An evolutionary approach to Function

Phillip Lord
School of Computing Science, Newcastle University, Newcastle-Upon-Tyne, United Kingdom, NE1 7RU

ABSTRACT

The distinction between function and role is a vexed and difficult one. While the distinction appears to be useful, in practice it is hard to apply; this can be even worse when applying this distinction to biology. In this paper, I take an evolutionary approach, considering a series of examples, to develop and generate definitions for these concepts. I test them in practice against work performed on the Ontology for Biomedical Investigations (OBI). Finally, I give an axiomatisation and discuss methods for applying these definitions in practice. Availability: `http://homepages.cs.ncl.ac.uk/phillip.lord/function/`

1 INTRODUCTION

Large parts of modern biology are aimed at answering the question: what is the function of this? Much of the Gene Ontology deals with molecular function (Ashburner et al., 2000). In dealing with the social aspects of science, roles are similarly important. It is clear, therefore, that function and role are important concepts in biomedical ontologies and are prime candidates for inclusion in an upper ontology. A coherent, consistent and shared definition for function and role is likely to decrease the effort required to integrate independently developed ontologies.

A number of groups are currently using one upper ontology, the Basic Formal Ontology (BFO)[1] which includes a definition of function. However, the definition is not naturally applicable to biology; it is not clear, for instance, that GO's molecular functions are also BFO functions, as discussed in Section 2.1. An alternative framework is available within the General Formal Ontology (GFO), which provides an ontology of function (Burek et al., 2006). This ontology provides a more extensive framework for describing functions but, in itself, does not define biological function. Despite this, there is a reasonable degree of informal agreement among biologists as to the meaning of the word. Conversely, while the definition for role seems clear, many people have difficulties in applying it in practice.

In this paper, I address two key issues relating to modeling function and role coherently for biomedical ontologies. These are: how we unify the definitions of function as they apply to artifacts and to life; and, how do we differentiate between roles and functions? I do this by considering illustrative examples where the answers are reasonably clear and evolve a definition from them. I then consider examples of the application of current definitions from BFO in a practical biomedical use. Finally, I offer a axiomatisation in OWL, stemming from our definitions and consider briefly how these definitions could be applied in practice.

[1] `http://www.ifomis.org/bfo`

2 FUNCTION

2.1 Biological Function

First, consider the current definition of function provided by BFO (see Def: 1).

DEFINITION 1. *Function is a realizable entity the manifestation of which is an essentially end-directed activity of a continuant entity being a specific kind of entity in the kind or kinds of context that it is made for*

As a simple example, a hammer (*the continuant entity*) was made to hammer nails (*the function*) in a hammering process (*the end-directed activity*). This definition is problematic for biological systems. The problem here is straight forward: most biological systems were not made or designed for any purpose.

Although, it has not been incorporated into BFO yet, there is a potential definition for "Biological Function" which would become a new child of Function (Def: 2) (Arp and Smith, 2008).

DEFINITION 2. *A biological function is a function which inheres in an independent continuant that is i) part of an organism and ii) exists and has the physical structure as a result of the coordinated expression of that organism's structural genes.*

As an example, a foot (*the independent continuant*) is part of an organism, exists, and developed in a controlled way as a result of gene expression.

There are some difficulties with this definition. Terminologically, we would normally expect coordination to have a coordinator. More substantially, consider the following examples: a differentiated tumour has its structure through coordinated expression of its genes, it exists and it engages in end-directed activity. A male ant has its structure as a result of gene expression, and engages in end-directed activity; however, it is not part of an organism[2]. From this, we conclude that a differentiated tumour does, indeed, have a function (growing). A male ant, however does not have a function. We can also consider molecular function: the physical structure of a protein is independent of the expression of an organisms structural genes – only its presence depends on this. A protein, therefore, does not have a function, by this definition. So, there are two key problems: the definition does not work for entities above or below a certain size. Also most biological entities have their structure as a result of coordinated expression. I offer the following alternate definition (Def: 3).

DEFINITION 3. *A biological function is a realizable entity that inheres in a continuant which is realized in an activity, and where*

[2] Or at least not a proper part of

the homologous structure(s) of individuals of closely related (or the same) species bear this same biological function.

The definition given uses the notion of homology; evolution is key to our understanding of biology and it is appropriate that it should be used to define biological function. If a biological structure has a function, then this function will have evolved along with the structure; so, other structures with a common evolutionary descent will display the same behaviour. This definition also mirrors closely normal biological practice; the most common way to determine the function of an unknown structure is to look for function of a homologous structure.

It should be noted that this definition of biological function is *not* circular, although it has itself as part of its definition; rather it is recursive; a chimp hand and a human hand can have the same function because of each other. It does require that a structure must have a homolog for it to have a function. It does not require that these homologous structures be extant.

Applying this to our examples: the tumour now has no function because it has no homologs (different tumours arise as independent events and share no common ancestor). Likewise, the activity of the male ant now clearly is a function, as many different, related organisms behave in a similar way. Finally, a protein may have a function depending on the activity of its homologs. In short, this definition results in the same conclusions as our biological understanding.

2.2 Relating the Functions

Next we consider the relationship between function and its biological, artifactual subclasses. Taking an illustrative example, consider the sole of my foot and the sole of my shoe[3]. They appear to operate to: provide a frictional surface to enable motion; provide padding to *reduce shock* to everything above; be tough enough to resist abrasion. They would appear to have the same function; indeed, like many artifacts, we would guess that the shoe owes much of its design to mimicry of biology. It would seem, therefore, sensible that instances of *Shock Resistance* could be either a biological or artifactual function. The alternative would be to duplicate many functions under both subclasses ("biological shock resistance", and "artifactual shock resistance") with very similar definitions. Only functions such as reproduction have to be a biological function, while liquifying iron can never be. In short, whether an instance of function is biological or artifactual should be determined from the nature of the entity in which it inheres, rather than the process by which it is realized. We therefore offer a simple definition of function which reflects this (Def 4).

DEFINITION 4. *A function is a realizable entity which is a biological function or an artifactual function.*

For the purposes of this paper, I note that the definition given for function earlier (Def 1), can serve as a reasonable definition for artifactual function.

It is also interesting that this definition covers some unusual but non-pathological examples. Take a bacteria whose colonies change colour depending on the presence of a toxin and which was produced using synthetic biology techniques. The components have all evolved, but the organisation has not. Is the detection of the toxin then a biological or artifactual function? This is clearly a difficult, if uninspiring, question but given Def 4, one we can avoid by simply describing it as a function; alternatively, we can describe it as both a biological or artifactual function, suggesting strongly that these two classes should not be disjoint.

3 ROLES

Next, having considered the definition of function and its applicability to biology, we consider the issue of roles. The current definition (Def: 5) is complex. More simply, it suggests that the entity having that role can be involved in an activity but that it was not necessarily intended for, nor necessarily has the structure for this.

DEFINITION 5. *A realizable entity the manifestation of which brings about some result or end that is not essential to a continuant in virtue of the kind of thing that it is but that can be served or participated in by that kind of continuant in some kinds of natural, social or institutional contexts.*

Consider the relationship between role and function. Again, we shall consider a simple biological example, in this case of a man walking on his hands. By our earlier definition (Def: 3), "to walk on hands" is *not* a function. While the homologous structure is, indeed, used for walking on in all closely related species, most humans do not walk on their hands. It would, therefore appear to be a role. In this context the hand has a role of *Shock Resistance*. This realizable entity also appears in the hands of many other primates; in this case, however, it would appear to be a function of the primate hand, as it is a function of their feet.

We are left with a similar conclusion as previously. As well as concluding that Shock Resistance maybe either a biological or artifactual function, we must also conclude that the individuals of the same class can be a role, depending on the nature of the relationship between the independent continuant and realizable entity.

3.1 OBI as a case study

So far, in this paper, we have considered a number of illustrative examples and used these to draw conclusions about definitions for functions and roles. This methodology is appropriate, but has the limitation that the choice of other examples may have led us to different conclusions. In this section, therefore, I will consider the use that OBI (Ontology for Biomedical Investigations) has made of function and role [4]. I choose to use OBI as it was built after BFO and with knowledge of it; many of the ontologies available from OBO were started without this use or knowledge.

Considering first the functions; OBI has 48 different functions. Can these functions be fulfilled by an entity which was not designed for the purpose? As shown in Table 1, most of them can, often by considering a highly generic device (like a computer) or organism (like a human). Some are highly generic in themselves and can be performed by many things (heat for example). Some can only be functions but because the definitions are truistic (for example, Device Function; Device is defined as something "designed [..] to

[3] The author acknowledges that neither rank among his best features

[4] Analysis was performed on OBI take from subversion, Revision: 1369

Function Bearer	OBI Function
Human	Perturbe, Measure (2), Separation (4), Sort
Computer	Information Processor (4), Consume Data
Highly Generic	Produce Data (2), Freeze, Heat, Environment Control, Mechanical, Record, Contain (2), Transfer, Cool, Connection, Synthesizing, Excitation (2), Ionization
Distant Galaxy	Magnify
Truisms	Device (9), Energy Supply (2)
Out of Scope	Molecular Function (6)

Table 1. OBI Functions considered as Roles.I provide suggestions for entities that could engage in the same end-directed activity, but without being designed for the purpose. "Function" has been omitted from the OBI term names. Numbers indicate child terms which have been omitted for brevity. I do not consider "Information Processor" to be the function of a computer, as the definition is more specific than the term suggests.

perform a specific function"). I consider "Molecular Function" to be out-of-scope for this paper as it is not clear whether it fulfils the BFO definition of function.

Next, we consider OBI roles. There are many more roles than functions – 100+ in fact. Due to the size, here we consider whole branches of the role hierarchy.

First, many classes relate to human or societal processes; for example, the regulatory role (in the sense of ethical rather than transcriptional regulation), or study personnel (clinical laboratory worker, principle investigator). Most of these, do indeed appear to be roles and can only ever be roles. I note, however, that even here there are counter examples; robots are common place in laboratories now, and often perform the tasks previously fulfilled by humans, even the task of designing the next experiment (King et al., 2004). These robots are normally designed specifically for a given purpose, which means they are fulfilling their function; note that as technology advances, robots may become more generic; now they will be fulfilling a role, again.

Second, there are a set of roles relating to reagents and their states. While label role (defined as a reagent role realized in a detection of label assay) seems sensible, is S^{35} CTP not manufactured specifically for this purpose? It certainly does not occur in nature.

Finally, let us consider reference roles (a role which can support the observation of similarities, differences, relative magnitude or change). In many cases, biological assays use a reference which is not manufactured. However, consider λ-HindIII fragments, a calibration standard or the international kilogram prototype. These would all appear to function as an reference and have been produced specifically for this purpose.

As a result of this analysis, we suggest a modification to the current definition of role (Defn 5). Both roles and functions can become apparent (*realized*) in natural, social or institutional contexts. That such a context exists does not provide a clear differentation between role and function; the critical distinction relates to whether the entity in question was designed or has evolved to be engaged in a given process. I suggest, therefore, this alternate and simpler definition for role (Defn 6).

DEFINITION 6. *A realizable entity the manifestation of which brings about some result or end that is not essential to a continuant because of its kind.*

In short, from this case study, I conclude that the role/function distinction is not clear. While OBI has a specific intent in mind with its application of the distinction (broadly and not exhaustively, social or experimental roles, device or instrument functions), this distinction is not the distinction made in the current definitions of role and function; further given that most functions could also appear to be roles, and many roles appear to also be functions, I suggest that the distinction made in the current definitions is not useful in the context of OBI.

4 AN AXIOMATISATION FOR FUNCTION AND ROLE

I have produced an axiomatisation in OWL of functions and roles as defined in this paper; due to space considerations we report here key differences between this and the BFO axiomatisation; it is available as described in the abstract.

- There is an explicit relationship between `RealizableEntity` and `Process`. Subclasses use a more specific relationship. So, a `ToAbsorbShock` function may only be realized by a `AbsorbtionOfShock` process, if it is realized at all.
- `Function` and `Role` are defined classes. Stating that an instance of `ToAbsorbShock` `is_function_of` instance of `FootSole` implies, therefore, that the former is a function.
- Most leaves of `RealizableEntity` are direct children of `RealizableEntity`, with a few exceptions (`ToReproduce` is a child of `BiologicalFunction`).

The definitions could be extended further; for simplicity, we have not added classes to differentiate between organisms and artifacts. These could be added to automate the population of `Biological` and `ArtifactualFunction`.

At the current time, it remains an open question whether `Role`, `Function` and its children should be disjoint. The key example of the function of a synthetic biological organism suggests that `Biological` and `ArtifactualFunction` function should not be disjoint (as it appears to be both), but I have no example which suggests whether `Role` and `Function` should be disjoint. In axiomatisating the examples given, these disjoint statements make no practical difference.

Many ontologies are built using the OBO format; while this has a slightly weaker semantics than OWL it is possible to represent much of OWL in a OBO (Golbreich et al., 2007). The axiomatisation presented here can be represented using "union" and "intersection of" to describe classes, which is usually translated as a definition. The universal link between `RealizableEntity` and `Process` has no natural equivalent. However, as this link its own specific relationship which is restricted to this use, problems caused by the lack of an inexact semantic equivalent are likely to be relatively minor.

The axiomatisation presented here is related to that produced by others (Dumontier, 2008; Burek et al., 2006); the first of these focuses more on roles, while the later provides for a more complex representation, which covers issues such as preconditions.

5 APPLYING THE DEFINITIONS IN PRACTICE

Finally, we consider whether these definitions are *applicable*; for a given set of entities how do we decide whether we have a function (of either subclass) or a role.

The definition of an artifactual function easily allows its application: first, we determine whether the entity in question is an organism or part of one (which it shouldn't be); second, we could ask whoever produced the entity what it was designed for. Of course, the second may not always be possible, in which case, we can guess from its design what its purpose is. In most cases, these will provide a clear answer.

For biological function, the situation is less clear. Whether an entity is an organism or part of one is, in practice, likely to be straightforward for extant entities; otherwise, we are can apply palentological techniques. Likewise, identification of closely related species, and homologous structures is well known as it forms the basis of taxonomy. While developing an exact definition for "closely related" is outside the scope of this paper, it is possible.

The definition that I introduce for Function in this paper (Defn 4) is conjunctive; it is either biological or artifactual. Here I have given little evidence that these are the only kind of function. Fundamentally, these two arise from very different mechanisms. There could be other appropriate subclasses of function; the most obvious possibility would be Chemical Function. However, artifacts are designed by humans who understand, mimic and improve on biology by building tools. It is this mimicry that we wish to reflect with a common definition joining biological and artifactual function; this is not true for Chemical Function.

To determine whether something is a role, it is possible to make a determination on the basis of whether the context is optional (Arp and Smith, 2008). However, this optionality is a difficult criteria; firstly, all RealizableEntity's are optional in the sense that they might never be realized and, secondly, the optionality can depend on how specifically we define the bearer. A hammer is not designed to hammer nails, as claimed earlier, it is designed to hit things; a nail hammer is designed to hit nails, a toffee hammer to hit toffee, a warhammer to hit anyone who irritates you. In practice, a role can be considered to be a negative definition; if, there is a continuant and an end-directed activity that the continuant can be involved in, and this involvement is known not to fulfil the definition of either function, then, we have a role.

In this paper, we have considered OBI and found that the distinction between role and function is hard to apply; this is not true for all ontologies. For example, consider the Gene Ontology. In many cases, the homology will be considered as a standard part of the operating procedure[5] in determining the function of a gene product; regardless the evidence codes would allow us to make the distinction. We can conclude, therefore, that when GO is used to annotate a protein, this describes a biological function rather than a role.

[5] `http://www.geneontology.org/GO.evidence.shtml`

6 DISCUSSION

Here, I have taken an evolutionary approach to function and role, by considering examples and using this to derive definitions which are as consistent as possible with current use within biomedical sciences. These have been encoded in an axiomatisation which should enable the use of these definitions in a machine interpretable way.

The applicability of these definitions is a key advantage; the current distinction being made between function and role is a hard one to understand and apply. Our definition distinguishes between the two based on the nature of the relationship to the independent continuant in which they inhere. I suggest that it is very hard to make the distinction at the class level; our study of OBI shows that very few of the functions and roles clearly fall into one category or another. For an individual continuant bearing a realizable entity, this distinction appears to be much more straight forward.

I also provide a definition of biological function, something that is currently lacking from BFO. I have paid close attention to current biological usage; the definition is close to the process used to determine function. Moreover, it is highly applicable; all parts of the definition are measurable. It is for this reason, for instance, that I have avoided a definition based on the outcome of selective pressure; this is hard to test in most circumstances, requiring expensive evolutionary studies, and impossible for extinct species. Serendipitously, it also avoids difficult questions about artificial selection; we can state clearly that cows do not have a function of producing beef, though this is the outcome of selection. Importantly, my definition of biological function works across multiple levels of granularity: organisms and organism part through to genes and molecules; this is not true of previous definitions (Arp and Smith, 2008), which cover only anatomy. While the coverage of the definition is useful, I tend to side with Dumontier (2008) who suggests that for molecules the (biological) function/role distinction may be redundant.

In summary, I believe that the definitions and axiomatisation given in this paper make a significant contribution to the use of role and function in biomedical ontologies. They should enable a consistent use of these classes, because they consider current usage of the terms and the applicability of these definitions. I seek not to change current use but to formalize it.

REFERENCES

R. Arp and B. Smith. Function, role and disposition in basic formal ontology. In *The BIo-Ontologies Workshop (at ISMB 2008)*, 2008.

M. Ashburner, C. Ball, J. Blake, D. Botstein, H. Butler, J. Cherry, A. Davis, K. Dolinski, S. Dwight, J. Eppig, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–9, 2000.

P. Burek, R. Hoehndorf, F. Loebe, J. Visagie, H. Herre, and J. Kelso. A top-level ontology of functions and its application in the open biomedical ontologies. *Bioinformatics*, 22(14):e66–e73, Jul 2006. doi: 10.1093/bioinformatics/btl266.

M. Dumontier. Situational modeling: Defining molecular roles in biochemical pathways and reactions. In *OWLED 2008*, 2008.

C. Golbreich, M. Horridge, I. Horrocks, B. Motik, and R. Shearer. OBO and OWL: Leveraging semantic web technologies for the life sciences. *Lecture Notes in Computer Science*, 4825:169, 2007.

R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427(6971):247–252, Jan 2004. doi: 10.1038/nature02236.