) in the circulatory system (S) and the analytical account (A) appropriately and adequately accounts for the circulatory system's (S's) capacity to transport food, oxygen, and waste products (ψ) by, in part, appealing to the capacity of the heart (X) to pump (ϕ) in the circulatory system (S).

Craver (2001) adopts and elaborates Cummins' account of function in the context of mechanistic explanation, restricting the notion of a system to that of a mechanism, and making the ascription of functions dependent on the manner in which an entity's activity is *organized* within a mechanism (see Craver 2001, 59–62). In Craver's account (2001), an entity's activity can only be ascribed a function relative to how it is organized within a mechanism, and in virtue of which it contributes to an overall activity of a mechanism. Craver (2001, 61) thus writes:

> Attributions of mechanistic role functions describe an item in terms of the properties or activities by virtue of which it contributes to the working of a containing mechanism, and in terms of the mechanistic organization by which it makes that contribution

Mechanistic role functions thus refer to activities that make a contribution to the workings of mechanisms of which they are a part (Craver 2001, 2007, 2012a). Mechanistic organization is key. Whereas organization is treated very loosely in Cummins' account, referring to something that can be specified in a program or a flow chart (1975), spatial, temporal, and active features of mechanisms are vital in Craver's (2001) account for the ascription of functions. For instance, in the context of explaining the circulatory system's activity of "delivering goods to tissues", the heart's "pumping blood through the circulatory system" is ascribed a function relative to organizational features such as the availability of blood, and the manner in which veins and arteries are

[5] I use Craver's (2001, 55) notation here, which differs slightly from Cummins' (1975).

spatially organized (Craver 2001, 64). Change any of these features, say, the spatial relations between the entities, and a mechanism would not have that overall activity, nor would its entities have their activities and functions.

In short, mechanisms are functionally individuated in terms of mechanistic role function ascriptions, which, in turn, crucially depend on mechanistic organization. The organization of a mechanism is specified in terms of those features—temporal, spatial, active—of the mechanism that are crucial for production of the phenomenon to be explained.[6]

This perspective on the general structure of mechanistic explanation and the functional individuation of mechanisms finds widespread support in the literature among authors that focus their analyses on fields like (cell) biology and neuroscience (Craver 2001, 2007; Bechtel 2006; Darden 2006; McKay Illari and Williamson 2010), psychology (Wright and Bechtel 2007; Bechtel 2008a), and cognitive neuroscience (Bechtel 2008b; Kaplan and Bechtel 2011). Frequently, in these analyses, mechanisms of technical artifacts, such as clocks, mousetraps and car engines, are invoked as metaphors to elucidate features of biological mechanisms (Craver 2001; Calcott 2009; Piccinini and Craver 2011), and features of mechanisms in general (Craver and Bechtel 2005; Glennan 2005, 2010; Darden 2006; McKay Illari and Williamson 2012). And also the mechanistic concept of role function, and its utility in the functional individuation of mechanisms, has been explicated in terms of mechanisms of technical artifacts (Craver 2001).

Yet, as I will argue, in engineering science technical systems/mechanisms are not individuated functionally in terms of the concept of role function *simpliciter*. Rather, different notions of engineering function are invoked to individuate technical systems and to explain their (internal) workings. Given this diversity in function notions and explanations, the general structure perspective on mechanistic explanation needs to be adjusted in order to capture explanatory practices in engineering in detailed fashion.

I concur with McKay Illari and Williamson's (2010) (opening) statement that: "There has been great progress in understanding mechanistic explanations in particular domains, but this progress needs to be extended to cover all sciences" (2010, 279).[7] In order to extend this progress to engineering science, analysis of engineering concepts of function and the way(s) in which they figure in technical systems/mechanisms individuation and explanation is required.

In the next section I give this analysis, focusing on reverse engineering explanation and malfunction explanation. In section 4, I turn to the current state of play in thinking about the explanatory power of mechanistic explanations, and argue that the engineering science analysis gives means to add rigor and precision to the current debate on the explanatory power of mechanistic explanations.

[6] Function ascription is also considered vital for the hierarchical nesting of explanations: lower-level explanations are nested in/related to higher-level ones if the phenomena explained by lower-level explanations—say, the heart's "pumping blood through the circulatory system"—have been identified as having a function in mechanisms described by higher-level explanations—say, the circulatory system's mechanism for "distributing oxygen and nutrients to parts of the body" (Craver 2001).

[7] Their analysis focused in particular on protein synthesis and natural selection.

3 Engineering notions of function and functional decomposition

3.1 Function and functional decomposition in engineering

Function is a key term in engineering (e.g., Chandrasekaran and Josephson 2000; Stone and Chakrabarti 2005). Descriptions of functions figure prominently in, for instance, design methods (Stone and Wood 2000; Chakrabarti and Bligh 2001), reverse engineering analyses (Otto and Wood 2001), and in diagnostic reasoning methods (Bell et al. 2007).

Despite the centrality of the term, function has no uniform meaning in engineering: different approaches advance different conceptualizations (Erden et al. 2008), and some researchers use the term with more than one meaning simultaneously (Chandrasekaran and Josephson 2000; Deng 2002).

This ambiguity led to philosophical analysis of the precise meanings of function involved. Vermaas (2009, 2011) regimented the spectrum of available function meanings into three '*archetypical*' engineering conceptualizations of function[8]:

- *Behavior function*: function as the desired behavior of a technical artifact
- *Effect function*: function as the desired effect of behavior of a technical artifact
- *Purpose function*: function as the purpose for which a technical artifact is designed

The concept of behavior function is advanced in several engineering design and reverse engineering methods (Stone and Wood 2000; Chakrabarti and Bligh 2001; Otto and Wood 2001). In these methods, a function is described as a conversion of flows of materials, energy, and signals, where input flows and output flows in the conversion (are assumed to) match in terms of physical conservation laws (see Otto and Wood 2001). For instance, the function "loosen/tighten screws" of an electric screwdriver is then represented as a conversion of input flows of "screws" and "electricity" into corresponding output flows of "screws", "torque", "heat", and "noise" (see Stone and Wood 2000, 364). Since these descriptions of functions are specified such that input and output flows match in terms of physical conservation laws, they are taken to refer to specific physical behaviors of technical artifacts (see Otto and Wood 2001; Vermaas 2009; van Eck 2011).

Effect function descriptions are also used in design methods (Deng 2002), as well as in diagnostic reasoning approaches (Bell et al. 2007). There, functional descriptions refer to only the technologically relevant *effects* of the physical behaviors of technical artifacts: the requirements are dropped that descriptions of these effects meet conservation laws and that matching input and output flows are specified (Vermaas 2009; van Eck 2011). The function of an electric screwdriver is then described simply as, say, "loosen/tighten screws", leaving it unmentioned what the physical antecedents are of this effect. Behavior function descriptions thus refer to the 'complete' behaviors

[8] The term 'archetypical' here refers to 'most common'; the three conceptualizations of function are not meant to be exhaustive. For instance, some engineers use 'function' to refer to intentional behaviors of agents (see van Eck 2010), and others have recently explored the idea that 'function' might be a Wittgensteinian family resemblance concept (Carrara et al. 2011). In reverse engineering analyses, 'function' refers to actual or expected behavior, without the normative connotation 'desired'.

involved, including features like thermal and acoustic energy flows, whereas effect functions refer to subsets of these behaviors, i.e., desired effects.[9]

Purpose function descriptions are also employed in engineering design (Deng 2002). Such descriptions refer to intended states of affairs in the world that are intended by designers, and which are to be created by the physical behaviors and effects of the technical artifact concerned (Vermaas 2009; van Eck 2011). The function of an electric screwdriver is then described as, say, "having connected materials", referring to a state of affairs outside the artifact and to be achieved by (manipulation of) the artifact.

Behavior and effect conceptualizations of function thus refer to features of artifacts, behaviors and effects of behaviors, respectively, whereas the concept of purpose function refers to states of affairs external to artifacts.[10]

Engineering descriptions and explanations of the workings of technical artifacts and artifacts-to-be-designed often are constructed by breaking down/functionally decomposing functions into a number of other (sub) functions. The relationships between functions and sets of their sub functions are often graphically represented in functional decomposition models. Like the concept of function, such models come in a variety of flavors. Elsewhere, I have regimented this diversity in three archetypical engineering conceptualizations of functional decomposition (van Eck 2011)[11]:

- *Behavior functional decomposition*: a model of an organized set of behavior functions;
- *Effect functional decomposition*: a model of an organized set of effect functions;
- *Purpose functional decomposition*: a model of an organized set of purpose functions.

The use of functional decomposition is ubiquitous in engineering science. Stone and Wood (2000) use behavior functional decompositions in, for instance, the conceptual phase of engineering design to analyze the desired functions of some artifact-to-be, and in the reverse engineering of existing artifacts for archiving functional descriptions of these artifacts and their components. Otto and Wood (1998, 2001) also use behavior functional decompositions in reverse engineering tasks to determine the organized

[9] Behavior and effect functions thus have a partly common semantic structure: certain aspects or features of behaviors that they both refer to. They are dissimilar in the sense that behavior function descriptions refer to additional behavioral aspects, not referred to in effect function descriptions, so as to make these descriptions accord with physical conservation laws. The relation between behavior and effect function is asymmetrical in the sense that effects, being subsets of behaviors, are straight forwardly derivable from behaviors, but not vice versa. From a given effect one cannot automatically derive the behavior of which the effect is a part. Cars that run on gas operate by means of different energy conversions than cars than run on electricity, yet both display the same effects, say, delivering acceleration. The semantic structure that they partly have in common creates the possibility and need to be pluralist about mechanistic role functions, i.e., different ways to conceive of the role functions of mechanisms, in the context of engineering science. I defend this pluralism about mechanistic role functions in section 3.2. To be sure, I am thus not advocating a pluralist view about functions of mechanisms with a completely dissimilar semantic structure. I thank an anonymous referee for pressing me on these points.

[10] In design methodologies that advance effect and/or purpose functions the concept of behavior is typically introduced as well, alongside function. By invoking behavior descriptions of technical artifacts alongside functional descriptions, physical conservation laws are taken into account.

[11] Single functional decomposition models in which different concepts of function are described are rare in engineering.

components and sub functions (behaviors) of artifacts—their mechanisms—, by which they produce their overall (behavior) functions, and in redesign tasks to identify components that function sub-optimally and require improving. Bell et al. (2007) use effect functional decompositions for explaining malfunctions of artifacts, and Deng (2002) uses purpose functional decompositions in the conceptual phase of engineering design.

This usage of different notions of engineering function and functional decomposition in different design and explanatory tasks is (currently) not accounted for by the view that the construction of mechanistic explanations proceeds via the functional individuation of mechanisms in terms of role function ascriptions. In order to capture the specifics of functional individuation of technical systems/mechanisms in engineering science, i.e., to progress our understanding of mechanistic explanation in this field, the mechanistic concept of role function needs to be regimented into behavior and effect interpretations of engineering function. Cases in point, to be discussed below, are reverse engineering explanations which use elaborate behavior functions and functional decompositions, and malfunction explanations which use less detailed effect functions and functional decompositions. (The notion of purpose function does not figure in this regimentation of mechanistic role function. I discussed it here to give a 'comprehensive' overview of the meanings of function used in engineering practice, and because the notion is relevant for understanding the specifics of malfunction explanation discussed later on.)

The upshot of this discussion is twofold. First, by regimenting the mechanistic concept of role function into two sub types, i.e., behavior and effect interpretations of engineering function, we gain an empirically-informed understanding of mechanistic explanation in engineering science. Second, I argue, in terms of engineering explanatory practices, that what are taken to be competing perspectives on the explanatory power of mechanistic explanations, in fact, are not, and that the endorsed desiderata of both perspectives are not desiderata for malfunction explanation: a specific combination of them is required to accommodate this explanatory context. I further argue that also behavioral explanations have explanatory power (to some extent) in contexts where one aims to explain contrasts between functioning and malfunctioning technical systems/mechanisms (section 4).

3.2 Reverse engineering explanation (and redesign)

In engineering science, reverse engineering and engineering design go hand in glove (e.g. Otto and Wood 1998, 2001; Stone and Wood 2000). In Otto and Wood's (1998, 2001) method, a reverse engineering phase in which reverse engineering explanations are developed for existing artifacts, precedes and drives a subsequent redesign phase of those artifacts. The goal of the reverse engineering phase is to explain how existing artifacts produce their overall (behavior) functions in terms of underlying mechanisms, i.e., organized components and sub functions (behaviors) by which overall (behavior) functions are produced. These explanations are subsequently used in the redesign phase to identify components that function sub optimally and to either improve them or replace them by better functioning ones. Otto and Wood (1998, 226) relate explanation and redesign as follows: "the intent of this [reverse engineering] process step is to fully understand and represent the current instantiation of a product. Based on the resulting

representation and understanding, a product may be evolved [redesigned], either at the subsystem, configuration, component or parametric level".

In the reverse engineering phase, an artifact is first broken down component-by-component, and hypotheses are formulated concerning the functions of those components. In this method, functions are behavior functions and represented by conversions of flows of materials, energy, and signals. After this analysis, a different reverse engineering analysis commences in which components are removed, one at a time, and the effects are assessed of removing single components on the overall functioning of the artifact. Such single component removals are used to detail the functions of the (removed) components further. The idea behind this latter analysis is to compare the results from the first and second reverse engineering analysis in order to gain potentially more nuanced understanding of the functions of the components of the (reverse engineered) artifact. Using these two reverse engineering analyses, a behavior functional decomposition of the artifact is then constructed in which the behavior functions of the components are specified and interconnected by their input and output flows of materials, energy, and signals (Otto and Wood 2001). Such models represent parts of the mechanisms by which technical systems operate, to wit: causally connected behaviors of components. They are the end results of the reverse engineering phase and are subsequently used to identify sub-optimally functioning components and so drive succeeding redesign phases. Examples of an overall behavior function and behavior functional decomposition of a reverse engineered electric screwdriver are given in Figs. 1 and 2, respectively.

In the model in Fig. 2, temporally organized and interconnected behaviors are described. Components of artifacts are described in Otto and Wood's method in tables, what in engineering are called 'bills of materials', together with a model, called 'exploded view', of the components composing the artifacts. Taken together, these component and behavior functional decomposition models provide functional individuations and representations of mechanisms of artifacts.

After the reverse engineering of a technical artifact, aimed at providing detailed understanding of the mechanism(s) by which it operates, the redesign phase starts by identifying components that *function sub-optimally*, and, thereby, cause artifacts to manifest their overall functions in sub-optimal fashion. Redesign efforts are subsequently directed towards designs with improved functionality of these components (Otto and Wood 1998, 2001). Otto and wood (1998) discuss an example of redesigning an electric wok. The (reverse engineered) artifact's desired behavior to "deliver a uniform temperature distribution across the bowl" failed to be achieved due to the fact

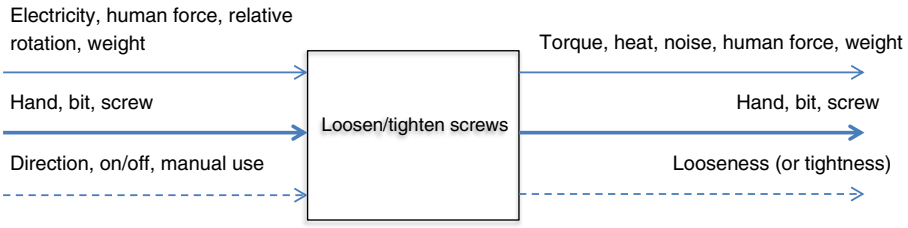

Fig. 1 Overall behavior function of an electric power screwdriver. *Thin arrows* represent energy flows; *thick arrows* represent material flows, *dashed arrows* represent signal flows (adapted from Stone and Wood 2000, 363, figure 2)

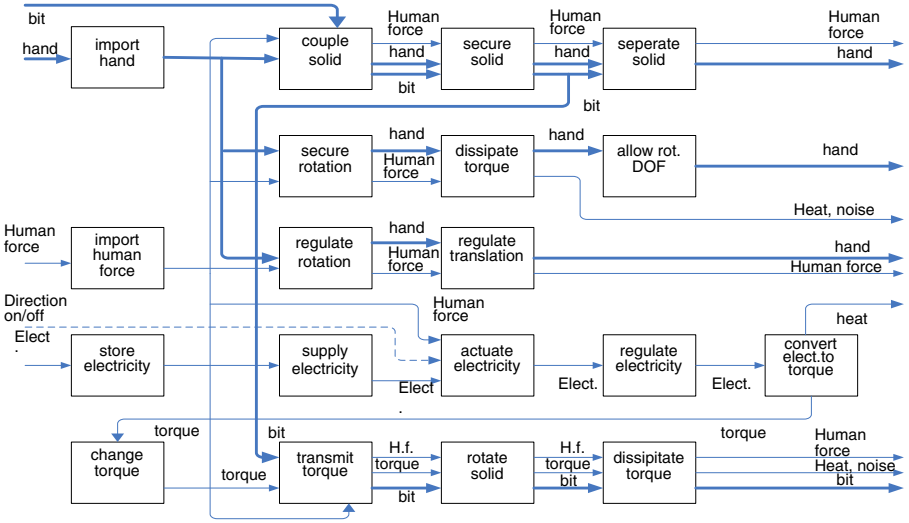

Fig. 2 Behavior functional decomposition of an electric power screwdriver. *Thin arrows* represent energy flows; *thick arrows* represent material flows, *dashed arrows* represent signal flows (adapted from Stone and Wood 2000, 364, figure 4)

that the electric heating elements of the wok, such as a bimetallic temperature controller, were housed in too narrow a circular channel (Otto and Wood 1998, 235). Redesign efforts were subsequently directed towards a design with improved functionality of the heating elements, inter alia resulting in a design with a thicker bowl and different shape than in the reverse engineered electric wok.[12] In sum, a reverse engineering—mechanistic—explanation of the operation of an existing electric wok was used to identify sub optimal functioning components—in this case, electric heating elements—which resulted in modifications to these components.

For this reverse engineering context, the choice to employ behavior functions and functional decompositions is the optimal one (van Eck and Weber 2014). Behavior functional decompositions in which behavior functions of components are specified and interconnected by their input and output flows of materials, energy, and signals, provide the most elaborate information on temporal and spatial relationships between behaviors and components, i.e., on the workings of mechanisms. Effect function descriptions omit relevant details and purpose function descriptions do not refer to the internal mechanisms of artifacts, since they describe state of affairs in the world to be realized by the behaviors of artifacts.

Also for the subsequent redesign phase, behavior function descriptions are the most useful for at least two reasons. Firstly, these contain the most detail and hence the most information to assess the performance of components and make comparisons between components. Say, returning to the wok example, a novel halogen heat lamp that fulfills the function of 'converting electricity to radiation' in a better way than the wok's heating coil since the halogen lamp produces less heat or noise, or both (see Otto and Wood 1998, 236). Secondly, in replacing components one needs to take the structural

[12] This redesign step involves a lot of mathematical modeling, use of physical and technological principles, and/or prototype building (Otto and Wood 1998, 2001). These details need not concern us here.

configuration of the reverse engineered artifact into account, i.e., how the to-be replaced component is organized with other components, in order to ensure that the novel component can indeed be placed in this configuration. Descriptions of behavior functions, and sequences thereof as specified in behavior functional decomposition models in which the behavior functions of the components are specified and interconnected by their input and output flows of materials, energy, and signals, provide the most elaborate information on structural configurations.

In malfunction explanation, this detail in mechanistic models is however not required: less detailed effect functions and functional decompositions there do a better explanatory job.

3.3 Malfunction explanation

When an artifact does not serve a function which we expect it to do, explanation-seeking questions of the following format arise:

> Why does artifact x not serve the expected function to ϕ?
> For instance: why does this electric screwdriver fail to drive screws?

Such questions are *contrastive*: they contrast the actual situation with an ideal and expected one (see Lipton 1993). Now, in the engineering literature, malfunction explanations that answer contrastive questions list different and fewer mechanistic features than reverse engineering explanations which answer questions about plain facts, such as explanations of why an artifact displays a certain behavior (e.g., an electric screwdriver's behavior of driving screws).[13] Contrastive malfunction explanations, as developed in engineering by, for instance, Price (1998), Hawkins and Woollons (1998), and Bell et al. (2007), pick out only a few features of mechanisms, i.e., those causal factors that are taken to make a difference to the occurrence of a specific malfunction. Malfunctioning components or sub mechanisms are specified, yet most information about their structural and behavioral specifics are left out, as well as how they are organized with other components and their behaviors. So judged by the information listed in these explanations, a more complete description which would include the structural and behavioral specifics of malfunctioning components and/or sub mechanisms, and their organization with other components and behaviors of an artifact, is overkill for malfunction explanation.[14] For instance, when a system level malfunction occurs, say the failure to drive screws of a power screwdriver, malfunction explanations refer to the malfunctioning components or sub mechanisms taken to

[13] This is not to say that the explanandum in reverse engineering explanation, or mechanistic explanation in general, cannot be described in contrastive fashion. For instance, the request for explanation may concern why an artifact x exhibits behavior b with value y rather than behavior b with value z. However, this is a different contrast than the one drawn in the explanandum of why artifact x does *not* exhibit behavior b rather than displaying behavior b.

[14] An anonymous referee rightly pointed out that the 'omission of specifics', as in the case of malfunction explanation, highlights a general principle of relevance that holds for most contexts in which one aims to explain an isolated feature of a system's behavior. However, such a general principle of relevance can be interpreted in different ways depending on the desiderata for mechanistic models one adopts. In section 4 I argue that two recently proposed desiderata for mechanistic models pull in opposite directions in the context of malfunction explanation, signaling the need for a novel one. I elaborate such a desideratum in this paper.

underlie this system level malfunction, say, a failing sub-mechanism for the conversion of electricity into torque, yet not to the components and operations that are similar in normally functioning screwdrivers and this particular dysfunctional one, say components that store electricity, supply electricity, and insulate heat and noise (see Fig. 2), and neither to the structural and behavioral specifics of the failing components and/or sub mechanisms.[15],[16]

Consider, by way of example, a methodology for malfunction analysis and explanation, called Functional Interpretation Language (FIL), developed by Bell et al. (2007). In FIL, the representation of a function consists of three elements: the *trigger* of a function, its associated and expected *effect*, and the *purpose* that the function is to fulfill. Triggers describe input states that actuate physical behaviors which result in certain (expected) effects. So triggers are the input conditions for effects, i.e., functions, to be achieved. Purposes describe desired states of affairs in the world that are achieved when a trigger results in an expected effect (Bell et al. 2007, 400). For instance, with FIL, the function of a stop light of a car is described in terms of the trigger "depress_brake_pedal", the effect "red_stop_lamps_lit", and the purpose "warn_following_driver" (p. 400). This description is a summary of some salient features of (manipulating) such artifacts; depressing the brake pedal will, if the system functions properly, result in the lighting of the stop lamps, which in turn supports the warning of fellow drivers that the car is slowing down.

According to Bell et al. (2007) such trigger and effect representations serve two explanatory ends in malfunction analyses: firstly, they *highlight* relevant behavioral features, i.e., effects, and, simultaneously, provide the means to *ignore* less relevant or irrelevant behavioral features, i.e., physical behaviors underlying these effects, of a given artifact; secondly, they support assessing which components are malfunctioning (pp. 400–401).

For instance, the trigger-effect representation "depress_brake_pedal"-"red_stop_lamps_lit" highlights the input condition of a pedal being depressed, and the resulting desired effect of lighted lamps, yet ignores the structural and behavioral specifics of the brake pedal and stop lamps, such as the pedal lever and electrical circuit mechanisms, as well as the energy conversions—e.g., mechanical energy conversions into electricity—that are needed to achieve this effect. Such representations only highlight those features that are considered explanatorily relevant to assess malfunctioning systems, and omit reference to physical behaviors/energy conversions by which desired effects are achieved.

There is another way in which the use of trigger-effect descriptions is considered an explanatory asset in highlighting explanatorily relevant features in malfunction

[15] This is the most straightforward scenario. If there are backup systems that are intended to prevent malfunction, and modern technology is replete with them, failing backup systems should be referred to as well, of course, in explaining system level malfunctions.

[16] That is, structural and behavioral characteristics are considered irrelevant in a first round functional analysis of malfunction. After this analysis, more detailed behavioral models of components and their behaviors are used for identifying specific explanatorily relevant structural and behavioral characteristics of malfunctioning components/sub mechanisms (Bell et al. 2007). However, immediately specifying these details in functional models is taken to result in listing a lot of irrelevant details (see the FIL methodology described in this section).

explanation: comparing normally functioning technical systems with malfunctioning ones (Bell et al. 2007). Trigger-effect descriptions support assessing whether the expected effects in fact obtain, and, if not, which and how components are malfunctioning (Bell et al. 2007). A normally functioning artifact, say the car's stop lights, has both a trigger and an effect occurring; the brake pedal is depressed and the stop lights are lit. Trigger-effect descriptions support analysis of two varieties of malfunction. First, a trigger may occur, yet fail to result in the intended effect. Say, the brake pedal is depressed, yet the stoplights are not on. Second, a trigger may not be occurring, yet the effect is nevertheless present. Say, the brake pedal is not depressed, yet the stoplights are on (see Bell et al. 2007). Such analysis of the actual states of triggers and effects allows one to focus on the most likely causes of failure (Bell et al. 2007). Say, if the pedal is depressed and the lights fail to ignite, first likely causes to investigate may be whether the electrical circuits in the lights are broken or the 'on/off' connection between the brake and electrical circuitry (connected to the lamp) is damaged. On the other hand, if the pedal is not depressed and the lights are lit, a first likely cause to investigate may be whether the 'on/off'connection between the brake and the electrical circuitry is damaged. To support more detailed malfunction analyses, functions are often decomposed into sub functions in FIL. An example of a functional decomposition of a two-ring cooking hob is given in Fig. 3.

Descriptions of functions and functional decompositions as used in FIL refer to desired effects of behaviors and effect functional decompositions (see van Eck 2011). This choice is the optimal one, given that function descriptions are used to black-box or suppress reference to unwanted behavioral and structural details. Effect function descriptions only highlight the relevant difference making properties with respect to malfunctioning artifacts, whereas more elaborate behavior function descriptions include irrelevant details such as, say, the thermal energy generated when lamps are lit. Purpose function descriptions provide a useful yardstick to assess whether desired states of affairs obtain—and if that is not the case signal a malfunction of some sort—yet have no utility in describing internal difference making factors, since they refer to states of affairs external to artifacts.

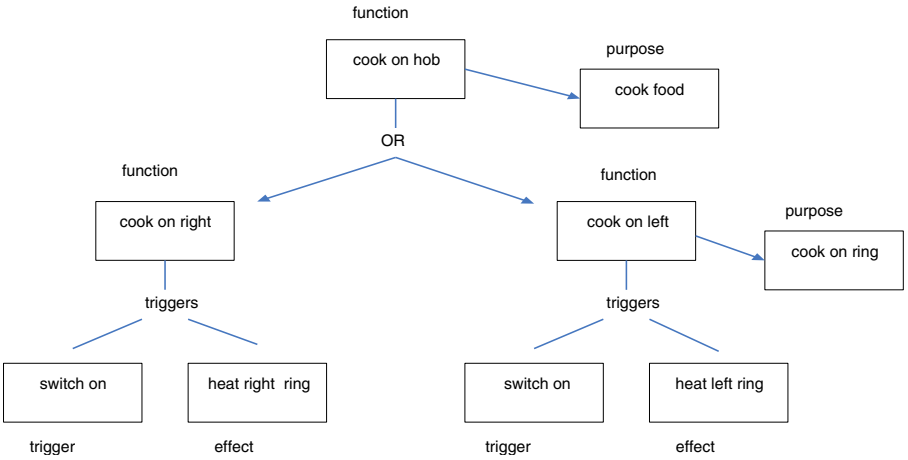

Fig. 3 Effect functional decomposition of a two-ring cooking hob (adapted from Bell et al. 2007)

3.4 Capturing mechanistic explanation in engineering science: pluralism about mechanistic role functions

As we can see, explanations in engineering are furnished relative to explanatory objectives and, importantly, the level of detail included in these explanations hinges on specific concepts of technical function. Engineering scientists simplify or increase the details of explanations—functional decompositions—depending on the explanatory purpose at hand, and these adjustments are made using specific concepts of technical function. In reverse engineering explanation, elaborate or 'complete' descriptions of mechanisms are provided, in terms of behavior functions and functional decompositions, to answer the question how a technical system exhibits a given overall behavior. In malfunction explanation, less elaborate 'sketches' of mechanisms are provided in terms of effect functions and functional decompositions, referring only to some mechanistic features, namely those difference making factors that mark the *contrast* between normal functioning and malfunctioning technical systems. So, depending upon explanatory context, mechanisms are individuated in different ways using different conceptualizations of function in engineering science. Neither function conceptualization in itself accommodates both ways in which mechanisms are functionally individuated in engineering science. Behavior and effect function ascriptions are (and need to be) invoked to individuate mechanisms in different ways depending on the task at hand.

However, this distinction in functional individuation, and the function concepts on which it hinges, remains opaque, when seen from a perspective that conceives of mechanism individuation and mechanistic explanation in terms of mechanistic role function ascription simpliciter. The concept of mechanistic role function, an activity that makes a contribution to the workings of a mechanism of which it is a part, admits of two interpretations in the context of engineering science: behavior function on the one hand and effect function on the other. Specifying the contribution of, say, a power screwdriver's motor in terms of the behavioral description 'converting electricity into torque' is different from specifying the motor's contribution in terms of merely its effect 'produce torque': the level of detail with which such contributions are specified, and hence the level of detail with which mechanisms are individuated, is relative to explanatory objectives (compare e.g. Figs. 2 and 3). So in order to arrive at empirically informed understanding of explanatory practices in engineering, and at consistency of the general structure of mechanistic explanation with these practices, regimenting the concept of role function into domain-specific engineering concepts of behavior and effect function, i.e., sub types of role function, is required.[17]

In doing so, we meet the descriptive goal of the mechanist philosophy to adequately capture explanatory practices in specific domains (see Machamer et al. 2000; Bechtel

[17] Note that behavior and effect descriptions of function describe, in different ways, the contributions of components to mechanisms of which they are a part. The distinction between behavior and effect function thus is not to be conflated with the distinction between a mechanism description and a description of a mechanisms' overall activity. Neither is the behavior-effect function distinction to be conflated with the distinction between 'isolated' and 'contextual' descriptions of an entity's activity (Craver 2001): isolated descriptions describe activities without taking into account the mechanisms in which they are situated; contextual descriptions describe activities in terms of the mechanistic contexts in which they are situated and to which they contribute. Both behavior and effect functions are of the contextual variety, describing contributions of components to the mechanisms of which they are a part.

and Abrahamson 2005; Craver 2007; McKay Illari and Williamson 2010), as well as the implicit norms by which scientists evaluate their explanations (Craver 2007). Significant headway has already been made on these issues in the context of biology, cognitive, and neuroscience. By regimenting the mechanistic concept of role function in two sub types, i.e., behavior and effect interpretations of engineering function, we also meet this goal with respect to engineering science. The intricacies of mechanistic explanation in the engineering domain can now also be accounted for, and seen to be consistent with the general framework for mechanistic explanation.[18]

Our analysis of engineering explanations has further ramifications. It shows that two currently discussed perspectives on the explanatory power of mechanistic explanations, 'completeness and specificity' and 'abstraction', are not competitors and, in addition, that other desiderata are required to accommodate malfunction explanations in satisfactorily manner. Spelling this out is the topic of the next section.

4 Explanatory power of mechanistic explanation

4.1 Explanatory power: current state of play

What makes mechanistic explanations explanatorily powerful? Surprisingly, only in recent years is the topic of explanatory power starting to get explicit attention in the mechanist literature (Craver 2007, 2012b; Illari 2013; Gervais and Weber 2013; Levy and Bechtel 2013). Craver's (2007) account of mechanistic explanation provides the first in-depth treatment of this issue.

Craver (2007) stipulates one core requirement or "central criterion of adequacy" (p. 139) that mechanistic explanations ought to meet: mechanistic explanations should

[18] Although my preferred conception of mechanistic explanation is an epistemic one, both advocates of epistemic (e.g., Wright and Bechtel 2007; Wright 2012) and ontic conceptualizations (e.g., Salmon 1984; Glennan 2005; Craver 2007, 2012b) of mechanistic explanation can sign up to this project. On an epistemic reading, explanations are explanatory texts that procure understanding, whereas on an ontic reading mechanisms in the world are explanations. Illari (2013) recently argued that the ontic-epistemic dispute currently going on in the literature is (or should be understood as) moving away from analysis of the term 'explanation' per se, to elaborating ontic and epistemic constraints on good explanations. Good mechanistic explanations should describe (ontic) mechanisms in the world in such a fashion that (epistemic) understanding of their workings is procured. If one subscribes to this view, as I do, then acknowledging the behavior-effect function distinction is relevant to capture good mechanistic explanations in engineering science. Depending upon explanatory context, (ontic) mechanisms are individuated in different ways using different conceptualizations of function in engineering science. These different functional individuations, in terms of different function conceptualizations, are tailor-made or 'engineered' to the task at hand, so as to highlight the (epistemic) relevant features of technical systems-mechanisms. So whether one takes features of mechanisms, behaviors and effects, as providing (ontic) explanations or descriptions of these features as doing the (epistemic) explanatory work, acknowledging the behavior-effect distinction is relevant to both camps for understanding good explanations in engineering science. Moreover, functional individuation of mechanisms is also explicitly endorsed by authors who defend an ontic view on explanation, at least by Machamer et al. (2000). They write: "mechanisms are identified and individuated by the activities, and entities that constitute them, by their start and finish conditions, and by their functional roles. Functions are the roles played by entities and activities in a mechanism. To see an activity as a function is to see it as a component in some mechanism, that is, to see it in a context that is taken to be important, vital, or otherwise significant." (Machamer et al. 2000, 6) So, function ascription is not only an epistemic factor affecting how we describe mechanisms: it is also an individuation condition of mechanisms (see McKay Illari and Williamson 2010).

be "complete" (p. 113) in the sense that they (ideally) describe all the entities, activities, and organizational features of mechanisms that are *constitutively relevant* for the multiple features of the phenomena to be explained (see Machamer et al. 2000). Briefly, on Craver's (2007) account, an entity's activity is considered constitutively relevant to the behavior of a mechanism as a whole if that entity's activity is a spatiotemporal part of the mechanism, and contributes to the behavior of the mechanism as a whole. This is taken to be established if one can change the overall behavior by intervening to change the entity's activity, and if one can change the activity of the entity by intervening to change the overall behavior. Such relationships of *mutual manipulability* are taken to provide a "sufficient condition for interlevel relevance" (p. 141), knowledge of which enables one"to know how the phenomenon changes under a variety of interventions into the parts *and* how the parts change when one intervenes to change the phenomenon" (p. 160). On this mutual manipulability account of constitutive relevance, constitutive relevance relationships between activities of entities and explananda phenomena are explicated in terms of an adapted version of Woodward's manipulability theory (2003), so that an explanation of the behavior of a mechanism as a whole can be procured by pointing to those activities and entities within the mechanism that make a difference to that overall behavior, and vice versa.[19]

Levy and Bechtel (2013) have recently taken issue with this perspective. They pitch their "account of abstraction" (p. 242) against the what they call "completeness and specificity" (p. 242) account that they associate with Craver's (2007) system, and more generally with the work of Machamer, Darden, and Craver (Machamer et al. 2000; Darden and Craver 2002; Darden 2006; Craver 2007). They make the point that in the work of Machamer, Darden, and Craver, structural features of entities always seem to get assigned explanatory (constitutive) relevance. However, Levy and Bechtel (2013) argue that in the explanatory context of explaining how organization impacts the behavior of a mechanism, often, skeletal models that suppress reference to structural aspects of components explain better than more elaborate models in which structural features are described. In this context, models that solely describe causal relations between components are best equipped to "explain temporal properties of mechanisms" (p. 241). Structural details are not needed. Hence, they argue that their 'abstraction' account provides a "significant corrective" (Levy and Bechtel 2013, p. 242) to the 'completeness and specificity' view.

Their case in point is work on what are called network motifs in genetics and cell biology, which concerns gene expression in bacteria and yeast. Levy and Bechtel (2013) back up their claims by stating that abstract, skeletal models here:

> "track those features of the system *that make a difference* to the behavior being explained" (p. 256). And that: "these models highlight the features of that specific system that make a difference in it—namely, its patterns of internal causal connections. Thus, we see the claim that abstract description stands in need of filling in as incorrect even with respect to explaining particular behaviors of

[19] Constitutive relevance is explicated in terms of an *adapted version* of Woodward's theory, since Woodward advances his manipulability theory as an account of *causal* explanation, while constitutive relevance is explicitly defined as a non-causal relationship (see Craver 2007, 153–154; Craver and Bechtel 2007, 552–554).

> particular systems. That is not to say that details are unimportant. In some contexts they surely are. But for some explanatory purposes, especially those having to do with organization, less is more" (2013, 259) (my italics).[20]

The claim here is that the omission of structural details makes salient those causal factors that make a difference to the phenomenon being explained, i.e., (functionally described) components and their causal relations specified in terms of components' causal roles (see Levy and Bechtel 2013).

4.2 Competing perspectives? Making sense of difference making

At first glance, there seem to be substantial differences between the two accounts of explanatory power. One is taken to always emphasize the explanatory relevance of structural details of components, whereas the other takes it that omitting reference to these features in some contexts gives mechanistic models more explanatory traction (see Levy and Bechtel 2013). Yet, is the abstraction account really in disagreement with the 'completeness and specificity' view? Like Levy and Bechtel (2013), Craver (2007) also speaks about difference making (p. 144, pp. 198–211) and is very explicit that complete models should (ideally) refer to all and only those elements that are constitutively relevant for the phenomenon to be explained. So it seems that if structural details are not constitutively relevant, they will not be referred to in mechanistic models. Nothing seems to preclude Craver from saying here that his account leads to the same results as the abstraction account: in both perspectives, models (ideally) only depict those factors that make a difference to the explanandum phenomenon.

Nevertheless, Levy and Bechtel (2013) take it that there is a significant difference and attempt to further spell it out in terms of the distinction between mechanism schemata and complete descriptions of mechanisms (see Machamer et al. 2000; Craver 2007). In Machamer, Darden, Craver lingo, mechanism schemata are abstract descriptions of mechanisms that can be filled in with details of entities and activities. Importantly, the filling in is what turns a schema into an explanation.[21] Craver (2007), similarly, treats schemata as descriptions in-between sketches and complete descriptions of mechanisms, and argues that explanatory progress is made as one moves along this axis toward the completeness endpoint. Now, Levy and Bechtel argue that their abstract models are similar to schemata and thus take it that their view opposes the idea that schemata are not bona fide explanations (2013, 258) or at least incomplete explanations. In this construal, Levy and Bechtel (2013), tie

[20] To be sure, as the quotation makes clear, Levy and Bechtel acknowledge that in other explanatory contexts structural details can be important.

[21] This is Levy and Bechtel's (2013) interpretation of Machamer et al. (2000). I think that their reading is correct, for Machamer et al. (2000) write: "We introduce the term "mechanism schema" for an abstract description of a type of mechanism. A *mechanism schema* is an abstract truncated description of a mechanism that can be filled with descriptions of known component parts and activities." (Machamer et al. 2000, 15). They elaborate: "When instantiated, mechanism schemata yield mechanistic explanations of the phenomenon that the mechanism produces." (p. 17) […] "Sometimes a sketch has to be abandoned in the light of new findings. In other cases it might become a schema, serving as an abstraction that can be instantiated as needed for the tasks mentioned above, e.g., explanation, prediction, and experimental design." (p. 18) So, instantiation—filling in details of entities and activities—is what turns a schema into an explanation according to Machamer et al. (2000).

Machamer et al. (2000) and Craver (2007) to the view that the filling in of structural details is what makes an explanatory description complete and hence good or better. Yet, their interpretation seems to rest on a mistake. If we follow Craver (2007) along the lines of constitutive relevance, complete models need not necessarily refer to structural details. As said, it seems entirely consistent with his view that in some cases constitutively relevant elements/difference making factors are only functionally specified components and their causal relations. What Levy and Bechtel (2013) take to be schemata, which they regard as explanatorily superior in some contexts, may count on Craver's (2007) perspective as complete descriptions in those contexts. There then seems little substantive disagreement; substantial differences and 'significant correctives' rather are chimerical, resulting from misinterpretation of terminology.

However, there is another way in which the perspectives can be pitched against one another and a relevant difference then does come out between the accounts (which might be the way Levy and Bechtel (2013) envision as well, yet they do not spell this out). I argue that, rather than being competing perspectives, 'completeness and specificity' and 'abstraction' constitute different explanatory virtues with respect to specific explanation-seeking questions or contexts. To see this, we need to have a closer look at the notions of difference making underlying the work of Craver (2007) and Levy and Bechtel (2013).

Levy and Bechtel do not spend much ink on the precise notion of difference making that they have in mind, but they do briefly refer to Strevens' (2004, 2008) work on explanation in defending their abstraction perspective. If we elaborate the abstraction perspective along the lines of Strevens' 'kairetic' account of (causal) explanation (2004; 2008) than a difference does emerge with Craver's (2007) system. On Strevens' account, explanatory models should refer only to those features that are large enough to make a difference to the *occurrence* of specific explananda phenomena.[22] This is a stringent constraint in the sense that factors that merely influence the precise manner in which the explanandum phenomenon manifests itself are to be omitted from an explanation. Weisberg (2007, 651) makes a useful distinction in this context between "primary causal factors" that make a difference with respect to occurrence and "higher order causal factors" that only affect the manner of occurrence. What Levy and Bechtel (2013) seem to have in mind, given their reference to Strevens' system, is that abstract mechanistic models should refer only to primary factors that make a difference to the occurrence of explananda phenomena. They write:

> "Strevens (2008) argues that good explanations are those that abstract to the least detailed causal model that enables one to demonstrate the causes of the explanandum […] Strevens is on to an important idea: oftentimes, omitting detail

[22] To be sure, in Strevens' (2004, 2008) system, explananda phenomena may comprise many things, such as events, properties, and regularities. What is crucial in Strevens' system is that explanatory models should only refer to those factors that are crucial for the explanatory target to obtain, i.e., to occur, whether it be a specific event, a regularity, a property, or something else. Strevens (2004, 158) himself puts it thus: "the explanatorily relevant parts of any causal network are the elements that made a difference to whether or not the explanandum occurred. It is important to note the *whether or not*. To be explanatorily relevant, a causal factor must not merely make a difference to *how* the explanandum occurred; it must make a difference large enough to bear on whether or not it occurred at all." Note that explananda are not restricted to events, properties, regularities, and the like, occurring or not but, rather, that explanatorily relevant factors must make a difference large enough for explanatory targets to obtain.

> permits one to distinguish those underlying factors that matter from those that do not […] the resultant explanation is better because—a la Strevens—it depicts those aspects of the system that make a difference." (Levy and Bechtel 2013, 256)

Since factors that matter for Strevens are "elements that made a difference to whether or not the explanandum occurred" (Strevens 2004, 158), we may interpret Levy and Bechtel as endorsing the position that abstract mechanistic models are considered suitable for a specific type of explanation-seeking question, to wit: 'which features of a mechanism make a difference to the *occurrence* of a mechanisms' overall behavior'. Explanations are then procured by listing those factors that make a difference in this sense.

Craver (2007, 152), in contrast, hitches his mutual manipulability account of constitutive relevance to Woodward's (2003) account of (causal) explanation. Drawing upon Woodward's (2003) interventionist framework, Craver specifies two conditionals (CR1, p. 155, and CR2, p. 159) that together comprise mutual manipulability:

> "(CR1) When ϕ is set to the value of $\phi 1$ in an ideal intervention, then ψ takes on the value f($\phi 1$)"
> "(CR2) When ψ is set to the value of $\psi 1$ in an ideal intervention, then ϕ takes on the value f($\psi 1$)"

These conditionals cover both scenarios in which interventions change the manner in which ψ or ϕ occur, i.e., their value, as well as ones that lead to the occurrence or elimination of ψ or ϕ (see Craver 2007, 149). In the latter case, ψ or ϕ would take on the value '1' or '0', respectively. So mutual manipulability relations comprise both constitutive relevance relations with respect to the occurrence of explananda phenomena, and relations concerning the precise manner in which explananda phenomena occur. This tracking of both 'primary' and 'higher order' constitutive factors, likely, relates to his "central criterion of adequacy for a mechanistic explanation" (p. 139), according to which an explanation:

> "should account for the multiple features of the phenomenon, including its precipitating conditions, *manifestations*, inhibiting conditions, *modulating conditions*, and nonstandard conditions" (Craver 2007, 139) (italics mine).

Here the request for explanation is different than an inquiry into 'which features of a mechanism make a difference to the *occurrence* of a mechanisms' overall behavior'. The explanatory request concerns multiple features of an explanandum phenomenon including, in addition to its 'manifestation', factors that 'modulate' the phenomenon, i.e., higher order constitutive factors that affect the manner in which the phenomenon manifests itself. Hence, it makes sense why mutual manipulability tracks both primary and higher order constitutive factors, and why models are more complete than abstract ones. Moreover, structural features often will be important in this explanatory context and thus referred to in more complete models. Consider, for instance, Craver's (2007) example of the mechanism(s) for the action potential. Spatial details are here key for understanding this mechanism; the *size* of ion channels affects the flow of ions and their

fit in small patches of membrane, their *shape* matters for functioning as channels and gating ion flows in the right fashion (Craver 2007, 137). Phrased in 'primary versus higher order' parlance, ion channels, functionally defined, have the function of gating ion flows and fulfillment of this function is, amongst a host of other processes, required for action potentials to occur; the structural specifics of ion channels, such as their size and shape, make a difference to the precise way(s) in which action potentials occur. Contingent on such structural features, more or less ion flows occur, resulting in greater or lesser voltage potentials, respectively. (within certain limits of course, if size or shape are outside a certain range, action potential mechanisms may shut down).

So rather than competing perspectives on explanatory power, abstract, skeletal models and more detailed ones that refer to structural features, have explanatory traction in different explanation seeking contexts. In other words, 'abstraction' and 'completeness and specificity' are explanatory virtues or desiderata in different explanatory contexts. That said, Craver (2007, 111) advances his account of explanatory power as a "regulative ideal for explanation". The above analysis debunks this perspective, for it depends on the explanatory request whether abstract or more complete models are optimal.

The relationships between explanatory requests and explanatory desiderata immediately invites the questions how the virtues of 'abstraction' and 'completeness and specificity' fare in the context of reverse engineering explanation of overall behaviors of technical systems-mechanisms, and in the context of malfunction explanation. I argue that in this latter context these desiderata pull in opposite directions and that a novel desideratum is required for malfunction explanation: *'local specificity and global abstraction'*.

4.3 Malfunction explanation: local specificity and global abstraction[23]

In the context of reverse engineering explanation presented here, engineers take details to matter: elaborate behavior functional decompositions, and related component models, are constructed to describe the mechanisms of artifacts, via the breaking down of artifacts component-by-component and assessing the effects of single component removals on their overall behaviors. This perspective agrees with the 'completeness and specificity' view on mechanistic explanations and the mutual manipulability account for establishing (evidence for) constitutive relevance that underlies it.[24] In the model of the reverse engineered electrical screwdriver in Fig. 2, for instance, both factors that make a difference to the occurrence of the screwdriver's overall behavior are listed, such as 'supply electricity' and 'convert electricity to torque', as well as factors that affect the way in which this behavior is manifested, such as 'dissipate torque' into 'heat' and 'noise' flows, and 'allow rotational degrees of freedom' (the latter concerns controlling the movement of materials along a specific degree of freedom (Stone and Wood 2000), here appropriate hand positions for correct functioning of the screwdriver).

[23] Parts of this section draw on (van Eck 2014).

[24] The notion of monitoring the effects of single component removals on overall behaviors of technical systems as is done in reverse engineering (see section 2), corresponds to the bottom-up condition of the mutual manipulability account of changing the overall behavior by intervening to change an entity's activity (see section 3.1).

Such primary and higher order details matter given that the reverse engineering explanation ultimately is in the service of redesign purposes: identifying components that function sub-optimally in a reverse engineered artifact and subsequent optimization in redesigned artifacts. The manner in which a technical system exhibits a given piece of behavior then becomes important. For instance, in the earlier discussed example of the electric wok redesign, structural features of components were relevant to the precise manner in which temperature distribution was manifested and to optimize temperature distribution across the bowl; the electric heating elements of the wok, such as a bimetallic temperature controller, were housed in too narrow a circular channel and optimized in the redesign phase (Otto and Wood 1998).

Yet, reverse engineering explanation is also used to build design knowledge bases in which (configurations of) components and their functions are archived (e.g., Stone and Wood 2000; Kitamura et al. 2005), which are used for other design purposes than the redesign of reverse engineered artifacts, like routine or innovative design.[25] The knowledge bases of the Kitamura-Mizoguchi lab, for instance, contain more skeletal, abstract models and seem to list only primary factors (see Kitamura et al. 2005).[26] The abstraction perspective thus also is in agreement with certain reverse engineering contexts.

At first glance it seems that the abstraction perspective also captures malfunction explanation. In that context, as we saw, engineers advance the maxim that 'less is more' when it comes to adequate explanations. Closer inspection however reveals that in this explanatory context 'abstraction' and 'completeness and specificity' pull in opposite directions.

To see this, consider that in order to understand how a malfunctioning component or sub mechanism makes a difference to the *occurrence* of a specific system level malfunction, one needs to know how the failing component or sub mechanism is situated within a mechanism that underlies normal functioning. That is, malfunctions are identified against a backdrop of normal mechanism functioning (see Thagard 2003; Moghaddam-Taaheri 2011). This is required to explain the contrast drawn in the explanandum—why malfunction, rather than normal function. This also happens in FIL, in which function descriptions and functional decomposition models in terms of trigger-effect descriptions are used to specify normal functioning, and to provide the context against which to assess specific malfunctions, such as a trigger that occurs yet fails to result in an expected effect—say, a cooking hobs' switch that is on but does not result in the heating of a ring (Bell et al. 2007). Such contrastive factors that explain the contrast drawn in the explanandum, i.e., make the difference, between malfunction and normal function are primary ones that underlie the occurrence of the specific system-level malfunction in question. Say, in the above example, the electrical circuitry connected to the ring that is damaged as a result of which the ring does not heat, and food cannot be heated. Also the details on normal functioning that are needed to understand why the factor(s) cited in the explanans, e.g., a broken electrical wiring, is a contrastive one, concerns primary factors that underlie normal functioning. Since

[25] There is debate in engineering on the precise meanings of notions like innovative and routine design (e.g., Stone and Wood 2000; Chakrabarti and Bligh 2001). This need not concern us here.

[26] Not incidentally, the concept of function employed in the Kitamura-Mizoguchi methodology is that of *effect* function.

fact and foil in the contrastive explanandum concern the occurrence of malfunction and function, respectively, the factors needed to understand which part(s) of the mechanism malfunction and which ones function normally should be primary ones as well. Information on the precise manner in which mechanisms normally manifest their functions is irrelevant here. Knowing that rings of cooking hobs normally heat when switches are thrown is sufficient to understand that when this trigger-effect relation does not obtain, a malfunction occurs.

Also, it suffices to describe properly functioning parts of mechanisms in abstract fashion, i.e., in terms of functionally characterized components and their functions, since their job is only to highlight where in the mechanism a malfunctioning component or sub mechanisms is located. Listing structural features, such as size and shape, is irrelevant here for what matters is knowing what these components/sub mechanisms (normally) do. I here label the constraint to specify common features of functioning and malfunctioning mechanisms in terms of functionally characterized components and their functions, *'global abstraction'*. However, the contrastive factor(s) that makes the difference to the occurrence of a specific system-level malfunction often will have to be described in more elaborate fashion and its description will, in addition to functional characteristics, also refer to structural features. The manner in which a component is, say, broken or worn often does make a difference to the occurrence of a system level malfunction. A rupture in the electrical wiring of the cooking hob, for instance, which leads to failure of the ring to heat. Here specificity with respect to structural features is needed as well. I label this constraint to describe both functional and structural characteristics of contrastive difference makers, *'local specificity'* (both to set it apart from 'global abstraction', and from 'completeness' in the sense of specifying both primary and higher order factors; 'local specificity' as I understand it here concerns primary factors only).[27]

Malfunction explanations thus require a format in between 'completeness and specificity' and 'abstraction': they require *local specificity* with respect to descriptions of malfunctioning components/sub mechanisms and *global abstraction* with respect to descriptions of the mechanisms in which the component/sub mechanism failures are placed. This analysis extends current thinking about the explanatory power of mechanistic explanations by spelling out a novel desideratum for malfunction explanations. The lesson is that in this context, explanations that contain local specificity and global abstraction are better than either complete or abstract mechanistic explanations. And, as we saw, in the context of engineering science, depending on the richness that is required of explanations, specific concepts of technical function and functional decomposition are invoked. The examples of reverse engineering explanation analyzed here use behavior functions and functional decompositions, whereas malfunction explanations are procured in terms of effect functions and functional decompositions.

A further question emerges: is 'local specificity and global abstraction' a desideratum only for malfunction explanations of technical systems, or does it also apply to malfunction explanations in other scientific domains, like biology? I argue below that

[27] This is in keeping with engineering practice. After a first round functional analysis of malfunction, more detailed behavioral models of components and their behaviors are used in FIL for assessing specific structural characteristics of malfunctioning components (Bell et al. 2007).

explanations of biological malfunctions also best exhibit 'local specificity and global abstraction'.

4.4 Malfunction explanation in biology

Also in the case of explaining biological malfunction, I take it that explanations that are locally specific and globally abstract are the optimal ones. Consider, for instance, impaired blood circulation in the circulatory system.[28] Malfunction explanations, of course, should single out those steps—entities engaging in activities—in the circulatory system's mechanism(s) that cause the circulation of blood to be impaired, i.e., make a difference to whether or not impaired blood circulation occurs. In the case of impaired blood distribution, the cause may be that blood transport is disrupted in particular vessels as a result of thrombosis in those vessels. The description of these contrastive factors—damaged vessels due to thrombosis—often will have to be described in elaborate fashion, i.e., in terms of both functional and structural specifics. In our example, it is relevant to know that the damaged vessels fail to perform their function of transporting blood. Yet the manner in which those vessels are damaged, and thus fail to perform their function(s), also makes a difference to the occurrence of impaired blood circulation. When the vessels are only slightly damaged they may still perform their function of transporting blood, so it is relevant to know the nature of the damage, i.e., the manner in which structural features of the vessels are deformed. Here, deformations due to thrombosis. Local specificity thus applies to descriptions of such contrastive difference makers.

And, again, to explain the contrast drawn in the explanandum—why malfunction, rather than normal function—one also needs to know how the failing component or sub mechanism is situated within a mechanism that underlies normal functioning, since malfunctions are identified against a backdrop of normal mechanism functioning (see Thagard 2003; Moghaddam-Taaheri 2011). However, descriptions of the relevant properly functioning parts of mechanisms can be given in abstract terms—functionally characterized components and their functions—since their job is only to highlight where in the mechanism a malfunctioning component or sub mechanisms is located. It suffices to know that, say, the cardiac muscle engages in coordinated contraction, that blood is ejected from the ventricles into the aorta and the arterial system, etc. Further detailing of structural specifics, say, the precise shape or size of the cardiac muscle has no added value for locating the fault(s) in the mechanism. So, the desideratum of 'local specificity and global abstraction' is not restricted to malfunction explanations of technical systems, but applies more broadly to malfunction explanations in the biological domain as well.

4.5 Contrastive power of behavioral explanations

Another moral can be drawn from our analysis of malfunction explanations in engineering science. There is a widespread assumption in the mechanist literature that all models, in order to be explanatory at all, should refer to the underlying mechanisms by which phenomena are produced (opinions diverge, as we saw, with which level of

[28] I adapt this example from (Nervi 2010)

detail to describe those mechanisms). On this view, mere (input–output) descriptions of overall system-level behaviors do not explain, but are in need of explanation (Glennan 2005; Craver 2007; Kaplan and Craver 2011). I agree that specification of underlying mechanisms, which depending on explanatory context should be rich or limited in detail, greatly increases the explanatory power of explanations. Yet, there are explanatory contexts in which input–output descriptions of system-level behaviors, i.e., behavioral explanations, also have explanatory power to some extent.[29] One such context is malfunction explanation. As we saw, explananda in malfunction explanations, here of technical systems, are drawn contrastively: why does a given technical system not exhibit a function that we expect it to serve? In such a request for explanation, a contrast is drawn between technical systems of a certain type that do function as expected and a system of that type that does not function as expected.

Against the view that system level input–output descriptions do not explain, I submit that functional descriptions of system-level functions as given in the FIL methodology for malfunction analysis (see section 3) already provide a course-grained answer to such contrastive questions.[30] Recall that functions in FIL are represented in terms of trigger-effect pairs, which highlight input conditions and effects that obtain when technical systems function properly. For instance, a car's stoplight function that is described in terms of the trigger "depress_brake_pedal" and the effect "red_stop_lamps_lit". In the case of normal functioning, if the brake pedal is depressed, the stoplights will be on. If either or both of these behavioral states do not obtain, this signals a malfunction of some sort. What we thus see here is that an input–output description, a trigger-effect representation, can be invoked to explain a contrast, albeit at a coarse-grained system-level, between normal function and malfunction. If either one of these trigger or effect states or both do not obtain, this marks a contrast between normal function and malfunction. Say, the brake pedal is depressed, but the lights are not on, signaling a component malfunction of some sort. Here, input–output system level descriptions do more than merely characterize explananda, since they, in addition, refer to a difference between systems that function as expected and ones that do not. They explain, albeit in a course grained fashion, some aspects of the contrast. And, moreover, they are heuristically useful for constructing more elaborate explanations for system-level malfunctions (as is precisely what happens in the FIL methodology when system-level functions are decomposed into sub-functions).

5 Conclusions

Mckay Illari and Williamson started their analysis on the general structure of mechanistic explanation by stating that: "There has been great progress in understanding mechanistic explanations in particular domains, but this progress needs to be extended to cover all sciences" (2010, 279). In this paper I have advanced this project further by applying the mechanistic account of explanation to engineering science. I discussed

[29] I adopt the term 'behavioral explanation' from Matthewson and Calcott (2011).

[30] I am not claiming that input–output descriptions are always explanatory. Yet, the manner in which input and output are specified in FIL-functional descriptions does confer explanatory traction on such descriptions, since they provide course-grained answers to the question 'why malfunction, rather than normal function?'.

two ways in which this extension offered further development of the mechanistic view. First, empirically-informed understanding of explanation in engineering science: functional individuation of mechanisms in engineering science proceeds by means of two distinct sub types of role function, *behavior* function and *effect* function, rather than a single role concept of function. Engineers simplify or increase the details of explanations depending on the explanatory purpose at hand, and these adjustments are made using specific sub types of role function. Second, it offered refined assessment of the explanatory power of mechanistic explanations. It was argued, using a case of reverse engineering explanation, that two allegedly competing views on the explanatory power of mechanistic explanations, in fact, are not in competition, but emphasize different explanatory virtues that hold in different explanation-seeking contexts. In addition, it was argued that in the context of malfunction explanation of technical systems, two key desiderata for mechanistic explanations endorsed by these perspectives, 'completeness and specificity' and 'abstraction', pull in opposite directions. I elaborated a novel explanatory desideratum to accommodate this explanatory context, dubbed 'local specificity and global abstraction', and further argued that it also holds for mechanistic explanations of malfunctions in the biological domain. I argued for these claims in terms of reverse engineering and malfunction explanations in engineering science. The analysis presented here explicitly relates specific explanatory desiderata on mechanistic explanations to specific explanation-seeking contexts. I hope that this insight proves to be a useful for analyzing the topic of explanatory power in other explanation-seeking contexts and domains besides the ones addressed here.

Acknowledgments I am grateful for helpful comments by Conor Dolan, Phyllis Illari, Bert Leuridan, and Huib Looren de Jong. Comments by two anonymous referees proved particularly rewarding.

References

Barros, B. (2009). Natural selection as a mechanism. *Philosophy of Science, 75*(3), 306–322.

Bechtel, W. (2006). *Discovering cell mechanism; The creation of modern cell biology*. Cambridge: CUP.

Bechtel, W. (2008a). Mechanisms in cognitive psychology: what are the operations? *Philosophy of Science, 75*, 983–994.

Bechtel, W. (2008b). *Mental mechanisms: philosophical perspectives on cognitive neuroscience*. London: Routledge.

Bechtel, W., & Abrahamson, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441.

Bell, J., Snooke, N., & Price, C. (2007). A language for functional interpretation of model based simulation. *Advanced Engineering Informatics, 21*, 398–409.

Calcott, B. (2009). Lineage explanations: explaining how biological mechanisms change. *British Journal for the Philosophy of Science, 60*, 51–78.

Carrara, M., Garbacz, P., & Vermaas, P. E. (2011). If engineering function is a family resemblance concept: assessing three formalization strategies. *Applied Ontology, 6*, 141–163.

Chakrabarti, A., & Bligh, T. P. (2001). A scheme for functional reasoning in conceptual design. *Design Studies, 22*, 493–517.

Chandrasekaran, B., & Josephson, J. R. (2000). Function in device representation. *Engineering with Computers, 16*, 162–177.

Craver, C. F. (2001). Role functions, Mechanisms, and Hierarchy. *Philosophy of Science, 68*, 53–74.

Craver, C. F. (2007). *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.

Craver, C. F. (2012a). Functions and mechanisms: A perspectivalist account. In P. Huneman (Ed.), *Functions* (pp. 133–158). Dordrecht: Springer.
Craver, C. F. (2012b). Explanation: The ontic conception. In A. Hutteman & M. Kaiser (Eds.), *Explanation in the biological and historical sciences*. Berlin: Springer.
Craver, C. F., & Bechtel, W. (2005). Mechanisms and mechanistic explanation. In S. Sarkar & J. Pfeiffer (Eds.), *The philosophy of science: an encyclopedia* (pp. 469–478). New York: Routledge.
Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563.
Craver, C. F., & Darden, L. (2001). Discovering mechanisms in neurobiology. The case of spatial memory. In P. Machamer et al. (Eds.), *Theory and method in the neurosciences*. Pittsburgh: University of Pittsburg Press.
Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72*, 741–765.
Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge: Cambridge University Press.
Darden, L., & Craver, C. F. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in the History and Philosophy of the Biological and Biomedical Sciences, 33*, 1–28.
De Ridder, J. (2006). Mechanistic artefact explanation. *Studies in History and Philosophy of Science, 37*, 81–96.
Deng, Y. M. (2002). Function and behavior representation in conceptual mechanical design. *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 16*, 343–362.
Erden, M. S., Komoto, H., Van Beek, T. J., D'Amelio, V., Echavarria, E., & Tomiyama, T. (2008). A review of function modeling: approaches and applications. *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 22*, 147–169.
Gervais, R., & Weber, E. (2013). Plausibility versus richness in mechanistic models. *Philosophical Psychology, 26*(1), 139–152.
Glennan, S. (2005). Modeling mechanisms. *Studies in the History and Philosophy of the Biological and Biomedical Sciences, 36*(2), 375–388.
Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntniss, 72*, 251–266.
Hawkins, P. G., & Woollons, D. J. (1998). Failure modes and effects analysis of complex engineering systems using functional models. *Artificial Intelligence in Engineering, 12*(4), 375–397.
Illari, P. (2013). Mechanistic explanation: integrating the ontic and epistemic. *Erkenntnis*, online first. doi:10.1007/s10670-013-9511-y.
Kaplan, D. M., & Bechtel, W. (2011). Dynamical models: an alternative or complement to mechanistic explanations? *Topics in Cognitive Science, 3*, 438–444.
Kaplan, D., & Craver, C. (2011). The explanatory force of dynamical and mathematical models in neuroscience: a mechanistic perspective. *Philosophy of Science, 78*, 601–627.
Kitamura, Y., Koji, Y., & Mizoguchi, R. (2005). An ontological model of device function: industrial deployment and lessons learned. *Applied Ontology, 1*, 237–262.
Levy, A., & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of Science, 80*, 241–261.
Lipton, P. (1993). Making a difference. *Philosophica, 1*, 39–54.
Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 57*, 1–25.
Matthewson, J., & Calcott, B. (2011). Mechanistic models of population-level phenomena. *Biology and Philosophy, 26*(5), 737–756.
McKay Illari, P., & Williamson, J. (2010). Function and organization: comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences, 41*, 279–291.
McKay Illari, P., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*, 119–135.
Millikan, R. (1989). In defense of proper functions. *Philosophy of Science, 56*, 288–302.
Moghaddam-Taaheri, S. (2011). Understanding pathology in the context of physiological mechanisms: the practicality of a broken-normal view. *Biology and Philosophy, 26*, 603–611.
Nervi, M. (2010). Mechanism, malfunctions and explanation in medicine. *Biology and Philosophy, 25*, 215–228.
Otto, K. N., & Wood, K. L. (1998). Product evolution: a reverse engineering and redesign methodology. *Research in Engineering Design, 10*, 226–243.
Otto, K. N., & Wood, K. L. (2001). *Product design: techniques in reverse engineering and new product development*. Upper Saddle River: Prentice Hall.

Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese, 183*, 283–311.

Price, C. J. (1998). Function-directed electrical design analysis. *Artificial Intelligence in Engineering, 12*(4), 445–456.

Salmon, W. (1984). Scientific explanation: three basic conceptions. *Proceedings of the Biennial Meeting of the Philosophy of Science Association, 2*, 293–305.

Skipper, R., & Milstein, R. (2005). Thinking about evolutionary mechanisms: natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 327–347.

Stone, R. B., & Chakrabarti, A. (2005). Guest editorial. Special issue: engineering applications of representations of function, part 2. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 19*(3), 137.

Stone, R. B., & Wood, K. L. (2000). Development of a functional basis for design. *Journal of Mechanical Design, 122*, 359–370.

Strevens, M. (2004). The causal and unification approaches to explanation unified—causally. *Noûs, 38*(1), 154–176.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Cambridge, MA: Harvard University Press.

Sustar, P. (2007). Neo-functional analysis: phylogenetical restrictions on causal role functions. *Philosophy of Science, 74*, 601–615.

Thagard, P. (2003). Pathways to biomedical discovery. *Philosophy of Science, 70*, 235–254.

Van Eck, D. (2010). On the conversion of functional models: bridging differences between functional taxonomies in the modeling of user actions. *Research in Engineering Design, 21*(2), 99–111.

Van Eck, D. (2011). Supporting design knowledge exchange by converting models of functional decomposition. *Journal of Engineering Design, 22*(11–12), 839–858. doi:10.1080/09544828.2011.603692.

Van Eck, D. (2014). Validating function-based design methods: an explanationist perspective. *Philosophy and Technology*, online first. doi:10.1007/s13347-014-0168-5.

Van Eck, D., & Weber, E. (2014). Function ascription and explanation: elaborating an explanatory utility desideratum for ascriptions of technical functions. *Erkenntnis, 79*, 1367–1389. doi:10.1007/s10670-014-9605-1.

Vermaas, P. E. (2009). The flexible meaning of function in engineering. *Proceedings of the 17th International Conference on Engineering Design (ICED 09), 2*, 113–124.

Walsh, D. M., & Ariew, A. (1996). A taxonomy of functions. *Canadian Journal of Philosophy, 26*(4), 493–514.

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy, 104*(12), 639–659.

Woodward, J. (2003). *Making things happen*. Oxford: Oxford University Press.

Wright, C. D. (2012). Mechanistic explanation without the ontic conception. *European Journal for the Philosophy of Science, 2*, 375–394.

Wright, C., & Bechtel, W. (2007). Mechanisms and psychological explanation. In P. Thagard (Ed.), *Philosophy of psychology and cognitive science* (pp. 31–79). Amsterdam: Elsevier.