From Functions to Means: Practical Consequences of Artifactual Functions

Jesse Hughes

March 20, 2007

Abstract

A functional ascription tells one what an artifact is "for". There is practical content to such ascriptions: they tell one that the artifact can be used to achieve certain goals. This is an essential part of the natural language discussions of function. Consequently, one expects that a functional ascription produces certain practical consequences, and in particular means-end relations. Nonetheless, the derivation of means from functional ascriptions is surprisingly subtle. We provide here a sketch of the relevant features of functional ascriptions that give reason to accept associated sufficient means-end relations. We also relate artifact failure and malfunction to these means-end relations.

1 Introduction

One of the defining characteristics of artifacts is that they have functions; they are *for* something. Many, perhaps most, non-philosophical discussions of artifacts and their functions have a decidedly practical component[1]. In general, if one knows an artifact's function, she knows that the artifact can be used to realize certain goals. Knowing the function provides one with means to related ends she may wish to pursue. We interpret such means-end claims as suggested by Georg Henrik von Wright [1963]: an end is a state of affairs one wants to realize and a means is an action—something to be done[2]—that can bring about the end.

That artifact functions yield means-end relations seems obvious and unproblematic at first glance. For instance, it seems clear that if we accept

"Staplers are for fastening stacks of papers." (1)

and if t is a stapler and s is a stack of papers, then we may infer the following means-end relation:

"One can use t to fasten s together." (2)

But whether this is true or not depends on the stack of papers s—it should not be too thick—as well as the stapler t—it should be in good working order. Moreover, different kinds of staplers have different capabilities—some can staple thicker stacks than others—but they all seem to satisfy (1). Finally, the purported means "use t" is too vaguely expressed to motivate action. It fails to express *how* t should be used. There must be more to a functional ascription than (1) if we are to make sense of the step from function to

[1]Maarten Franssen even suggests in [2006] that evaluations of artifacts ("This is a good/poor screwdriver.") are really judgments about instrumental utility.

[2]Sometimes, one speaks of an object as a means, as when we say that a bridge is a means to crossing a river. We regard such locutions as shorthand for some action involving the object: in this case, walking across the bridge is the means we have in mind. Regardless of whether every object-as-means can be reduced in this way, we are interested in means-end relations in von Wright's sense here.

means-end relation.

We claim that plausible ascriptions of artifact functions implicitly include four features that allow one to identify associated means-end relations. Such features include: the type of artifact involved, the end to which the function aims, the use plan that can realize that end and a specification of normal contexts in which the artifact can be used. Each of these features is a natural product of the act of function creation for artifacts. Suppose Joe claims that he has designed a function-bearing artifact or discovered a new function for an existing type. We would doubt his claim if he could not answer: "Which type? To what end? How should it be used? When and where should it be used?" Houkes even claims one has a "socio-epistemic right to ask" questions like these [2006]. In this sense, functional ascriptions presuppose these characteristic features, but more often than not, examples are given in the terse language of (1).

Our primary thesis can now be stated: any plausible functional ascription implicitly involves the four characteristic features. These features serve to propose associated means-end relations. That is, from the four characteristic features, one can construct propositions of the form, "In situations $\mathbf{s}$, α is a means to φ." An agent who accepts the functional ascription will have a defeasible epistemic reason to accept the proposed means-end relations. Thus, we see how technical knowledge—knowledge about artifact functions—yields practical consequences—reasons to accept certain means-end relations. This is *not* a step from facts to oughts, however: means-end relations are not normative. Nonetheless, means-end relations hold a special place in practical (in particular, instrumental) reasoning—a means-end relation together with appropriate desire yields a reason to perform the means—and so this development helps clarify the practical role of technical knowledge[3].

[3] [Houkes 2006] also discusses the relation between function and practical reasoning, but we believe our account is more specific on this point. His primary aim is to distinguish technological knowledge from other forms and so he does not give much detail about how functions yield reasons to believe certain means-end

Furthermore, we show that the resulting means-end relations can be used to define certain functional terms, including artifact failure and malfunction. This latter term has played an important role in the literature, with discussions in [Millikan 1984; Neander 1995; Davies 2000; Schurz 2001; Franssen 2006; Vermaas and Houkes 2003] and elsewhere. But the connection between malfunction and instrumental reasoning has been largely ignored[4]. We show that our account of function can provide a natural distinction between these two essential functional concepts.

The paper is organized as follows. In Section 2, we argue that practical consequences of artifactual functions are essential: more often than not, we are interested in these functions for the means they provide. Section 3 introduces the four characteristic features of (artifactual) functional ascriptions that systematically induce means-end relationships. By the end of the section, we give a preliminary sketch of how the four features produce expectations about means-end relations. We revisit this sketch in Section 4, adding normal tokens to the account. We put this analysis to work in Section 5, showing how failure and malfunction can be defined in terms of induced means-end relations and normal tokens.

2 Functional explanations and practical consequences

Much of the recent philosophical interest in functions has come from philosophy of biology and functions in evolutionary theory. Beginning with Larry Wright's seminal article [1973], the primary role of biological functions was fairly clear: they explained the presence, prevalence or persistence of certain biological features in evolutionary terms. There were exceptions to this view, notably Robert Cummins [1975] explicitly denied that functions *could*

relations.

[4][Franssen 2006] is a notable exception, although he gives a reason-based definition of malfunction rather different than our ability-based analysis.

explain the presence of features in a system—rather, functions explain system capacities. In either case, the explanatory role of functions has dominated their discussion in philosophy of biology.

The situation is different for artifactual functions, however. Engineers and users are primarily interested in the practical consequences of artifactual functions, rather than teleological explanations. More often than not, when one considers the function of an artifact, he is interested in how it can be used to realize his ends—or ends he may adopt in the future. Artifact users are less often interested in teleological stories about how the artifact came to be where it is or why a particular artifact type is so common. After all, users *use* artifacts and engineers design them to be *used*. Usage is an inherently practical term and functions are about how an artifact can be used, so we expect that functions have a clearly practical component. In particular, artifact use is goal-directed and so we expect a connection between functional ascriptions and means-end relations.

Of course, *some* discussions of artifact function have explanatory roles, just as some discussions of biological function include means-end relations. Examples of the former are common in archeology and reverse engineering, for instance. When one points to an unrecognized component in a complex system (say, a fuel injector in an automotive engine) and asks, "What's that for?" he is seeking a functional explanation. Similarly, a physician may combat an illness by using his patient's natural immune system. In this case, he is interested in the biological function of the immune system for its practical consequences. Nonetheless, we believe that artifactual functions are more closely and systematically connected to means-end relations than biological functions.

In fact, *usage* is not a characteristic feature of biological functions. The function of a human heart is to pump blood, but a properly working heart is not *used* in order to realize this function. Hearts just do what they are supposed to more often than not and it is the rare exception that a particular action by some agent is required for a heart to realize its function. Thus the heart's function does not yield clear means-end relations, since a means is something one *does* in order to realize an end. That is not to say that there are no means-end relations involving hearts, but rather that biological functions do not systematically induce means-end relations.

But it appears that some biological items are *used* in a way roughly analogous to the way artifacts are used. Aren't fingers are used to grasp objects, after all? In fact, this sense of usage seems rather different than the analogous use of artifacts—finger-bearers know how to use their fingers rather intimately—but let us ignore this difference. Our point remains: we may accept that *some* biological functions have associated uses, but this is more an exception than a rule. Usage is not characteristic of biological functions.

Artifacts, on the other hand, are intimately associated with uses. Indeed, following [Vermaas and Houkes 2006], we consider use plans to be an essential part of function ascriptions. If an artifact is *for* some end, like fastening papers, then there should be some way to *use* the artifact to realize this end. Of course, this claim requires argument, but let us postpone the issue for the following section, when we expand on the concept of "use". For now, let us summarize our contrast between biological and artifactual function. Hereafter, we will focus our attention on functions of artifacts.

	Primary relevance	Usage
Artifactual function	practical consequences	every function includes use plan
Biological function	teleological explanation	no/few functions involve usage

3 The features of functional ascriptions

In this section, we will take the initial steps to explaining the connection between functions and means-end relations. In particular, we will describe four features that are characteristic of functional ascriptions and which account for their practical consequences. Because these features are essential to our theory, we will spend some effort clarifying and describing each. We will also justify that each of these features really *is* characteristic of artifactual functions, indeed, that they are products of the act of function creation. Thus, our account will apply to artifactual functions broadly.

Our basic approach is inspired by [Houkes 2006]. Houkes argues that use plans are an essential part of artifact functions, on the grounds that functional ascriptions presuppose usage. For the claim "Artifacts of type T are for X," to be plausible, one must assume that there is in fact a way to use T-tokens to realize X. We adopt a similar strategy, but add an analysis of contexts of use missing from Houkes's account.

We should be clear on the sense in which function ascriptions presuppose use plans, context specifications, and so on. We do *not* mean that whenever an agent knows that, say, the Hubble space telescope is for viewing distant astronomical objects, he also knows how and when to use the telescope. Certainly, one can have a justified belief about Hubble without detailed understanding—he may have been told about the function by a reliable source. But if he believes that this is what Hubble is for, then he must also believe that someone knows (or once knew) how to use Hubble appropriately. In particular, Hubble's designers must have had an idea how it can be used. Thus, as in [Houkes 2006], we will derive the characteristic features of artifact functions by examining the act of function creation[5].

Function creation is the act of initially ascribing of a particular function to an artifactual type (or, sometimes, a token). The creation can be a product of either design or discovery

[5]This act is called "design" in [Houkes 2006], but we prefer the more neutral term of function creation. Proper functions are designed while accidental functions are more often discovered. We aim for an account that includes both kinds of function.

or some combination. It is an intentional act[6], including activities such as imagining, selecting, planning, manufacturing. It may be a lengthy process, as when a new automobile is designed, or a nearly instantaneous recognition, as when a passerby notices that a fallen log makes a good footbridge. An artifact may have many functions (both "proper" and otherwise) and so may have many distinct function creations. Functional ascriptions may be independently re-discovered, so that a single function has several independent "creations". Functions can also subtly change over time, so that one cannot easily identify a single point of creation. Nonetheless, we expect that for each function ascribed to a type, there was *some* point at which some user or designer realized the artifact could be used for that function.

Each ascribed function has certain characteristic features that are products of the function creation. A designer that claims an artifact has a particular function must be able to tell us a bit about what the artifact should do and how and when to use it. In other words, the function creator should be able to answer the following questions.[7]

(a) What is the aim of the function?

(b) How should the artifact be used to accomplish the aim?

(c) When and where should the artifact be used for this end?

(d) What artifacts are you talking about?

The first three questions correspond to the following features.

[6]Some historical theories of function may deny that all functions are created by intentional acts. For such theories, it may even be that the manufacturers and users of an artifact are entirely unaware of its "real" function—the feature that explains its reproductive success. But such hidden functions are not the sort that have obvious practical consequences for the user and our analysis does not apply to these.

[7]By focusing on the function creator, our account seems to be biased towards "backwards-looking" theories of function, in the sense of [Perlman 2004]. On the contrary, even for forward-looking theories, we expect that a function ascription is plausible only if *someone* can provide answers to questions (a)–(d).

(A) a functional goal,

(B) a use plan,

(C) a set of normal contexts of use.

More often than not, the answer to (d) will consist of an artifactual type, although some accidental functions may apply to particular tokens rather than types. Token-level ascriptions are certainly an important exception, but we will postpone consideration of these kinds of functions. Instead, we will focus on those ascriptions that include

(D) an artifact type.

These four features correspond roughly to the features of Cummins systems described in [Millikan 2002, p.120]. The artifactual type corresponds to the Cummins system and the functional goal to the output capacity being analyzed. The use plan is roughly analogous to the "allowable inputs" to the system and the contexts of use appears as Millikan's "allowable conditions of operations". Our account differs from her discussion primarily by placing emphasis on the level of artifact types and their attributed functions rather than on the analysis of system capacities.

We will discuss each of the features (A)–(D)in turn.

Functional goals

Every artifactual function includes a *functional goal*: a condition which can be realized by properly using the artifact. A functional ascription asserts that the artifact is good for *something*, specifically, for bringing about some particular state of affairs. We use φ to denote functional goals, since such states of affair are commonly expressed as propositional functions. But we do not assume that such goals are Boolean: some goals (such as, "drying hair in a timely manner") can be satisfied to greater or lesser extent. We also do not assume that the user can reliably evaluate the degree to which the goal has been satisfied, but we *do* assume that there is a fact of the matter involved.

It is worth emphasizing here that we view functional goals as *propositional functions* and not simple sentences[8]. In logical terms, they include (typed) free variables: staplers are not for fastening this or that stack of papers, but for fastening stacks of papers. This functional goal may be crudely represented as **Fasten**(x), where x is a variable ranging over stacks of papers of appropriate thickness. A particular application of the stapler will involve a particular stack s of papers and will be successful just in case, after the use, **Fasten**(s) is true.

In this respect, our account is similar to the discussion of *derived proper functions* in [Millikan 1984]. According to Millikan, if **Fasten**(x) is the direct proper function of staplers and s is a stack, then **Fasten**(s) is a derived proper function. The terminology seems a bit awkward—are staplers really for fastening s, even derivatively? Normally, functions are stated in more general terms than this. Our contexts of use will play a similar role, but without Millikan's terminology.

Use plans

The connection between functional ascription and the related goal is so intimate that we colloquially confuse the two. We say that "Staplers are for fastening papers together" is a functional ascription, but there is nothing explicit in that sentence aside from an artifact type (stapler) and a functional goal (the fastening together of papers). Nonetheless, if this *is* a functional ascription, it must include certain other implicit claims. In particular, one

[8]Millikan expresses the same observation when she discusses *relational* functions. See [Millikan 2002, 1984]

can reasonably infer that there is a standard way of *using* a stapler for fastening papers together. It is surely not the case that one requires a novel, *ad hoc* procedure to fasten each pile of papers. Functional ascriptions are about particular ways of using an artifact.

Thus, every artifactual function explicitly or implicitly includes a *use plan* [Houkes 2006; Vermaas and Houkes 2006], which we denote α. Credible claims about artifact capabilities assume prescriptions for how the artifact should be used to realize the functional goal. Such prescriptions are constructed or discovered during function creation. It would be absurd to claim that one has designed an artifact for a particular function but does not know how the artifact should be used to do this task. Similarly, discovery of an accidental function must involve some plan for how to realize the functional goal. For a more thorough discussion along these lines, see [Houkes 2006].

A use plan is a prescription for how one should manipulate the artifact and related objects in the context of use in order to realize the goal. We allow for broad differences in the expression of such plans. They may be explicit and detailed ("Tighten the filter one quarter turn after initial contact.") or vague and broad ("Always obey local traffic laws."). They may include conditional actions ("If the stapler is empty, load it."). But in each case, they describe what one should *do*. Thus, as we will see, use plans provide the *means* for our means-end analysis.

But some artifacts passively realize their goals and so it is unclear whether they have associated use plans. For instance, components in a complex system (such as the master brake cylinder in a car) do not seem to come with user prescriptions. Similarly, how does one "use" a retaining wall? In each of these cases, one may argue that there is no *use* evident. The artifact just does what it is supposed to in the right circumstances, much like the human heart.

On the contrary, the user does indeed have something to do in each of these cases. He has to ensure the cylinder or wall is installed properly and thereafter he should perform regular inspections and maintenance. This is the use plan for these artifacts. A retaining wall will not prevent erosion unless it is installed properly and is in good working condition. If one wants to prevent erosion, then, she must make sure that the wall is installed and maintained (perhaps by delegating the task to a competent party). Similarly, the driver should be sure that a qualified mechanic installs and maintains the brake cylinder as part of the general maintenance of the vehicle. Unless she does this, she cannot reasonably expect the end she desires, namely, that the brakes slow and stop the car when needed.

These use plans seem superficially similar to prescriptions that heart-bearers should follow ("Schedule regular medical checkups."), but the connection between use plans and artifactual functions is stronger than the biological analogues. The development of such plans is a necessary step in artifactual function creation, but the discovery of biological functions does not involve any analogous plans in general. The scientist who discovers the function of a frog's brain may not know how to keep it in good working order, but the engineer who designs brake cylinders must be able to say how they should be installed and what physical features must be maintained to ensure reliable performance.

We clearly distinguish use plans from functional goal but many authors appear to conflate the two. For instance, Larry Wright defines "Z is a function of X" as the conjunction:

1. X is there because it does Z.

2. Z is a consequence (or result) of X's being there.

In this definition, it is not clear what type Z has. In (1), it appears that Z is an activity, something which is *done*. In (2), on the other hand, Z is evidently a condition or state of affairs, the sort of thing that can be a consequence of another state of affairs (X's being there). There seems to be a type confusion here.

We avoid such confusion by explicitly distinguishing actions from goals. Loosely, things which are *done* are part of the use plan while *end states* toward which the action aims are part of the functional goal. Unfortunately, natural language expressions can confuse the distinction. For example, “Staplers are for fastening papers together,” appears to identify an activity (“fastening papers together”) as the functional goal, but this is just a misleading oddity of language. One does not use staplers because he desires to experience the *act* of fastening papers, but rather because he wants the end result, namely, that the papers are fastened[9].

Contexts of use

Functional ascriptions should express not only *how* but *when* one should use an artifact. That is, ascriptions presuppose a *specification of contexts* in which the artifact can be used to realize the functional goal. Such specifications are again products of function creation: design and selection includes identifying situations in which the item can be employed. These specifications include descriptions of users and other objects involved. For instance, staplers are applied to small stacks of papers and cars are operated by persons who know how to drive.

Specification of contexts serve three distinct roles.

(i) They limit the situations in which an artifact is expected to perform its function. A car should not be expected to provide reliable transportation if its operator does not know how to drive (does not have *operational knowledge*, in the terminology of [Houkes 2006]).

[9]Sometimes, the experience of some activity *is* the end, of course. Roller skates are for those that enjoy skating, for instance, and here there seems to be some overlap between the use plan and the functional goal. But the action “skate” is distinct from the propositional function “is skating”. In order to realize the end, that is, in order to be skating, one must perform the action, that is, he must skate.

(ii) They provide parameters for the use plan and functional goal. One uses a stapler by inserting a stack of papers and pressing down on the mechanism. Like functional goals, use plans include free variables: $\mathbf{Insert}(x); \mathbf{Press}$ where x is a variable stack of papers. A particular usage includes choosing an appropriate stack s and executing the use plan with s. Thus, contexts serve as a bridge from the general (use plans and functional goals) to the specific (applications of use plans and evaluation of goals in contexts).

(iii) Success can be context dependent. Brakes should stop cars on both wet and dry pavement, but we expect shorter stopping distances on dry pavement.

Like the other features we've identified, the contexts of use can evolve over time. It is unlikely that early designers of ballpoint pens explicitly considered a gravitational field part of the normal context of use. Nonetheless, the fact that such pens don't work in zero gravity is relevant at least to NASA, if not to the rest of us.

Contexts of use have been largely ignored in the literature. They play no obvious role in [Vermaas and Houkes 2006], for instance. On the other hand, Ruth Garrett Millikan's *Normal conditions* [1984] seem very similar to our specification of contexts, at least regarding role (i), and her derived proper functions play roles similar to (ii) and (iii). Also, user abilities and circumstances are an explicit part of reasons to use an artifact in[Franssen 2006], essentially role (i).

Artifactual types

Lastly, we assume that functional ascriptions apply to specific artifact *types*, denoted T. But some ascriptions of accidental functions appear to apply only to particular tokens rather than types. For instance, when one needs to retrieve his keys from a grating into which they've fallen, he may apply some chewing gum to the end of a stick and try to pull them up by attaching the sticky gum to the keys. It may seem implausible that this device is part of a larger artifactual type, even though it is a well-known solution to the problem.

Token-level function creation is an interesting topic, but we save it for later work. The restriction to type-level function creation allows a richer development of our theory. In particular, Section 5 deals with malfunction and this concept relies on a sense of normal tokens of similar type. In the rare case that a token is truly novel and not an instance of a larger type, concepts like malfunction may not apply, since malfunction is (as we will argue) a comparative term.

In fact, it seems plausible that artifactual functions *do* primarily refer to artifact types and only derivatively to tokens. One may argue that, even for novel artifacts, the proper subject of a functional ascription is an artifact type, albeit a type instantiated by a single token. Some realists may object to the proliferation of types, however, and we prefer to avoid such ontological distractions. So let us restrict our attention to functions that *do* apply to types and avoid this controversy.

Artifact types can be broad or narrow. Because our functional ascriptions include use plans, we require that our functional types are narrow enough so that a common use plan can apply to each token. A broad type like "bottle opener" can be realized in many different ways and with many different use plans, and so is too broad for our needs here. Instead, means-end relations are introduced by narrower subtypes like "corkscrew".

These four features—functional goal, use plan, contexts of use and artifact type—are products of the design and discovery processes that yield functional ascriptions. Admittedly, it is sometimes hard to discern each of these features in informal functional ascriptions. We often speak tersely about functions, giving only the type and the functional goal. But such terse ascriptions cannot generate clear means-end relations unless the missing features of use plan and context are assumed to be implicit.

These features are fairly modest and certainly fall short of characterizing functions. Issues regarding what counts as an artifactual function or how functional ascriptions are justified are outside the scope of our work here. Similarly, we do not use these features to distinguish proper function from accidental. We regard this flexibility as appropriate: both proper and accidental functions yield means-end relations (but we follow the generally accepted tradition that malfunction claims apply only to proper functions).

In particular, most intentional accounts of function will be compatible with (A)–(D). This includes [Vermaas and Houkes 2006], Neander's [1991] discussion of intentional selection and Christopher Boorse's [2002] goal-contribution theory. On the other hand, the theories of Ruth Garrett Millikan [1984] and Robert Cummins [1975] seem less compatible with our analysis, but these theories do not share our aims. Millikan claims to be introducing a technical notion, so it is not clear that her notion of "function" *should* come with clear practical consequences. Cummins, on the other hand, uses function to explain a system's propensities in terms of its components. It is not clear that this sense of function generates means-end relations in the way we have in mind here.

Let us turn to the means-end relations constructed from features (A)–(D). At this point, it is apparently straightforward to see how expectations are generated. Let us suppose we have a functional ascription with functional goal φ, use plan α, contexts of use C and artifact type T. Let $\mathbf{s}$ be some context in C, that is, some situation that is a normal context of use. We write $\varphi_{\mathbf{s}}$ for the result of substituting variables in φ with the instances in $\mathbf{s}$ and $\alpha_{\mathbf{s}}(t)$ for: follow use plan α in context $\mathbf{s}$ with token t. We tentatively claim:

> In situation $\mathbf{s}$, one has reason to believe that a token t of type T can be used as prescribed by α in order to realize φ. In other terms, in such situations, one expects that $\alpha_{\mathbf{s}}(t)$ is a means to $\varphi_{\mathbf{s}}$. (**ME-1**)

Note that the means-end relation here is that of a sufficient means to an end. Functions do not provide necessary means, but instead suggest one way to realize the functional goal.

4 Normal tokens

The step from features (A)–(D) to (**ME-1**) seems deceptively simple. The features were admittedly chosen to fit into an appropriate means-end relation, although each was justified in terms of the process of function creation. But even with these features at hand, our work is not complete. The remaining difficulty comes in interpreting the expectations expressed in (**ME-1**).

The expectations in (**ME-1**) are apparently in terms of universal quantifiers, but this seems implausible. A functional ascription does not entail that *every* token of the appropriate type will reliably realize its end in appropriate contexts. Tokens do not always behave as one would like. Sometimes this comes as a surprise, but other times this is predictable. Some tokens are visibly defective and one should not expect such tokens to realize their functional goals.

Nonetheless, one may defend the universal interpretation of (**ME-1**) by taking our context specification C to be sufficiently narrow and include conditions on the token t. That is, we could require that C includes a description of what a "normal" T-token is. Since (**ME-1**) applies only for contexts satisfying the specification C, our expectations apply only to those tokens that meet certain structural requirements. This approach is again similar to Millikan's Normal conditions [1984], since the condition of the artifact plays a role in the Normal explanation of its function.

We avoid this alternative for two primary reasons. First, with complicated artifacts, the user is unlikely to know what structural features are relevant for use or whether a particular token possesses such features. An engineer may know how the wires inside my television should be connected, but I surely do not know this and I am unable to easily confirm that they are properly connected in any case. So this precondition for using my television cannot be part of my understanding of its function. The second reason for keeping a notion of normal token separate from conditions of use is discussed in Section 5 when we give a preliminary analysis of malfunction. For this, we must distinguish environmental abnormalities in applications from abnormalities in the token used.

Thus, we amend (**ME-1**) so that it applies to "*normal*" tokens of type T, rather than *every* T-token. We propose:

In every situation $\mathbf{s}$ realizing some context in C and for every *normal* T-token t, one has reason to expect that $\alpha_{\mathbf{s}}(t)$ is a means to $\varphi_{\mathbf{s}}$. (**ME-2**)

This introduction of normal tokens is not intended as a definition of normality, but a clarification of the expectation in (**ME-1**). Such expectations may fail, for instance, when an artifact type is poorly designed and incapable of realizing its functional goal. One needs an account of such normal tokens, then, that is independent of (**ME-2**)[10]. But first, a few words about normality.

We claim that normal tokens approximate natural reasoning about artifacts and their functions. Functional ascriptions express expectations about how artifact tokens will be-

[10]See [Schurz 2001] for an alternative account of normality. His evolutionary account of normality may give a different refinement of (**ME-1**), that T-tokens can *normally* be used to realize φ.

have, but *which* tokens? Since functions apply primarily to types, the expectations cannot be about the particular token at hand. When I am holding an obviously broken stapler, I would still agree that staplers are for fastening papers, but *this* one won't do the job. If asked why not, I would point out the ways in which it is structurally different than working staplers, i.e. the reasons why it is not a *normal* stapler in respect of this function. The most natural way of interpreting these expectations is in terms of normal tokens.

Normal tokens are abstractions that represent our expectations regarding the structure and behavior of typical instances of an artifactual type. The origins of our beliefs depend strongly on our understanding of the physical structure of an artifactual type as well as our beliefs regarding its intended operations and experiences with similar artifacts. An average user gains certain expectations about the behavior of a television from his previous interactions with tokens of that type. He has learned that the buttons on the front panel are an interface for changing the channel, volume, and so on. He also knows that a television needs electricity to work, typically from an external plug. He expects that, if he toggles the power switch, the tube will display an image. These expectations come from previous experience with such sets, technical manuals (and other communications) describing how to operate a TV, inferences regarding the designer's intentions and so on. Taken together, these form a collection of beliefs regarding the behavior of normal televisions and also some diagnostic tools to distinguish a broken set from a working set (a television with a cracked tube is unlikely to behave like a normal TV).

The electrical engineer has more detailed and precise knowledge regarding televisions. He understands better how a television is supposed to realize its function and so can more reliably distinguish normal tokens from abnormal tokens. The engineer and user both form expectations by referring to beliefs about normal tokens. The difference between the two is that the engineer relies more on knowledge regarding design, physics, and so on, and less on naive induction from previous uses[11].

Because types are sometimes badly designed, the reason to accept the means-end relation in (**ME-2**) must be defeasible. One can accept a functional ascription and still deny the derived means-end relation. Suppose crazy Uncle Al builds a perpetual motion machine. We may accept that the function of the machine is to create energy—this is what it is for, why Al constructed it. Nonetheless, we do not believe that this device or even normal tokens of the same type can perform this function. Our beliefs about physics defeat the reason in (**ME-2**). In this case, the artifact is badly designed: even normal tokens cannot do what they are supposed to do.[12]

We still must take the step from abstractions to actual tokens to complete our explanation of (**ME-2**). So far, we have discussed (**ME-2**) in terms of "normal tokens", but users are not usually interested in the behavior of such abstract entities. They want to know about the tokens at hand and their utility in bringing about desired ends. How does (**ME-2**) explain expectations regarding actual tokens?

Suppose our type is well-designed, so that the normal tokens can do what they are supposed to do. Then an actual token should fulfill the means-end relation in (**ME-2**) if it is sufficiently similar to a normal token, i.e. if it is relevantly similar in structure to some normal T-token. Our abstractions regarding normality include both behavioral and structural features. To the extent that our actual token shares those features, we are justified in supposing that it is capable of fulfilling its function. Of course, our beliefs about normal tokens may be mistaken—we may be wrong about what structural features explain relevant behavior or whether an actual token is relevantly normal. There are many judgments that

[11] This difference is also discussed in [Houkes 2006].

[12] Some authors would balk at accepting the functional ascription to begin with. Beth Preston suggests in [1998] that bug zappers are not for reducing the number of mosquitoes in the area but for giving the appearance of doing so. In that case, the expectation in (**ME-2**) comes with a much stronger reason: normal tokens *should* (at least sometimes) be able to do what they are supposed to.

can go wrong in our step from functional ascription to practical consequence, but that is a feature about reasoning generally and practical reasoning in particular. It is tough to get the premises right, but the principle in (**ME-2**) seems nonetheless correct.

There is clearly more to say about the use of normal tokens in means-end relations derived from functional ascriptions. The sketch we give here is preliminary but sufficient for this introduction to practical features of functional ascriptions. In the following section, we put the concept of normality to work.

5 Failure and malfunction

Malfunction serves as a kind of litmus test in [Millikan 1989]. She writes, "an obvious fact about function categories is that their members can always be defective... hence unable to perform the very functions by which they get their names." Etiological theories can account for malfunction while system capacity theories like [Cummins 1975] cannot[13], and hence (according to Millikan) the former are preferable for this reason. The importance of malfunction is echoed in [Neander 1995; Schurz 2001; Vermaas and Houkes 2003] and elsewhere.

The actual meaning of "malfunction" is rarely discussed in detail. A brief definition is suggested in Millikan's above quote and also in Preston's [1998] paraphrase of Millikan: "If you can say what a thing is supposed to do, then you can also say when it is failing to do something that it is supposed to do, that is, malfunctioning." In [Neander 1995], we find that "a biological part functions properly when it can do what it was selected for and malfunctions when it cannot[14]." Presumably, the case is similar for artifacts. Perhaps the

[13] Paul Davies rejects this claim in [2000]. He argues that etiological theories are no more capable of dealing with malfunction than Cummins's theory.

[14] Neander refines this rough definition later, by specifying that "what it was selected for" should be interpreted as the "lowest level of description" applicable.

clearest definition comes from [Franssen 2006]: " 'x is a malfunctioning K' expresses the normative fact that x has certain features f and that because of these features, a person p has a reason not to use x for K-ing." But these reasons seem vague and broad. A tie is for appearing attractive, but because I am wearing a striped shirt, I have a reason not to wear a plaid tie—according to some fashion theories, anyway. Surely the tie is not malfunctioning, but it appears to satisfy Franssen's terms.

We see two features common to each of these accounts. First, malfunction is about ability to perform. It is not about whether the artifact did what it was supposed to on some particular occasion, but what it would do *if* used[15]. Second, malfunction applies to individual tokens, not types. Types may be poorly designed, so poorly that—like Uncle Al's perpetual motion machine—no token can do realize its functional goal, but types do not malfunction[16].

Failure is similar to malfunction in this respect: it applies to tokens rather than types. But failure, unlike malfunction, is about individual performances. An artifact token t either *fulfills* or *fails to fulfill* its function in a particular application. This judgment applies to a specific instance in which the token was used according to use plan α in some situation realizing a context in C. We suggest the following definition:

> A token t *fulfills its function* in a particular application just in case the functional goal $\varphi_{\mathbf{s}}$ was realized as a result of that application. Otherwise, it *fails* to do so. (**Fail**)

Fulfillment and failure may be a matter of degree, since $\varphi_{\mathbf{s}}$ may be realized to greater or lesser degree.

[15] Preston's paraphrase seems an exception, since she speaks of performance ("failing to do") rather than ability. We will follow Millikan and Neander on this point, however.

[16] See [Franssen 2006] for a distinction between bad design and malfunction.

On this view, a token t fulfills its function if the goal is realized *due* to the application of the use plan with respect to t. Thus, the application must be causally relevant to the realization of $\varphi_\mathbf{s}$, but even this requirement is perhaps a bit liberal. In practice, we may also require that the application realizes the goal *in the intended way*, rather than by accident. A gun that misfires, striking a branch that falls on and kills the rabbit at which one was aiming, has failed to perform as intended even though the shooter's goal was realized. Let us ignore that complication at present and take (**Fail**) as our working definition of fulfillment and failure.

While failure is about a particular application, malfunction is a broader claim. A well-functioning token may fail to realize its goal: a perfectly good anti-aircraft missile may miss its target, for instance. Thus, failure is a negative evaluation of a particular outcome, but it does not mean that the token used was somehow bad. A malfunctioning token, by contrast, is one that cannot be expected to reliably or effectively bring about the functional goal. Malfunction is about hypothetical applications, what would happen if the token were used.

We should stress that there are other senses of malfunction which we do not address here. As Sven Ove Hansson has pointed out (private correspondence), some malfunctions involve unintended consequences rather than inability to bring about an intended outcome. A television that presents a crisp image while filling the room with deadly radiation is surely malfunctioning, even though it is realizing its primary function. This kind of malfunction has been ignored in the literature so far[17]. We will focus on malfunction-as-inability for now, since it is more closely related to means-end relations than malfunction-as-undesirable-propensities.

The concept of ability needs some attention, however. Suppose one has a pistol with a

[17]Note that Franssen's definition *does* include this sense of malfunction.

weak spring in its firing mechanism. Most times the trigger is pulled, the hammer does not strike hard enough to fire the primer, but sometimes the gun fires as desired. The pistol is therefore able to do what it is supposed to, but it does not do so reliably. Nonetheless, it seems clear that this pistol is malfunctioning.

Perhaps we can tweak our functional goal to include reliability. Maybe the functional goal of a pistol is to discharge a bullet every time (or almost every time) that the trigger is pulled. But this fix comes with a price: we would lose our definition of failure. Failure is about particular applications, but "discharge a bullet every time" quantifies over a range of applications. It makes no sense to ask whether a particular application realized the goal "discharge a bullet every time the trigger is pulled."

Instead, we propose putting the concepts of reliability and effectiveness into the definition of malfunction.

> A token t is *malfunctioning* with respect to a proper function if it is unable to *reliably* or *effectively* realize φ in some situations $\mathbf{s}$ satisfying C when used according to α, i.e. if $\alpha_{\mathbf{s}}(t)$ is not a reliable or effective means to $\varphi_{\mathbf{s}}$ in such situations. (**Mal**)

Thus, like Millikan, Neander, et al., we define malfunction in terms of ability, but with additional considerations of reliability and effectiveness. We justify our beliefs about the ability of tokens by referring to their physical characteristics, causal laws and past experience with the token itself.

Our definition (**Mal**) contains two undefined terms, namely "reliably" and "effectively". The concepts of reliability and effectiveness are inherited from the semantics of means-end relations. A means to an end may be more or less likely to realize its end. That is, in a particular context $\mathbf{s}$, there is some probability that doing an action $\alpha_{\mathbf{s}}$ will realize $\varphi_{\mathbf{s}}$. We call this probability the *reliability* of α as a means to φ in $\mathbf{s}$. *Effectiveness* is the degree to which the functional goal $\varphi_{\mathbf{s}}$ would be realized as a result of $\alpha_{\mathbf{s}}$. Of course, the situation can be considerably more complicated than this, since a means may have low probability of very effectively realizing its end and a high probability of a less effective outcome. The two measures are thus not orthogonal in practice, but let us ignore these complications for this initial sketch of malfunction.

Definition (**Mal**) requires a comparative notion, however: a token malfunctions if it is not reliable or effective, but compared to what? Here again, we rely on our intuitions regarding normal tokens. A token is malfunctioning if it is considerably less reliable or effective at fulfilling its function *than normal tokens of the same type*[18]. Thus, malfunction involves a comparison between the behavior of an actual token and the behavior one expects of normal tokens.

But normal tokens are relative to an artifactual type, and types have a hierarchical structure. A particular phone may be a token of its model type, of wireless phones, of push-button phones, and of telephones *simpliciter*. A user will typically have different beliefs about the normal tokens of each of these types, and this may yield conflicting judgments about whether the token is malfunctioning. If the wireless phone loses connection with its base outside of a short radius, then it may be behaving less effectively than normal wireless phones. But suppose that early models of wireless phones had shorter ranges than later models. If this token is an early model, then it is behaving as designed. Should one say that this phone is malfunctioning?

Claims regarding malfunction should involve normal tokens of a *suitably narrow type*. "Suitable narrowness" is unfortunately difficult to specify precisely. We would like to

[18]Some design specifications explicitly include constraints on reliability and effectiveness. These constraints may affect our expectations about normal tokens, but we must allow that normal tokens are unable to meet these requirements due to bad design.

compare our old wireless phones with wireless phones of similar age, for instance, but should we restrict our attention to phones of the same model? What criteria determine suitable narrowness for this purpose[19]? Suppose that telephones made on Monday mornings are typically less reliable than telephones made at other times during the week. Should we compare our Monday morning phone to "normal" Monday morning phones before declaring that our token is malfunctioning?

We should not expect a rigorous definition of "suitable narrowness". Clearly, malfunction claims are context-sensitive and so we may expect that the relevant artifactual type from which we draw our normal tokens is similarly context-sensitive. Consider a token t of a particular model of stapler, say Acme. Suppose that this model is not particularly reliable: it often mangles staples rather than closing their tines, and so fails to fasten papers in situations in which better designed staplers work well. Our token t may behave as well as expected for T-tokens, but more poorly than normal staplers. In many situations, we would say that t is malfunctioning, and in other situations, that t is functioning as expected but is poorly designed. We should expect our semantics for malfunction to accommodate both judgments: to allow that in some situations, normal behavior for the type "stapler" is relevant and in others, normal Acme-behavior is what matters to us.

Let us be satisfied with loose guidelines, then, that allow one to exclude too-broad and too-narrow artifactual types from consideration. For broadness, we will use the same criteria we discussed in Section 3. The artifactual type must be relevant for the function discussed and so it must be narrow enough that a single user plan suffices for every token of that type. For narrowness, we suggest the following: the narrowest "suitably narrow" type is the type defined by the *design* of the token at hand. That is, the token at hand is

[19]One may also ask what criteria determine our typology for artifacts. Does the collection of blue staplers form an artifactual type?

the result of some design or selection process and that process defines a relevant artifactual type. Thus, for typical manufactured goods, we will sometimes consider normal tokens of the same *model* in order to make malfunction judgments, but we will not consider narrower types like "Monday morning telephones". We may sometimes want to distinguish bad design from malfunction and for this, we need design-defined artifactual types. But other historical features are not relevant to malfunction judgments. That a particular token was badly manufactured, for instance, explains rather than excuses its malfunctioning.

Thus, our constructed means-end relations provide a natural analysis of failure and malfunction. The latter also involves a comparison between an actual token and normal tokens of the same (suitably narrow) type. This comparative character has been largely omitted in the literature, because the notion of ability has lacked consideration of reliability and effectiveness. But these concepts are a natural part of our means-end analysis, a conceptual side benefit in addition to our primary aim: explaining the practical consequences at the heart of most discussions of artifactual functions.

References

Boorse, Christopher. 2002. A rebuttal on functions. In *Functions: New essays in the philosophy of psychology and biology*, ed. André Ariew, Robert Cummins, and Mark Perlman, 63–112. Oxford University Press.

Cummins, Robert. 1975. Functional analysis. *Journal of Philosophy* 72(20):741–765.

Davies, Paul Sheldon. 2000. The nature of natural norms: Why selected functions are systemic capacity functions. *Noûs* 34(1):85–107.

Franssen, Maarten. 2006. The normativity of artefacts. *Studies in History and Philosophy of Science* 37:42–57.

Houkes, Wybo. 2006. Knowledge of artefact functions. *Studies in History and Philosophy of Science* 37:102–113.

Millikan, Ruth Garrett. 1984. *Language, thought and other biological categories*. The MIT Press.

———. 1989. In defense of proper functions. *Philosophy of Science* 56:288–302.

———. 2002. Biofunctions: Two paradigms. In *Functions: New essays in the philosophy of psychology and biology*, ed. André Ariew, Robert Cummins, and Mark Perlman, 113–143. Oxford University Press.

Neander, Karen. 1991. The teleological notion of function. *Australasian Journal of Philosophy* 74:261–301.

———. 1995. Misrepresenting & malfunctioning. *Philosophical Studies* 79(2):109–141.

Perlman, Mark. 2004. The modern philosophical resurrection of teleology. *The Monist* 87(1):3–51.

Preston, Beth. 1998. Why is a wing like a spoon? A pluralist theory of function. *The Journal of Philosophy* 95(5):215–254.

Schurz, Gerhard. 2001. What is 'normal'? An evolution-theoretic foundation of normic laws and their relation to statistical normality. *Philosophy of Science* 68(4):476–497.

Vermaas, Pieter E., and Wybo Houkes. 2003. Ascribing functions to technical artefacts: A challenge to etiological accounts of functions. *British Journal for Philosophy of Science* 54:261–289.

———. 2006. Technical functions: a drawbridge between the intentional and structural natures of technical artefacts. *Studies in the History and Philosophy of Science* 37:5–18.

von Wright, Georg Henrik. 1963. Practical inference. *The Philosophical Review* 72(2): 159–179.

Wright, L. 1973. Functions. *Philosophical Review* 82:139–168.