Darwinian Populations and Natural Selection

Reviews by Daniel Dennett, Kim Sterelny, and David Queller, with replies by Peter Godfrey-Smith

For *Biology and Philosophy*, 2011

Homunculi Rule

Daniel Dennett

1. Down with Essentialism

This is the best, most thought-provoking book in the philosophy of biology that I have read in a long time. It is rigorously argued, deeply informed, full of wonderful examples, and it has more novel ideas within its pages than any two other recent books in the field that I can think of. It has opened my mind on several key issues and changed my mind as well; it is also provoked me to come up with what I hope are better defenses of some of the points of my position criticized in the book.

One of the most important things we have learned from Darwin (but some philosophers, alas, still don't get it), is that essentialism is simply a mistake. There is no mystery about why many philosophers resist this verdict: their method, going back to Socrates, demands exceptionless definitions and self-evident axioms, from which deductive consequences can be made to flow. Philosophers are not alone in their weakness for essentialism. Even evolutionary theorists have often succumbed to the temptation to define the essence of Darwinism, the necessary and sufficient conditions for natural selection to occur. For one thing, this makes for effective pedagogy. But a good Darwinian should attempt to honor Darwinism about Darwinism itself—in Glenn Adelson's apt phrase—and Peter Godfrey Smith (henceforth PGS) shows us how to think about all the many marginal cases—the semi-Darwinian processes, proto-Darwinian phenomena, quasi-Darwinian relationships—without *marginalizing* them. Therapsids, after all, were just as real as their ancestor reptiles and their mammalian descendants, and the question of whether viruses are alive (or not—do we *have* to fish or cut bait?) is less important than seeing which Darwinian features viruses share with bacteria, with aspen groves, with us—and why. PGS sets out with a deliberately neutral, maximally latitudinarian, basic category, *Darwinian populations.* Literal populations of organisms are populations, of course, but so are all the cells in a multicellular organism, and the set of chromosomes in a cell, and (maybe) all the songs on the Hit Parade. Some populations are just more “Darwinian” than others, for a variety of reasons. The penumbral cases are just as real, and just as important to understanding biology, as the parade cases, the paradigms that unmistakably exhibit all the key features.

2. Two contributions

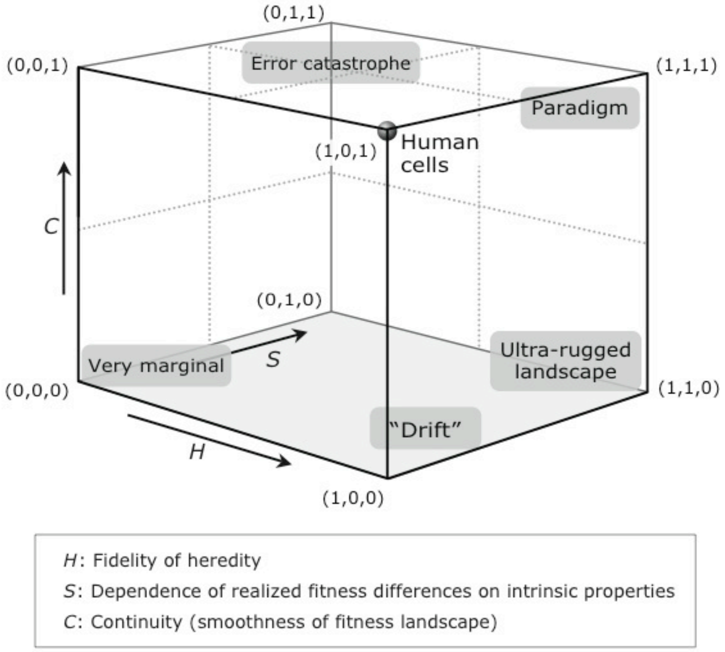


Figure 3.1. A Darwinian Space

The most valuable contribution of the book, in my opinion, is PGS’s introduction of a “spatial tool” (p8) a 3-d representation of one or another “Darwinian Space” (p63), in which Darwinian populations can be located, permitting us to see—at a glance—the relationships between key variables such as *fidelity of heredity (H), abundance of variation (V), competitive interaction with respect to reproduction (a), continuity, or smoothness of fitness landscape (C)* and *dependence of reproductive differences on intrinsic character (labeled S—for supervenience, roughly).* As figure 3.1 (p64) makes clear, we can comfortably represent only three variables at a time in a standard 3-d projection, but we can compose diagrams for any three other variables of interest *ad lib.*, and “In most cases, the way to think about Figure 3.1 [for instance] is to assume that everything in the graph has high values on various unseen dimensions, so we can focus on the difference-making role of a few key factors.” (p63) So if H is really low, for instance, (paradigmatic) evolution cannot occur because adaptations cannot accumulate faster than they dissipate due to mistaken transmissions, and we get the “error catastrophe" that must somehow be avoided in the early days of replication. If S is really low, selection is as good as random, not "for cause"—for any intrinsic merit, in effect, and you get drift. If C is really low, there are no traversable hill-climbing paths to the optimal arrangements, so populations will get stuck on inefficient or ineffective arrangements.

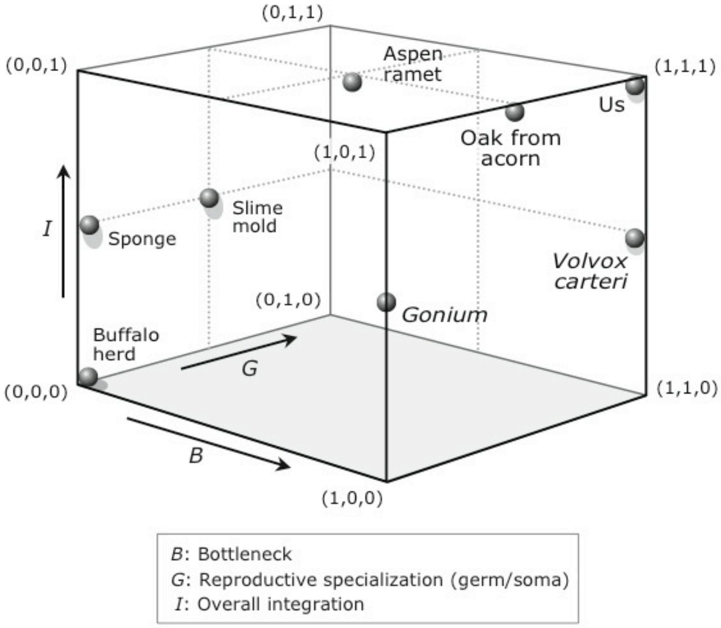


Figure 5.1. A space with three reproduction-related dimensions (*B, G, I*)

Another diagram, Figure 5.1 (p95) plots *bottleneck (B), reproductive specialization (G, for germ/soma)* and *overall integration (I).* Note that we humans sit in the paradigm corner, sharing important affinities with oaks and Volvox that we don't share with aspens and buffalo herds. Such a visual display of the options "makes vivid the role of gradients and partial similarities between cases. Second, it gives us new ways of thinking about processes with a *non*-Darwinian character." (p65) Exactly. I expect to use these PGS Spaces a lot in the future. The basic schema is so powerful that it even extends beyond PGS's use of it, permitting us to notice omissions and biases in his own account. For instance, in spite of his salutary opposition to essentialist thinking, he gives short shrift to at least two dimensions which—I will argue—are just as important to a complete evolutionary account as the dimensions he discusses so illuminatingly: the gradient between "formal" and "material" replication, and the gradient between differential *replication* and differential *persistence*. It is not that he overlooks these entirely; he notes instances of purely "formal" parent-offspring relations (prions, retroviruses and LINE transposons) and he has careful discussions of the murky boundary between differential growth/persistence (aspen groves, tribal growth) and differential replication (oaks, Hutterite colonies (Sober and Wilson, 1998)), but there are candidate generalizations of these observations that are nicely revealed for our consideration when we plot these on PGS Spaces, generalizations that point to possible extensions of Darwinian thinking at two of its most notorious—and "opposite"—boundaries : the origin of life from pre-biotic phenomena at the early end and the origin of culture and language at the most recent end. More on this below.

Vying for the laurels of most valuable contribution is PGS's concept of *de-Darwinizing,* which can be visualized in a PGS Space as the *motion* of some Darwinian populations*,* over evolutionary time, away from the far-upper-right paradigm corner to a more marginal position in the space*.* A vivid case is how the transition from unicellularity to multicellularity resulted in the partial de-Darwinizing of somatic-cell reproduction: these cells, a Darwinian population in their own right, are no longer in such competitive interaction (the *a* dimension) as their protist ancestors were, except when they get cancerous. Moreover, thanks to our high value on the G dimension (the germ/soma distinction) the difference in fitness between germ line and somatic cells is "location," not any intrinsic property, so our somatic cells are also Darwinian populations with low S, which is why their 'evolutionary activities' are not paradigmatically Darwinian. (p56)

The concept of de-Darwinizing permits us to see clearly the similarities between eukaryotic cells, multi-cellularity and eusociality, for instance (p122). PGS stresses that this motion in the space of Darwinian possibilities is not itself purposive (e.g., p128), even though it responds to deep constraints.

3. Two problem themes

These two excellent ideas, PGS Spaces and de-Darwinizing motion within them, are in tension with two less valuable—indeed mildly obstructionist—themes in PGS's book:

(1) his reluctance to characterize any Darwinian processes in terms of *information,* and

(2) his disapproval of "agential" description and explanation.

In some regards, there is nothing wrong with this tension, and it is even to be welcomed. The best bandwagon should have a good set of brakes, and it is a good thing that PGS wants to limit his tolerance, lest his Darwinism about Darwinism dissolve into pluralistic mush. But the limits he sets need to be motivated and defended, and here I find some serious blind spots in his vision. His two pet peeves are related, of course, and can be found united, not surprisingly, in the work of Richard Dawkins (and John

Maynard Smith, and George Williams, and David Haig, and myself, among others). The fact that most of the outrage expressed against Dawkins' talk of "selfish genes" is utterly benighted (e.g., Mary Midgley, 1979, 1983) should not lull us into thinking that there are no informed and sophisticated criticisms, and PGS sets out to distill the best objections to the widespread enthusiasm for thinking of genes as agents of sorts, with agendas, pursuing strategies, probing for weaknesses in their opponents, etc. What sets his critique apart from others is his uniting of this theme with a steadfast opposition to the position, most forcefully articulated by George Williams (1992) that genes, like poems and recipes, are properly seen to be the *information carried* by one physical vehicle or another, translatable into different media, and properly (and unproblematically) abstract. Don't ask what a poem weighs, or what physical elements it is made of, and don't make the parallel mistake of thinking that genes *are* strands of nucleotides; they are the information carried by those sequences of codons. I had thought that this perspective was widely accepted as enlightened, but PGS will have none of it, so I will deal with his mistrust of information first, and then turn to agents.

For PGS genes are 'treated in a materialist way, as small parts of organisms" (p9). The language here is interesting. Those of us who view genes, like poems, as abstract, as informational items that depend on one physical vehicle or another, do not view ourselves as departing in the slightest from materialism any more than we regard our appreciation and use of the concept of *software* as a step towards a Cartesian view of computers. PGS, in contrast, does see something ominous, or at least untoward, in such abstractions. *But he never says why*. Or at least, so far as I can see, this is an unmotivated and misplaced abstemiousness in his brand of materialism. To check my own reading of his reluctance, I composed a thought experiment for him to consider:

> Herb and Alice want to have a baby, but here's how they do it:
>
> 1. They both have their genomes sequenced.
>
> 2. Then they write a little computer program that does the meiosis algorithm on both their genomes, creating virtual sperm and eggs, which are then (randomly) united *in silico*, to create a new genome specification (which passes all bioinformatic tests for being the specification of the DNA of an offspring of Herb and Alice). (So far this is all done literally in terms of A,C,G.T, a purely computational process of string-rewriting.)
>
> 3. This specification is then used to construct, codon by codon, an actual DNA implementation of the entire genome
>
> 4. Which is then implanted in the nucleus of a human egg (if it matters to you whose egg this is, let me know), and becomes a "test tube baby" in one of the usual ways.

> Now you might want to compare the resulting infant, Hal, with Swamp Infant, which is, thanks to Cosmic Coincidence (yuck) genetically identical to Hal. **Is either one of them the offspring of Herb and Alice?** I would say Hal is, and Swamp Infant is not (but Swamp Infant is impossible in every sense that matters—logical possibility is just not interesting here.) I'm not trying to pin any essentialism on you; I'm just trying to see how, given your antipathy to informational views, you would characterize this (quite possible) mode of . . . reproduction. It seems to me to be a case where Hal is very clearly the biological offspring of Herb and Alice, because what matters is information, and the causal transmission of information (in the form of ASCII code for "A", "C", "G" and "T", not in the form of molecules (codons). The causal link might, for instance, pass through telecommunication satellites, instead of taking the more direct, biochemical routes. (personal correspondence, 4/26/2010)

And here is the response from PGS:

> I accept that Hal is the offspring of Herb and Alice.
>
> I guess the closest real-world case of something like this would be reproduction by retroviruses, where the RNA in the virus gets transformed into DNA and then back into RNA in the new viral offspring. Your case is a grander case of the same sort of thing. There is structure in the parent, and by a somewhat tortuous causal process we end up with similar structure in the offspring.
>
> What I am not so sure about is what role 'information' has in this case. You say:
>
> "It seems to me to be a case where Hal is very clearly the biological offspring of Herb and Alice, **because what matters is information**, and the causal transmission of information (in the form of ASCII code for "A", "C", "G" and "T", not in the form of molecules (codons)."
>
> What matters is the re-creation of structure in a new generation. Computers are information-processing devices, sure. And we use the computer to represent the genomes of Herb and Alice. I don't have any problem with the idea that somewhere in the memory of the machine there is a structure being used to represent Herb's genome. But that does not make inheritance an informational process. Forgetting genes, we could take a representation of Herb's facial features and a representation of Alice's facial features, and get the computer to do a meiosis-style mixing of them, and then build a robot with the blended face. Here we use an information-processing device to represent some structure in Herb and Alice, and also to create a new structure which is based in a particular way on the old structure.

> In your Hal case a form of artificial inheritance is devised, in which an information-processing device is used at some crucial steps. But that does not show that biological inheritance is a matter of transmission of information. It shows that information-processing *can* be used in the re-creation of structure across generations. Most of the time (in normal humans), something other than information-processing is used. (personal correspondence, 5/3/2010)

I am bemused by this response. The "recreation of structure" would seem to me to be close to a defining power of information. When a sender has a structure (a message) and succeeds in getting a receiver to recreate that structure, this just *is* successful information-transmission, however it is accomplished. I think it is obvious that we want, or even need, a perspective that permits us to theorize, to make predictions and consider explanations, about phenomena that succeed in "recreating structure" via substrate-neutral paths. We learn that our enemy has somehow *obtained information* about the design of our new submarine. Did a spy copy the blueprints onto pieces of paper and smuggle them across the border, or encode a recipe for the design into patterns of dots and dashes that were sent by radio, or memorize the blueprints and walk across the border, or . . . . .? We might be able to discover that the information moved from Bethesda to Beirut to Bern to Bonn to Baghdad and still not know what physical medium or what code was involved in each leg of the journey. PGS grants that "artificial inheritance" can be accomplished by interposing an information-processing device (and nothing else) in the normal process. To me, that is as good as a proof that the normal process of reproduction is fundamentally an information-transmission process. In a similar vein you can replace the ropes and pulleys and drums connecting the helm to the rudder with radio links (together with the transducers and effectors needed to give the skipper a sense of the resistance, and to actually move the rudder), showing that steering is fundamentally an information-transmission process or control process, as Norbert Wiener recognized when he created the term *cybernetics* from the Greek verb for *steering*.

A few days after this exchange, Craig Venter announced the creation of his novel bacterium with hand-crafted genome, inspiring PGS to send me another email message:

> Here is little Hal.
>
> http://news.bbc.co.uk/2/hi/science_and_environment/10132762.stm
>
> Though actually he has a formal parent and a material parent. (personal correspondence, 5/21/2010)

Just so; it is now possible to take the information and use it to construct a new vehicle for that information that can be read just fine by the organism that contains it.

4. "Darwinian paranoia"?

Now let me turn to his anti-agent theme, which is developed in much more detail, and is, in fact, as good as the punch line of his book.

According to PGS, "the agential perspective on evolution has always been an uneasy mix of the metaphorical and the literal. . . . . but all talk of benefits and agendas comes with a peculiar psychological power." (p10) The peculiar power is, of course, the power of the intentional stance (Dennett, 1971, 1978), which enables us to think strategically about all manner of phenomena, from our fellow human beings and animals, to computers and even to evolutionary processes—evolutionary game theory, the tragedy of the commons, and the question I have claimed to be central to all evolutionary thinking, *cui bono?* (who or what is the beneficiary of this competition?). Let me note first that I agree with PGS that there is a variable mixture of the metaphorical and the literal here, but I view this as a feature, not a bug; it is just another instance of Darwinian anti-essentialism: drawing a "principled" dividing line between *genuine* belief-talk or agent-talk and mere *as if* belief-talk and agent-talk is the sort of task Jerry Fodor insists on, pounding his fist on the table, not a methodological maxim any Darwinian should have any truck with.

PGS cites David Haig as describing this as "a way for us to be smart when we think about evolutionary problems" but PGS doesn't "think that this is such a good thing." (p10) Why not? Because it leads to what Richard Francis (2004) has called "Darwinian paranoia", defined as "the tendency to think of all evolutionary outcomes in terms of reasons, plots, and strategies." (p10). And what is wrong with that? The coinage by Francis is a brilliant propaganda stroke—right up there with "Just So Story" (Gould and Lewontin, 1972) and "Darwinian Fundamentalism" (Gould, 1997)—but on closer inspection I find that, once again, rhetorical branding has taken the place of calm, objective criticism. I cannot recall a more mean-spirited book on evolutionary thinking than Francis (2004)—spare yourself the experience and read Herb Gintis's (2004) accurate and critical review—and I hope that PGS will jettison Francis's term as unworthy to express his own critique in the future. My own experience discussing the term and the issues it raises with evolutionary biologists during the last year has shown that it gets in the way of serious communication. For instance, a biologist working on molecular evolution who sees the deep value of strategic formulations in considering the arms race between transposons and the genetic elements that combat them is likely to have the charitable reaction that "Darwinian paranoia" must refer to some pathological practice far removed from her bailiwick, and be frankly incredulous when told that she does indeed stand diagnosed as suffering from this disturbance. The plain fact is that agential talk is not at all restricted to the embattled terrain of evolutionary psychology—though this is where it is most often deplored by critics[1]—and is *literally* ubiquitous (that is, it can be found *everywhere* in evolutionary theorizing, from origin of life research and molecular genetics to cell biology, and yes, even evo-devo biology of the sort Francis champions as the alternative to it).

Ironically, PGS himself finds it irresistible to engage in agential talk when explaining and defending his own best ideas. Here are a few examples among many[2], in PGS's discussion of de-Darwinizing: "Their independent evolutionary *activities* are c*urtailed, constrained*, or *suppressed* [my italics] by what is happening at the higher level" (p122), "Another way of dealing with this *problem* [my italics] would be for one member of the collective to *prevent* [my italics] reproduction altogether by other individuals . . . " "This can be put more explicitly in terms due to Calcott (2008). For a transition to occur there must somehow be both the *generation of benefit* and the *alignment of reproductive interests*." (p124) So Calcott is praised for putting it "more explicitly" instead of condemned for indulging in such metaphorical agential talk. A double standard is clearly visible, and PGS has not yet shown how to tell the good guys from the bad guys. (I am not asking, Fodor-like, for the *essence* of good—or bad – agential talk, but just some serious discussion of when and why it is good and bad.) PGS does say:

> . . . there is a style of selectionist thinking in biology that I think does not involve any form of paranoia. This is the kind of investigation where someone asks: suppose a population was like this, and such-and-such a mutation appeared, what would happen to it? Thinking this way does not require the idea that genes are "ultimate beneficiaries" of anything. (p145)

---

[1] It is worth noting that evolutionary biologists confidently hypothesize historical events—horizontal gene transfers, for instance, that occurred billions of years ago, give or take a few hundred million years, or speciations and migrations that must have occurred at some point, ill-defined in space, time and causation—without fear of being chastised for indulging in Just So Stories. It is pretty much only hypotheses about human evolution that are held to a higher—conveniently unattainable—standard of evidence, by the critics of sociobiology or evolutionary psychology. Yes, there are egregious cases of hypotheses being defended solely on grounds of their plausibility, given the few facts available, but they shade into entirely reasonable cases—across biology, so far as I can see—with no clear boundaries. Much of the progress in evolutionary biology consists in the confirmation or disconfirmation of bold hypotheses that started out as plausible guesses—Just So Stories.

[2] See also, *e.g.,* p133, on homing endonuclease genes, and p134: "The driving chromosome has a 'resistant' element at the place in the genome that the killer targets." See also PGS's reflections on why he doesn't accept the gene's-eye view as fundamental, pp134-5.

Of course it does not "require the idea"; one is entitled to theorize with both hands tied behind one's back. The question is whether couching this thinking in terms of the genes as ultimate beneficiaries is enlightening or leads to serious misdirection. Given PGS's expression of strong disapproval, I was expecting a parade of Bad Examples, shocking or embarrassing instances of agentialists led on a wild goose chase, or blinded to a simpler truth. But I found none in the book, and when I turned to Francis, I found the parade all right, but most if not all of the cases were instances of what I call *rathering*:

> A rathering is a rhetorical move much beloved by some of the other ideologues of biology, such as Steven Rose and the late Stephen Jay Gould. The general form of a rathering is "it is *not* the case that blahblahblah, as orthodoxy would have you believe; it is *rather* that suchandsuchandsuch–which is radically different." Some–not all–ratherings are little more than sleight-of-hand, since the word "rather" implies–without argument–that there is an important incompatibility between the claims flanking it. (Dennett, forthcoming)

When encountering a rathering, one should always ask whether instead one can have one's cake and eat it too. In the case of Francis, the oft-implied claim is that one should be an evo-devo theorist *instead of* an adaptationist/selectionist/agentialist. Why not be both? Like PGS, Francis is quite content to help himself to doses of Darwinian paranoia when it is just too obvious that they are illuminating. (For the details, see the Gintis review.)

One part of the motivation for opposing agentialism has always been the political hunch that it gives aid and comfort to the enemy, creationism. It is thought by many that talking about the design or purpose or teleology evident in the workings of nature is playing into their hands since, as they like to insist, there is no design without a designer. The alternative political hunch, which I share, is that it is much better to attack the supposed principle (no design without a designer) as a pre-Darwinian illusion, like the obsolete principle that atoms are, by definition, unsplittable. PGS addresses this division of opinion: "Sometimes Darwinism is seen as demolishing the last elements of a teleological outlook, but at other times Darwinism is seen as constructively domesticating these ideas, showing that they have a limited but real application to biological processes." (p12) I have always thought that the latter view is much more perspicuous, much more defensible, when explaining evolution to laypeople. It is just so obvious—to them, to everyone really—that organisms are breathtakingly cleverly designed, all the way down to the macromolecules. I was disconcerted to overhear some medical students talking in a bar recently. One exclaimed: "How could anybody believe in evolution after learning about the intricacies of the DNA replication machinery?" To the extent that well-meaning evolutionists had inadvertently convinced them that Darwinians are eager to gloss over or deny these facts, this is evidence that the political tactic of denying teleology root and branch is apt to be self-defeating. PGS goes so far as to countenance a "very deflationary sense of 'supposed to'" (p13) and then draws back:

> But these [teleological] intuitions are part of a set of habits and ideas that steered us wrong for centuries, as far as theorizing is concerned, and had to be overcome to develop the Darwinian view. The *feeling* that some particular way of looking at things yields understanding should not always be taken at face value, is not the end of the matter. (p13).

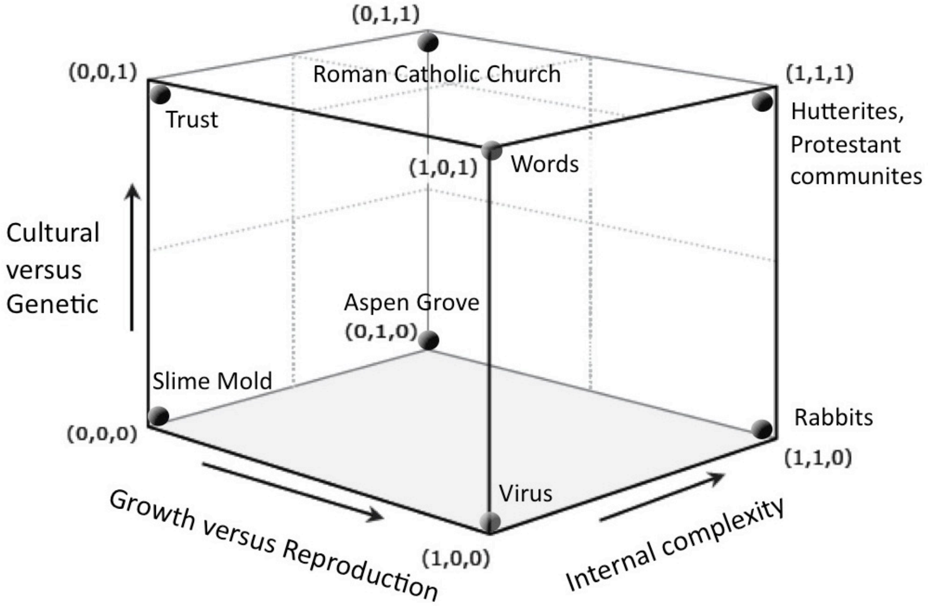
If he means that in the centuries before Darwin, they steered us wrong by convincing us that there was a Designer who created "all creatures great and small," he is right, of course, but if he means that Harvey was wrong about the function—the purpose—of the heart or the circulatory system, that naturalists were wrong about the purpose of male combat and display in competing for mates, or about the bargains made between plants and pollinators, he is distorting the history of science badly. It was after all Paley's mind-boggling catalogue of the teleological wonders of nature that inspired Darwin's curiosity in the first place. We can all be grateful that pre-Darwinian naturalists were not fanatically abstemious about postulating functional descriptions and explanations of the phenomena they observed, and were largely right to trust their "*feeling"* that they were encountering features with purposes.

There is no doubt that there is an all too inviting slide from pre-Darwinian thinking via agential thinking into mistaken thinking about evolution, but it isn't obvious that agential thinking is the culprit. Exactly how is that supposed to work? PGS notes on p143 that Williams, Hamilton, and Maynard Smith developed the gene's eye point of view as an antidote to naïve group (and species) selectionism. But then, he says, it took on an illicit power of its own as a "special kind of agential narrative" (p144). Perhaps, but show us, please. He gives us a list of suspects: "demonic possession narratives, the sub-personal creatures of Freud's psychology (superego, ego, id), and selfish genes and memes." (p144) But you can't banish this whole flock of ideas with an epithet, lumping demonic possession with selfish genes, as if it were obvious that they were equally dubious ideas. You have to demonstrate that these are harmful ways of thinking, and that has not been done. Homuncular functionalism is alive and well in many guises, even in PGS's book: "The cells in a human are a bit like a parliament, or an orchestra—but one whose members reproduce as part of their orchestral activities." (p149)

As the reader can see, PGS has not made a dent in my unrepentant enthusiasm for the design stance in biology, for seeing biology as reverse engineering at every level. Indeed, I emerge from his critical review more confident than ever that this is the best way of thinking biologically, largely because his persistent and eloquent resistance to the idea comes up with so little that perplexes or discomfits me.

5. What about memes?

When PGS turns to cultural evolution, his discussion, as usual, is more open-minded, nuanced, and imaginative than one generally finds (among meme-skeptics, in particular) and he makes some valuable original points. He proposes, contra Sperber, for instance, that the focus should be on "how various mechanisms that might be found in human culture relate to the Darwinian framework in principle, rather than to say which mechanisms are the most empirically important ones." (p147) And he sees that the rejection of an evolutionary approach to culture by Fracchia and Lewontin (1999) is not conclusive: 'an initial populational mode of interaction (which would be Darwinian in various regards) may give rise to something else." (p149). Indeed it may; citing Reisman (2005), he suggests that cultural items are less like a population when the society becomes more structured, where power relations are more asymmetrical. This is—though he doesn't note this—a good place in the discussion for another PGS space, along these lines:



Here we see de-Darwinizing that parallels the multicellular organism case, in which the structure of relations between the elements (one is tempted to say the *organic* structure) render large "memeplexes" (Blackmore, 1999) , such as theories and religions, less amenable to (most) Darwinian perspectives.

He even notes the limited applicability of dual pathway (genetic and cultural) inheritance, where the fitness in question is always considered to be the fitness of the humans. There is a second way of being a Darwinian about culture, the memes way: "cultural variants as making up their *own* Darwinian population." (150) This, however,

> does not, in principle, require replicators. Here, as before, I understand replicators as members of a Darwinian population that reproduce asexually with high fidelity, "preserving structure" over many generations of copying. (p151)

PGS does not like the replicator concept, and offers, in addition to the (inconclusive) objection that it is "agential," an uncharacteristically opaque argument against it. He says that the replicator concept isn't necessary and, more important, doesn't handle all cases. The argument he gives for this is a thought experiment about an imaginary possibility of evolution occurring in the absence of any replicators at the lower level. (p33) But such a mere "conceptual" possibility, methinks, may depend, like so many philosophical thought experiments about what is possible or impossible in principle, on not looking deeper at the causal structure of this imagined heritability. What PGS needs to offer is a positive sketch (it can be equally imaginary) of a mechanism in which there is *clearly* no replicator lurking, but in which the molar effects (tall breeds tall, etc.) are more or less guaranteed.

He also notes that the replicator concept typically breaks things into "types" (alleles at a locus, really) and that this type-casting is "not necessary." He "in contrast, will make much of the simple phenomenon described above: the possibility of evolution in a situation where *everything is unique*." He ties this to 'epigenetic' inheritance (Jablonka and Lamb, 1995). (p34) What is curious about this objection is that it threatens to violate his own anti-essentialism. How is it an *objection* that there are (or, really, *might* be) cases in which no clear replicator is found but which still count as (marginal) cases of evolution? Can't the replicator folks say "Yes, there are these marginal phenomena in which replicatorhood is close to zero" ? PGS notes this possibility (p39) but I don't see that his later discussion dismisses it. (I think I may well be misunderstanding PGS on this point; on my reading, his arguments fall clearly short of meeting their burden.)

His discussion of Sober and Wilson's treatment of the Nuer and Dinka notes that in this instance, in place of tribal *reproduction* (with daughter tribes spawned somehow) tribal *growth* is the analogue or stand-in, and this is marginal in PGS's terms, but he goes on to observe: "A group that persists does have the opportunity to transform itself into another kind of group—in that sense persistence is linked to novelty." (p152) Yes, and here is where PGS might have looked at the other end of the biological spectrum, the pre-biotic world, and seen that much the same applies there. Before there were replicators in a strong sense, there was differential persistence and differential growth (of metabolic cycles or networks, of proto-membranes, etc.), and these provide an arguably stable (stable enough) framework for the development and maintenance of novel features. As he says, in discussing the origin of novelty early in the book, “Natural selection can reshape a population in a way that makes a given variant *more likely to be produced* via the immediate sources of variation than it otherwise would be.” (p43) This is the key to the Baldwin Effect, of course, but also can be seen to be critical in some prebiotic processes: they raise the probability of certain otherwise extremely improbable combinations coming into existence—without requiring full-blown replicators to set the stage. Just as the Tufts undergraduate *student body* of 2006 becomes the student body of 2007 and then 2008, with departures and arrivals as well as returning students, preserving some characteristics while transforming, gaining and losing others, so the pre-biotic world could contain phenomena that explored the space of possible characteristics in a way that improved the prospects for the rise of novel adaptations. PGS acknowledges that persistence can substitute for replication, and then adds: “But it is a weaker link than the one seen when there is differential reproduction.” (p152) Perhaps , but he gives no grounds for this. Is it true? And even if it is, he has shown us how to embrace marginal, weak cases of Darwinian phenomena, and almost certainly such proto-Darwinian processes paved the way for the paradigm processes that took over so fecundly.

PGS finds the memes approach “strange” and adds “It *is* a strange idea, though this fact can be obscured by patterns of description that reify cultural traits and make things like ‘ideas’ sound more concrete than they are.” (p152). What is this problem of concreteness? He has already cited prions as bona fide cases of parent-offspring that are entirely “formal” not material, and adds: “So formal reproduction can be a basis for Darwinian evolution.” Right there he is granting the main point of Dawkins’ discussion of memes. (See Szathmary, 1999, for more on the parallel between prions and memes.)

Now we just have to look at the details, seeing how memes are like and unlike other Darwinian populations. PGS offers a nice example of using a turntable as a musical instrument.

> A few people see or hear you, and do the same thing. The behavior spreads. It is not quite accurate to say that your individual behavior was the parent of theirs, but something close to this is true. (p152)

As he says, they acquire a “disposition.” How concrete is that? The induction of change (e.g., prion protein refolding) will do for formal reproduction (p153), so the induction of disposition acquisition should count as well. We could make good use of another PGS space with abstract-concrete on one of the dimensions. Another dimension could be number of parents. As PGS observes (p153), having a single parent is not essential. Words, for instance, are spread from person to person by the joint action of a variable number of “parent” utterances. The first time a new token is perceived by a child, it is, perhaps, just a distinctive sound; the second time, it is a rather familiar distinctive sound;

the third time, it is a sound plus some recognizable context, and perhaps the accumulated effect of those three parent tokens is enough to install a copy in the brain of the child, where it will either soon die without offspring of its own, or else provoke its own rehearsal, yielding offspring and grand-offspring of its own. Usage and pronunciation—and, of course, meaning—can then be pruned and shaped by further encounters with conspecifics. Not the replication system of bacteria or bears, but still recognizably a process in which structure (characterized at a very abstract level!) is recreated by a causal process that depends on the structure of the "parents" (so that mutations—of meaning or pronunciation, for instance) get transmitted).

***

There are a bounty of other thought-provoking claims and suggestions in the book. One highlight: there are three varieties of reproduction: basic or simple, collective, and scaffolded. The asexual replication by fission of single cells is the paradigm of simple reproduction; our variety is actually an instance of collective reproduction, and viruses, chromosomes, mitochondria and (for those who take them seriously) memes are instances of scaffolded reproduction. Another is the very interesting discussion of Haig and Grafen, and the idea that "Genes, roughly speaking, are late-comers." (p140). The idea is that (re-)identifiable sequences of genetic material that can have "an evolutionary role" arose as a product of the evolution of crossing-over; "And the evolution of crossing-over is what set small genetic elements free as evolutionary players." (p141). There are many others.

So PGS has enlightened me, and clarified my thinking on these topics. We are close to harmony on the question of the potential role of memes: not all of cultural evolution (in the broadest sense of the term) is illuminated by being considered a Darwinian process taking one Darwinian population (of memes) to a successor population, but there is plenty of room for such a perspective, and work for it to do, particularly in the earliest days of culture, and in the evolution of the simplest units of cultural transmission (e.g., words, the paradigm memes). What discord remains is largely due to PGS's continuing streak of puritanism about the use of the intentional stance in biology, and a concomitant reluctance to talk in informational terms. I view the resulting tension as another one of nature's salutary opponent processes: I tug in my favored direction and he tugs in the opposite, and we tend to correct each other's overstatements.

References

Blackmore, Susan, 1999, *The Meme Machine*, Oxford: Oxford Univ. Press.

Dennett, D. C., 1971, “Intentional Systems,” *J. Phil.* **68**, pp.87-106.

Dennett, D. C. 1987*, The Intentional Stance,* Cambridge, MA: MIT Press/A Bradford Book.

Dennett, D. C. forthcoming, “Shall we tango? No, but thanks for asking,” (commentary on Evan Thompson, *Mind in Life*) in *J. Consciousness Studies.*

Francis, Richard, 2004, *Why Men Won’t Ask for Directions: The Seductions of Sociobiology*, Princeton, NJ: Princeton University Press.

Gintis, Herb, 2004, review of Francis, 2004, *Evolutionary Psychology*, [on line]

Gould, S. J. 1997, *“Darwinian Fundamentalism”* (a review of *Darwin's Dangerous Idea*), in *The New York Review of Books*, June 12, 1997,

Gould, S. J., and Lewontin, R., 1979, "The Spandrels of San Marco and the Panglossian Program: A Critique of the Adaptationist Programme," *Proc. Royal Soc.* (London), B205:581-98.

Jablonka, Eva and Lamb, M. J. 1995, *Epigenetic Inheritance and Evolution: the Lamarkian Dimension*. Oxford and New York: Oxford University Press.

Midgley, Mary, 1979, "Gene-juggling," *Philosophy*, **54**, pp439-58.

Midgley, Mary, 1983, "Selfish Genes and Social Darwinism," *Philosophy*, **58**, pp.365-77.

Sober, E. and Wilson, D.S. (1998): *Unto Others: the Evolution and Psychology of Unselfish Behavior,* Cambridge, MA: Harvard UP.

Szathmary, Eörs, 1999, "Chemes, Genes, Memes: A Revised Classification of Replicators," *Lectures in Mathematics in the Life Sciences*, **26** , American Mathematical Society, pp1-10.

Williams, George C., 1992, *Natural Selection: Domains, Levels, and Challenges*, Oxford Univ. Press.

Darwinian Spaces: Peter Godfrey-Smith on Selection and Evolution

Kim Sterelny

1. Darwinian Populations.

Darwinian Populations and Natural Selection is a superb work, introducing an important analytical technique, and applying it to a range of difficult and contested issues within evolutionary theory. The core idea of the book is that of a Darwinian Population (and, derivatively, a Darwinian individual). In developing this concept, Peter Godfrey-Smith builds on a long-standing tradition within evolutionary theory, that of developing a spare, bare-bones specification of the machine of selective change, both to better articulate the core causal processes of biological evolution, and to explore the possibility that selective regimes explain change in other domains (see also (Okasha 2006). Godfrey-Smith's ingredients are familiar: Darwinian Populations are populations of interacting individuals that (potentially) reproduce; that vary from one another in ways that sometimes influence their reproductive potential; and when reproduction does take place, offspring resemble their parents.

Nothing new here, but Godfrey-Smith handles these ingredients in a distinctive way. First, everyone recognises that heritability — the tendency of offspring to resemble their parents — comes in degrees. Godfrey-Smith explicitly extends this recognition of gradients to the other constituent features of Darwinian Populations. He shows that there are clear and less clear cases of reproduction. For example, with asexual clonelines, the distinction between growth and reproduction is complex and contested (Dawkins 1982a). Similarly, with social groups, there is no obvious dividing line between cases in which the individuals within the group have reproduced, and cases in which the group itself has founded colonies. So Godfrey-Smith devotes two splendid chapters to exploring and systematising some of these difficult cases. Fitness, and fitness differences are equivocal too, both because fitness differences can be counted in different ways, and, arguably, because differences in actual reproductive success only count as fitness differences if they have systematic and repeatable causes. Finally, there are unmistakable and marginal cases of a group being a population at all. Individuals are members of a single population in virtue of their between-member interactions, and those interactions vary in their intensity and consistency. So there are clear and marginal cases of populations.

Second, Godfrey-Smith explicitly distinguishes between clear, but minimal Darwinian Populations, and paradigm Darwinian Populations. Over the last four billion years, evolution in Darwinian Populations has resulted in a fabulously disparate biota, and, often, an astonishingly complex and intricately adapted biota. But not all Darwinian Populations have the potential for extensive change, or change resulting in complex adaptations. For example, the cells of a worker ant constitute a Darwinian Population. They interact. They reproduce, and daughter cells resemble their parent. They vary from one another (including genetically, as a consequence of somatic mutations), and some of those variations affect reproductive potential. Yet the potential for change is very limited. Paradigm Darwinian Populations, then, are those that possess the additional characteristics that allow selection-driven evolution to be fecund; the characteristics that make the evolution of disparity and complexity possible. A major project of the work is to identify these additional characteristics. While Godfrey-Smith does not use the term, in part <u>Darwinian Populations and Natural Selection</u> is an essay on evolvability.

This is a very fruitful way of conceptualising the field. The focus on gradients takes pointless hair-splitting about particular, difficult cases off the table: we do not have to decide (say) that a wolf packs meets some threshold so the pack itself counts as reproducing. But it does much more: it focuses our attention on the dynamics of evolvability. In a lineage of socially interacting agents, like our wolves, the degree of integration, of reproductive specialisation, of ancestor-descendant similarity need not stay fixed over time. A population of packs might be a very marginal case of a Darwinian Population at one point of time; a clear but minimal case at another; a paradigm Darwinian Population at a third. Populations can both Darwinise and de-Darwinise. Godfrey-Smith (following up (McShea 2002)) points out that as Darwinian individuals become parts of larger, paradigmatically Darwinian individuals, they themselves tend to become simpler, and less obviously members of Darwinian populations. An ant is a more obvious member of a Darwinian Population than an ant cell. The gradient conception of Darwinizing characteristics thus focuses our attention on the causes and consequences of such changes in population-level characteristics; on the flow and ebb of evolvability.

Moreover, theories of evolvability have tended to focus on characteristics of individual agents; typically, on features of their developmental system. But while agent level traits are clearly relevant to evolvability, it is a feature of populations and lineages of populations, not of individual agents. So Godfrey-Smith frames the project asking the right questions, focusing on the features of *populations* that explain the potential for rich evolutionary change. Finally, Godfrey-Smith treats these gradient characteristics of Darwinian Populations as dimensions, and that enables him to represent these cases spatially. For example, the volvocacean aquatic algae are mostly single-celled eukaryotes. But a number of multi-celled lineages have evolved, and these differ in complexity, reproductive specialisation and internal differentiation. Some of volvocaceans seem to be clear cases of complex Darwinian individuality; others are perhaps more naturally seen as temporary alliances between simpler, single-celled individuals (Kirk 2005; Michod 2011). Godfrey-Smith's spatial treatment is heuristically very valuable here and elsewhere, because it enables him to exhibit these lineages' similarities and differences clearly and compactly. A major virtue of this book is Godfrey-Smith's development of a clear and user-friendly way of highlighting and visualising the key phenomena.

2. Darwinian Populations and Gene Selection

In building his models of the Darwinian machine, Godfrey-Smith builds on a tradition in evolutionary theory beginning with Darwin himself. But its best known expression is in the work of Richard Lewontin (Lewontin 1970). This tradition takes individual organisms forming a population as the base case, and generalises from that case, abstracting away from details of agent and population. Godfrey-Smith sees the importance of this project, but is acutely sensitive to its limits. He points out that these abstract general recipes for evolution by natural selection cannot deliver on all their ambitions, for there is a fundamental tension between generality and simplicity. Simple, general formulations cannot capture all of their intended targets. So, for example, he points out that Lewontin's recipe does not capture cases in which the members of a population vary only in the rapidity with which they produce offspring (rather than varying in the number they produce). This is not logic-chopping. While no real population is likely to vary only in this characteristic, arguably selection for maximising the rate of reproduction is the most pervasive from of selection on bacterial populations: see (Lane 2005). Even if Lane overstates his case, if As reproduce more rapidly than Bs, As will increase their representation in the population over time. There is no simple, literally true, specification of the necessary and sufficient conditions of evolution by natural selection.

Godfrey-Smith suggests that we do better in seeing formulations like those of Lewontin as models: as a descriptions of an ideal, simple system that is strikingly and importantly similar to many real cases. Thus while seeing the limitations of classical recipes for selective change, Godfrey-Smith suggests ways these recipes can be both enriched and reinterpreted to lay bare the structure of uncontroversial but difficult cases of Darwinian machines in action (for example, evolution amongst the multicelled plants), and to identify more cryptic Darwinian machines in action. He is markedly less charitable to an alternative, more recent attempt to specify a general recipe for evolutionary change, a recipe built around the idea of a replicator. George Williams, Richard Dawkins, and David Hull have argued that evolution is best seen as competition between lineages of active replicators (Williams 1966; Dawkins 1982a; Hull 1988a; Hull 1989). These are structures that (i) pass on their organization to descendants with high fidelity, through some copying process; and (ii) influence their own prospects of being copied. Active replicators exert power over their local environment, and the nature and extent of that power determines their fitness. Over much of the history of life, the default strategy of successful replicators has been that of forming ensembles which guide the construction of vehicles (or "interactors", in Hull's terminology). These vehicles interact with the environment in ways which, if successful, lead to a further cycle of replication and vehicle building. So replicator lineages normally grow through the co-operative construction of biological machines. But while this is their default strategy, it is not the only route to replicator success.

Just as the classical recipe takes organisms in competition to be the base case, obviously this replicator-centred view is abstracted from the gene. Godfrey-Smith is deeply sceptical. As a former defender of this view, I shall attempt a defence of my former time slice. After some preliminary sparing, the core argument is developed in chapter 7, and I read that argument as follows:

> 1. Godfrey-Smith begins with the basic Lewontin formulation: evolution by natural selection is a consequence of populations in which we find reproduction, heritability and fitness differences. But Godfrey-Smith refines, revises and enriches that formulation in developing his ideas of minimal and paradigm Darwinian Populations.
>
> 2. He shows that this conception is theoretically fruitful, by showing how we can use it to organise and systematise a menagerie of difficult cases involving putative multi-level selection, reproduction and individuality. There is, for example, a terrific discussion of plant reproduction. Plants are appallingly behaved (Clarke 2011). Many reproduce by some form of vegetative propagation, so physiological and genetic criteria for individuality come part. They do not have a germ-line that develops early and is isolated from the rest of the organism, so mutations in the course of growth can result in morphological modules which are both genetically different from other regions in the plant, and which have the capacity to reproduce independently. So they are often chimeras. Many ferns reproduce via a haploid stage which grows into a complex multi-celled form, before producing gametes which disperse and, if lucky, fuse with another gamete. Counting plants, and hence estimating plant fitness, is a horrible problem. Godfrey-Smith decomposes the intuitive, folk biological concept of reproduction into three factors: the integration of the putative individual; the extent to which it is genetically homogenous; the extent to which it develops through a narrow bottleneck. In doing so, he gives us a framework to compare the different cases.
>
> This theoretical productivity step in the argument is crucial: the replicator-vehicle framework was formulated an alternative to, and replacement for, organism-centred conceptions of evolution. To the extent that a Darwinian Population is an idealised, abstracted model of organisms in interaction, of course it does not fit interacting

gene lineages. The gene-selection model could not be an advance, if it did, and the replicator-vehicle recipe would not be saying anything new.

3. He shows that the genes in an organism (or in a population of organisms) do not form a clear, unmistakeable Darwinian Population. In those cases in which genes are well individuated in terms of their physical organization and action on the world (for example, coding for a specific protein), they have a shared fate. For genes are well-individuated in prokaryotes and only prokaryotes. But in those prokaryotes, genes succeed or fail together. So we do not see the fitness differences characteristic of unmistakeable Darwinian Populations. Meiosis and cross-over in eukaryotes releases genes from their shared fate, but such genes are much less well-individuated in terms of cohesive physical structure and phenotypic action. The eukaryote challenge is the real driving point at this stage of the argument: a replicator-vehicle model can take the single bacterial chromosome to be a single replicator. But Godfrey-Smith argues that the eukaryote genome does not consist in a temporary alliance of clearly identifiable, autonomous replicators.

4. Godfrey-Smith then suggests that many of the standard gene's eye cases are better seen as selection on individual organisms, but for *genetic properties* of those organisms. Supposed selection on giraffe-inhabiting genes for long necks is really selection in favour of giraffes that carry that gene. Thus he points out that even gene selectionists count fitness by counting giraffes with the long-neck gene. Even by gene-selection lights, a long-neck gene does not increase its fitness by increasing its copy-number, because long-necked animas need to grow more neck cells.

5. Outlaw genes — meiotic drivers, sex ratio distorters and the like — are appropriately described as cases of gene selection. For there is variance in the fitness of germ-line genes in individual organisms. But these are exceptional, cases of Darwinian populations. They may even be marginal cases, for though the phenotypic upshot of selfish genetic elements is determinate, their boundaries are often indeterminate, and, as with other eukaryotic genes, heritability is compromised by cross-over.

The bottom line: selection between genes is not even one paradigm case of evolution by natural selection, let alone the base case that should organise our conception of evolution. This is a powerful line of thought. But it does not give replicators their due.

Extended Phenotypes. One idea behind the replicator-vehicle model is that genes — replicators — do not always leverage themselves into the next generation via their effects on the phenotypes of well-defined Darwinian individuals. Dawkins' The Extended Phenotype introduced this idea, and it has been further developed (in different terminology) in the niche construction literature (Odling-Smee, Laland et al. 2003). To borrow Godfrey-Smith's term, there is a menagerie of such cases. These include cases in which organisms physically modify their environment; they include genes whose adaptive function is to establish stable patterns of social relationships (like those found in many species of co-operatively breeding Australian birds). They include symbiotic alliances, in which each partner is selectively important to the others, but where the lineages have not fused, like eukaryotic cells, into single individual. Bullhorn acacias, for example, have physical structures that house and feed ants, their symbiotic partners, and those ants engage in guarding and anti-herbivore patrols that are adaptive for the ants only via their benefit to their symbiotic partners (Janzen 1966). Yet we still clearly have independent lineages here. They also include the seriously spooky cases of parasite control of host phenotypes (Combes 2005); for example, the parasitic barnacles that feminise their hosts morphology, and hijack their behaviour (Gould 1996a).

These extended phenotype cases have two distinctive features. First: they are all cases of selection acting powerfully to shape the biological world. These extended phenotype adaptations are not produced in minimal, low-powered, or marginal Darwinian regimes. Second, if we recast these evolutionary episodes into the framework of Darwinian Populations, we have to choose between two suboptimal representations. One represents these evolutionary histories as the histories of paradigm Darwinian Populations, but the actual adaptive effect that drives change is suppressed: it is represented by its proximal shadow. We represent the ants' behaviours, but not their effect on acacia growth, health and productivity, and hence their indirect effects on the ants' home; we represent the chemical messengers that the parasitic barnacle produces, but not its effects on the host phenotype, effects which lead to host care of the barnacles' own eggs. Alternatively, we represent the adaptive levers, but the lever is acting on aberrant or marginal cases of

Darwinian individuals. Our Darwinian Populations will consist of rabbits plus their burrows; ant-acacia partnerships; crabs feminised by their barnacle parasites.

The extended phenotype menagerie shows that paradigm cases of selective shaping do not line up 1:1 with selection driving change in paradigm Darwinian individuals in paradigm Darwinian Populations. Marginal and unclear cases of Darwinian Populations are not always cases where the Darwinian machine is running less powerfully. So if the point of distinguishing between minimal and paradigm populations just is to distinguish those regimes in which the generation of novelty, complexity, or additional disparity is possible, the distinction is only partially doing its job. Novelty and complexity is sometimes not novelty and complexity in a paradigm Darwinian individual. The replicator-vehicle model captures that fact.

Replication and Cumulative Selection. I also doubt that Godfrey-Smith does justice to the links between replication, high fidelity inheritance, and evolutionary potential. It is a truism of evolutionary biology that adaptive complexity can be built only by cumulative selection of small variations on existing phenotypes. A large, undirected change from an existing phenotype is almost certain to be maladaptive. But a small variation on an existing phenotype has some reasonable chance of improving it. Complexity is built incrementally. In *Climbing Mount Improbable* (especially), Dawkins connects this truism to the replicator-vehicle framework. Cumulative selection requires high fidelity inheritance, as small variations from existing types must be preserved if they are to be available as a basis for further improvement (Dawkins 1996). High fidelity in turn depends on replication; a discrete, perhaps even digitalised, template copying process (Dawkins 1983). Replication, Dawkins argues, is the only mechanism that could ground high fidelity inheritance. It is fundamental to the evolution of complex, antecedently improbable systems.

I am not sure where Godfrey-Smith disagrees with this line of analysis. He rightly points out that defenders of the replicator-vehicle model are just mistaken when they claim that replication is essential for any form of evolution by natural selection, and equally mistaken in thinking that evolutionary change depends on the persistence of types across generations. Height in a population can evolve over time even if no two individuals ever have the same height; it suffices that offspring resemble their parents more than they resemble others (p33). It is also true that we can indeed imagine paradigm Darwinian Populations with no replication. Godfrey-Smith imagines one in which genes direct protein synthesis, but once the organism develops, the genes are metabolised, but reverse translated from proteins immediately prior to reproduction. In such a biology, there would then be no unbroken chain of DNA copies from parent to offspring. We can indeed imagine such a case, but what has imagination got to do with it? Godfrey-Smith agrees that the imagined system is probably not biochemically possible. The fact that we do not see DNA to protein to DNA cycles is no mere quirk of evolutionary contingency.

So the Williams-Dawkins Conjecture is this: (i) richly evolvable selective regimes depend on high fidelity inheritance; (ii) all actual and feasible cases of high fidelity inheritance depend directly or indirectly on replication. Indirect cases are scaffolded by Dennett's cranes (Dennett 1995): complex adaptations whose evolution *does* depend directly on replicator-based inheritance. As Godfrey-Smith notes, high fidelity cultural inheritance in the human lineage probably does not depend on replication (a point that Dan Sperber has made forcefully: (Sperber 1996)). But it does depend on complex psychological adaptations whose evolution did depend on replication. Any form of high fidelity inheritance seems to require some combination of a benign environment and complex machinery; machinery unlikely to be available for free. That is why the origin of life presents such a brutally difficult chicken and egg problem. Hi fidelity without replication exacerbates that problem; you need even more machinery. Using a structure as input to a template copying process must be simpler that using a *representation* of that structure as input; template copying stores information in the world. The machinery which replicates DNA is complex enough. Imagine what you would need to add, if as well you needed molecular machines which would translate unfolded, complex three dimensional protein shapes into a 2D linear order of amino acids, before these were reverse translated into DNA. If the Williams-Dawkins Conjecture is right, even though

Godfrey-Smith is indeed right that some evolutionary regimes do not involve replication, replication is fundamental to evolution in a way his framework does not capture.

Suppose Godfrey-Smith was to accept much of this. What then? The importance of replication could be built into his picture very simply. In graphing paradigm Darwinian Populations in selection space, one dimension represents heritability. Godfrey-Smith could simply stipulate that high heritability values depend on high fidelity replication, thus explicitly recognising the pivotal role of replication.

Capturing the extended phenotype zoo is less straightforward. But the model-based strategy defended in the book does not demand exclusivity. Godfrey-Smith writes:

> "One option is to give a literally true description of the important features of all cases of the phenomenon. The other option is to give a detailed description of one class of cases, usually a relatively simple case, and use it as a basis for understanding the others. … Understanding is achieved via a similarity relation between the simple cases we have explicitly analysed and the more complex ones" (p26).

This meta-theoretical perspective does not require a single choice of model. So Godfrey-Smith could argue that while many central episodes of evolutionary history are best understood as evolutionary change in, from, or to paradigm Darwinian Populations, when the adaptive effects of genes and gene changes are distributed over several biological individuals, or when they affect coupled organism-environment systems, the Replicator-Vehicle model is more illuminating. We would get a more complex picture and we would need some account of the interaction of the two kinds of case. But the move is available.

3. Selection Spaces.

A crucial part of Godfrey-Smith's analysis is his distinction between marginal, minimal and paradigm Darwinian Populations. These distinctions are drawn by locating populations in an abstract, five-dimensional selection space. The dimensions are (i) H: the fidelity of reproduction; (ii) V: the supply of variation; (iii) α: the intensity and ubiquity of competition for reproduction through the population; (iv) C: the smoothness of the fitness landscape; and (v) S: a parameter that represents the extent to which fitness differences in the population depend on differences in the intrinsic features of the individuals in the population. The importance of V and H is obvious. The role of C has also been identified in the literature. Cumulative evolution requires high fidelity inheritance. But it also requires similar phenotypes to be similar in fitness. If the fitness landscape is rugged, with very similar individuals have very different reproductive fates, there will be fitness trenches between (say) rudimentary and more complex eyes (Kauffman 1993). Thus Nilsson's and Pelger's famous model of eye evolution assumes that small increments in visual resolution corresponds to small increments in fitness (Nilsson and Pelger 1994).

Godfrey-Smith's α takes up a neglected issue is evolutionary theory, that of identifying populations. The possums on Black Mountain (in Canberra) interact only occasionally with those on nearby Mount Ainslie, because a chunk of urban Canberra divides them. So do we have one population or two? Parameter α is a first-pass attempt to address this issue. A group of individuals is more paradigmatically a population, the more reproductive success for one reduces the prospects of the others. High α accentuates potential fitness gradients. Suppose a new variant in the population is unusually fit. If α is high, interaction is pervasive, and that distributes the effect of the advantage across the population. But as Godfrey-Smith acknowledges, this is at best a first pass attempt at capturing a very complex set of possibilities. So, for example, it does not capture interaction chains: common when the radius of individual influence is small compared to the suitable habitat through which individuals are spread. A tree in an extensive forest competes with its neighbours, and those neighbours compete with their further neighbours, and so on. But our first tree, and one growing a kilometre away, have negligible effects on each others' prospects. The same will be true of many of the small invertebrates living on and by the tree. Sometimes evolution acts in networks or webs rather than relatively discrete populations, and α does not capture this difference. So we should see α as a temporary placeholder for the complex and varying ways in which a species is typically composed of a variable and overlapping mosaic of populations, quasi-populations and networks of interaction.

The importance of H, V, and C is widely recognised. As just noted, there has been much less work exploring the different ways populations are integrated, and Godfrey-Smith is right to move that issue to centre stage. He is also innovative in arguing for the special importance of intrinsic fitness differences. But this is an innovation I resist. Intrinsic properties contrast to relational ones, and Godfrey-Smith argues that while relational properties make a difference to fitness, and can be inherited, evolutionary possibilities based on relation-based fitness differences are limited. His example is location. Location matters: an animal's exposure to lighting strikes, storms, and other unpredictable disturbances will often depend on location. In many animals, the distinction between germ-line and somatic cells, likewise, depends on accidents of initial location, and this dampens down the effect of within-organism evolution in the population of cells. A variant that increases its frequency in the cell population does not thereby increase its representation in the sample of cells from which new organisms derive. Moreover, location can be inherited. Many organisms reproduce close to the place they were born. That is true even of relatively mobile animals. The birthplace of a brushtail possum will resemble that of its parents.

However, Godfrey-Smith points out that while the location of possums can evolve, evolutionary possibility is very limited. If a fortunate possum finds itself in a favourable area, the demographic focus of the population can change; a greater proportion of the population comes to live hither rather than yon. A population can explore its physical environment, concentrating in favourable areas; thinning out in less favourable ones. But that is about all: "if extrinsic features are most of what matters to realised fitness — if intrinsic character is not very important — then other than this physical wandering, not much can happen (p55)." Thus basing the germ-soma distinction on location within the developing embryo powers down Darwinian processes at the level of cells. Fitness differences depend on extrinsic traits of the cell, and cell-level evolution goes nowhere (pp102-103).

I am not convinced. First, I wonder whether the concept of an intrinsic trait is well-defined. What of those traits whose development is richly dependent on environmental input? The food preferences of a rat, for example, depend on maternal imprinting and early experience, as well as the rat's own genetic endowment. So are preferences intrinsic? Think too of traits that guide interaction with the environment. Consider a gerygone's disposition to build a camouflaged, pouched nest. If the gerygone's nest building were powered by an internal template that specified location, construction technique, materials and the like, the nest building disposition would be an intrinsic trait. But what if, as is likely, gerygones "store information in the world". The inner template guides only an initial starting point and choice of materials. The growing nest then plays an essential role in stimulating and guiding further steps in the process. Elaborate courtship displays, like those of the great crested grebe, are quite likely to be like this too. Neither sex has a mental template of the whole routine. Rather, each bird has a linked set of responses triggered by the previous step of its partner. These are cases in which the agent in question has a disposition or capacity that can be *specified* without reference to its environment. But the categorical basis of the disposition is not internal to the agent, as the categorical basis of the disposition of a crystal glass to shatter is internal to the glass. Rather, it depends on the way the agent is embedded in its environment. If this is how displays are organised and nests are built, are these dispositions intrinsic traits?

Second, I think there is an important difference between relations that merely happen to an agent, and relations that an agent actively makes and maintains. So the extended phenotype cases that fuel my sympathy for the replicator-vehicle model make me sceptical of the special importance of intrinsic difference makers for fitness. Termites live in elaborate physical structures, and in networks of complex social relationships. These are not intrinsic properties of the termite (they might be intrinsic properties of termite mounds, but even if a mound is a Darwinian individual, so is a termite). Yet these are clearly the result of elaborate, rich evolutionary trajectories. Some ancestral proto-termites were fitter than others in their population, because of their social relations with other termites, and through their physical organization of their surrounds. In general, when one Darwinian individual is part of a larger Darwinian individual, the inside individual's fitness will depend heavily on evolved and elaborate relational properties. The same is true of other extended phenotype standards: beaver dam and lodge systems; parasite control over host behaviour and morphology. There might not, for example, be much current, heritable variation in beaver dam and lodge systems. But if these evolved, as they surely did, by cumulative selection, then some proto-beavers built less good dam and lodge systems, and others built better ones, and were fitter as a consequence. They were fitter because they lived in lodges that resisted attack better, or because their dams were better sited, deeper or more secure. Living in a deep, well-protected lodge and dam system is not an intrinsic feature of a beaver. But such differences were fitness difference makers, and they lead to important evolutionary changes.

That said, the parameter S gets at something: the connection between developmental control and heritability. Cumulative selection requires high heritability, and high heritability requires fine-grained control of the developing phenotype. The more development is affected by environmental noise and disturbance, the less reliably offspring resemble their parents. The boundary of the agent is often the boundary of reliable developmental control: a homeostatically buffered, predictable inner environment gives way to a wild and intractable outer one. So relational properties will develop reliably and predictably — and hence can be tuned and modified by selection — only in special cases; cases in which agents can actively manage and protect them. But the niche construction-extended phenotype literature suggests that these cases are neither rare, marginal or inconsequential.

Here is another way to put the idea. Godfrey-Smith is right to identity the intrinsic-extrinsic difference as important to evolutionary potential. But the distinction is important when we explain the *cause* of parent-offspring similarity, not to the similarities themselves. When traits are heritable largely because of extrinsic factors that happen to impinge on genealogical lineages within a population in repeated ways — where they happen to be born; disturbance regimes that regularly act on the phenotypes of some members of the population but not others — nothing much will happen. Evolutionary potential is much greater when parent-offspring similarity is largely due to factors intrinsic to parent and offspring. As we see above, intrinsic versus extrinsic causes of heritable differences roughly line up with intrinsic versus extrinsic traits. But the correlation is rough. So while S tracks an important control on the supply of heritable variation, I am yet to be persuaded that intrinsically is important in itself.

In brief: I have reservations about some of the uses to which Godfrey-Smith puts his analytic apparatus. But in many cases, he uses it very illuminatingly: for example, in the analysis of difficult cases of reproduction and individuality; and in the analysis of multi-level systems. Most importantly, I have nothing but admiration for the toolbox he has built for all of us.

References

Clarke, E. (2011). Plant Individuality and Multilevel Selection Theory Major Transitions in Evolution Revisited. B. Calcott and K. Sterelny. Cambridge, MIT Press: 227-250.

Combes, C. (2005). The Art of Being a Parasite. Chicago, University of Chicago Press.

Dawkins, R. (1982a). The Extended Phenotype. Oxford, Oxford University Press.

Dawkins, R. (1983). Universal Darwinism. Evolution from Molecules to Men. D. S. Bendall. Cambridge, Cambridge University Press: 403-425.

Dawkins, R. (1996). Climbing Mount Improbable. New York, W.W. Norton.

Dennett, D. C. (1995). Darwin’s Dangerous Idea. New York, Simon and Shuster.

Gould, S. J. (1996a). "Triumph of the Root-Heads." Natural History 105: 10-17.

Hull, D. (1988a). Interactors and Vehicles. The Role of Behaviour in Evolution. H. C. Plotkin. Cambridge Mass., MIT Press.

Hull, D. (1989). The Metaphysics of Evolution. Albany, State University of New York Press.

Janzen, D. (1966). "Coevolution of mutualism between ants and acacias in Central America." Evolution 20(3): 249-275.

Kauffman, S. A. (1993). The Origins of Order: Self-Organisation and Selection in Evolution. New York, Oxford University Press.

Kirk, D. (2005). "A Twelve Step Program for Evolving Multicellularity and a Division of Labor." BioEssays 27(3): 299-310.

Lane, N. (2005). Power, sex, suicide: mitochondria and the meaning of life. Oxford, Oxford University Press.

Lewontin, R. C. (1970). "The Units of Selection." Annual Review of Ecology and Systematics 1: 1-14.

McShea, D. W. (2002). "A complexity drain on cells in the evolution of multicellularity." Evolution 56: 441-452.

Michod, R. (2011). Evolutionary Transitions in Individuality: Multicellularity and Sex. Major Transitions in Evolution Revisited. B. Calcott and K. Sterelny. Cambridge, MIT Press: 169-198.

Nilsson, D.-E. and S. Pelger (1994). "A Pessimistic Estimate of the Time Required for an Eye to Evolve." Proceedings of the Royal Society of London, B 256: 53-58.

Odling-Smee, J., K. Laland, et al. (2003). Niche Construction: The Neglected Process in Evolution. Princeton, Princeton University Press.

Okasha, S. (2006). Evolution and The Units of Selection. Oxford, Oxford University Press.

Sperber, D. (1996). Explaining Culture: A Naturalistic Approach. Oxford, Blackwell.

Williams, G. C. (1966). Adaptation and Natural Selection. Princeton, Princeton University Press.

A gene's eye view of Darwinian populations

David C. Queller

Abstract

Biologists and philosophers differ on whether selection should be analyzed at the level of the gene or of the individual. In Peter Godfrey-Smith's book, Darwinian Populations and Natural Selection, he argues that individuals can be good members of Darwinian populations, whereas genes rarely can. I take issue with parts of this view, and suggest that Godfrey-Smith's scheme for thinking about Darwinian populations is also applicable to populations of genes.

Evolutionary biologists are sometimes viewed with a bit of suspicion by other biologists. It's not that these other biologists question the Darwinian paradigm. Instead, it is because they feel that evolutionary biology is in some way softer than other branches of biology. It is not as reductionist and seems to make less appeal to the "hard" sciences. It is intrinsically historical, and much of what is important happened in the past and cannot be subject to controlled experiment. Yet looked at another way, evolutionary biology is doing better than its mechanistic counterparts. The evolutionary process essentially comes down to selection and inheritance, and we have an excellent understanding of both of those. We know pretty well what selection is and how to measure it and we know how genes are transmitted from one generation to the next. We can model these two phenomena quite effectively, and any biologist who thinks evolution is not based on hard rigorous mathematical theory is just ignorant. There is something missing though. There is a gap between how genes get transmitted and how selection works on phenotypes; we need to understand how genes translate into phenotypes. It is the problem of development, though not just conventional development in multicellular organisms. It is the whole problem of genes being turned on and off to produce products that then interact in very complex ways. This is, of course, the primary domain of exactly those mechanistic biologists who may view evolutionary biology as soft. So we could respond that we've done our bit – actually our two bits – and we're just waiting on you mechanists to do their part.

This is a bit unfair of course. Mechanists helped us understand our bits, as when they figured out how DNA underlies heredity. And they do in fact know quite a bit about how genes translate into phenotype – just not enough for a full understanding of how we get an organism. The biggest hole in our understanding is how the body gets built, and evolutionary biologists are divided about where they choose to hide this embarrassing body problem. Peter Godfrey-Smith (2009) follows the traditional choice in this respect (although in often interesting and novel ways). Along with most evolutionary biologists over the years, he puts phenotypic selection on the ground floor of his evolutionary edifice, and stashes the embarrassing body in the basement. Phenotypic selection is relatively easy to understand and then, by some mysterious process operating below that level, down in the cellar, favored traits are more or less faithfully rebuilt in the next generation. Of course we understand how the genes get passed on, but to really complete the cycle and get back to phenotypes, we need to breathe life into the body hidden in the basement. More often, in our models, we ignore the issue, perhaps looking guiltily away from the cellar door.

The alternative approach, pioneered by Richard Dawkins (1976, 1981), puts genes and heredity on the ground floor. This doesn't get rid of the body problem; it just stashes the body up in the attic. We understand quite well what happens on Dawkins' ground floor too; genes are passed faithfully from one generation to the next via rather simple rules. But a complete understanding of evolution requires looking upward, where genes build bodies that get selected.

The two cases seem quite symmetrical, differing only in which door causes you to avert your eyes. Dawkins sometimes seems to accept this, but at other times argues that his approach of focusing on genes or, more generally, on immortal replicators, is a better way to view evolution. Godfrey-Smith takes the opposite view, implicitly saying that the body is better kept in the cellar than the attic.

Godfrey-Smith (2009) spends most of his stimulating book on tidying up the main floor of his evolutionary house, working out the nature of Darwinian populations, and only a minority of it criticizing Dawkins' housekeeping, which is where I want to focus my remarks. This is bit unbalanced, so I will begin with a quick look around Godfrey-Smith's place, where there is much that is pleasing.

Starting from traditional formulations of selection involving heritable variation that affects reproduction, Godfrey-Smith isolates five conditions that must be fulfilled to be a paradigm Darwinian population. Each is given a letter designation, as follows:

H: Fidelity of heredity
V: Abundance of variation
α: Competitive interaction with respect to reproduction
C: Continuity, or smoothness of the fitness landscape
S: Dependence of reproductive differences on intrinsic differences

The meaning of the first two will be pretty clear to everyone. The third, α, refers to exchangeability or replaceability; if your lineage could never displace mine, or vice versa, then we aren't part of the same population. The fourth, C, utilizing Wright's adaptive landscape metaphor, posits that small changes in phenotypes should lead to small changes in fitness. And the fifth, S, requires that there be some kind of causal link between the intrinsic traits of the individuals and their fitness.

We could still have selection evolution occurring without some of these. Make S zero and we have either drift or selective change due to another character. Make H zero and we can have selection, but no response to selection. Make C zero and we get selection that may lead to changes not clearly related to the selection. What Godfrey-Smith is after here is therefore not just selection, and not just evolution, but adaptive evolution – selection that leads to complex traits with the appearance of design. I think that is just the right focus for populations called "Darwinian".

Godfrey-Smith asks us to imagine scoring each of these five principles from zero to one, with high values usually being more Darwinian. When we combine these variables, they define a kind of Darwinian population space. "Paradigm" Darwinian populations tend to fall near the corner of the space where all the variables approach 1 (although, as Godfrey-Smith points out, it is possible to have too much fidelity of heritability).

Some will quibble that these variables are hard, or even impossible to measure, a point that Godfrey-Smith acknowledges. But I agree with him that this difficulty does not make them useless. The point is not measurement, although measurement would be nice, but to give us some tools to use in thinking about selection. Joan Strassmann and I have recently used a very similar strategy to think about what an organism is, with the two axes in our space being the amount of cooperation between the units comprising the organism candidate, and the amount of conflict among those units (Queller and Strassmann 2009; Strassmann and Queller 2010). One thing the exercise made clear to us is that there is no single clear-cut dividing line for organisms versus non-organisms. Paradigm cases grade into marginal cases that possess some, but not all, of the desired characteristics, or possess them to intermediate degrees. Godfrey-Smith's selection space makes an identical point about Darwinian populations.

This recognition of gradations, of partial or marginal cases, extends through the rest of Godfrey-Smith's book. The treatments of reproduction, of the individual, and of cultural memes all come benefit from this sensitivity to gradations and variation. This is partly because Godfrey-Smith is familiar with more than the canonical organisms, and can therefore consider the challenges posed by the different lifestyles of plants, fungi, slime molds, bacteria and viruses. But it also probably owes something to a certain openness to ideas. In contrast to some works that touch these themes, I do not sense a strong agenda behind Godfrey-Smith's work beyond the desire to make sense of the range of Darwinian phenomena.

If there is an exception to this openness, it is in Godfrey-Smith's rejection of replicators, or of the gene's eye view of evolution. This seems odd given the effort put into working out the kinks of the standard view. Why, I wonder, is an apparent stumble by Dawkins taken as fatal to his approach, when the rest of Godfrey-Smith's book is all about setting other stumblers on their feet? Dawkins, perhaps because of his own pugnacious style, doesn't get cut much slack by his critics. That's not to say that Godfrey-Smith puts no thought into these issues or that the treatment is completely onesided. There are arguments given and, even here, he points out some of the benefits of the replicator view he ultimately rejects. But, because I disagree with the conclusion, here is where I want to focus my attention.

Godfrey-Smith's first argument is that Dawkins' highly faithful replicators are not required for evolution. Imagine that organismal heredity is as it is – taller individuals tend to give rise to taller, and shorter to shorter – but that this rather sloppy heritability is not caused by genes or by any other faithfully replicating unit. Evolution and adaptation could still occur. On this point I agree with Godfrey-Smith. I have never understood why Dawkins could not view individuals as loose but good-enough replicators. They make copies of themselves, not especially faithful copies, but faithful enough that selection on parents produces some change in the offspring generation. Of course in our real biological world we know that this heredity is ultimately due to replication of genes, but in the universe of possible types of evolution, this need not be true. But conceding this point about immortal replicators does not necessarily reduce the utility of a gene' eye view in our biological world where genes are the foundation of heredity.

Godfrey-Smith's second argument is that Dawkins' replicator view leads to a kind of agential thinking that is dangerous. Whether we think about genes or individuals, it is possible to view selection as striving to maximize some criterion, specifically Hamilton's inclusive fitness. From there it is a short step to thinking of genes as agents, making decisions that are optimal for their own perpetuation, as if they have purposes, strategies, and agendas. In reality of course, genes don't strategize; selection simply acts to increase the frequency of genetic alleles that act as if they made the right strategic decisions. This is a metaphor, and as with all metaphors, the question is whether it informs more than it misleads.

Godfrey-Smith's main evidence that it misleads is Dawkins' view, discussed above, about the necessity of long-live replicators. Assuming that Dawkins' view was indeed due to agential thinking, it does not follow that there is any necessary connection. I use agential thinking all the time, and yet I can still side with Godfrey-Smith on the point that highly faithful replicators are not, in theory at least, an essential feature of all selection and evolution.

Godfrey-Smith further criticizes agential thinking for leading to particular ways of thinking, such as teleology. But none of this is new or unique to genic replicators; long before Dawkins taught us to imagine genes in this way, biologists have used optimality and agential metaphors at the organismal level. Agential thinking is in fact much more deeply rooted at the organismal level, so if it is pernicious, one could argue that it might be better for us to shift our thinking down to the genic level. So is there something especially pernicious about a genic agent metaphor? Godfrey-Smith thinks that genic agents are worse than organismal ones because there is something particularly misleading yet psychologically attractive about tiny agents – whether imps and ids or genes and memes. These agents are viewed as undermining our true designs, and thinking about them engenders a kind of "Darwinian paranoia". In contrast, larger agents – gods, Gaia, and groups – are more associated with cooperation and mutual benefit. But it can't really be true that the former are more seductive; gods seem to have insinuated themselves into our psychology much more deeply than imps have. Instead I suspect the real reason is not the seductive appeal of the message, but the message itself: genic agents are seen as orchestrators of bad behavior. But selfish genes do not necessarily build selfish organisms, and even if they did, the better response is to point to the chasm between "is" and "ought". Behavior, whether animal or human, is what it is, a complex mixture of both selfish and cooperative elements, and this mixture isn't determined by the theory or metaphor we use. Instead it is the task of the theories to account for this full range of behavior. Selfish behavior is rather easy to explain. But it was W.D. Hamilton's thinking about genic effects that gave us our first rigorous way of understanding selection for altruistic behavior (Hamilton 1964).

At any rate, I would want to see a much deeper analysis of the flaws of gene agential thinking, and probably one that goes beyond psychology, before I would give it up as a working biologist, because I can attest to its utility in generating hypotheses. It is much easier to think about "agents" maximizing their spread than to think out a population genetic model. Quick – will selection favor an allele that gives one offspring to a full sibling while costing three offspring to a half sibling? Hamilton's inclusive fitness rule provides an immediate answer, whereas a population genetic model is going to take you a while. You can always use the optimality thinking first and then check the reasoning with a population genetic model, but in practice that may not be necessary because the metaphor is good, or its limitations well understood.

In saying that the full model is not necessary, I do not mean that viewing genes as agents is a simple way to get every evolutionary question right. There are at least two major reasons why it isn't. First, optimization models are used to address questions of adaptation, but the full evolutionary picture is broader than that. Optimization won't help you much with genetic drift or other non-adaptive processes; drift doesn't have anything like an agenda. So, in using the agent metaphor, one should keep in mind that adaptation may be compromised if the population is sufficiently small or if mutation rates are huge, or if there is migration from a differently selected population.

The second reason why the metaphor may not produce proper results is the body in the attic. The ways that genes and environment produce phenotypes may be more complex than our strategic thinking allows. Suppose, for example, we study selection on humans in Africa. Resistance to malaria is good, so it ought to spread. Maintaining nicely rounded functional blood cells is also good. But in fact we find a lot of people who are not resistant to malaria, and a lot of others whose blood cells sickle under oxygen stress. You can think about optimal strategies until the cows come home and you won't get these two traits right unless you understand that they happen to be tied together genetically. But this is really a failure of knowledge, not a failure of the metaphor. Once we understand that, at the sickle cell locus, one homozygote gets malaria and the other gets sickle cell disease, whereas the heterozygotes suffer from neither defect, we can think correctly about gene agents. Each allele experiences two broad genetic environments, where the relevant environment is its allelic partner. When its environment includes another copy of its own allele, it does poorly; when it includes a copy of the other allele, it does well. It is then easy to reason out the best strategies, which depend on the frequencies of the two allelic environments, not to mention the malarial one. If we do not have that knowledge about how those alleles translate into phenotypes, any method will get the answer wrong, including a full population genetic model. Burying the body in the basement is no better than stashing it in the attic.

What this case really illustrates, I think, is that if you are going to use strategic thinking, it can be *better* to do it at the level of genes, not worse as Godfrey-Smith argues. It seems to me that you cannot apply agential or strategic reasoning to this case at the level of individual organism. Is there any way to view this as optimizing some individual property?

Though such cases exist, what makes optimality arguments so useful is that they allow us to reason about adaptation even when we don't know the details of how genes translate into phenotypes. Traits like sickle cell may fool us, but what is remarkable is how often optimality methods seem to work.

But these two issues – Dawkins' view about immortal replicators, and agential thinking – are something of a sideshow. Godfrey-Smith's book is mainly about Darwinian populations and the individuals that populate them, so the most relevant issue about the gene's-eye view is how it might fit into this scheme. Godfrey-Smith says that it fits badly, that genes are at best marginal Darwinian individuals, except perhaps in special cases where truly selfish genes operate against the interests of the organism in which they reside. Some of the reasons hinge on Dawkins' view that evolution requires highly faithful replicators and, as I don't share that view, I need not address these arguments.

One possible general response is to argue that, as Darwin knew nothing about genes, any scheme that adheres closely to his initial organism-centered theory might not be expected to be easily transferable to the genic level. Perhaps a gene's eye view requires a view of selection formulated in a radically different manner. But I am going to take the opposite tack and argue that Godfrey-Smith's framework is applicable, at least in some form.

So let's work a little with the idea of genes as Darwinian individuals in Darwinian populations. The obvious place to start is with Godfrey-Smith's criteria. First, do genes have heritability? Yes, much more strongly than individuals. Second, are they variable? Yes. Third, do they reproductively compete with each other? Yes. Fourth, do they have smooth fitness landscapes? Yes, to a similar degree as the individuals they are a part of, since the individual fitness landscape derives from its genes. Finally, do their reproductive differences depend on intrinsic character? Yes, alleles cause effects that are selected.

So the problem with genes appears not to lie in the five criteria for Darwinian populations, but Godfrey-Smith sees several other issues. First, he argues that genes are uncountable, which means their populations can't really even be specified. I disagree. Godfrey-Smith worries about a larger total gene count, both because cell number, and hence DNA copies, may vary across individuals, and also because each individual has many genes. On the cell number point, one could do the accounting at the cell level, with genes in most cells maximizing their inclusive fitness by helping other cells, but it is often easier in clonal multicellular organisms to view all copies of an allele as a single unit, which gives the population count for alleles in a diploid population as 2N. With respect to whether we should count all genes in the genome, I believe we should not. Each locus can be viewed as belonging to its own separate Darwinian population, with the other loci forming a part of its environment. Against this, Godfrey-Smith notes that some genes, especially transposable elements, can copy themselves to other loci. But the way to view this, especially for the great majority of genes that duplicate only rarely, is that the migrating genic individual has entered a distinct population. This meshes perfectly with Godfrey-Smith's α-requirement for Darwinian populations – competitive interaction with respect to reproduction. Once an allele copy jumps, it is no longer a competitor with other alleles at its original locus, and therefore not a member of that genic population. There is the possibility that is might jump back, but this is exceedingly rare for most genes. In other words, we can treat migrating genes much like we treat migrating individuals. Any problems of migration of between populations are probably more pervasive at the level of individuals, but present no insuperable barriers to using population concepts.

Assuming we have solved this counting problem, there remains the issue of defining the entity to be counted. The word "gene" has notoriously been used for many purposes. I think the meaning that matters most here a population geneticist's gene: a site that varies, especially if it varies in a way that makes a phenotypic difference that selection could work on. Let's make it the minimal such site, which will often be a single nucleotide difference, unless it is a larger deletion or insertion.

This decision leads to some nice consequences, and also some peculiar ones. The genome is seen to consist of a very large number of genic populations. It is nice that we have pretty clear borders on our gene, and that they are essentially non-overlapping, a problem with the alternative view of the gene as a larger unit held together by sufficiently low recombination. A stranger consequence is that some regions of the genome, those that do not currently vary, do not form good Darwinian populations at all (low V in Godfrey-Smith's scheme), though they may become good populations the instant that mutation produces a new allele. This seems odd, but I see nothing illogical with this view.

Since I am not defending the immortal replicator view, I need not be concerned about how much recombination will maintain an immortal piece of DNA. But recombination, or lack of it, does raise a related issue. My minimal genes occur linked to other minimal genes, and this can be seen as raising a problem with respect to Godfrey-Smith's S-criterion – the dependence of reproductive differences on intrinsic character. If my focal locus has a small effect on fitness, its best allele may lose out because of an unfortunate linkage association with a bad allele at a more important locus. Thus the reproduction of alleles at the focal locus depends on the extrinsic qualities of other loci. At the extreme, in a strictly asexual species, all loci are linked, and at any given time a single locus may dominate selection. One way to view this is that only the dominant locus is a good genic population during that time and the other loci are temporarily out of the game. Another way, which I think I prefer, is to recognize that the other loci are actually still making their own intrinsic contributions to reproduction, and are still respectable Darwinian individuals subject to selection, but that they suffer unfortunate associations, which can be considered as part of their environment. An analogy at the individual level might be associations with location or habitat in metapopulations. Some individuals are born into "sink" populations, which may persist for considerable time, but do not contribute much to the larger metapopulation. Individuals in those sink populations are not really non-Darwinian, but they are doomed by their association with an aspect of their environment, even if they might otherwise be as fit or fitter than the fortunate individuals in source populations.

A final problem with viewing genes as Darwinian individuals, according to Godfrey-Smith, is that they are "scaffolded" reproducers; they cannot reproduce on their own but instead require the cellular machinery from elsewhere. Viruses are a familiar example. He acknowledges that all individuals acquire materials and energy from elsewhere, which could seem to make them scaffolded reproducers, but argues that there is something special about using some one else's reproductive machinery. But is there really? Consider a cowbird that lays its eggs in the nests of another species, and requires the foreign parents to successfully raise its young. It has scaffolded a major part of its reproduction. Does this mean it is significantly less Darwinian than a falcon which depends on other species of birds for raising its young, but in a more conventional gastronomical fashion? Similarly, figs depend critically on fig wasps for their reproduction, and vice versa, but both form perfectly good Darwinian populations that evolve complex adaptations.

One might argue that these still retain their truly fundamental cellular reproductive machinery, and that's true, but let's do a couple of thought experiments to see whether that should be considered essential. First, recall Godfrey-Smith's thought experiment, used against Dawkins' replicator view, concerning hypothetical organisms that somehow reproduce without genes. If that's fair game – and I have granted that it is – then we should be free to also consider a hypothetical organism that somehow reproduces without cellular machinery. Would it not still be perfectly Darwinian, assuming it met the other standard requirements? Second, imagine an organism that, instead of breaking down its food to make building blocks, is able to directly incorporate and use components of its prey's reproductive machinery, for example ribosomes or centrioles. I don't know if this ever happens in reality, but that is not the point. The question is whether such a scaffolding of reproduction would result in significant loss of evolutionary potential. I think the answer is no, except for the evolution of the captured organelles themselves, which will continue to evolve to benefit members of their home species. Actually there is an extremely common and important case that approaches this. In sexual species, male reproduction typically relies very heavily on the cellular reproductive machinery that females have sequestered in the egg. Yet, while their parasitic nature has not escaped notice, no one has suggested that males are illegitimate Darwinian individuals. In short, while scaffolded reproduction is an interesting concept, I don't think it separates the Darwinian from non-Darwinian. I am happy to install viruses and genes alike in the Darwinian pantheon, perhaps in a dedicated impish chapel.

That something is awry with the various reasons given for genes not being Darwinian individuals seems evident from Godfrey-Smith's acceptance of a special case: selfish genetic elements that act against the interests of the rest of the genome. These must be accepted, because they evolve in ways that cannot be attributed to the higher-level unit of individuals. Yet, they partake of all of the supposed non-Darwinian sins of other genes. Their borders no better defined (as recombinational units). They are no easier to count (harder, in the case of transposable elements that show considerable horizontal transmission). They are scaffolded reproducers, fully dependent on cellular machinery not of their making. They fall into about the same region in the state space of Darwinian populations as other genes. What really separates them from other genes is not anything on the list of important Darwinian properties, but the additional simple fact that they *cannot* be folded into a strictly organismal view. To me, that constitutes grounds for *accepting* other genes as Darwinian individuals as well.

In short, though I may appear to be criticizing, I am actually arguing that Godfrey-Smith's scheme, which gives us interesting ways to think about selection, may be broader than he thinks it is. It gives us interesting ways to think about genic selection too. I think that is almost inevitable, given the symmetry in the two ways of looking at selection. The organismal one puts selection on the ground floor and the genic one puts heredity there, but each requires that genes get built into organisms. Godfrey-Smith's suggestion that "a lot of talk about the evolutionary role of genes is really talk about organisms, genetically characterized" (p. 131) is true enough, but one could equally say that a lot of talk about the evolutionary role of organisms is really talk about genes, organismally characterized. Perhaps a better integration of these two views will come only when the mechanistic biologists get their act together and bring the body up from the basement, or down from the attic, to the main floor.

References

Dawkins R (1976) The Selfish Gene. Oxford University Press, Oxford.

Dawkins R (1976) The Extended Phenotype. Oxford University Press, Oxford

Godfrey-Smith P (2009) Darwinian Populations and Natural Selection, Oxford University Press, Oxford.

Hamilton WD (1964) The genetical evolution of social behaviour. I-II. J Theor Biol 7:1-52.

Queller DC Strassmann JE (2009) Beyond society: the evolution of organismality. Phil Trans Roy Soc 364:3143-3155. The social organism: congresses, parties, and committees. Evolution 34:605-616.

Agents and Acacias: Replies to Dennett, Sterelny, and Queller

Peter Godfrey-Smith

The commentaries by Dennett, Sterelny, and Queller on *Darwinian Populations and Natural Selection* (*DPNS*) are so constructive that they make it possible to extend and improve the book's framework in several ways. My replies will focus on points of disagreement, and I will pick a small number of themes and develop them in detail. The three replies below are mostly self-contained, except that all my comments about genes, discussed by all three critics, are in the reply to Queller. Agential views of evolution, discussed by Queller and Dennett, are addressed in my reply to Dennett.

Dennett

Dennett locates two main points of disagreement. One is over the role of information in biology, the other over concepts of agency. I'll begin with agents.

DPNS opposes a view of evolution which sees it as a contest between agents pursuing goals. The opposition here is not to selection, adaptation, or optimality, but only to the idea that evolutionary processes themselves are contests between agents. Dennett and Queller both see themselves as defending the agential approach. Some of their arguments are defenses of optimality and adaptation, which are not under attack, but the book was probably not clear enough about what the target is. Once the target has been made clear, I think that Dennett will want to defend it. So clarification does not leave everyone on the same page. Dennett's insistence that a "Cui bono?" question is "central to all evolutionary thinking," for example, is something I am against.

The agential view is related to several other controversial forms of description in biology. One is the idea that organisms are akin to products of design, akin to artifacts (Lewens 2004). Another is a picture of organisms as adapted to their environments in a lock-and-key manner (Lewontin 1985). A third is the view that organisms will approximate optimal solutions to problems posed by their ways of life (Grafen 2007). These ideas all make subtly different contributions to biology, and probably have different mixes of good and bad effects. The view that is opposed in *DPNS* is the view of evolutionary processes themselves as contests between underlying agents that persist through the material flux. This sort of agential description shades into other members of this family, and into ordinary selectionist thinking.[3]

One possible defense of any of these views is purely pragmatic: "this method has worked so far, so let's keep using it." One cannot give a principled argument against a heuristic understood in this way. If the results are good, that is all that matters. Both Dennett and Queller do appeal to pragmatic arguments, and a response would need a thorough accounting of good and bad outcomes in empirical work guided by the heuristic, with comparison to results gained through alternative methods. *DPNS* does not, as Dennett says, contain such an accounting.[4] But I think the agential perspective is usually seen as closely connected to theoretical considerations, including foundational work. Then more general arguments can be given against it.

To see the distinctive role of agential thinking, we need to look for cases where it takes us beyond other kinds of selectionist thinking and makes its own contribution. Here is an example. *DPNS* discusses the evolution of altruistic and cooperative behavior. One summary idea I accept is that costly forms of altruism can survive if the benefits of altruistic acts fall mostly on those with a tendency to pass the behavior on. Suppose this general idea is embedded within an agential view of evolution. Then the idea becomes something like this: altruistic acts can survive in an evolutionary context because there is

---

[3] Dennett says "Ironically, PGS himself finds it irresistible to engage in agential talk when explaining and defending his own best ideas," but I think Dennett casts this net too broadly, and includes adaptationist description and sometimes ordinary causal talk. Here I have inserted comments after his italicized examples. "Here are a few examples among many, in PGS's discussion of de-Darwinizing: "Their independent evolutionary *activities* are c*urtailed, constrained*, or *suppressed* by what is happening at the higher level" (p. 122), [ordinary causal description, with some metaphor, but not agential in the relevant sense] "Another way of dealing with this *problem* [point granted] would be for one member of the collective to *prevent* [ordinary causal description] reproduction altogether by other individuals . . . " "This can be put more explicitly in terms due to Calcott (2008). For a transition to occur there must somehow be both the *generation of benefit* and the *alignment of reproductive interests.*" (p. 124) [point granted, and I do think Calcott's terms are helpful here] So Calcott is praised for putting it "more explicitly" instead of condemned for indulging in such metaphorical agential talk."

[4] Dennett compares the accounting project in Francis (2004), which I think much more highly of than he does.

an underlying selfish beneficiary which persists and hoards the benefits of these acts. The existence of this beneficiary answers Dennett's "Cui bono?" question.

But altruistic behavior can persist without an underlying selfish beneficiary. An illustration is seen in a recent paper by Jeff Fletcher and Michael Doebeli (2009). Assume haploid asexual organisms varying at two loci at which there are alleles causing production of a useful chemical which is made available to others in the neighborhood. The producer alleles are A and B, with a and b as non-producing free-riders. Assume that the AB combination is lethal, as the two pathways that produce the chemical interact destructively. So the population has ab, Ab, and aB individuals. Ab and aB individuals pay a cost for producing the public good. But assume assortment of types, such that Ab and aB interact with each other and ab meets itself. A and B will both proliferate, even though "the help which carriers of A provide never goes to other carriers of A and the help that carriers of B provide never goes to other carriers of B" (Fletcher and Doebeli 2009, p. 16).[5] If a person thought that for altruism to survive, there must be a selfish underlying entity, here they would encounter no selfish beneficiary, and would expect altruism not to survive. But assortment of the two altruist types solves the problem.

A response might be that this is a case of "reciprocal altruism" (Trivers 1972), of the sort also seen in symbioses. This is often regarded as part of the same package of ideas as the agential approach (Dawkins 1976). Here we again reach the problem of working out the distinctive contribution of agential description as opposed to its relatives. I see reciprocity as part of ordinary selectionist thinking, not the special kind which involves the introduction of an underlying agent. A person might also say that both A and B are selfish beneficiaries, or the partnership of A and B is. But to say the former is just to re-express the fact that both A and B do well. To call them a "partnership" is to add nothing that is true to the claim that there is assortment in the population that brings A and B together and each does something that benefits the other.

Turning to the second of Dennett's main themes, Dennett advocates an information-based approach to many issues in biology. Via George Williams, he holds that genes are not "strands of nucleotides" but "the information carried" by those material

---

[5] I follow Fletcher and Doebeli's description here though talk of alleles at loci is problematic in the absence of sex.

objects. This affects his view of reproduction, inheritance, and evolutionary processes. In contrast, I do see genes as strands of nucleotides. More exactly, *if* genes are real enough to be Darwinian individuals, then it is strands of nucleotides that are Darwinian individuals – the status of genes as objects will be discussed in my reply to Queller. When one gene is a copy of another, that is a similarity between two separate Darwinian individuals. The same goes with cells, organisms, and reproducing groups. As Darwinian individuals, these are all particular material things. Organisms and cells are not abstract, not made of information, and genes are treated the same way.[6]

Dennett regards reproduction as an informational matter. He gives an argument in which ordinary human reproduction is compared with an artificial case. Suppose we sequence two human genomes, scramble their sequences in a meiosis-like manner, and then construct a physical genome with the results and use it to create a new person. For Dennett, the two people with the genomes we started with are parents of the new individual. I accept this conclusion (adding that the egg provider is a parent too). I also accept some informational or quasi-semantic descriptions of gene action. There is a reasonable sense in which genes do *code for* proteins. But reproduction *per se* is not an informational matter. Dennett thinks that once one accepts that "'artificial inheritance' can be accomplished by interposing an information-processing device (and nothing else) in the normal process..., that is as good as a proof that the normal process of reproduction is fundamentally an information-transmission process." I reply that the fact that something *can* be achieved with the aid of information-processing machinery is not proof that the task is intrinsically an information-related one. We might use information-processing machinery to do something that in other cases has a different character.

There is a weak sense of "information," used in information theory, in which anything in the world that can take on various different states is a *source* of information, and any other variable whose state is correlated with the first one *carries* information about it. In this sense, clouds carry information about impending rain and footprints carry

---

[6] Sometimes genes *seem* to be more abstract than this because they are not being treated as Darwinian individuals, or other objects, at all. Genetic *properties* are often used to characterize organisms. When this is being done, organisms are being treated as particular objects, perhaps as Darwinian individuals, and genetic features, such as *being AA*, are used to identify a similarity between them. This is not treating genes as Darwinian individuals.

information about who has walked by. That does not make clouds into informational objects, abstractions as opposed to collections of water droplets, or even into messages. The same applies to genes. Due to the biology of inheritance, parents carry information about features of their offspring. Inheritance is an informational matter in *that* sense, but so is almost any other causal process.

How might we tell whether inheritance is informational in a stronger sense? Dennett, I as understand him, thinks that certain kinds of results can only be achieved by the transmission of some kind of representation. That I take to be part of the moral of the story he tells about spying and submarines.

> We learn that our enemy has somehow *obtained information* about the design of our new submarine. Did a spy copy the blueprints onto pieces of paper and smuggle them across the border, or encode a recipe for the design into patterns of dots and dashes that were sent by radio, or memorize the blueprints and walk across the border, or . . . . .? We might be able to discover that the information moved from Bethesda to Beirut to Bern to Bonn to Baghdad and still not know what physical medium or what code was involved in each leg of the journey.

Dots and dashes are one way to do it, but there may be others. Suppose someone knows that our submarine's design is strongly constrained by a particular technical problem. Many other features will be determined by the way it is solved. Their agent might steal blueprints, but might instead smuggle out a crucial *piece* of the structure, which can be used to infer the whole. ("This looks like a triple-ported Aston-sprocket... but it's been modified... Aha!") In one sense the agent has provided information to our enemy. He has made it possible for them to "re-create structure." But he did not do that using a representation or "code".

Dan Sperber and Deirdre Wilson, in *Relevance* (1986), their classic in the philosophy of language, argue that work on language has been dominated by the "code model," in which everything that a hearer gets out of a sentence must have been encoded inside it. People tend to assume that this is the only format that is even possible for a theory of language. But this is not the only option. Communication sometimes works by speakers making an inference problem more tractable for hearers. Sperber and Wilson think this is distinct from a code-based explanation (though often working alongside it). In an earlier paper (2000) I imagined a biological inheritance system in which samples of proteins, rather than DNA specifications of proteins, were sent over generations. The offspring runs off new proteins from the samples received from the parents. The device-smuggling case above, the inheritance thought experiment in my earlier paper, and the inference-based linguistic phenomena described by Sperber and Wilson all have in common that they are ways of achieving a result that *can* be achieved by sending encoded messages, but *need* not be achieved that way.

Dennett regards me as a "puritan" about information. History certainly shows how hard it is to avoid being either puritan or libertine.[7] I think that recently it has started to become clear how one might make good naturalistic sense of a large family of semantic phenomena. I have in mind the development of what can be called "sender-receiver" models (Lewis 1969, Millikan 1984, Skyrms 2010). The aim is to give a theory of the ways in which signs operate in the coordination of two kinds of entities, senders and receivers (in broad senses of those terms). Not every sender-receiver interaction involves representations, as the examples above show, so more has to be said. But even when those problems have been solved, this model brings with it a constraint: a sign is only as real as its reader or user. This applies to internal representations as well as to external ones.

A relevant discussion of this constraint can be found in an earlier paper by Dennett, written at a crucial stage in the history of cognitive science.

> The only psychology that could possibly succeed is neo-cognitivist, which requires the postulation of an internal system of representation. However, nothing is intrinsically a representation of anything; something is a representation only *for* or *to* someone; any representation or system of representations requires at least one *user* of the system who is external to the system. Call such a user an exempt agent. Hence, in addition to a system of internal representations, neo-cognitivism requires the postulation of an inner exempt agent or agents – in short, undischarged homunculi. Any psychology with undischarged homunculi is doomed to circularity or infinite regress, hence psychology is impossible. (Dennett 1978, p. 101).

---

[7] There is a sense in which Dennett is both at once, given his interpreter-dependent treatment of semantic properties.

Dennett's response to this dilemma is to accept that "perhaps the *prima facie* absurd notion of self-understanding representations is an idea whose time has come, for what are the "data structures" of computer science if not just that: representations that understand themselves?" (p. 102). I think this is the wrong interpretation of the message of computer technology. In standard computers the distinction between representation and reader is present in the hardware of the machine, in the distinction between memory and processor. The message of computer science for psychology is not the possibility of self-reading representations, but the power of simple read-write devices attached to an organized memory (Gallistel and King 2010).

This framework can also be employed when thinking about genes as carriers of information; this case, too, can be addressed within a sender-receiver model. That is being attempted by Shea (2007) and by Bergstrom and Rosvall (2011). One problem (which I outline in my commentary on Bergstrom and Rosvall, 2011) is that identifying a receiver for a genetic message when describing inheritance across generations is not straightforward. If that view can be defended, genes as they operate in inheritance would be information-carrying devices in the stronger sense that fits a sender-receiver model. If not, the language of "information" could still be applied to genes and the relations across generations, but only in the weak sense of this term that can be used in the description of any causal process.

Sterelny

DPNS contrasts "classical" and "replicator" analyses of natural selection. The classical view holds that evolution by natural selection will take place in any population in which variation, heredity, and fitness differences are found (Lewontin 1970). The replicator analysis holds that two roles must be occupied in any evolutionary process. There must be *replicators*, things which are copied in a reliable way, and *interactors* (or *vehicles*), things which interact with the environment in a way that determines the rate at which different replicators are copied (Dawkins 1976, Hull 1980). On the classical view there is no such bifurcation of roles. There is just one kind of entity needed, which I call a "Darwinian individual." These are the things which exhibit variation, heredity, and fitness differences.

I criticized existing versions of both analyses, but then built on the classical view. Sterelny thinks that the replicator/vehicle model is better than I realize. It need not be seen as a complete account, but the framework of *DPNS* allows for a pluralist approach, in which some cases are handled with a refinement of the classical view, and others are handled with the replicator/vehicle view. (I will mostly follow Sterelny's terminology here in talking of "vehicles" rather than "interactors," though I think Hull's concept and terminology are better.)

In *DPNS* the attention was on the replicator side of the Dawkins-Hull framework; once we understand why replicators are not needed, there is no motivation for the bifurcation of roles that introduces a vehicle or interactor concept. Sterelny thinks that I underestimate the importance of both replicators and vehicles. My reply to Queller will discuss replicators. In this reply to Sterelny I look at the vehicle side. Sterelny argues that we need this concept because of the evolutionary role of various entities that are not Darwinian individuals, but do fit the concept of a vehicle well.[8]

Sterelny holds that "extended phenotype" cases are the source of the problem. The cases on his list that I see as especially important are those that involve cooperation between members of different reproducing populations. One example is the symbiotic association between ants and bullhorn acacia trees. The trees build hollow structures which house ants that guard the tree. In some cases the trees also makes structures that feed the ants. Sterelny claims that the combination, tree-plus-ants, is an adapted unit, built by evolution. But these ant-tree combinations are not Darwinian individuals. They do not stand in parent-offspring relations, except perhaps in a marginal sense. Trees have offspring trees, and ants (and ant colonies) have offspring ants (and colonies), but an ant-tree combination does not reproduce as a unit. It might sometimes be that the ants in a particular tree are the descendants of ants who lived in a parent of that tree, but that would be accidental. The most we can apparently say is that one ant-tree combination, X, is the offspring of another ant-tree combination, Y, if either the ant part or tree part of X (or both) is the offspring of the ant part or tree part of Y. Even this makes the association sound tighter than it often is, as one colony may be supplanted by another and in some

---

[8] A view of this kind is also being developed by Austin Booth (forthcoming).

cases colonies of more than one species may occupy a tree at the same time (Wheeler 1913, McKey 1984, Palmer et al. 2000, Raine et al. 2004).

Another relevant case here is the squid-vibrio symbiosis, in which colonies of bacteria provide luminescence for bobtail squid (Nyholm and McFall-Ngai 2004). The initial bacterial colony is picked up from the seawater while the squid is young and generally contains a half dozen to a dozen cells.[9] So as with acacias and ants, one squid-vibrio combination is not the parent of another, though squid beget squid and vibrios beget vibrios.

Sterelny holds that "if we recast these evolutionary episodes into the framework of Darwinian Populations, we have to choose between two suboptimal representations." One option tracks the evolution of the clear Darwinian individuals – the ants and the acacias – and does not treat the combination as an entity evolving through natural selection. The other option is to treat the combination as an evolving unit. Then the problem is that we have a complex and adapted structure produced by a marginal Darwinian process. "Novelty and complexity is sometimes not novelty and complexity in a paradigm Darwinian individual. The replicator-vehicle model captures that fact."

I agree that these cases raise special issues and an advocate of the replicator-interactor view has what looks like a good way of describing them. An ant-acacia combination can be seen as an interactor or vehicle. Vehicles need not be reproducers in their own right. They may arise from the combined actions of members of many replicator lineages, and can serve replicators of several kinds. How might such cases be handled within the Darwinian populations framework?

Focusing on the ant-acacia case, I choose the first option Sterelny offers. Ants are Darwinian individuals, as are ant colonies. So are acacias. Evolution has shaped both sides in such a way that ant-acacia combinations are reliably produced, even though there is not *re*production by ant-acacia combinations. We have an adapted unit that was not produced by a Darwinian process acting on those units.

---

[9] Ruby and Asato (1993) did find that a single bacterium can suffice to establish the symbiosis in a young squid. Wollenberg and Ruby (2009) conclude that the usual number is six to twelve, one or two for each of six "crypts" in the developing light organ.

In *DPNS* I distinguished *origin explanations* and *distribution explanations*. When we give a distribution explanation we assume the existence of a set of variants in a population, and explain why they have the distribution they do. An origin explanation is directed on the fact that a population has come to contain individuals of a particular kind *at all*. In the light of Sterelny's commentary, it is possible to distinguish two kinds of origin explanations that are relevant in an evolutionary context. I will call these *direct* and *co-evolutionary* origin explanations. In the direct case, a single Darwinian population changes over time in such a way that new kinds of individuals arise. In a co-evolutionary case, the entities whose origination we are explaining are not Darwinian individuals at all (at least at first). Rather, two or more Darwinian populations produce Darwinian individuals which interact in some new way that yields the combined or collective entities we are explaining.

This process may involve the evolution of new kinds of individuals within those Darwinian populations, or it might merely involve changes in the distributions of features in those populations, where these changes in distribution lead to novel combinations of entities, drawn from each population, becoming likely to appear.

In Sterelny's ant and acacia case, the unit we encounter is the product of evolution in at least two different Darwinian populations, acacias and ants. (I will be vague from here about whether ants, ant colonies, or both are the relevant Darwinian individuals on the ant side.) No direct origin explanation of the kind discussed in *DPNS* is possible. But a co-evolutionary origin explanation can be given. Explanations can be given separately for the ant traits and acacia traits that, when combined in the right way, give rise to the ant-acacia combination. In these explanations, each population operates as part of the environment for the other.

A co-evolving conglomerate at one time can give rise to a Darwinian individual at another. A clear example is the eukaryotic cell, and a less clear example is lichen. This could happen, in principle, with the ants and acacias. Suppose the ants in an ant-acacia complex start to locate and protect the offspring trees of their current mature tree, detecting these offspring trees by chemical means. When the new tree is large enough, an offshoot of the mature ant colony moves in. This would be the beginning of an "egalitarian transition," in Queller's terminology (1997). (This would sound completely far-fetched if ants were not doing equally far-fetched things already, such as herding aphids for food and grooming the aphid eggs in their nests: Matsuura and Yashiro 2006). If the two parties became tightly fused in a way that included mutual reproductive dependence, then when the combination continued to evolve new traits, the origin explanations that would be given would be direct rather than co-evolutionary. We might expect a system like that to produce more adaptive complexity than the current ant-acacia arrangement. Indeed, some ants are not very kind to their acacias: "*C. nigriceps* Emery... has the peculiar behavior at our study site of removing virtually all axillary buds except at swollen thorns, effectively sterilizing the tree" (Palmer et al. 2000).

So I accept that these cases initially seem to fit the replicator/vehicle model well. But the Darwinian populations approach, with additions prompted by Sterelny's objection, can handle them. What is the remaining difference between the analyses? One difference is that the replicator/vehicle analysis adds that an ant-acacia combination is a unit of selection, something that can do well or badly in a Darwinian context, and in the same sense that applies to sexually reproducing (but non-replicating) organisms like us. As in *DPNS*, I hold that it is a mistake to see things that do not reproduce as units of selection.

Once the distinction between direct and co-evolutionary origin explanations has been made, it starts to make an appearance in many different places. Another topic Sterelny discusses is the role of parameter *S*. This parameter represents the extent to which differences in reproductive output in a population are due to intrinsic differences between individuals rather than extrinsic ones. (The intrinsic properties of an object are those that do not depend on the existence and arrangement of other objects.) I argue that in paradigm Darwinian processes, *S* is high. Reproductive differences are due largely to intrinsic differences. For an example of a reproductive difference that is a low-*S* affair, consider two internally very similar individuals, one who happens to be struck by lightning while the other goes on to have a large family.

Sterelny has many doubts about the role of *S*. He is not even sure that "the concept of an intrinsic trait is well-defined," at least in some contexts where it matters:

> What of those traits whose development is richly dependent on environmental input? The food preferences of a rat, for example, depend on maternal imprinting and early experience, as well as the rat's own genetic endowment. So are preferences intrinsic?

In reply, a preference is an intrinsic trait. It might be a causal consequence of some features that are not intrinsic, and this may affect its heritability, but it is still intrinsic.

Another of Sterelny's examples is more awkward for me and leads to some interesting points. Consider a bird species with an elaborate courtship display that includes behaviors from both sexes. Sterelny speculates that the following might often be true: "Neither sex has a mental template of the whole routine [produced by that sex]. Rather, each bird has a linked set of responses triggered by the previous step of its partner." Suppose we have a case like this. Is the *ability to produce the male courtship display* an intrinsic property of a male bird? We might initially have thought that this was an intrinsic trait, but if the responses made by X's partner are wrong, X cannot perform the display. All X has is the ability to follow a set of conditionals: if I see move 1, I do move 2; if I see 3, I do 4. The other bird follows a different and interlocking set of rules. Provided that one side can get the process started, all the moves will follow in the right order. If it is true that the male has *the ability to produce the male display*, this is not true in virtue of the male's intrinsic properties.

What follows from this? Is it a problem for my view because the male's capacity is a *bona fide* trait that I must treat as questionable? I think, instead, that when we learn these things about the bird's display, it leads us to re-think our description of the trait and its evolution.

The total courtship sequence characteristic of the species, a process involving moves from both partners, is clearly something whose origin explanation is not direct. In that way it is like the ants and acacias. Evolution builds a pair of traits, one expressed in males and one in females. No individual produces both. When the two traits are expressed together, the result is the coordinated courtship display. One might think that at least *one* side of the birds' courtship display, the sequence of behaviors produced by one of the two sexes, is a trait that has a direct origin explanation. But in the case Sterelny is describing, each sex does not have the capacity to produce their side of the display without the prompting of each move by the partner. The temporal ordering of each partner's moves is dependent on the other partner. Then what we think of as "the male's display" does not have a direct origin explanation either; instead it is one side of a co-evolved unit that is not a trait of any Darwinian individual.

Queller

Queller looks closely at genes and the "gene's eye view" of evolution. He agrees that replicators of the sort Dawkins describes are not necessary for evolution by natural selection. For Queller as for me, a simple in-principle argument suffices: if reproduction exhibits some degree of heritability, there can be a response to selection and hence adaptation, whether or not the system contains replicators.

Queller also offers a reconciliation of the "classical" view (based on heritable variation) and the replicator view: "I have never understood why Dawkins could not view individuals [organisms] as loose but good-enough replicators. They make copies of themselves, not especially faithful copies, but faithful enough that selection on parents produces some change in the offspring generation." I don't think this move is easily available to Dawkins. In Dawkins' view of evolution it is central that there are persisting entities that compete with each other. Material objects come and go, but replicators may live on in the form of their copies. Whether or not reproduction is high-fidelity, this picture is much more natural if reproduction is asexual. Asexual reproduction creates a tree in which every individual is at the end of a unique path going backwards in time. When a parent has many offspring asexually, it is possible to regard each offspring as "more" of the stock of the parent. When reproduction is sexual, however, every individual is a meeting of two lineages – actually more than two, because multiple lineages came together in each of its parents. It is not at all straightforward to regard the new individual as a mere continuation, a case of more-of-the-same, with respect to *both* of its parents, if the parents themselves differ. So organisms like us cannot be "loose but good-enough replicators," to use Queller's phrase, unless the link between replication and persistence is broken. Organisms can be *reproducers*, however, because reproduction, in the sense used here, is not a matter of making a continuation of the parent, but making something new, and making something new can be a joint effort.

Genes are a contingent part of the evolutionary machinery, Queller accepts, but undeniably an important part, and the core of his commentary is an application of the framework of *DPNS* to genes. I argued in the book that genes are generally marginal Darwinian individuals. Queller is not convinced. He runs through the tests that I argue should be applied to any candidate Darwinian individual. Genes reproduce, and do so with high heritability (*H*). They vary (*V*). They engage in reproductive competition in the sense measured by α. Is their fitness landscape reasonably smooth (*C*), and are fitness differences due in large part to intrinsic features (*S*)? Yes in both cases. DNA sequences affect organisms' phenotypes, and genes' fitness landscapes should be smooth "to a similar degree as the individuals they are a part of, since the individual fitness landscape derives from its genes."

So far so good, but we have not yet worked out what genes in the relevant sense *are*. If genes are Darwinian individuals, we should know at least roughly where one ends and another begins. We should be able to work out, at least in principle, how many of these individuals are present in a given system. In *DPNS* I asked how we might determine how many individuals are present in a Darwinian population of human genes, and argued that there is no answer. There is a number of human nucleotides, a number of chromosomes, a number of cells, and of organisms, but no number of genes in the evolutionarily relevant sense. When I say there is "no number," I do not mean that no count can be given; the problem is more that too many can. Williams (1966) and Dawkins (1976) say that a gene, for evolutionary purposes, is any stretch of DNA which is subject to positive or negative selection pressure of a degree "several or many times its rate of endogenous change," where endogenous change includes both mutation and crossing-over. Some stretches of DNA will pass a "several" test but not a "many" test, and when the environment changes, shifts in selection pressure will instantaneously wipe out some genes and create others. Many genes will overlap, or exist inside others. A gene that codes for an important protein will contain many sites with slightly different degrees of selective pressure bearing on them.

Queller offers a response to this problem. The way to recognize genes as Darwinian individuals is to treat them, in general, as single nucleotides.

> The word “gene” has notoriously been used for many purposes. I think the meaning that matters most here a population geneticist’s gene: a site that varies, especially if it varies in a way that makes a phenotypic difference that selection could work on. Let’s make it the minimal such site, which will often be a single nucleotide difference, unless it is a larger deletion or insertion.

This resolves ambiguities over counting. Queller's genes have definite boundaries and no overlaps. His next move is to hold that separate Darwinian populations exist for each locus, where, again, a locus is the site occupied by a single nucleotide.

I will spend some time looking at the consequences of this view. I will set aside the role of insertions and deletions, and look at the possibility of viewing nucleotides as Darwinian individuals.

The first thing to note is that there are only four character states possible for members of each population: C, A, T, and G.[10] There are no traits of Darwinian individuals that are combinations of a number of variable characters. Mutation can replace a Darwinian individual of one kind with another, and the frequencies of the four types can go up and down, but no other kind of change is possible in such a population. This is Darwinian atomism.

DPNS distinguished between *origin explanations* and *distribution explanations*, where the former explain the appearance of new variants in a population and the latter explain why some variants are common and some are rare. In a genetic Darwinian population of Queller's kind, no origin explanations can ever be given, (except in the earliest stages). Once all four nucleotides are present, no new traits ever arise. Larger pieces of genetic material – such as cistron-sized pieces – can evolve in a gradual way, with each substitution building on its predecessors and new combinations arising from old. But those larger entities are made up of large numbers of Darwinian individuals, each located in a different population. In my reply to Sterelny, I added to the framework of *DPNS* by distinguishing *direct* from *co-evolutionary* origin explanations. The aim was to handle cases in which novel structures arise through evolution in several different Darwinian populations. Ant-acacia symbioses are examples. In Queller's framework,

---

[10] Or five if methylated C is a different option from unmethylated C.

almost all origin explanations are co-evolutionary; even the ability to code for a single protein requires the coordination of many Darwinian individuals.[11]

This does not mean that direct origin explanations for traits like eyes and immune systems cannot be given at all. They can be given by treating organisms as Darwinian individuals. Queller does not argue for an *exclusively* gene's eye view of evolution. Population genetics is very much a science of distribution explanations.

I do not see the points above as arguments against Queller's view yet, especially given that the deficiencies in a gene-level description can be made up by moving to organisms. The next issues I will raise are more of a problem. At an initial stage in Queller's discussion he said that genetic Darwinian individuals will often have reasonably high scores for *S* and *C*. But his decision to identify genes in most cases with single nucleotides changes this. Queller notes this fact himself for the case of *S*. When one genetic individual reproduces and another does not, this may be partly due to an intrinsic difference between them – being A rather than T – but a huge number of realized fitness differences will be due to extrinsic factors. Setting aside "environmental" factors in the usual sense, whenever one individual does well and one badly because of what is going on at other loci (including the nucleotides next door), realized fitness is affected by extrinsic matters.

We also need to look again at *C* (for "continuity," which I italicize here to distinguish it from the C nucleotide). *C* is a measure of whether small changes to individual character lead to small changes in fitness. Queller said initially that genes will often have respectable *C* scores, but this changes once genes are the size of nucleotides. Then it is hard to even ask the question about continuity, because a genetic population has just four kinds of individuals, with no intermediates between the different states. There is one continuity-related feature that might be recognized. The purines (A and G) could be seen as more similar to each other than each is to the pyrimidines (C and T). Pyrimidines would also be internally similar. Whether this has consequences for fitness

---

[11] I am not sure how Queller intends his analysis to apply to asexual organisms, such as bacteria, where the entire genome is the usual unit of replication. Here I assume there are not separate Darwinian populations for each nucleotide-sized locus, as facts about competition at a locus are dependent on meiosis.

depends on the structure of the genetic code and on what various proteins do. Suppose transitions (moves between purines, or between pyrimidines) generally have fewer phenotypic consequences and hence fewer fitness consequences than transversions (movements between a purine and a pyrimidine). That would make for some degree of continuity. But that seems to be just about the only question that can be even asked about *C* for this Darwinian population. This is because of the atomism of Queller's version of the gene's eye view.

I will address two other topics in this area. One is the only issue on which there is some misunderstanding. Genes, it is clear, are scaffolded reproducers. Their reproduction depends on much machinery that is external to them. Queller attributes to me the view that cases where reproduction is scaffolded are "less Darwinian," and he disagrees with this. That claim about scaffolded reproduction is not part of my view, however. HIV is a scaffolded reproducer, for example, and it is a paradigm evolver (p. 152). Scaffolded reproducers are special because their reproductive capacities have a kind of fragility. But that does not make their evolutionary processes marginal when the scaffolding they require is present.

Finally, Queller notes that I accept a Darwinian description of gene action for some special cases: selfish genetic elements.

> Yet, they partake of all of the supposed non-Darwinian sins of other genes.... They fall into about the same region in the state space of Darwinian populations as other genes. What really separates them from other genes is not anything on the list of important Darwinian properties, but the additional simple fact that they *cannot* be folded into a strictly organismal view. To me, that constitutes grounds for *accepting* other genes as Darwinian individuals as well.

This is indeed a problem. If these special cases are properly described in terms of selection on genes then they should be different *in their own right* from the ones where I resist the gene's eye view in favor of an organism-level description. The question should not be which description works "best," so that some cases receive a gene-level description because rival descriptions are especially bad. The question should just be how well genes fit the Darwinian criteria in each case.

Queller's objection might be met in at least some cases. Meiotic "driving" is often the capacity of a large complex on a chromosome that is protected from crossing-over by an inversion (Burt and Trivers 2006). That gives them more integrity than other stretches of genetic material. LINE transposons, another example used in the book, also have clearer boundaries than most genetic elements, because they travel as units. If something of this kind is not applicable in each case, then we have a situation where the cases that seem to *need* a Darwinian description at gene level do not "partake of the same sins" as the cases where I resist or deflate that description.

Queller's exploration of the gene's eye view is illuminating. A trade-off emerges. One option is to resolve problems of counting and boundaries by regarding the genetic Darwinian individuals as the minimal units, nucleotides, which are indivisible for these purposes. Then each population has only four kinds of individuals, problems arise with S and C, and selection within a single Darwinian population never explains origins. Another option is to return to a view in which evolutionary genes are larger units, perhaps roughly the size of a cistron or promoter. Then the conclusions regarding S and C that seemed appealing at the start of Queller's discussion can be restored, and direct origin explanations can be given for the shaping of genes with special causal capacities. This view must then confront the problems that genes of this kind are very dubious entities. Defenders of this version of the gene's eye view have often talked about genes as if they were objects that could be counted at least in principle (allowing some vagueness), and which could be treated as comprising entities in a population. *DPNS* argued that this is an illusion. In the light of Queller's discussion, I think that this second option is nonetheless the more promising direction for the defender of the gene's eye view to pursue. This is partly because of the special cases – meiotic drivers, transposons, and other selfish genetic elements. Insofar as they force us to recognize a version of the gene's eye view, it is a view of this second kind rather than Queller's atomism.

References

Bergstrom, C. T. and M. Rosvall (2011). "The Transmission Sense of Information." *Biology and Philosophy* **: ***-***.

Booth, A. (forthcoming), "Selection, Symbiosis, and Individuality."

Burt, A. and Trivers, R. (2006). *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge, MA: Harvard University Press.

Calcott, B. (2008). "The Other Cooperation Problem: Generating Benefit," *Biology and Philosophy* 23: 179–203.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dennett, D.C. (1978). "A Cure for the Common Code," in *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge MA: Bradford Books, pp. 90-108.

Fletcher, J.A and M. Doebeli (2009). "A Simple and General Explanation for the Evolution of Altruism." *Proceedings of the Royal Society B* 276 (2009): 13–19

Francis, Richard, 2004, *Why Men Won't Ask for Directions: The Seductions of Sociobiology*, Princeton, NJ: Princeton University Press.

Gallistel, C.R., & King, A. (2009) *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience.* New York: Blackwell/Wiley.

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

Godfrey-Smith, P. (2011). "Senders, Receivers, and Genetic Information: Comments on Bergstrom and Rosvall." *Biology and Philosophy* ***: ***-***.

Grafen, A. (2007). "The Formal Darwinism project: a Mid-term Report." *Evolutionary Biology* 20: 1243–1254

Hull, D. L. (1980). "Individuality and Selection." *Annual Review of Ecology and Systematics* 11: 311–32.

Lewens, T. (2005). *Organisms and Artifacts: Design in Nature and Elsewhere*. Cambridge MA: MIT Press.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge MA: Harvard University Press.

Lewontin, R. C. (1970). "The Units of Selection." *Annual Review of Ecology and Systematics* 1: 1–18.

Lewontin, R. C. (1985). "Adaptation." In R. Levins and R. C. Lewontin (eds.), *The Dialectical Biologist*. Cambridge, MA: Harvard University Press, pp. 65–84.

Matsuura, K. and Y. Toshihisa (2006). "Aphid Egg Protection by Ants: a Novel Aspect of the Mutualism Between the Tree-Feeding Aphid *Stomaphis hirukawai* and its Attendant Ant *Lasius productus*." *Naturwissenschaften* 93 (2006): 506-510.

McKey, D. (1984). "Interaction of the Ant-Plant *Leonardoxa africana* (Caesalpiniaceae) with its Obligate Inhabitants in a Rainforest in Cameroon." *Biotropica* 16: 81-99

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

Nyholm, S. and M. J. McFall-Ngai (2004). "The Winnowing: Establishing the Squid-Vibrio Symbiosis." *Nature Reviews Microbiology* 2: 632-642.

Palmer, T., T. Young, M. Stanton, and E. Wenk (2000). "Short-Term Dynamics of an Acacia Ant Community in Laikipia, Kenya." *Oecologia* 123: 425–435

Queller, D. C. (1997). "Cooperators Since Life Began (Review of *The Major Transitions in Evolution*, by J. Maynard Smith and E. Szathmary)." *Quarterly Review of Biology* 72: 184–8.

Raine, N., N. Gammans, I. MacFadyen, G. Scrivner, and G. Stone (2004). "Guards and Thieves: Antagonistic Interactions Between Two Ant Species Coexisting on the Same Ant-Plant." *Ecological Entomology* 29: 345–352

Ruby, E. and L. Asato (993). Growth and Flagellation of *Vibrio fischeri* During Initiation of the Sepiolid Squid Light Organ Symbiosis. *Archives of Microbiology* 159: 160-167

Shea, N. (2007). "Representation in the Genome, and Other Inheritance Systems." *Biology and Philosophy* 22: 313-331.

Skyrms, B. (2009). *Signals: Evolution, Learning, and Communication*. Oxford: Oxford University Press.

Sperber, D. and D. Wilson (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Trivers, R. L. (1971). "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46: 35–57.

Wheeler, W. M. (1913) Observations on the Central American Acacia Ants (ed. by K. Jordan and H. Eltringham). *International Conference of Entomology, Oxford, August 1912, Vol. 2 Transactions*. London: Hazell Watson and Viney, pp. 109–139.

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought.* Berkeley, CA: University of California Press.

Wollenberg, M. and E. G. Ruby (2009). "Population Structure of *Vibrio fischeri* Within the Light Organs of *Euprymna scolopes* Squid from Two Oahu (Hawaii) Populations." *Applied and Environmental Microbiology* 75: 193-202.