Final Draft – January 20, 2006

The (Alleged) Inherent Normativity of Technological Explanations

Jeroen de Ridder

Delft University of Technology, Department of Philosophy

Jaffalaan 5, 2628 BX Delft, The Netherlands

phone: +31 15 2785141, fax: +31 15 2786233

email: g.j.deridder@tbm.tudelft.nl

Abstract

Technical artifacts have the capacity to fulfill their function in virtue of their physicochemical make-up. An explanation that purports to explicate this relation between artifact function and structure can be called a technological explanation. It might be argued, and Peter Kroes has in fact done so, that there is something peculiar about technological explanations in that they are intrinsically normative in some sense. Since the notion of artifact function is a normative one (if an artifact has a proper function, it *ought* to behave in specific ways) an explanation of an artifact's function must inherit this normativity.

In this paper I will resist this conclusion by outlining and defending a 'buck-passing account' of the normativity of technological explanations. I will first argue that it is important to distinguish properly between (1) a theory of function ascriptions and (2) an explanation of how a function is realized. The task of the former is to spell out the conditions under which one is justified in ascribing a function to an artifact; the latter should show how the physicochemical make-up of an artifact enables it to fulfill its function. Second, I wish to maintain that a good theory of function ascriptions should account for the normativity of these ascriptions. Provided such a function theory can be formulated — as I think it can — a technological explanation may pass the normativity buck to it. Third, to flesh out these

abstract claims, I show how a particular function theory — to wit, the ICE theory by Pieter Vermaas and Wybo Houkes — can be dovetailed smoothly with my own thoughts on technological explanation.

Keywords

technical artifacts, explanation, mechanisms, normativity, proper function,

1. Introduction

To introduce the topic of this paper, here are two observations about technical artifacts. First, technical artifacts have proper functions; that is the very reason behind our designing, making, and using them. They have their functions partly in virtue of their physicochemical make-up. One cannot reasonably ascribe the function to *f* to an artifact, which one knows to have an utterly inappropriate physicochemical constitution — a pencil cannot function as a laptop computer. Hence, there must be some sort of explanatory link between an artifact's function and its physicochemical make-up (or, for short, its 'structure'). When one wants to understand how it is that artifact *x* has the function to *f*, there will be mention of *x*'s structure at some point.

Second, the notion of proper function is a normative one. It makes sense to say of an artifact that it *ought to* exhibit certain behaviors, namely those associated with its function. Such claims do not make sense for normal physical objects, such as stones, solar systems, or sugar molecules.[1] There can be discrepancies between an artifact's proper function and its actual behavioral capacities. An artifact can have the function to *f* even though it cannot *f*. A

[1] At least not in as strong a sense as for artifacts. Of course we can express our (sometimes strongly) inductively supported beliefs about the behavior of physical objects in terms of normative 'ought to'-claims, but it is not as if we have some sort of *right* to expect physical objects to behave as we desire — as *is* the case for technical artifacts (cf. Franssen 2006). More on this in sections 4 and 5.

broken television set still is a television set and the proper function of a worn-out light bulb still is to provide light.[2]

If we combine these two observations we arrive at the conclusion that there is something peculiar about *technological explanations* — i.e. explanations that account for an artifact's function in terms of its structure. Since (1) there must be an explanatory link between an artifact's function and its structure, and (2) the notion of artifact function is normative, it seems to follow that technological explanations are special by being inherently normative.

Or so Peter Kroes (1998; 2001) argues. It is my aim in this paper to scrutinize this argument for I think it runs together a couple of different points about artifact functions and explanations. In the next section, I will present Kroes's arguments in more detail. Section 3 contains internal criticism of his arguments, and in section 4 I will argue that there is a more fundamental confusion underlying Kroes's arguments and I will show how we can dispose of this confusion by analyzing his endeavor in two separate projects. We need to distinguish between a theory of artifact functions on the one hand and an account of technological explanations on the other. The former should deal with the normativity of functions, so that the latter can then pass the buck. The rest of the paper serves to flesh out this reply; in section 5 I will present a specific theory of artifact functions and show how it can be combined with an account of technological explanation in the way I envisaged in section 4. Section 6 contains the conclusion.

---

[2] There are limits here; one would be hard-pressed to still call a television set that has been smashed to a thousand pieces with a jackhammer a television set.

2. Kroes on Technological Explanations and Normativity[3]

To argue his point about the peculiarity of technological explanations, Kroes first observes that technical artifacts have a dual nature. They are physical objects, but they also have intentional or functional properties essentially. As a result, we can give both functional and physicalistic descriptions of artifacts, with either description partially or wholly black-boxing the other. A clock is any time-keeping device, whatever its exact physicochemical make-up and, alternatively, someone without any experience with pencils cannot deduce that a 6-inch hexagonal elongated piece of wood with a lead inside is for writing (though she might discover that it can be used for writing). The two descriptions are logically independent and, as a result, it is impossible to deduce function from structure or the other way around. Standard deductive-nomological explanations are barred.

Next, he presents an example of a technological explanation that involves the Newcomen steam engine. The main function of Newcomen steam engines was to drive water pumps. They did so by means of the up-and-down movements of their great beam. The great beam itself was driven by the actual steam engine that consisted of a boiler, a steam valve, and a cylinder with moving piston (see Figure 1). Roughly, the explanation of these engines has three ingredients.

(1) Physical laws or phenomena, e.g., that steam occupies a much larger volume than does water, that rapid condensation of steam in a closed vessel creates a partial vacuum, that atmospheric pressure exerts a force on the piston.
(2) The physical make-up and configuration of the engine, e.g., the boiler, steam valve, movable piston, and great beam.

[3] This section summarizes sections 4 and 5 of (Kroes 1998).

(3) Dynamic behaviors and causal interactions of the components, e.g., heating and expansion of water and steam, opening and closing of the steam valve, injection of cold water, condensation of water, creation of a partial vacuum, and movements of the piston.

Kroes rightly observes that it does not follow from an explanation along these lines that the function of the steam engine is to drive pumps, nor that it is to move the great beam up and down. All that follows is that the steam engine can be used to drive pumps, that it is a means to that end, or that it has the capacity to drive pumps. It is impossible to get the normative *explanandum* containing the ascription of a proper function from the purely descriptive *explanans*. He concludes that the explanation as presented is not a technological explanation since it does not properly account for the steam engine's function in terms of its structure.

Kroes (2001: 38-9) contains a sketchy possible repair. Perhaps, says Kroes, the relation between *explanandum* and *explanans* can be conceived in terms of pragmatic rules of actions that are grounded in causal relations. For example, if one's goal is to drive a water pump, and a steam engine has the capacity to do so (i.e., something like the following causal conditional holds: If the steam engine is put to use properly in appropriate circumstances, it will drive a water pump), then one can infer the following rule of action: To drive a water pump, use a Newcomen steam engine. In this context of action, the steam engine is a means to an end and acquires a function. The engine's physical structure still figures indirectly, since the rule of action is formulated on the basis of a causal conditional that was derived from the engine's structure. Kroes concludes: "A technological explanation, therefore, is not a deductive explanation, but it connects structure and function on the basis of causal relations and pragmatic rules of action based on these causal relations." (2001: 39). So, in sum, Kroes's points are: (1) a technological explanation must account for an artifact's function in terms of

its structure, (2) 'standard' explanations (along the lines of the D-N model or a somewhat loosened version of it, as in the example above) cannot accomplish that task, and (3) using the notion of action rules, it seems possible to construe a more adequate account that does connect structure and function in the desired manner.

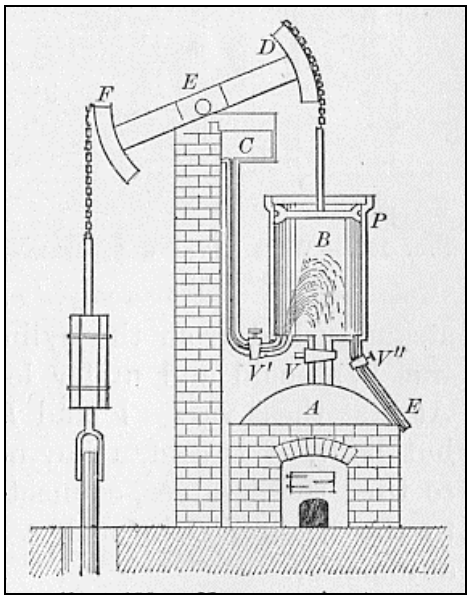

Figure 1 – Newcomen's steam engine

3. Kroes's Arguments Reconsidered

In my opinion, there is something seriously wrong with these arguments. I will argue that Kroes's arguments do not show what they purport to show, even on their own terms, and, in the next section, that his construal of technological explanations runs into trouble because it conflates two rather different projects. As a result, Kroes's effort has to satisfy a set of inconsistent requirements and is doomed to fail.

I can be relatively brief about the first point. It is not clear which of the following claims Kroes aims to establish:

(1) A technological explanation of Newcomen's steam engine does not fit the mold of the D-N model of explanation.

(2) Most or all technological explanations do not fit the mold of the D-N model.

(3) Most or all technological explanations do not fit any of the currently available models of explanation.

I think he should be interested in (3), because that would be a good reason to think that there is something truly peculiar about technological explanations. If the currently available accounts of explanation (such as the D-N account, unification accounts, and causal accounts) are capturing important aspects of what it is to be an explanation, and if technological explanations do not conform to any of these accounts, then they might represent an interesting new species of explanation worthy of philosophical attention. Unfortunately, however, the only claim Kroes establishes with some plausibility is (1). To be fair, I should add that if (1) is correct and the explanation of the steam engine is a representative example of technological explanations in general, the truth of (1) lends inductive support to (2). So to the extent that this inductive argument is compelling, the plausibility of (2) is established as well. But the plausibility of (2) does very little to prove (3). For that, it would have to be shown that technological explanations fit none of the currently available accounts of explanation, e.g. Friedman's and Kitcher's unification accounts, Salmon's, Woodward's, and other causal accounts, Van Fraassen's pragmatic account, and Cartwright's *simulacrum* account. Even accounts of intentional explanation might be relevant if one thinks artifact functions are intrinsically related to agents' intentions. Or accounts of social explanation, if one is of the opinion that artifact functions are inherently social phenomena. I am not saying that this cannot be done, but Kroes has certainly not done it. He has only shown that technological explanations cannot be construed as D-N explanations. While that may be perfectly true, it is

hardly a reason for distress, since for many the D-N model has by now been relegated to the domain of philosophical relics. In fact, as I will make clear in due course, there is every reason to think that his construal of technological explanations suffers from internal inconsistencies to such an extent that no account of explanation ought to fit it, on pain of being inconsistent itself.

As far as I can see, the suggestion to construe the relation between *explanandum* and *explanans* in terms of action rules is not successful either. The step from a causal relation to a rule of action is relatively unproblematic: If one wants a certain effect and one knows one or more sufficient cause(s) of this effect, then, given the usual *ceteris paribus* clauses for causal relations and some hedging assumptions about the proportionality of the means in relation to the end, it is perfectly rational from a practical point of view to bring about this effect by bringing about one of these sufficient cause(s). If Newcomen's engine can drive a water pump if it is operated properly, one could use it to pump water if one wants so, but — and this illustrates the chief difficulty — in the same vein we can add that, if it can be used to tear stuff apart, one could use it to tear stuff apart if that is what one wants. One can use an electric guitar to play licks, and if one so desires, it would be rational to use it for that purpose, but if one is in a rockstar-type of mood, a guitar can also be used to smash loudspeakers, and it would be no less rational to use it to that end. None of this, however, goes to show that Newcomen's steam engine is for tearing stuff apart or that smashing loudspeakers is an electric guitar's proper function.

Although the fact that something has a number of capacities that can be expressed in terms of causal conditionals warrants inferences to various rules of action (under the assumptions mentioned), nothing supports one of these rules in particular as the *proper* one, and neither does the artifact considered in isolation give you any reason to suppose that one of these causal capacities is the artifact's *proper* function, as opposed to an accidental or system

function, i.e. just something it can do. While causal knowledge may underpin rules of action, I do not see how it could sustain proper function ascriptions. In the end, the suggested repair is not much of an improvement over Kroes's initial proposal. All that can be inferred from action rules is that if a certain artifact can be used to accomplish some end, then it is rational to use it to that end, but that follows virtually analytically (again, given some background assumptions) from the fact that it *is* a means to that end, and that was already established in the initial proposal.

4. Functions: To Ascribe and to Explain

Given that Kroes's project leads to a dead end considered by its own lights, let us now take a step back and turn to the second point. I will argue that there is a more fundamental confusion vexing the project. Unearthing this confusion will also enable us to see why his project really was a non-starter. Kroes stipulates that a technological explanation is an explanation that accounts for an artifact's proper function in terms of its physicochemical make-up. This construal is, I think, seriously misguided because it runs together two rather different projects, to wit (1) that of giving an account of proper function ascriptions and (2) that of explaining how, in virtue of its physicochemical make-up, an artifact can fulfill its function. The result of (1) is a set of necessary and jointly sufficient conditions for the truth or assertibility of claims like 'artifact *x* has proper function *f*.' It is fairly obvious that this set will contain more conditions than just those related to the *x*'s physicochemical make-up — that is in fact the negative result of Kroes's argument: claims about proper functions cannot be deduced solely from information about the artifact's physicochemical make-up. But it is not so obvious that something like a highly detailed account of *x*'s workings must be among these conditions, for that would mean that no one except highly knowledgeable engineers could ever be justified or correct in claiming that an artifact has a proper function. Project (2), on the other hand, provides an account of how an artifact's physicochemical make-up enables it to exhibit the

behaviors required for its proper functioning and here the notion of mechanistic explanation immediately springs to mind. What Kroes tries to do, however, is to get the results of both project (1) and (2) while drawing exclusively on the means for project (2). That is an impossible task.

An analogy will help to clarify the reason why. Suppose we want to explain why the function of the heart is to pump blood, or, more precisely, to determine whether the proper function of the heart is to pump blood. Surely, an elaborate scrutiny of hearts and their behavior by itself will not allow us to conclude that their proper function is to pump blood, yet this is the only option open to us on an extrapolated version of Kroes's proposal, since he seems to be thinking that an item's proper function could be determined just by looking at its physicochemical make-up. Instead, we should distinguish the project of spelling out the truth or assertibility conditions for "The function of the heart is to pump blood", from that of explaining how the heart is able to pump blood. Accounting for the fact *that* the function of the heart is to pump blood is not the same as accounting for *how* it can pump blood. The first project will involve more than just the heart's 'intrinsic' properties. Biological function theories disagree on exactly what more; some suggest synchronic relational properties such as the heart's current contribution to organism fitness (Walsh 1996; Lewens 2004), others look at diachronic relational (historical) properties such as the heart's ancestors contribution to ancestor fitness (Millikan 1984, 1993; Neander 1991a, 1991b). The outcome of the second project, however, will look more like Kroes's proposed *explanans*. It will explicate how the physicochemical make-up of the heart and its constituent parts in their particular configuration leads to dynamic behaviors that, in the appropriate environment, add up to pumping blood.

The crucial point is that accounting for an item's proper function, on the one hand, cannot be done without taking the item's environment into account, be it its current ecological niche,

its history, its ancestors, its users, its designers, or their intentions and/or (justified) beliefs. Proper functions are not among the intrinsic properties of an item and therefore they cannot be discovered by solely looking at the item itself, isolated from its environment. An item's capacities and its behaviors, on the other hand, are among its intrinsic properties and can be explained by looking just at the item's physicochemical constitution and mereological make-up. The two projects are largely independent. One can be justified, even correct, in ascribing proper functions to organs or artifacts without knowing how they are able to perform that function, and, alternatively, one can explain how it is that organs or artifacts (or their parts) have the capacities they have or show the behaviors they show without knowing that one of these capacities or behaviors is associated with a proper function. Of course, one is typically interested in an explanation of how an organ or artifact can perform the behavior associated with its proper function, since that tends to be its most interesting feature (that computers can function as paperweights is not the reason people buy them).

What I have said so far should not be taken to imply that the projects are entirely unrelated; I have only argued that it is unwise to try and tackle them in one fell swoop. I now want to look at possible connections, two in particular. The first one is that an explanation of how something is able to perform its function might pop up in the justification for its having that function. Roughly, the intuition is that function ascriptions must have something to do with the actual behavioral capacities an object has, at least for paradigm exemplars of the object. In order to justify the claim '*x* has proper function *f*' (where *x* is a normal exemplar of its type) there must be evidence that *x* can in fact *f*, and an adequate explanation of how *x* can *f* would be very good evidence, albeit not the only permissible type of evidence. Naïve theories of artifact functions overlook this intuition. Consider a theory that defines the function of an artifact to be what the designer intended the artifact to do. Such a theory lacks the evidence-requirement and thereby fails to link claims about proper functions to (evidence of) actual

capacities. As a result, it allows for crazy function ascriptions. A mad designer's *intention* to build a spacecraft from a bunch of matchsticks does not warrant the conclusion that the result he produces *is* a spacecraft, for there is no way in which matchsticks could ever compose a spacecraft, at least not by current scientific lights. So the first way in which the two projects are related is by way of justification. An explanation of how something can fulfill its function can be among the justificatory grounds for the claim that an artifact has that proper function.

The second connection appears in malfunction cases: situations where an item still *has* a proper function, even though it cannot *perform* that function. I assume that such cases do exist, both in biology and technology; malfunctioning hearts are still for pumping blood, and the proper function of a worn-out light bulb still is to provide light.[4] For malfunction cases, the second project I identified takes on a slightly different form, since the question of how the artifact can perform its function is obsolete when we know that it cannot perform its function. What can be explained, however, and what is not obsolete, is how the artifact was *supposed* to perform its function. An answer to that question will look a lot like the answer to the original explanatory question, except that it will be phrased in normative or counterfactual terms. It explicates how the various parts *ought* to be configured, behave, and interact, or how they would have been configured and how they would have behaved and interacted, were the artifact to function properly.[5] Even if one does not think that this answer is valuable in and of

---

[4] One might argue over whether cases of worn-out artifacts properly belong under the heading of malfunction. For example, light bulbs are apparently designed so as to stop working after a certain amount of burning hours. I can see that one might interpret this as evidence that wearing out is in fact part of the proper function of a light bulb. For brevity's sake I will ignore this terminological quibble while taking it to be uncontroversial that a worn-out artifact still has a proper function.

[5] Establishing the truth of counterfactual claims is a notoriously troublesome issue, which I cannot hope to address to any satisfactory extent here. I rely on an intuitive way of thinking about it, but will

itself, it should be obvious that it has instrumental value as background knowledge for determining the causes of malfunction. Only in contrast to how the artifact was supposed to work will it become possible to find out how it malfunctions.

Unlike scientific explanations of natural phenomena, technological explanations can inherit the normativity of function ascriptions. Although we might claim that photons 'ought' to behave as particles, this only goes so far as the theory from which we infer this claim has been inductively supported or as our previous experiences lend inductive support to such a claim. Such claims merely express inductively supported expectations about phenomena. Technological explanations, however, can incur an extra and stronger type of normativity in that there are independently ascertainable and objective facts of the matter as to how the artifact and its components ought to behave. These facts are grounded in the justified beliefs, intentions, and communicative actions of the designer(s) who devised the artifact or in the beliefs, intentions, and actions of the (group of) users who put the artifact to a new use that has gradually become widespread standard use.[6] Under the assumption that she is *competent*, i.e. broadly rational and skillful and in possession of appropriate justification for the beliefs upon which she acts, a designer objectively determines an artifact's proper function. That fact entitles us to objective claims about what this proper function is, even in the face of malfunction. Note that the competence assumption is essential: only if designers tend to have

---

add one important qualification. The possible worlds taken into account must be limited to those close to our own with roughly similar laws of nature. Without this constraint, it may be possible to think of worlds where materials and artifacts have very different properties and capacities so that, say, a bunch of matchsticks could compose a spacecraft. If this brief remark does not satisfy the reader, my advice is to forget about the counterfactual reading altogether and focus on the normative reading.

[6] For brevity's sake, I will ignore such user-imparted proper functions for the rest of this section, but a story very similar to the story I am about to tell can be told about them.

correct beliefs about the workings of the artifacts they devise, skillfully build the artifacts they devise (or see to it that this gets done), and truthfully communicate about functions, will we have additional reasons, beyond mere past experience or other inductive support, for claiming that artifacts ought to behave such-and-such.

Looking at the kinds of justification involved can further bring out the difference. The justification for ought-claims about malfunctioning artifacts differs in kind from the sorts of justification we might have for normative statements about the behavior of natural objects. Of course, designers base their beliefs on scientific theories or practical experience with the materials they use and in this sense their knowledge about artifacts parallels the sort of knowledge scientists have about natural phenomena. An engineer's claim that an iron bar ought not to buckle under a specific pressure does not differ in kind from the claim that a photon ought to behave as a particle; both are supported by normal scientific evidence. For non-designers, however, another story must be told. Provided the competence assumption mentioned above is warranted — as it certainly seems to be in our society — they can take the designer's word[7] as support for claims about proper functions and hence about how the artifact and its components ought to behave. What is more, because of social, economical, and legal arrangements in our society, users have legal rights vis-à-vis designers with regard to claims about what artifacts ought to do. Warrant for the competence assumption is officially institutionalized, so to speak. Designers are expected to be trustworthy and reliable in what they do, i.e. they are expected to be competent. Failing these expectations leads to sanctions. Because of all this, non-designers are entitled to objective normative claims about the proper functions of artifacts. The justification for such claims consists of beliefs about what the designers wanted an artifact to do. These beliefs screen off other types of justification, such as

[7] Or something derived from that through a chain of communication, e.g. what the label on the box says, or the salesperson, or your sister who just bought the artifact.

experience with the artifact, testimony about successful artifact use, or even a theory about the artifact. Of course, non-designers sometimes also have these latter justifications for a claim that artifact *x* ought to *f*, but my point is that they do not *need* it in order to be justified in claiming so. All they need for that — still under the competence assumption — is knowledge or justified belief that the designer intended *x* to be for *f*-ing. This screens off other types of justification and provides just the extra normative force that adheres to proper function claims and that can be inherited by explanations of how a malfunctioning artifact ought to function.

To round off this lengthy excursion about the normativity of proper function claims, let me give an illustration. Say I have been commuting happily in my car every day for the past year, but then one morning when I turn the key it will not start. I want to claim that my car still has its proper function (say, personal motorized transportation) and that it ought to start if I turn the key, even though it presently malfunctions. What sort of evidence do I have for this claim? Obviously my past experience with the car, but that is not crucial. What is more important is that I have every reason to believe that my car was designed and built by competent engineers with the purpose of designing and building something that has the proper function of providing personal motorized transportation. Therefore, the normative claim that my car ought to start carries with it an extra and stronger normative force beyond that offered by the mere past experience induction. If I were to not have had that experience, I would still have been entitled to claim that my car ought to start. Compare this with my successfully and regularly lulling a child to sleep by the monotonous sound of driving. If one day the child will not go to sleep, I might also claim that my car ought to lull the child into sleep, but this claim clearly carries a smaller normative force because it lacks the additional support of claims related to proper functions.

What does all this mean for the (alleged) inherent normativity of technological explanations? Let me spell out the ramifications of what I have said.

(1) Proper function ascriptions have a normative force in that they can be correct of an artifact even when that artifact cannot perform its proper function. Malfunctioning artifacts still have proper functions.[8]

(2) An adequate theory of artifact functions — the result of what I dubbed project (1) — should account for this normativity, i.e. it should reproduce the better part of our intuitions about which malfunctioning artifacts nonetheless have proper functions.

(3) An account of technological explanation — the result of project (2) — may pass the normativity buck to the theory of artifact functions and need not account for normativity itself.

(4) Technological explanations can inherit the normativity of function ascriptions. If an artifact functions properly, a good technological explanation truthfully explains how it does so. If an artifact malfunctions and still has a proper function, a good technological explanation explains — in equally truthful ought-claims or counterfactuals — how the artifact was supposed to function, had it not been malfunctioning.

The obvious question is whether the two projects I have outlined are feasible. It is one thing to formulate a set of requirements that a theory of functions and an account of explanation ought to satisfy, but quite another thing to show that these requirements can be met. That is why I will use the next section to sketch a theory of functions and an account of technological explanation — the former borrowed, the latter of my own making — that, for all I can see, satisfy the requirements I have submitted.

---

[8] One might argue over whether the notion of malfunction presupposes that of a proper function so that every malfunctioning artifact necessarily has a proper function. Nothing much depends on this for me, so I leave the question undecided.

5. Making It Work

Wybo Houkes and Pieter Vermaas have developed a theory of artifact functions which seems to me perfectly suitable for the present purposes (Houkes and Vermaas 2004; Vermaas and Houkes 2006). First, a bit of background. On this theory, artifacts are embedded in the action-theoretical notion of a *use plan*: a series of considered actions undertaken to realize a practical goal desired by an agent, in which at least one of the actions involves the manipulation of the artifact. By exercising one or more of its capacities an artifact contributes to the realization of the overall goal of the plan. Designing engineers devise use plans when they design artifacts, but users are free to invent their own alternative use plans, which may subsequently become new standardized uses. The theory itself, then, is a theory about when agents are justified in ascribing functions to artifacts. Here is what it says.

> An agent *a* [justifiably, JdR] ascribes the capacity to *f* as a function to an artifact *x*, relative to a use plan *p* for *x* and relative to an account *A*, iff:
>
> I. the agent *a* has the belief that *x* has the capacity to *f*, when manipulated in the execution of *p*, and the agent *a* has the belief that if this execution of *p* leads successfully to its goals, this success is due, in part, to *x*'s capacity to *f*;
>
> C. the agent *a* can justify these two beliefs on the basis of *A*; and
>
> E. the agents *d* who developed *p* have intentionally selected *x* for the capacity to *f* and have intentionally communicated *p* to other agents *u*.
>
> (Houkes and Vermaas 2004: 65, with slight notational adjustments)

A few remarks for clarification. First, on this account functions are relativized to use plans; the latter is the more fundamental notion. Having a function means for an artifact to be embedded in a use plan that privileges one (or a few) of its many capacities as special, i.e. as

its proper function(s). Secondly, the beliefs that *x* can *f* and that its doing so contributes to the realization of the use plan's goal need to be justifiable on the basis of an account *A* (which is itself subject to normal standards of justification). This account can take on a number of forms; for new, inexperienced users it can be simple testimony or observation (having heard that this contraption is a laser pointer, or having read the inscriptions on the package), for technically savvy users who enjoy taking apart their electrical appliances, it can be practical insight in their internal workings combined with experiential knowledge, and for engineers, it will typically be full-fledged technological and scientific explanations, often combined with practical experience from prototype tests. Thirdly, as foreshadowed in the previous sections, the notion of function turns out to be a relational one. To put it somewhat crudely, artifacts by themselves do not have functions; they acquire functions in a context of use plans, users and designers, and their justified beliefs, intentions, and actions.

Does this function theory account for the normativity of proper function ascriptions? Its creators think it does and I am inclined to agree with them. For brevity's sake, I will not laboriously go over a host of examples that the theory successfully covers, but limit myself to an outline of its general strategy for coping with the normativity of function ascriptions and a discussion of one worry.[9] Since agents only need justified beliefs, as opposed to knowledge, about the artifact's capacities, the theory allows for cases in which an agent's beliefs are defeated by later evidence. In this way, one can ascribe functions to malfunctioning artifacts. I may have every reason for believing that my phone has the appropriate capacities to allow me to call my mother and fulfill all the other conditions laid down by the theory and, by that token, be justified in ascribing the function of allowing for conversations at a distance to it,

[9] A more elaborate discussion of the theory can be found in (Houkes and Vermaas 2004, 2005; Vermaas and Houkes 2006).

but if — unbeknownst to me — a practical joker has removed the microphone from my phone, it will nonetheless malfunction.

This example, however, does raise a concern, for the theory seems to imply that once I have learned of my phone's malfunctioning, I can no longer ascribe the same proper function to it because I no longer have the belief that it has the capacity to transmit my voice to the other end of the line. That is a counterintuitive result. To deal with cases like these, we must modify condition I. The agent does not have to have the belief that *x* can *f* but may also have the overriding belief that *x* would have been able to *f*, had particular counteracting interferences not occurred, or that it ought to be able to *f* given what the designers communicated about *x*. In short, condition I should read: agent *a* has the belief that *x* has or *should* have the capacity to *f*, etc. (and, of course, *a* must be able to justify this belief too). With this modified condition in place, I can still ascribe the function of teleconversation to my phone after learning about the removed microphone, for I am justified in believing that it would have had that capacity, had someone not been playing this joke on me.[10, 11]

So far so good then. The next task is to see if this function theory matches up with an account of technological explanation in the way I envisaged. Not unexpectedly, I think it does. As I argued above, such an account of explanation must deal with explanations that explicate how artifacts are able to exhibit various behaviors, and the behavior associated with their function in particular. I think the resources for this are available in the literature on mechanistic explanation, although they have not always been clearly recognized and

---

[10] To be complete, I should add that for situations where an artifact malfunctions due to normal wear and tear, the I-condition must be modified to include something like 'agent *a* knows that *x* used to have the capacity to *f* but has now stopped having that capacity due to normal wear and tear'.

[11] The suggested modifications are in line with what Houkes and Vermaas say, but not explicitly theirs.

presented. I have given the contours of this account of explanation in another paper (De Ridder 2006) and I will briefly summarize my ideas here. To explain a particular piece of artifact behavior, there are two general strategies available, leading to two different complementary types of understanding (cf. also Bechtel and Richardson 1993: 18). I have tried to capture these strategies in the following descriptions.

> *Top-down strategy*: take the behavior to be explained and decompose it into more basic sub-behaviors, reiterate this step if possible — it should become clear how the complex behavior being explained is realized by simpler behaviors in a specific spatiotemporal configuration — and for all the sub-behaviors, indicate which component(s) take(s) care of them.
>
> *Bottom-up strategy*: identify the structural components of the artifact and give information about their physicochemical make-up and spatial configuration, show how their physicochemical features and configuration result in various behaviors and then describe how these behaviors, in their spatiotemporal configuration, together make up the behavior to be explained.

The first strategy focuses on behaviors; it explicates how a complex behavior is realized by ever-simpler sub-behaviors by decomposing the overall complex behavior in its constituent sub-behaviors. It provides purely *functional*[12] understanding, solely in terms of behaviors, thereby black-boxing the physicochemical make-up of the artifact and components that exhibit these behaviors. The second strategy opens up the black box; it starts from the structural decomposition of the artifact by identifying its component parts, and then describes

[12] Note that, in this context, the term 'functional' has the weak sense of 'having to do with input-output relations only'; it does not refer to the richer notion of proper function.

their relevant characteristics (morphological, physical, and chemical properties relevant for the behavior being explained), and how these characteristics enable particular behaviors (under appropriate circumstances). Finally — and here it overlaps the first strategy — it shows how these behaviors add up to the complex behavior being explained. The second strategy offers *structural* understanding of the artifact's workings. So while the first strategy starts from a decomposition of the behavior and subsequently indicates how the structural parts fit into this functional, or behavioral, decomposition, the second strategy starts from the structural decomposition and works its way upwards to the behaviors exhibited by the structural components, showing how the behavioral decomposition maps unto the structural decomposition. Although the two strategies are complementary I do not think they should be merged into one. The demands that this merged strategy would place on a good technological explanation are too strict. Explanations that only provide functional understanding would automatically be disqualified as incomplete, whereas I am convinced — although I will not argue the point here — that they are perfectly good explanations in many contexts, not just in the pragmatic sense of being acceptable to the person with the explanatory request, but also in the stronger sense of being an objectively good explanation.

I hope this brief sketch suffices to give the reader an impression of what this account of technological explanation looks like. Let us now move on to the last part of the paper and see if this account lives up to the standards I set for it in the previous section. The crucial question is whether it can grapple with the normativity issue for malfunction cases. If we ascribe a proper function to a malfunctioning artifact on the basis of the modified ICE conditions there are three options: (1) we have a false but justified belief that the artifact has the capacity to function, (2) we have a justified (and true) belief that the artifact should have the capacity to function (in the sense described earlier), or (3) we know that the artifact used to have the capacity to function, but that it is now worn-out. In all three cases, it seems to me perfectly

possible to give an explanation of how the artifact is believed to work, supposed to work or how it used to work. In each case and for both explanatory strategies, the explanans will be phrased in normative or counterfactual terms, e.g., this component should sit here and interact with that other one right there so that they would have shown such-and-so behavior, thus contributing to the proper functioning of the artifact. Or: this light ought to go on when I hit that switch. Or: this spring used to push that thing back. Like explanations of properly functioning artifacts, these explanations can be evaluated in terms of truth, justification, acceptability, or whatever else is deemed appropriate. I don't see any particular problems about normativity left here that an account of technological explanation could not pass on to a function theory.

Someone may worry about circularity, though. If part of the justification for ascribing a function is a technological explanation and if an account of technological explanation passes the normativity buck to the function theory, doesn't that land us in some sort of justificatory circle? Not if we look closer at the exact justificatory relations. Typically, professional engineers or technically savvy 'laypersons' will have at their disposal more or less elaborate technological explanations as justifications for (some of) the function ascriptions they make. That means that they will have justified beliefs about the physicochemical properties of the artifact and its components, the components' configuration and interactions, and their behavioral capacities. But the justification for these beliefs, and hence for the explanation, in no way depends on the function ascription; instead it is based on the normal justificatory mechanisms for beliefs about stuff in the world: observation, experiments, experience, and testimony. So if an engineer ascribes a function to a malfunctioning artifact, the normativity of this ascription is in the end epistemic, derivative of the normativity of epistemic justification. Although the justification for a function ascription will, for some persons, rely

on a technological explanation, the justification for this explanation in its turn does not rely on the function ascription and therefore there is no justificatory circularity here.

Persons lacking access to technological explanations who make function ascriptions justify these ascriptions by observation, experience, or testimony. In addition to the normative force of good justifications, these laypersons have an additional normative claim that entitles them to say that an artifact ought to have a certain proper function and fulfill it properly, as elaborated in the previous section. The epistemic division of labor in our society is such that professional engineers are entrusted with the task of designing properly functioning contraptions for various purposes. Laypersons have a legal and 'social-epistemic' right to expect engineers to have true beliefs about the workings and functions of the artifacts they make and to trust their testimony.[13] Whatever the details of this arrangement, we do not stumble on a justificatory circle here and that is the point I wanted to argue. The circularity worry is misplaced and the combination of the ICE theory of function ascriptions and my account of technological explanation can bear the burden. For all I can see, the two together deal adequately with the normativity of proper function ascriptions and technological explanations.

6. Conclusion

Against Peter Kroes I have argued that technological explanations are not necessarily special because they have to deal with the normativity of proper function ascriptions. Kroes's argument rests on a confusion of two rather different projects: that of giving a function theory and that of giving an account of technological explanation. The first project should grapple with the normativity of function ascriptions, i.e. it should explicate the conditions under which malfunctioning artifacts have proper functions. The second project can then pass the

[13] I owe this point to Wybo Houkes, cf. also (Houkes and Vermaas 2005) esp. chapter 8.

normativity buck to the first project. The principal reason for distinguishing these projects is that the property of having a proper function is relational, or extrinsic, whereas the property of having the capacity to exhibit a particular behavior is intrinsic. Consequently, accounting for the property of having a proper function must take the artifact's context into account, while accounting for the property of having a behavioral capacity can be done by looking just at the artifact itself. I have also argued that my way of framing the problem is more than wishful thinking, because Vermaas and Houkes's ICE function theory and my account of technological explanation do a good job in meeting the requirements I set out for the two projects. Besides, they fit together fine in the way I envisioned at the beginning of this paper.

References

Bechtel, William, and Robert C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

De Ridder, Jeroen. 2006. Mechanistic artefact explanation. *Studies in History and Philosophy of Science* (forthcoming).

Franssen, Maarten. 2006. The normativity of artefacts. *Studies in History and Philosophy of Science* (forthcoming).

Houkes, Wybo N., and Pieter E. Vermaas. 2004. Actions versus functions: A plea for an alternative metaphysics of artifacts. *The Monist* 87 (1): 52-71.

———. 2005. *Artefacts: From Functions to Plans of Use*. Book manuscript.

Kroes, Peter A. 1998. Technological explanations: The relation between structure and function of technological objects. *Technè* 3 (3): 18-34.

———. 2001. Engineering design and the empirical turn in the philosophy of technology. In *The Empirical Turn in the Philosophy of Technology*, edited by P. A. Kroes and A. W. M. Meijers. Amsterdam: JAI (Elsevier Science).

Lewens, Tim. 2004. *Organisms and Artifacts: Design in Nature and Elsewhere*. Cambridge, MA: MIT Press.

Millikan, Ruth. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

———. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Neander, Karen. 1991a. Functions as selected effects: the conceptual analyst's defense. *Philosophy of Science* 58 (2): 168-184.

———. 1991b. The teleological notion of 'function'. *Australasian Journal of Philosophy* 69 (4): 454-468.

Vermaas, Pieter E., and Wybo N. Houkes. 2006. Technical functions: A drawbridge between the intentional and structural natures of technical artefacts. *Studies in History and Philosophy of Science* (forthcoming).

Walsh, Denis M. 1996. Fitness and function. *British Journal for the Philosophy of Science* 47 (4): 553-574.