Abstractions, Idealizations, and Evolutionary Biology

Peter Godfrey-Smith

Philosophy Department Harvard University

October 2006/M

From the 2006 Paris Symposium "Making Up Organisms." To appear in A. Barberousse, M. Morange, and T. Pradeu (eds.), *Mapping the Future of Biology: Evolving Concepts and Theories.*

- **1. Introduction**
- 2. Idealization and Abstraction
- 3. Successes and Pitfalls
- 4. The Informational Gene

1. Introduction

Many philosophers agree that *idealization* and *abstraction* are key aspects of science, especially scientific work based on the construction and assessment of models. And while much of the initial philosophical work in this area was concerned with physics, it has become clear that biology, especially evolutionary biology, is another area in which these phenomena play a very significant role. But despite a consensus on the *importance* of idealization and abstraction, there is not much consensus on what these things *are*, and what exactly they achieve. Here I will take the opportunity to offer some general ideas about idealization, abstraction, and the relations between them, emphasizing applications to evolutionary theory.

2. Idealization and Abstraction

These two terms have become quite problematic, and surrounded by an unruly literature in which a number of closely related views have been developed. I will proceed initially by abandoning the terms and focusing on some underlying phenomena. Once these are clear, the terms will be re-applied.

In describing something, one may find oneself doing one or both of the following:

- 1. Leaving things out, while still giving a literally true description.
- 2. Treating things as having features they clearly do not have.

The former is a matter of ignoring things, ignoring detail. The latter involves an act of imagination; we imagine that something is different from how it actually is.

Initially, I will treat the latter as equivalent to imagining the existence of a fictional thing that is similar to the real-world object we are interested in. There is a general and important role for imagining non-actual objects, properties, and systems in science. Scientific model-building is in large part the investigation of fictions. Philosophers have often tried to see the models developed in theoretical science as abstract mathematical objects (Suppes 1960), but I suggest that it is often more natural to see them as having a status similar to that of fictions within literature. That is, I see model-builders as often trying to describe and understand systems that are only imaginary, but which would be *concrete if real* (Godfrey-Smith forthcoming). When a modeler says that he or she is trying to understand a particular kind of neural network, economy, or ecology, which is too simple to be found in the actual world, I take the modeler at face value. The modeler *is* trying to understand an imaginary neural network, economy, or ecology. He or she will often try to describe and understand it *using* mathematical tools, but that is not the same as saying the modeler is trying to describe an abstract mathematical tools.

Very often, these fictional systems are simpler analogues of real-world systems. And in discussing them, a scientist will frequently talk in a way that "aims" the description directly at the real-world system of interest; he or she will present ideas in the

form of a literally-false description of the real system, rather than a true description of a different thing, the imagined system.

Going back to my two-way distinction above: ignoring some features in a description of a system is *inevitable* to some extent in *any* description. The question is only how much is left out, and what is retained. The kind of imaginative act seen in the second operation, in contrast, is always optional. One could try one's best never to do it. One could try to always say only what one thinks is literally true.

I now connect these two phenomena to our vexed terminology. An *abstract* description of a system leaves a lot out. But it is not intended to say things that are literally false. An *idealized* description of a system is a description that *fictionalizes in the service of simplification*, in the way described above. The idealized description is often presented verbally *as* a description of a real system, but not a description that is literally true. It is also possible, I said above, to present this as a specification of an imaginary system that resembles the real system. But one thing that the usual way of talking does is make it clear *which* real-world system the fictional one is intended to function as a simplification of. I will return to the small but significant differences between these two ways of thinking of idealized descriptions below. (I should also note that the "more complex system" of interest might itself sometimes be fictional, but the usual case is that in which the other system is real, and I will only deal with this usual case.)

This treatment of idealization and abstraction is not intended to be completely faithful to the way scientists talk; it is intended to clean up a somewhat disorderly set of relations between the two concepts. Several previous philosophical treatments are at least close to this view, and Martin Thomson-Jones (2005) presents a view that is essentially identical, differing only in points of emphasis. Thomson-Jones also helpfully distinguishes this view from related positions in writers like Cartwright (1989) and McMullen (1985). Like me, Thomson-Jones notes that this treatment involves a certain amount of revision of scientific usage.

One thing that leads to confusion here is the fact that both idealization and abstraction can be described as "leaving things out." That phrase can refer to *simple omission* of features, or to an *imagining away* of those features. And crucially, it can be quite difficult to simply omit things, especially in a mathematical context; very often it is

necessary to imagine things away. If one wants to "leave out" friction, or genetic drift, and one wants an equation as the form of one's representation, this usually has to involve imagining things absent, even if one offers the verbal comment that one has merely "omitted" these factors.

On the view offered here, what is the relation between idealization and *approximate truth*? The latter is a topic that philosophers have made heavy weather over, especially when seeking to think of a statement's semantic content in terms of its logical implications. Using the more relaxed realist framework I am employing here, it is natural to say that idealized descriptions are in many cases approximately true. Here I understand approximation to truth as a simple matter of similarity between the circumstances depicted and the circumstances that obtain. So when we treat an idealization as a description of some target system, it may, depending on how it is expressed, be approximately true of that system. Further, degree of idealization will correlate inversely with approximate truth. But the relation between the two concepts is not straightforward, because of the evaluative dimension. *Goodness* of idealization is naturally understood in terms of the purposes the idealization serves for us. A description could be very highly idealized, hence scoring low on approximate truth, but be a good idealization in the circumstances, despite this. Conversely, a description could score high on approximation to truth, and be a bad idealization given our goals.

In the light of the ideas above, I can now say something about relations between three closely related scientific activities: offering a description known to be only *approximately true* of its object, offering an *idealized description* of a real object, and building an idealized *model* of that object. First, we saw that idealization involves a departure from reality in the direction of some kind of simplicity – not all approximations and not all models will have this feature. But suppose we concern ourselves only with simplifications. Then I said above that an idealized description of a real object will tend to be approximately true, and also that an idealization can be seen as the specification of a fictional relative of the real object. So once simplification becomes the goal, it seems that approximation, idealization, and modeling are very closely tied together – hard, in fact, to pull apart.

I see the differences between them as in large part practical or strategic. Talk of "approximation" is natural when the description is closely shadowing the real system (at least in intention), and there is little role for the deliberate and rich imaginative construction of non-actual features. Talk of "modeling" is most natural when the scientist's immediate focus is the fictional system itself, relations to the real system are secondary, and the differences between the two are substantial (see also Weisberg forthcoming). Talk of "idealization" can be natural within *either* of these kinds of activity. So I do not see the mere presence of idealization as an infallible indicator of a model-based style of science, though any idealized description can, *post hoc*, be considered as the specification of a model system that might be investigated in its own right.

To illustrate some of the ideas in this section, I will briefly look at the simplest formal models of evolution, such as one-locus population genetics. It is often said that these models are "abstract," but often said that they are "idealized." It is not really possible for a single description to be both idealized and abstract at once, with respect to the same features of the object being described. So which are they? As I see it, these models do exhibit both phenomena, but with respect to different features.

Suppose we consider a deterministic random-mating one-locus diploid model of natural selection of the simplest kind (Roughgarden 1979), and declare that we treat it *as* a representation of evolution in some real population in which evolutionary change is occurring at one locus. We take the descriptions specifying the model and treat them as descriptions of the empirical system. Then we get a mixture of idealization and abstraction, even assuming that the match with respect to the parameters is as good as it could possibly be for a model of that kind and a real-world system.

The description of the population in terms of gene frequencies at one locus is an abstraction – although a more unusual one than it looks, as I will discuss below. But the model makes an assumption of infinite population size; that is an idealization. The model also assumes random mating; that may perhaps be literally true and an abstraction, depending on the interpretation of probabilities, but it is much more likely to be an idealization. The model assumes no mutation; almost certainly an idealization. What about the fitness parameters? If they are constants in the model, then it is *almost*

inevitable that they are idealizations. They are being treated as independent not only of frequency change but also of exogenous environmental change, which will almost always occur to some extent. In many cases, but not in all, the assumption of non-overlapping generations will also be an idealization. So simple evolutionary models are both idealized *and* highly abstract, but in different respects. And strictly speaking, *any* degree of idealization is enough for us to say that the description as a whole is not literally true.

I will make a brief comment on two other technical tools. The "Price equation" is becoming more and more popular as a foundational treatment of evolutionary processes (Price 1972, Frank 1998, Okasha forthcoming). It is an interesting case. Any evolving population in which there is asexual reproduction at the lowest level (which includes copying of alleles) can be *abstractly* described by a Price equation. The partitioning operation achieved by the equation does not amount to an idealization, and the equation does not use an infinite population size assumption. The fitness parameters are said to hold only for one generation and to represent "realized" fitnesses. All of this can be treated as a literal description. Other theorists, especially in the recent European tradition, like to treat the "replicator dynamics" as a foundational model (Hofbauer and Sigmund 1988, Nowak and Sigmund 2004). Here we are very much in the realm of idealization, especially when assumptions of infinity are made. So some of the alternative foundational models in evolutionary theory at the present time represent slightly different choices with respect to idealization and abstraction.

3. Successes and pitfalls

Abstraction, as described above, is inevitable in all description. So there is no news in the idea that abstraction is essential to scientific work. Without it, communication itself would be impossible. That leaves a significant question, for each field and each question, of how *much* abstraction is appropriate. Idealization I see as near-enough essential to science also, but in a way that is more controversial and substantial. The practice of science shows that fictionalizing is often a crucially important strategy, and also something that people can do while not recognizing what they are doing, and perhaps while talking in misleading ways about what they are doing. But we can certainly

imagine a person policing their scientific practice in a way in which all literally false claims are avoided, purged, replaced as soon as possible. The embrace of idealization is a substantial matter, and not something inevitable.

Do these operations generate characteristic kinds of error or wrong turn? Abstraction I said is always present, so the idea of a *characteristic* error is not straightforward. But we might say that there are characteristic errors that can come with extreme or unobvious forms of abstraction. A clear possibility would be a process in which one deliberately leaves some features out, notes the usefulness of the representation that results, and infers that the things abstracted away from do not exist at all. But in this paper I want to emphasize a more subtle kind of error, the reification of the products of abstraction. It can be easy to describe the results of an operation of abstraction in terms that suggest the existence of an additional *entity*. Examples that might be given here will often be tendentious, because people often think this sort of move is justified, not a fallacy. But I would cite the case of propositions, as they figure in much philosophy of language. We note the way that many sentences can have a similarity of meaning, and then come to talk of the "the meaning" of those sentences. Really this is a grouping via abstraction. But the process can lead to a reification of the thing, the meaning, which is then called a proposition. This is seen as an abstract object that the sentences all have some special *relation* to. I think this is a wrong turn, and a kind of wrong turn that will figure again in the final section of my paper. As Thomas Pradeu suggested to me, the idea of such a pitfall also has links with Berkeley's criticism, in his Principles of Human Knowledge (1710/1965) of John Locke's treament of abstract ideas.

The "reification of abstractions" is sometimes seen as a characteristic error in some theoretical work in biology – Lewontin's critiques of ANOVA and heritability methods (1974) are examples. How does this fit into the framework used here? The situation alleged, in such a critique, is something like this. An abstract description yields a summary, such as a statistic. The statistic is then taken to pick out something like a *discrete causal factor*, a causal player of its own, something that is part of the machinery. The *grammar* of description appropriate to a discrete causal factor is applied. In the most problematic cases, it is treated as an entirely different kind of entity, with special relationships to the other, ordinary causal players.

Idealization certainly generates a characteristic error close to the first one discussed in the case of abstraction. We pretend that some features of a real system do not exist, for the purposes of simple and compact representation. Finding ourselves doing well with the resulting compact description, we infer that that there is nothing more to the system that what has been recognized in our idealized treatment. I have argued that this is a characteristic of some philosophical work in metaphysics (2006). Here I follow much older work by John Dewey, who described a similar problem in different terms, in Chapter 1 of *Experience and Nature* (1929/1958).

But to conclude this section, it is worth emphasizing how well things can go for a scientific discussion, when it features the right mix of highly idealized work and more concrete particularistic work. In evolutionary theory and related parts of biology, this sort of interaction has a long history. Darwin's Origin of Species (1859) is notable for the very slight role for idealization. His focus is squarely on the concrete nature of organisms, and the actual processes in which these organisms are involved. He does offer abstract claims, and entertains abstract hypothetical propositions from time to time. Darwin was also willing to formulate some principles in the form of "laws" (see the "Recapitulation" of the Origin). But in part because he does not try to formalize these principles mathematically, he does not find himself idealizing away from the actual nature of organisms very much, and the work is also much less abstract than much of the work that came later. It was partly this focus on the empirical and concrete that made the Origin so powerful and convincing when it appeared. But idealized work, including mathematical modeling, appeared quickly. Francis Galton's "Typical Laws of Heredity" (1877), for example, looks very much like a piece of model-based science. It was followed by similar work by himself and others within the "biometric" school (Provine 1971).

Many of the landmark works of the 1930s, such as R. A. Fisher's and Sewall Wright's work, have an entirely different character from that of the *Origin*. Many of Fisher's crucial contributions were made via his ability to pick the right simplified fictional system to think about, in order to fully understand how a process works in the more complex empirical domain. He thinks about populations whose size is effectively infinite, that are unstructured in space (or not affected by any structuring that may exist),

and whose members differ with respect to the simplest kinds of genetic features – usually one-locus genotypes.

People sometimes side with one style of science or the other. Fisher's approach might be lampooned as "bean-bag genetics" and seen as an exercise in mere mathematical fantasy. It is less common to disapprove of Darwin's style of biology, but one does hear occasional murmurs of this kind. Darwin's work is, to the formal mind, a great unruly mess of empirical detail. However, I intend this pair of cases to be an illustration of how things look when things go well. The two styles of evolutionary theorizing seen here are surely complementary; each did things that the other could not have done, and our understanding of evolutionary processes draws essentially on both.

I think we see some of the same kind of complementarity in some recent work on the "major transitions" in evolution. Leo Buss' *The Evolution of Individuality* (1987) is a Darwin-like work. Entirely non-mathematical and crammed with empirical detail, it spurred much of the later interest in the problem. Maynard Smith and Szathmáry's *The Major Transitions in Evolution* (1995) is a model-based book full of idealizations. So we have two books written around the same time, on the same topic, and of similar importance, where one book applies a modeling strategy and the other does not. This gives us a nice illustration of the contrasts between the *style* of science that uses models and the style (or one style) that does not. The Maynard Smith and Szathmáry book is also noteable because although it is a model-oriented book, most of the models are *not* presented mathematically. The exposition proceeds through a constantly shifting mixture of words, pictures, and a few equations here and there. So this book is a useful corrective to the view that model-based science has some essential connection with mathematical techniques.

Maynard Smith and Szathmáry's book can be contrasted in this respect with Rick Michod's book *Darwinian Dynamics* (1999). Michod is concerned with similar topics again. (He is close to Buss in his focus.) Michod's book is model-based, but in this case the models *are* given in explicit mathematical form. So recent work on the major evolutionary transitions shows what I see as a complimentary mix of model-based and non-model-based approaches, and the modeling work itself includes work that is primarily mathematical and work that is not.

4. The informational gene

In the rest of this paper I will use the framework outlined earlier to comment on what might seem to be an unlikely case, the informational treatment of genes in biology. I take it that this is a striking and philosophically interesting feature of recent genetics and evolutionary theory. Different aspects of the phenomenon can be distinguished. One is the claim that genes code for, or contain information specifying, whole-organism phenotypes or the course taken by developmental processes. Here, we see an unusual and philosophically controversial description of the causal role of genes, but the genes themselves are taken to be ordinary material objects. More recently, it has become quite common to insist that we should, further, see genes themselves as in some sense *made* of information, as informational objects, at least in the context of evolutionary work. So the informational mode of description is now not just being used in the specification of causal roles, but also in the specification of what kinds of things, what kinds of objects, figure in evolutionary theory. The most dramatic expression of this kind of thinking is seen in George Williams' 1992 book Natural Selection. There he claims that we have learned, via evolutionary theory, that information is one of the basic ingredients of the universe, along with matter and energy, inhabiting its own "domain."

In earlier work I have criticized the less metaphysical kind of informational description of genes – the use of semantic concepts in the specification of causal roles (Godfrey-Smith 2000). My argument has been that there are good reasons to take genes as coding for, or carrying information specifying, the primary structure of protein molecules, but there is no reason to use this semantic language to describe the role of genes in the causation of any product downstream of those protein molecules (see also Griffiths 2001). In this discussion I will comment on the more tendentious and, to my mind, extravagant version of the informational treatment of genes, the idea that there is such a *thing* as the informational gene, as well as the concrete DNA molecule, or that genes for evolutionary purposes informational *entities*. I think this is misconceived idea, and one that can be understood as the misapplication of a process of abstraction.

Let us first ask how the idea of an informational treatment of genes as objects arises in evolutionary theory. It arises in order to achieve a kind of *grouping* operation. Considering an actual-world population, we find it is composed of a collection of particular organisms containing a great mass of particular molecules. Theory has taught us that we can get substantial purchase on change in this population by attending to a particular kind of molecule in the cell nucleus (DNA). And our "attending" to it involves a peculiar kind of focus. We attend to individual chunks of this molecule that are not physically separated from other chunks and their surrounds. We group token chunks of this molecule into types according to base sequence, and then characterize the whole population by the frequencies of the types. Thus we get the evolutionarily important measure of a gene frequency at one locus. We then pay close attention to changes in these frequencies, by attending to the rates at which new molecules of a given kind are produced via template processes from old ones.

All this involves a more peculiar counting procedure than is often supposed. In a population with a germ line, we ignore almost all tokens of the molecules in our counting. We do the grouping of tokens into types by attending to the sequence of bases, but we are allowed to ignore some base differences ("silent" differences, for instance). The demarcation of "chunks" and other aspects of the counting is also dependent on the peculiar process of "crossing-over," occurring at a particular stage in the cell cycle.

What we have done is engage in a highly abstract description of both the organisms and the population. At the heart of this abstract description, again, is a *grouping* operation, one that groups token molecular structures separated by both space and time. A great many token molecular structures are treated as falling into a common type, in virtue of their sequence properties. There is nothing wrong with this, but it sets us up for a possible additional move that is mistaken. Rather that saying that all there is is the collection of molecules themselves, which we describe abstractly for certain purposes, we might be tempted to say there is an extra entity, an abstract additional thing, that the particular molecules have some important relation to. That is, we could be tempted to say that when we talk of genes for evolutionary purposes, we mean to describe the doings of an abstract entity, the "informational gene." For me, this is just like the unnecessary and misleading move that many philosophers writing about language make,

when they reify "propositions." They note the phenomenon of similarity or identity of meaning across sentences (type and token), and conceive of this in terms of relations between the sentences and an extra entity, the proposition, that is real but abstract.

In the biological case, the shift between the operation of grouping material objects for purposes of abstraction, and the introduction of an extra entity that bears some relation to the material things, is illustrated in this quote from David Haig (1997).

Gene ... can refer to the group of atoms that is organized into a particular DNA sequence — each time the double helix replicates, the gene is replaced by two new genes — or it can refer to the abstract sequence that remains the same gene no matter how many times the sequence is replicated. The *material gene* (first sense) can be considered to be a vehicle of the *informational gene* (second sense).

I don't think that Haig means to put anything like the same metaphysical weight on the informational gene that people like George Williams do. In practice, Haig regards this operation of grouping, and all the concepts used here, in very pragmatic ways. That is clear from his paper. But the road to an error of reification is shown very clearly in the language here. We seek to group a number of tokens, in virtue of a similarity of sequence: fine. But that does not require anything like the introduction of an extra entity for which the token material objects function as "vehicles."

Perhaps the point might be put by saying that in the case of *other* molecules, like sugars and lipids, we are perfectly able to group them into types without introducing an extra abstract thing that the material molecules are "vehicles of." There is no "informational lipid," grouping all the lipids of a given structure; there is just a collection of molecules that we have reason to group as a kind. Genes do play a different causal role from that of lipids, because of the importance of exact DNA sequence in protein synthesis, and the importance of DNA in inheritance. I know that this is supposed to mark a key difference between DNA and things like lipids. But those differences in causal role are not of a kind that motivate an entirely different ontological treatment of genes. They are just differences in the concrete causal processes in which the two classes of molecules figure.

Why does this matter? First, if one cares about ontology then it seems to simply be an erroneous move to treat genes in a de-materialized fashion. But there are further consequences as well. Once genes are de-materialized in this way, they are removed from the ordinary causal give-and-take, and can come to be seen as explanatory factors of a special, ultimate kind, telling us our true natures. But all that is real here is the DNA molecules, with their various properties. They are, indeed, extremely important, but important as material things that are part of the causal give-and-take.

* * *

Acknowledgment: I am grateful to Anouk Barberousse, Marie-Claude Lorne, and Thomas Pradeu for helpful comments on an earlier draft.

Berkeley, G. (1710/1965). *Principles of Human Knowledge*. Reprinted in D. Armstrong (ed.), *Berkeley's Philosophical Writings*. New York: Collier.

Buss, L. (1987). The Evolution of Individuality. Princeton: Princeton University Press.

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Darwin, C. (1859/1964). *On the Origin of Species*. (Facsimile of first edition.) Cambridge MA: Harvard University Press.

Dewey, J. (1929/1958). *Experience and Nature*. (revised edition). New York: Dover Reprints.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Frank, S. (1998). Foundations of Social Evolution. Princeton: Princeton University Press.

Galton, F. (1877). "Typical laws of heredity." *Proceedings of the Meetings of the Members of the Royal Institution of Great Britain* 8: 282-301.

Godfrey-Smith, P. (2000). "On the Theoretical Role of 'Genetic Coding," *Philosophy of Science* 67: 26-44.

Godfrey-Smith, P. (2006). "Theories and Models in Metaphysics." *Harvard Review of Philosophy* 14: 4-17.

Godfrey-Smith, P. (forthcoming). "The Strategy of Model-Based Science." To appear in *Biology and Philosophy*.

Griffiths, P. (2001). "Genetic Information: A Metaphor in Search of a Theory." *Philosophy of Science* 68: 394-412.

Haig, D. (1997). "The Social Gene," in J. Krebs and N. Davies (eds.), *Behavioural Ecology*, fourth edition. Oxford: Blackwell, 284-304.

Hofbauer, J. and K. Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.

Lewontin, R. (1974). "The Analysis of Variance and the Analysis of Causes." *American Journal of Human Genetics* 26: 400–11.

Maynard Smith, J. and E. Szathmáry (1995). *The Major Transitions in Evolution*. Oxford: Oxford University Press.

McMullen, E. (1985). "Galilean Idealization." *Studies in the History and Philosophy of Science* 16: 247-273.

Michod, R. (1999). *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton: Princeton University Press.

Nowak, M. and K. Sigmund (2004). "Evolutionary Dynamics of Biological Games." *Science* 303: 793-799.

Okasha, S. [forthcoming]: *The Levels of Selection Debate*. Oxford: Oxford University Press.

Price, G. (1972). "Extension of Covariance Selection Mathematics." *Annuals of Human Genetics* 35: 485-90.

Provine, W. (1971). *On the Origins of Theoretical Population Genetics*. Chicago: Chicago University Press.

Roughgarden, J. (1979). *Theory of Populations Genetics and Evolutionary Ecology: An Introduction*. Upper Saddle River: Prentice Hall.

Suppes, P. (1960). "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences." *Synthese* 12: 287-301.

Thomson-Jones, M. (2005). "Idealization and Abstraction: A Framework." In M. Thomson-Jones and N. Cartwright (eds.), *Idealization XII: Correcting the Model*. Amsterdam: Rodopi, pp. 173-217.

Thomson-Jones, M. (2006). "Models and the Semantic View." Forthcoming in *Philosophy of Science* (Proceedings of PSA 2004).

Weisberg, M. (forthcoming). "Who is a Modeler?" To appear in the *British Journal for the Philosophy of Science*.

Williams, G. C. (1992). *Natural Selection: Levels, Domains, and Challenges*. Oxford: Oxford University Press.