Informationally-Connected Property Clusters, and Polymorphism

Abstract

I present and defend a novel version of the *homeostatic property cluster* [HPC] account of natural kinds. The core of the proposal is a development of the notion of *co-occurrence*, central to the HPC account, along information-theoretic lines. The resulting theory retains all the appealing features of the original formulation, while increasing its explanatory power, and formal perspicuity. I showcase the theory by applying it to the (hitherto unsatisfactorily resolved) problem of reconciling the thesis that biological species are natural kinds with the fact that many such species are polymorphic.

1 Introduction

The *Homeostatic Property Cluster* account (also *HPC* henceforth – see Boyd 1988, 1991, 1999) provides an attractive way of unpacking claims of naturalness for kinds that lack intrinsic essences.[1] This, for example, allows the HPC theorist to defend that biological species are natural kinds (Assis 2009; Assis & Brigandt 2009; Boyd 1999; Franz 2005; Keller, Boyd & Wheeler 2003; Rieppel 2005; Wilson 1999; Wilson, Barker & Brigandt 2007) even if, as many claim, they do not have intrinsic essences (see, e.g., Okasha 2002).

According to Boyd, a homeostatic property cluster exists when, among other things,[2]

1. There is a family (F) of properties that are contingently clustered in nature in the sense that they co-occur in an important number of cases.

---

[1] "[T]he essence of a given natural kind is a set of intrinsic (perhaps unobservable) properties, each necessary and together sufficient for an entity being a member of that kind." (Wilson 1999, p. 188)

[2] *Among other things*: in Boyd's original presentation nine further clauses are introduced, some of them probably intended as further necessary conditions for the presence of an HPC. For our current purposes, the HPC view can be adequately characterised by providing analogues of the two conditions that follow.

2. Their co-occurrence is, at least typically, the result of what may be metaphorically (sometimes literally) described as a sort of homeostasis. Either the presence of some of the properties in F tends (under appropriate conditions) to favor the presence of the others, or there are underlying mechanisms or processes that tend to maintain the presence of the properties in F, or both. (Boyd 1999, p. 143, first presented in 1988)

For an example of what an HPC looks like, consider the kind *eukaryotic cell*: according to the HPC account, this kind is constituted by two entities. First, a cluster of properties. To a first approximation, the properties to be part of this cluster are those typically had by individual eukaryotic cells: *Having mitochondria*, *Having an endomembrane system*, *Having a nucleus*, *Having chromosomes*, among many others. In general, in the HPC theory, properties *had* by kind instances partially *constitute* the kind. There are also properties *had* by the kind (consider "The eukaryotic cell is the smallest living unit in eukaryotes"), but the HPC theory doesn't put those latter properties to use.

The second constituent of the eukaryotic cell HPC is the family of mechanisms that explain the fact that properties in the cluster tend to be instantiated together – tend to "co-occur", as clause 1 of the passage just quoted puts it; getting clear on what precisely we should mean by that is one of the main goals of this paper. So, for example, it's no accident that tokens of *Having mitochondria* make tokens of *Having a nucleus* more likely. In characterising these mechanisms one needs to mention, e.g., facts about individual cells that explain that they tend to maintain their integrity, so that, e.g., if *Having mitochondria* and *Having a nucleus* co-occur now at one corner of a Petri dish, they will co-occur later at the same location. One also needs to mention the processes of cell reproduction that explain that if cluster properties (tend to) co-occur wherever a certain cell in the root tip of an onion is, they will also (tend to) co-occur wherever the cells deriving from it by mitosis are.

Eucharyotic cells are *systems*: organised, complex, differentiated wholes; but not all HPCs are like this. For example, *biofilms* have significantly less structure[3], and for many purposes they can be satisfatorily described as homeostatic clusters of the properties *Being a microorganism* and *Being extracellular polymeric substance*. The HPC theory can acommodate kinds with very different degrees of structural complexity.

The HPC account is congenial to traditional elucidations of natural-kindhood in that HPCs are taken to be (or at least give rise to) groups of *similar things*: each sample/instance of an HPC will typically have a significant number of properties in common with any other sample or instance. This reliance on property overlap among kind instances makes the HPC account vulnerable to the accusation, put forward most forcefully by Ereshefsky & Matthen (2005), that it is unable to account satisfactorily for the existence of *polymorphism*, the phenomenon by

---

[3]At the relevant scale – strictly speaking they have all the structure of their component cells.

which many kinds (most prominently, many biological species) have members that fall under a small, stable number of different variants, dissimilar in relevant respects to one another.[4] In this paper I defend the HPC account from this accusation.

In section 2 I present the problem of polymorphism for the HPC account, and briefly discuss the two existing approaches to its solution, Boyd's own (1999) and Magnus's (2011). After that, I provide a first, informal pass over my own proposal: there are structural similarities between, on the one hand, polymorphic kinds (with, e.g., male and female instances), and, on the other hand, kinds whose instances have a degree of internal articulation (so that, for example, biofilm samples have at least two main components: bacteria and extracellular slime). Whatever theoretical resources are deployed in the characterisation of the latter kinds can be deployed in the characterisation of the former – or so the proposal goes. This characterisation can be carried out without the resulting theory incurring (with Boyd) a commitment to dubious counterfactual-conditional properties, or yielding (with Magnus) incorrect predictions as to the extension of most polymorphic kinds.

Section 3 spells out in detail how the HPC account should be ammended so as to make the description of internal structure (and hence polymorphism) possible: the notion of *co-occurrence*, used by Boyd in the explication of what it is for property instantiations to cluster, and (mostly uncritically) taken up by other HPC theorists, should be regimented. While it is often assumed that *co-occurrence* means something in the vicinity of *spatio-temporal contiguity*, internal articulation can only be adequately captured if co-occurrence is understood as an informational connection among property instantiations. After introducing these *informationally-connected property clusters* [also *ICPC* henceforth], I reconstruct the proposal made in section 2 about polymorphic HPCs in more perspicuous ICPC terms. Section 4 takes a closer look at the identification of species with ICPCs, and provides answers to two possible objections to this identification.

Finally, section 5 makes a more general case for the theoretical fruitfulness of ICPCs in the study of natural kinds. In particular, I present three arguments in their favour: first, spatio-temporal-contiguity HPCs are a special case of ICPCs, so the explanatory power of the HPC theory in the alternative offered here is at least equal to the explanatory power of the theory under its traditional understanding; second, its explanatory power is strictly greater: there are HPCs which *have* to be described the IC way; third, the introduction of the information-theoretic toolbox allows us, among other things, to characterise in clearer terms the degree of naturalness of a kind, and to provide a rigorous sense in which some properties are more criterial of the presence of a kind than others. Section 6 is the conclusion.

---

[4]For brevity, in this paper I use "polymorphism" to refer to both polymorphism and polyphenism. Section 2 provides a more careful characterisation of the phenomena that I will be calling "polymorphism".

2 Polymorphism

Many, perhaps most, species are polymorphic. Two prominent examples of this phenomenon are sexual dimorphism in mammals (i.e., the fact that they appear in male and female variants) and caste polyphenism in ants (i.e., the fact that they appear in, e.g., queen and worker variants). As I have suggested above, there is a sense in which the HPC theory upholds a traditional view of natural kinds according to which they are collections of similar things. This view, as Ereshefsky & Matthen (2005) point out, is in tension with the fact that often species members come in different, sometimes *very* different, variants.

To make clear which phenomenon it is that Ereshefsky and Matthen claim cannot be accommodated by the HPC account, it will be useful to consider first a straightforward, but unsuccessful, approach to polymorphic HPCs. According to this approach we should think of the different variants (the usual term is *morphs*) in a polymorphic species as simply exemplifying *imperfect homeostasis*: these morphs belong to the same species because they are sufficiently similar to (share a sufficient number of properties with) one another; and they are different morphs because the number of properties they do *not* share is significantly large (as the HPC theory happily allows it to be). Polymorphic species occupy the right zone in the continuum from total property overlap to no overlap.

There are at least two problems with this approach. The first problem is that it is probably empirically inadequate: consider, for example, the extreme sexual dimorphism of the triplewart seadevil, or the black and yellow garden spider. It is unlikely that the (comparatively small) intersection of male and female properties in these species will be able to uniquely single them out from other related species.

The second, more important problem concerns the explanatory power of the resulting theory. The imperfect-homeostasis approach would make polymorphism indistinguishable from, say, mere statistical phenotypic variation. But there seems to be a distinctive, interesting explanandum in phenomena of *stable*, *discrete* variability. Writing off, e.g., height variability in male humans as a mere failure of homeostasis is probably the sensible theoretical move; but writing off the phenotypical differences between male and female mammals, or between worker and queen ants, as a failure of the same type is theoretically less apt. As Ereshefsky & Matthen (2005) helpfully put it, species are not just homeostatic (in that species members tend to resemble one another), but also, and equally importantly, *hetero*static, in that the same kind of differences (such as those among sexual dimorphs, for example) consistently recur across a species. The imperfect-homeostasis approach conflates mere statistical variability with heterostasis, and this is unsatisfactory.

In what follows, then, I will be using "polymorphism" to refer to the kind of stable, discrete phenotypic variability which cannot be written off as a simple failure of homeostasis, and of which sexual dimorphism in mammals, seasonal

polyphenism in some butterflies and moths, or developmental polymorphism in many social insects are examples.[5] Before presenting my own proposal for the accommodation of polymorphism in the HPC theory, I will briefly discuss two other solutions to this problem: Boyd's own (1999) and (Magnus 2011).

2.1 Boyd: Conditionally-Specified Properties

> The fact that there is substantial sexual dimorphism in many species and the fact that there are often profound differences between the phenotypic properties of members of the same species at different stages of their life histories ... together require that we characterize the homeostatic property cluster associated with a biological species as containing lots of conditionally specified dispositional properties for which canonical descriptions might be something like, "if male and in the first molt, P," or "if female and in the aquatic stage, Q." (Boyd 1999, p. 165)

Boyd's idea here appears to be that every individual in a polymorphic species is, surprisingly, in fact similar to every other individual: regardless of how profoundly different their phenotypes are, they all have very nearly the same properties; it's only that many of these are "conditionally specified dispositional properties", and the antecedent in the conditional may or may not be true.

This is an unsatisfactory solution to the problem of polymorphism. On the one hand, it forces the theorist to accept the existence of properties along the lines of

$$\lambda x \left[\text{Being such that, had } p \text{ been the case, } x \text{ would have been an } F\right]$$

Metaphysical scruples about such counterfactual-conditionally specified properties are perfectly understandable; in the same measure, an accommodation of polymorphism that carried no commitment to such entities would be preferable.

---

[5]There is, though, one kind of phenomenon that fits the characterisation of polymorphism I have just given, which my proposal does not cover. A clear example is the following: many species of butterflies present mimicry-based polymorphim and, e.g., females of *Papilio dardanus* differ in wing colouration, in some fourteen (stable, discrete) varieties (see Joron & Mallet 1998 for details). The case of mimicry polymorphism is significantly different from any of the ones I have just enumerated. In the latter cases, one and the same homeostatic mechanism brings about all of the morphs; in the mimicry case, the different morphs are brought about by slightly different, but overlapping, homeostatic mechanisms. This is mirrored in the fact that, while there is no temptation of talking of, e.g., the male *Homo sapiens* as a subspecies of *Homo sapiens*, we do recognise subspecies for each of the morphs of *Papilio dardanus*.

While mimicry polymorphism falls outside the scope of the HPC regimentation to be developed and defended in what follows, this is just as it should. The relation of mimicry morphs to their species is entirely analogous to the relation of species to higher taxa, and it should be treated analogously: as depending on the partial overlap of homeostatic mechanisms among the different subspecies, rather than on one single homeostatic mechanism bringing about different morphs connected informationally.

And, on the other hand, even if one is prepared to follow Boyd in recognising the existence of such properties in general, it is unlikely that members of polymorphic species have the very properties that are needed for the proposed accommodation of polymorphism to be successful. Consider a certain female in the aquatic stage of the species Boyd is alluding to in the passage quoted above. It is supposedly a fact about this individual that it has the following property: if it had been a male and in the first molt, it would have been P. But, as Magnus (2011) points out, for many individuals in a polymorphic species it is very unclear that they have this kind of properties. It is, for example, very unclear what it would take for a certain female to have been male: it is not always possible to tease out facts leading to the existence of a certain concrete individual from facts leading to the determination of its sex. Down this route, we will need to start worrying about ways of distinguishing essential from contingent properties of individuals, and ways of allocating sex properties to one of the two groups. Such metaphysical excursions are often unproductive, and should be avoided whenever possible.

2.2   Magnus: Cutting the Homeostatic Mechanism Loose

Magnus (2011) reacts to the problem of polymorphism by advocating a radical departure from common HPC wisdom: the only role that property clusters should play in the individuation of HPCs is in identifying which one is the homeostatic mechanism responsible for the recurrence of such clusters. Once the mechanism is identified, it "goes it alone": whatever it brings about should be considered as part of the kind that it individuates:

1. We look around and find clusters of properties.
2. We look for the mechanisms responsible for these clusters.
3. We identify natural kinds by the scope of those mechanisms.
   (Magnus 2011, p. 863)

Magnus's hope is that in this way we will be able to recover every morph in a polymorphic species. This is how the process is supposed to work:

1. We look around and find a female domestic cat, say
2. We then look for the mechanism that is responsible for the fact that female domestic cats keep recurring: it is sexual reproduction between cats.
3. Finally, we calculate the "scope" of this mechanism: it produces both female and male cats.

The kind *domestic cat* should, then, be identified with the set of male and female cats. Magnus is, rightly, sceptical of solutions to the problem of polymorphism that attempt to recover morphs as departures from a theoretical, undifferentiated kind member – and both the imperfect-homeostasis solution I reviewed above, and Boyd's own conditionally-specified-properties solution are variations on this

theme. But the only alternative Magnus envisages is cutting the homeostatic mechanism loose: everything that the mechanism that produces one morph produces should be counted in as a member of the kind.

This is an overreaction: the solution to the problem of polymorphism cannot be to follow homeostatic mechanisms anywhere and everywhere, because homeostatic mechanisms can go places where we do not want species to go. In particular, homeostatic mechanisms, the very same mechanisms that maintain homeostasis in a species, are also involved in the generation of the differences that eventually lead to speciation events – they are, as Ereshefsky & Matthen (2005) point out in the passage quoted below, *heterotic*. So, for example, while it is true that "[t]here are females . . . because of ongoing sexual interactions between females and males" (Magnus 2011, p. 863), it is every bit as true that there are *mutants* because of the very same kind of ongoing sexual interactions between females and males.[6] If we count as member of a kind whatever it is that ongoing sexual interactions produce, we will be forced to count mutants, even those that go on to establish their own, independent populations, as part of the original species.

That is, abiding by Magnus's recommendation will, in many ocassions, imply following the homeostatic mechanism of sexual reproduction accross the species boundary. Cutting the homeostatic mechanism loose has, as Magnus rightly claims, the virtue of generating all of the morphs: but only at the price of generating also all of the mutants, and beyond, to newly evolved especies. In the context of an effort to provide the HPC account with a process to generate plausible species extensions, this is an unaffordable price.

The lesson to be drawn from Boyd and Magnus is that neither focusing on the property cluster of an ideal instance of a species (whence morphs derive via imperfect homeostasis, or actualization of conditional properties), nor abandoning property clusters completely, will do. There is a third option, though: when we are dealing with a polymorphic kind, the right property cluster to focus on is the one consisting on *all the morphs together with the informational connections that hold among them*. The following sections develop this idea.

2.3   Heterostasis is Widespread

Ereshefsky and Matthen's challenge to the HPC account of polymorphic species can be summarised as a demand for two additional mechanisms, besides the one that ensures similarity among kind instances:

> . . . [W]e need to recognize "heterotic" mechanisms that produce variation, and "heterostatic" mechanisms that maintain it. (Ereshefsky & Matthen 2005, p. 10)

---

[6]The same point can be made, of course, about the homeostatic mechanism that is asexual reproduction.

Now, the HPC theory allows that homeostatic mechanisms can be imperfect, and this is enough to ensure a degree of heterocity – novelty will crop up through the cracks created by failures of homeostasis. This is not where the main challenge is; the main challenge is in the possibility of heterostasis.

Boyd's appeal to conditional properties can be reconstructed as denying that heterostasis is needed after all: every instance of a kind is, in the relevant sense, similar to every other instance, and thus homeostasis is enough to keep a polymorphic HPC together. Magnum's proposal that we individuate HPCs by homeostatic mechanisms (and do without property clusters) trusts that homeostasis, left to its own devices, will be enough to recover all morphs. As I have argued above, it will, but at the cost of generating very implausible species boundaries, precisely because homeostasis is, by design, imperfect.

We may now note that heterostasis is, in fact, a very widespread property of natural kinds: most every kind has a degree of internal structure and articulation, whereby *different* parts are kept *apart* from one another. A couple of examples will make clear what I mean: In the HPC *eukaryotic cell*, it is not just that all instances are relevantly similar to one another; it is also the case that, consistently, each individual cell has dissimilar parts so that, for example, mitochondria are, consistently, nothing like endoplasmic reticula. Consider now the HPC *Pseudomonas aeruginosa biofilm*: again, all samples of this biofilm are similar in substantial respects to one other. But they are, at the same time, composed of substantially different parts – extracellular matrix and *P. aeruginosa* cells.

There is a close parallelism between these two examples and polymorphic species. In the latter, one can recognise larger units that recur along the history of a species: couples of male and female morphs; groups of caste-system morphs in some polyphenic species, etc. These larger units keep happening again and again, very similar to one another. One can also recognise smaller units that keep recurring as well, consistently *different* to one another: each of the morphs: males and females; queens, drones and soldiers, etc.

That is, both the former, central examples of natural kinds, those that the HPC account is designed to deal successfully with, and the latter, problematic polymorphism cases share the following structural pattern:

- A larger unit which is copied recurrently throughout the kind's extension.
- Smaller units (in a part/whole relation with the larger one), persistently different (and in persistently the same respects) to one another.

Our tendency to downplay this structural similarity, I speculate, depends on the fact that we single out as kind instances the larger units in non-polymorphic kinds (each individual eukaryotic cell, or individual biofilm sample), and the smaller units in polymorphic kinds (each individual morph). This perhaps makes us see natural kinds in the two groups under different *Gestalten*. In any event, the structural similarity is there.

This could be seen as aggravating the problem that HPC has with polymorphic kinds: heterostasis, it turns out, needs to be accommodated even for the central, non-polymorphic cases, since the mere "clustering", the mere agglomeration of property tokens will not be able to distinguish haphazard conglomerates of polysaccharides and bacterial DNA from *P. aeruginosa* biofilms; or describe the way in which organelles are arranged in an eukaryotic cell.

Or we can see the glass half full: it also points to the way in which the HPC theory has to be ammended, if it's to make room for polymorphism. What we need is a general way to describe internal articulation and structure – roughly, the arrangement of smaller units into larger ones. Once we have this, the solution to the problem of polymorphism will fall out of this general treatment of heterostasis. I submit that the way of ammending the HPC account is by regimenting the appeal to *co-occurrence* in clause 1 of the characterisation of HPCs quoted in the Introduction:

> There is a family (F) of properties that are contingently clustered in nature in the sense that they co-occur in an important number of cases.

Both HPC theorists and critics often tacitly assume that this *co-occurrence* stands for spatio-temporal contiguity. That is, what homeostatic mechanisms do is to make it the case that the properties in the cluster are tokened roughly at the same time, roughly at the same place. Evidence that this is Richard Boyd's own tacit understanding of HPCs is provided by the fact that his original discussion is framed in terms of "things displaying properties" (Boyd 1999, p. 143). While this is compatible with "things" being scattered entities, it also clearly evokes a picture of properties instantiated at one place (wherever the *thing* is). The very "clustering" metaphor evokes this picture.[7] This tacit assumption, in the case of species, translates into the idea that the locus of co-occurrence should be individual species members:

> Conspecifics share many features in certain combinations – the cluster of properties characterising an HPC kind. (Assis & Brigandt 2009, p. 250)

But spatio-temporal contiguity is too coarse-grained a notion to account for heterostasis. What we need is a way of describing how several different collections

---

[7]I do not wish to commit myself to the claim – indeed, I do not believe – that Richard Boyd's explicit, fully worked-out theory of homeostatic property clusters involves an explicit conception of co-occurrence as spatio-temporal contiguity. The fact that he is ready to count the good, knowledge, or feudalism as examples of HPC suggests that it does not. What I do claim is that his conception of HPC is underdeveloped in a crucial respect – that of the nature of co-occurrence – and that when he has tried to offer a explicit characterisation, it has leaned towards contiguity co-occurrence. His proposed solution to the problem of polymorphism – discussed above – is a case in point. I offer the regimentation of co-occurrence as informational connectedness presented in the sequel as that which Boyd should be taken as referring to.

of properties (morphs in a polymorphic species; organelles, parts, etc. in non-polymorphic ones) tend to happen at different, but interrelated, locations. The right relation to do this job is not contiguity, but *informational connectedness.* *Informational* is used here in the sense in which a certain tradition beginning with Shannon (1948), and made salient to philosophers by, among others, Dretske (1981), understands it. One way to present the central insight of this tradition is as follows:

> The natural way to measure the information in a signal is to measure the extent that the use of that particular signal changes probabilities. (Skyrms 2010, p. 8)

Thus, a signal carries information about a state insofar as it forces us to reevaluate the likelihood that the state will happen.[8] The proposal I am making is that we exploit the fact that instantiations of properties in the cluster of an HPC are *signals* that carry information about instantiations of other properties in the same cluster, in exactly Skyrms's sense: the instantiation of some (groups of) properties makes the instantiation of other (groups of) properties more, or less, probable. I will use *ICPC* (informationally-connected property clusters) to refer to this proposed regimentation of homeostatic property clusters, with co-occurrence understood as informational connectedness,[9]

We can use informational connections to describe the internal structure of a *Pseudomonas aeruginosa* biofilm: The extracellular polymeric substance consists in tight informational connections between three kinds of exopolysaccharides and extracellular DNA among other things (Ma et al. 2009). *P. aeruginosa* bacteria in the biofilm colony are to be identified with another set of informational connections between organelles, a single chromosome, and a membrane. Then biofilms will consist on slightly looser connections between extracellular polymeric substance and bacteria, so described. And the very same kind of strategy can be used in the description of polymorphic species. Indeed, while it is true that the properties of, say, a sexual dimorph of a certain mammalian species are not generally spatio-temporally contiguous to the properties of the other dimorph, it is also true that the homeostatic mechanism that makes, e.g., the male dimorph recur regularly – the mechanism that makes the density of instantiations of properties in the male cluster stay above a certain value accross a certain range

---

[8]Skyrms talks about the signal *changing* probabilities, but this should not be understood in any causal sense; the signal need not be, and often will not be, causally upstream from the state signalled.

[9]It should be noted that whatever is explaining the fact that probabilities move the way they do – whatever explains the informational connections in place –might change from case to case. It is overwhelmingly likely that, in biology, such explanations will always be causal. But a theory of natural kinds in general needs to make room for the possibility that for some kinds (most notably, kinds in physics), informational connections be rock bottom (Ladyman et al. 2007) or depend on something else entirely. This is one reason why ICPCs are characterised in informational, not causal, terms. The possibility of using information theory in the study of kinds is another.

of locations – also makes the density of instantiations of the properties in the *female* cluster stay above a certain value in related locations. For a very obvious example, the presence of a female cat here now, increases the probability (all the way to 1, really) of the presence of a female and a male cats in the comparatively recent past. The homeostatic mechanism of mammalian sexual reproduction makes the properties in the two dimorphs co-occur in the informational sense.

In general, polymorphic species can be satisfactorily described as ICPCs. Consider polyphenisms such as those resulting in insect caste systems: the presence of a queen here now raises the probability of the presence of workers and soldiers in the vicinity. Consider developmental polyphenism in holometabolous insects: the presence of a pupa here now increases the probability of the presence of an imago around here, in the near future, etc.

What kinds of homeostatic mechanisms are sustaining these informational connections in polymorphic species? This is to be discovered empirically, and such empirical research does not belong in the theory of HPCs, but in the theory of this or that particular HPC. Quite a bit is known about particular homeostatic mechanisms, though. Caste-system polyphenism, for example, often involves hormone-based developmental switches (Nijhout 2003), triggered as a result of a convoluted process that will frequently depend on environmental factors, such as, e.g., changes of temperature related to overwintering, or adult control of larval developmental trajectories (Anderson, Linksvayer & Smith 2008, p. 120). In sexual dimorphism, switches are mostly genetic (Williams & Carroll 2009 is an informative review).

In summary, the answer to the polymorphism problem doesn't involve reference to dubious counterfactual-conditional properties, as Boyd suggests. Just as we need informational connections to distinguish biofilms from haphazard collections of polysaccharides and eDNA, the very same kind of informational connections can be used to describe polymorphic species: their informationally-connected property clusters will be composed by the properties of each of their morphs, and the informational relations among them.

Pace Magnus, then, there are no reasons to shy away from similarity fetishism about *this* kind of cluster.[10] A species *is* a set of similar things; it's just that the things in questions are not individual organisms, but informationally connected groups of organisms.

---

[10]At least no reasons having to do with species polymorphism: the ICPC theory, though, is still an HPC theory, and still relies on the notion of similarity. While I, predictably, regard this to be an advantage of the view (which is designed to be as small a departure from common HPC wisdom as possible), other theorists will undoubtedly disagree. For example, the ICPC predictions regarding speciation will differ from those of, e.g., Population Structure Theory (Ereshefsky & Matthen 2005; Matthen 2009) in that the former allows for anagenesis, which is at least uncongenial to the latter. The merits and demerits of the ICPC theory in this and other respects are quite independent of the issue of polymorphism, and should be assessed independently.

3 Informationally-Connected Property Clusters, and Heterostasis

More explicitly, *informationally-connected property clusters* are sets of informational connections, each of which holds between the fact that a group of properties is instantiated at certain locations and the fact that another group of properties is instantiated at other locations, such that the locations in the second group is a function of the locations in the first group.

For example, consider the following particular aspect of the HPC *human musculo-skeletal system*: for a vast majority of us, the combined positions of the left shoulder and the left elbow severely constrain the position of the left wrist. Once the first two are fixed, the probability of the third's being in a certain portion of sphere is very much raised (and the probability of its being outside this region very much diminished). In this case, the function from locations to location is one that takes the location of left shoulder and left elbow, and outputs a distribution of probability of the location of the left wrist.

In general, an *ICPC* can be characterised by providing ordered pairs of the form $\langle P(p), P(p \mid q) \rangle$ where $P(p)$ is the unconditional probability of p and $P(p \mid q)$ is the probability of p conditional on q. q will be a substitution of the schema

F_i is tokened at location l_ι and ... F_n is tokened at location l_ν

And p will be a substitution of the schema

$F_{i'}$ is tokened at location $f_{i'}(l_\iota \ldots l_\nu)$ and ... $F_{n'}$ is tokened at location $f_{n'}(l_\iota \ldots l_\nu)$

where $f_{i'}$ is a function that takes locations of instantiations of properties $F_1, \ldots, F_n$ and outputs the (more or less probable) location of an instantiation of $F_{i'}$.

Although in all the examples given so far locations have been spatio-temporal regions, *location* is intended as a semi-technical term that might range over, e.g., temperatures, or pH values, so that, for example, there is informational connectedness in a certain HPC if the presence of a certain property in acidic media increases the probability of the presence of a certain, different, property in basic media. In general, any meaningful substitution of x in the schema

P is instantiated at x

where P is a property, is a location in the intended sense.

It is easy to show how an informationally-connected property cluster, so understood, can be polymorphic. Consider the following toy model:

Let us introduce the function $f : Locations \rightarrow Locations$ as the one that takes every location l to the region of points within 200 and 300 m of l. The toy model's property cluster is composed of substitutions, for any $l \in Locations$, of the following four duple schemas:

- $\langle P\left(A \text{ is tokened at } f\left(l\right)\right), P\left(A \text{ is tokened at } f\left(l\right) \mid C \text{ is tokened at } l\right)\rangle$
- $\langle P\left(D \text{ is tokened at } l\right), P\left(D \text{ is tokened at } l \mid C \text{ is tokened at } l\right)\rangle$
- $\langle P\left(C \text{ is tokened at } f\left(l\right)\right), P\left(C \text{ is tokened at } f\left(l\right) \mid B \text{ is tokened at } l\right)\rangle$
- $\langle P\left(A \text{ is tokened at } l\right), P\left(A \text{ is tokened at } l \mid B \text{ is tokened at } l\right)\rangle$

where, we assume, in each duple the unconditional probability is lower than the conditional probability. What the duples are saying is that, if property C is instantiated somewhere, the ICPC's homeostatic mechanism makes it more probable that property A is instantiated within 200 and 300 m of C; and also makes it more probable that property D is instantiated at the same location as C. Likewise, if property B is instantiated somewhere, this makes it more probable that property C is instantiated within 200 and 300 m of B, and also makes it more probable that A is instantiated at the same location as B.

The resulting situation is one in which there are two "morphs", one made of a token of A and a token of B; the other made of a token of C and a token of D; and the ICPC's homeostatic mechanism makes it probable that they are 200 to 300 m apart, across a certain set of locations. This homeostatic mechanism is heterostatic: it sustains the difference between the two morphs – A-plus-B is nothing like C-plus-D. It is also, well, homeostatic: it makes the larger unit consisting of [an A-plus-B morph and a C-plus-D morph, 200 to 300 m from one another] keep recurring. See Figure 1 for the result of a simulation of this ICPC at play.

In summary, the HPC theory (in its ICPC version) has an attractive way of understanding polymorphism: *A species is polymorphic when its property cluster is composed of two (or more) informationally connected groups of properties, such that the properties in each group are (roughly) contiguity co-occurrent with one another.*

4   Species as ICPCs

This concludes the announced accommodation of polymorphism in the HPC theory. While I have produced nothing like a full defence of the claim that species are ICPCs, the foregoing discussion plausibly provides some abductive argument for it: insofar as the original HPC account of species has some initial plausibility, one major obstacle to its adoption is removed once we show how HPCs (in the regimentation defended here) can satisfactorily deal with stable polymorphism and heterostasis in general.
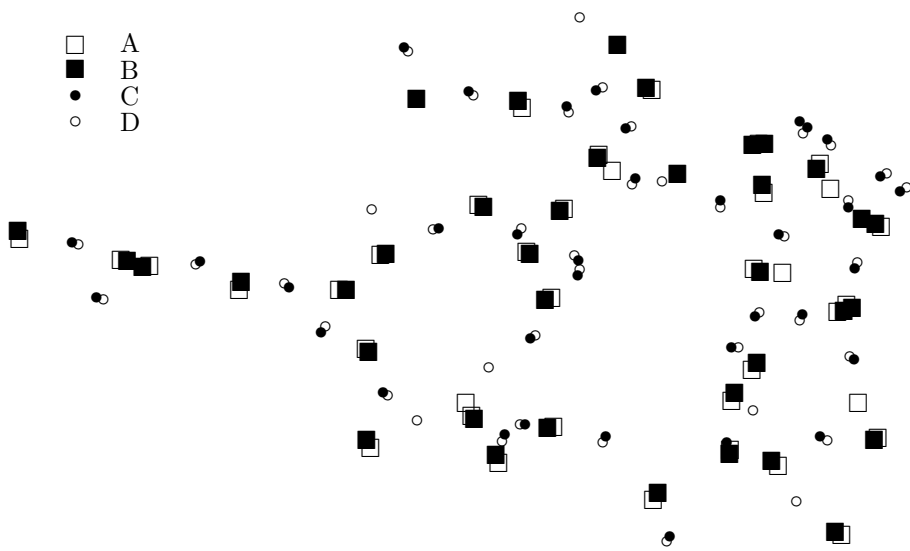
Figure 1: A toy model of a polymorphic kind with two "morphs": black-square-plus-white-square and black-circle-plus-white-circle. Polymorphism as presented here is still compatible with imperfect homeostasis – note the occasional single square or circle.

This treatment of polymorphism can be incorporated, as an independent module, to other aspects of the treatment of species as HPCs. One crucial such aspect is the identification of the homeostatic mechanisms peculiar to species:

> ... gene exchange between certain populations and reproductive isolation from others, effects of common selective factors, coadapted gene complexes and other limitations on heritable variation, developmental constraints, the effects of the organism-caused features of evolutionary niches, and so on – act to establish the patterns of evolutionary stasis that we recognize as manifestations of biological species. (Boyd 1999, p. 165)

The question of what mechanisms should count as affording species- (as opposed to other taxa-) homeostasis is a substantial question in the philosophy of biology, of course, but one that is largely independent of the pressure put on the HPC account by species polymorphism. The ICPC regimentation I am defending here follows the HPC account in remaining as uncommitted and liberal as possible regarding what should count as a natural kind of one or another type. The general metaphysical theory of natural kinds should be able to encompass different kinds of kinds, natural to different degrees.

The treatment of species as ICPCs, though, invites at least two objections, which I now discuss in turn.

4.1 Mechanisms and Population-Level Phenomena

I have been relying on a parallelism between inner structure in kinds such as *biofilm* or *eukaryotic cell* and species polymorphism. One might worry that talk of structure in relation to species is misplaced: species are populations, and populations are not *systems*; it is not the arrangement of parts into differentiated roles that make species tick (while, arguably, it *is* what makes eukaryotic cells tick). Species are simply not that internally organized.

The first thing to say is that 'structure' does not apply only to parts playing differentiated roles within a system. A less charged, but perfectly natural, sense sees structure in an HPC whenever these properties are not homogeneously clustered, but rather appear in discreet, repeatable chunks. Systems, indeed, show structure so understood: the repeatable chunks in question are there, interacting with each other to produce certain relevant effects. But structure can be merely phenomenological. The toy model of a polymorphic kind in figure 1 above is a case in point: the black-square-plus-white square morph and the black-circle-plus-white-circle morph are not parts of a mechanism, and do not interact in any way; they are, though, repeatably distinct chunks of the toy property-cluster, and thus the resulting ICPC qualifies as structured, and polymorphic, in the intended sense.

Thinking of HPCs as informational structures has the advantage of not encouraging a mechanistic reading: the existence of an informational connection among property tokens already presupposes that something or other is sustaining the connection – one is free to call whatever it is that is doing the sustaining a mechanism, but not much (certainly not the contention that HPCs must be systems) should be read into this. More to our current point, there are, indeed, instances of population-level mechanisms (in this minimal sense) which help maintain polymorphism: apostatic selection of *Catocala* moths (Bond & Kamil 1998), or crowding as a trigger of developmental switches in desert locusts (Nijhout 2003; Tanaka 2006). The fact that such mechanisms depend on population-level phenomena does not prevent the treatment of *Catocala* moths or desert locusts as ICPCs: the informational connections between morphs are very much in place and, for example, the presence of a certain individual of the dark morph of *Catocala relicta*, in the experiment described in (Bond & Kamil 1998), is a signal of the presence of other morphs of the virtual *Catocala* population elsewhere in the habitat they share with blue jays. Apostatic selection makes it so.

On the other hand, even if species are (as they surely are) populations, this does not mean that there are not proximal causes for the presence and maintainance of polymorphisms, that make informational connections among morphs even tighter: haplodiploidy, for example, makes it the case that females carry more information about the presence of males in the recent past than males do about females. There is no need to resort to population-level descriptions in order to explain why this is so; the very sex-determination system ensures that this will be so.

In summary: although species are not systems, but populations, they have structure in the minimal sense that properties in the species cluster appear in a small number of different, repeatable chunks. Such structure can, and often does, emerge as a result of population-level processes (apostatic selection; crowding effects) although it will often have more proximal causes (sex determination). The treatment of species as ICPC shows that, and how, natural-kind thinking is compatible with population thinking.

4.2   Associations of Organisms and the Species Boundary

Many organisms establish very close relation with other organisms: parasites, symbionts, commensals, etc. It will often be the case that the presence of, say, one of a pair of symbionts will be a signal of the presence of the other. Doesn't this force the ICPC theorist to count populations of such associations of organisms as species?

It does not. The thesis that species are informationally-connected property clusters leaves considerable latitude in the details of what makes any one such property cluster a species. I have suggested above that a way to fill in these details is to identify what is the mechanisms that maintain the informational

connections in species property-clusters. Boyd (1999), Rieppel (2005) and Assis & Brigandt (2009) among others have claimed that, for something to count as a taxon, *common descent* must be at least part of what is keeping these connections in place. This screens off most of the above problem cases: parasite and host will not usually stem from a common ancestor, nor will commensals or symbionts.

Some problem cases will likely remain, though: Godfrey-Smith (2013) discusses the symbiosis between aphids and *Buchnera aphidicola* (see also Jousselin, Desdevises & d'acier 2009) in which parent-offspring relations among aphid-Buchnera combinations are well defined, and there has been common descent for millions of years. Godfrey-Smith calls aphid-Buchneras "organisms". While populations of such organisms would seem to conform to the letter of many definitions of species, no theorist, that I am aware, recognises an aphid-Buchnera species. This perhaps suggests that the common-descent restriction on species homeostatic-mechanisms is tacitly coupled with the requirement that the organisms that enter in parent-offspring relations be monogenomic. The ICPC account is compatlble with this and other ways of filling in the details of its application to biological species.

On the other hand, the ICPC account of natural kinds allows a homeostatic property clusters to be considered a natural kind whenever "the homeostatic unity they (imperfectly) display is a causally and explanatorily important factor in the complex systems we study" (Keller, Boyd & Wheeler 2003, p. 105). Clusters of individual organisms belonging to different species often play such an important causal and explanatory important role, and we do have a number of kind terms that single them out: *microbial consortium* or *multispecies biofilm* are two examples.

In summary, nothing forces the ICPC theorist to count close symbiotic partners as conspecific. But many such aggregations will indeed be natural kinds.

5 ICPCs in the Study of Natural Kinds

I have claimed that the most theoretically fruitful regimentation of "homeostatic property cluster", beyond its role in the accommodation of polymorphism, is informationally-connected property clusters. I will now present three arguments for this claim. The first is that the traditional way of understanding co-occurrence, as spatio-temporal contiguity, can be captured without residue as informational connectedness; contiguity property-clusters are a special case of IC property clusters. The second is that there are kinds which have been, or should be, identified with HPCs in which co-occurrence can only be informational connectedness. The third is that thinking of HPCs in informational terms allows us to deploy information-theoretic tools in the study of HPCs, in an illuminating way.

5.1 Contiguity HPCs are ICPCs

Contiguity HPCs are, for some purposes, a good first approximation to the nature of many kinds (kinds of stuff, for example, or non-polymorphic species, such as bacteria[11]), so it is reasonable to worry whether, by moving wholesale to the IC understanding of HPCs, we are losing the ability to describe contiguity HPCs.

We are not. The expressive power of the theory of IC HPCs is a strict superset of that of the theory of contiguity HPCs. In particular, a contiguity property cluster is just an IC property cluster in which all informational connections are of the following kind:

$$\langle P\left(F \text{ is tokened at } l\right), P\left(F \text{ is tokened at } l \mid F' \text{ is tokened at } l\right)\rangle$$

Where l ranges over spatio-temporal regions.

That is, in contiguity HPCs all probability changes happen conditional on properties being tokened at the same spatio-temporal region. The informational connections are still there, but hold only between property instantiations that happen at the same time, in the same place. For example, it is a good first approximation to the structure of an eukaryotic cell to say that each individual is constituted by spatio-temporally contiguous tokens of properties in the cluster. One can make the exact same point by saying that each instantiation of the properties in the eucharyotic-cell cluster brought about by its homeostatic mechanism is a signal that carries information about instantiations of other properties in the cluster at the same place (namely, wherever the cell is). According to the definition above, then, contiguity HPCs *are* ICPCs.

On the other hand, of course, if all HPCs were of the contiguity kind, there would be no need for the additional structure that ICPCs introduce: for example, the appeal to locations would be otiose, as all informational connections would hold between properties tokened at the same location. The following section shows that using the HPC theory to account for the nature of several kinds which have been proposed, by Boyd or others, as falling within the scope of the theory, requires that property clusters be understood the IC way.

5.2 HPCs Which Must be Described the IC Way

Many homeostatic mechanisms, notably the ones that individuate kinds of processes, only ensure co-occurrence in the informational sense. A clear example: the onset of AIDS is separated from the original HIV infection by an asymptomatic phase that lasts, without therapy, ten years on average (Nowak 2006, p. 169). HIV infection and AIDS are, then, not contiguity co-occurrent, their locations

---

[11]Of course, once past the most coarse-grained level of description, bacteria are exquisitely complicated structures, and something like ICPCs must be used in their characterisation.

being ten years apart, but they *are* informationally connected in the intended sense: the presence of an HIV infection at, say, $\langle Sabrina, 1981 \rangle$ increases the probability of the onset of AIDS at $\langle Sabrina, f_{time}(1981) \rangle$, where f_{time} takes its argument to a time ten years into the future.[12]

Many other process-kinds such as, e.g., human pregnancy, presumably also HPCs, will have property clusters in which informational connections are much more central than contiguities – the presence of a morula raises the probability of the presence of a fetus ten weeks later, etc.

Of course, the main thesis of this paper is that polymorphic species are also HPCs which must be described as ICPCs.

5.3   HPCs and the Information-Theoretic Toolbox

As a final argument in favour or ICPCs, I offer the following two ways in which information theory can improve our understanding of natural kinds.

5.3.1   Quantity of Information and Criterial Properties of a Kind.

Each ordered pair of unconditional and conditional probabilities in an informationally-connected property cluster can be used to calculate the *quantity of information* that one of the groups of property instantiations carries about the other group. If, for example, the ordered pair is

$$\langle P \left(F \text{ is tokened at } l \right), P \left(F \text{ is tokened at } l \mid F' \text{ is tokened at } l' \right) \rangle$$

then *F'* being instantiated at *l'* carries

$$log_2 \frac{P \left(F \text{ is tokened at } l \mid F' \text{ is tokened at } l' \right)}{P \left(F \text{ is tokened at } l \right)}$$

bits of information about the state consisting of *F* being instantiated at *l*. Any one signal (for us, this is an instantiation of a group of properties at a location) might, and usually will, carry information about many different events (again, for us, these are instantiations of properties at locations).

Let us call the set of events that a certain signal S carries information about $PInst_S$. The *Kullback-Leibler distance* (see Skyrms 2010, p. 36; the original discussion is Kullback & Leibler 1951) is the weighted average of all of the quantities of information corresponding to each of the informational connections between S and members of $PInst_S$:

---

[12]For a defence that some diseases are HPCs see Williams (2011).

$$I\left(S\right) = \sum_{event \in PInst_S} P\left(event \mid S\right) log_2 \frac{P\left(event \mid S\right)}{P\left(event\right)}$$

Different signals will give rise to different Kullback-Leibler distances. Now, those properties the instantiations of which have the highest distances are the ones that carry the most information about other properties in the cluster. They are, that is, the properties such that, if instantiated, make it most probable that a larger part of the cluster will be instantiated.

For a simple example, consider the HPC *bird*. The property *has wings* has a higher Kullback-Leibler distance than the property *has eyes* (and both are members of the HPC): the former makes other properties in the bird cluster (say, *has a beak*, *lays eggs*, *has feathers*) much more probable than the latter does. Of the two properties, it is the more criterial for the presence of a bird.

5.3.2 Average Kullback-Leibler Distances and the Naturalness of a Kind.

Finally, we can take *all* signals in an ICPC, and calculate the weighted average of Kullback-Leibler distances for that kind:

$$\sum P\left(S\right) I\left(S\right)$$

The higher this summation is, the more properties, and more frequently instantiated, are highly criterial: more properties are like *has wings* for birds, and less like *has eyes*. The ICPC is more integrated, imperfect homeostasis is less frequent, and inductions, as a result, are succesful more frequently. This average distance can be, thus, suggested as one measure of the *naturalness* of the kind in question.

These information-theoretic elucidations of criteriality and naturalness are only available under the IC regimentation of the HPC account.

6 Conclusion

A homeostatic property cluster, once past the crudest level of idealization, very often needs to be characterised as a collection of informational links among property instantiations. As I have tried to show, individuating HPCs by these kinds of structures has several theoretical advantages: it allows us to accommodate the fact that many natural kinds are polymorphic; it helps with the description of the internal articulation and structure of kind members; it makes the deployment of information-theoretic tools in the description of natural kinds possible.

Furthermore, this represents no fundamental modification of the HPC account. For example, ICPCs are still based on similarity among groups of instantiations of properties in a cluster; and homeostatic mechanisms are still needed to explain the recurrence of these groups, the fact that they keep on happening time and again. It's only that property clusters are more interesting entities than we thought.

References

Anderson, KE, Linksvayer, TA & Smith, CR 2008, 'The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae)', *Myrmecological News*, vol. 11, pp. 119–132.

Assis, L & Brigandt, I 2009, 'Homology: Homeostatic property cluster kinds in systematics and evolution', *Evolutionary Biology*, vol. 36, no. 2, Springer US, pp. 248–255.

Assis, LCS 2009, 'Coherence, correspondence, and the renaissance of morphology in phylogenetic systematics', *Cladistics*, vol. 25, no. 5, Blackwell Publishing Ltd, pp. 528–544.

Bond, AB & Kamil, AC 1998, 'Apostatic selection by blue jays produces balanced polymorphism in virtual prey', *Nature*, vol. 395, pp. 594–596.

Boyd, R 1999, 'Homeostasis, Species, and Higher Taxa', in RA Wilson (ed.), *Species: New interdisciplinary essays*, Mit Press, pp. 141–185.

Boyd, R 1988, 'How to be a moral realist', in G Sayre-McCord (ed.), *Moral realism*, Cornell University Press, pp. 181–228.

Boyd, R 1991, 'Realism, anti-foundationalism and the enthusiasm for natural kinds', *Philosophical Studies*, vol. 61, no. 1-2, pp. 127–148.

Dretske, F 1981, *Knowledge and the low of Information*, The MIT Press.

Ereshefsky, M & Matthen, M 2005, 'Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory', *Philosophy of Science*, vol. 72, pp. 1–21.

Franz, NM 2005, 'Outline of an explanatory account of cladistic practice', *Biology and Philosophy*, vol. 20, no. 2-3, Kluwer Academic Publishers, pp. 489–515.

Godfrey-Smith, P 2013, 'Darwinian Individuals', in F Bouchard & P Huneman (eds), *From groups to individuals: Evolution and emerging individuality*, The MIT Press, pp. 17–36.

Joron, M & Mallet, JLB 1998, 'Diversity in Mimicry: Paradox or Paradigm?', *Trends in Ecology and Evolution*, vol. 13, no. 2, pp. 461–466.

Jousselin, E, Desdevises, Y & d'acier, AC 2009, 'Fine-scale cospeciation between *brachycaudus* and *buchnera aphidicola*: bacterial genome helps define species

and evolutionary relationship in aphids', *Proceedings of the Royal Society B*, vol. 276, pp. 187–196.

Keller, R, Boyd, R & Wheeler, Q 2003, 'The illogical basis of phylogenetic nomenclature', *The Botanical Review*, vol. 69, no. 1, Springer-Verlag, pp. 93–110.

Kullback, S & Leibler, RA 1951, 'On Information and Sufficiency', *Annals of Mathematical Statistics*, vol. 22, pp. 79–86.

Ladyman, J, Ross, D, Spurret, D & Collier, J 2007, *Every thing must go*, Oxford: Oxford University Press.

Ma, L, Conover, M, Lu, H, Parsek, MR, Bayles, K & Wozniak, DJ 2009, 'Assembly and development of the *Pseudomonas aeruginosa* biofilm matrix', *PLOS Pathogens*, vol. 5, no. 3.

Magnus, PD 2011, 'Drakes, Seadevils, and Similarity Fetishism', *Biology and Philosophy*, vol. 26, no. 6, pp. 857–870.

Matthen, M 2009, 'Chicken, Eggs, and Speciation', *Noûs*, vol. 43, no. 1, pp. 94–115.

Nijhout, HF 2003, 'Development and Evolution of Adaptive Polyphenisms', *Evolution & Development*, vol. 5, no. 1, pp. 9–18.

Nowak, MA 2006, *Evolutionary Dynamics*, Cambridge: Harvard University Press.

Okasha, S 2002, 'Darwinian Metaphysics: Species and the Question of Essentialism', *Synthese*, vol. 131, no. 2, pp. 191–213.

Rieppel, O 2005, 'Monophyly, paraphyly, and natural kinds', *Biology and Philosophy*, vol. 20, no. 2-3, Kluwer Academic Publishers, pp. 465–487.

Shannon, C 1948, 'A Mathematical Theory of Communication', *The Bell System Mathematical Journal*, vol. 27, pp. 379–423, 623–656.

Skyrms, B 2010, *Signals: Evolution, Learning & Information*, New York: Oxford University Press.

Tanaka, S 2006, 'Phase-related body-color polyphenism in hatchlings of the desert locust, schistocerca gregaria: Re-examination of the maternal and crowding effects', *Journal of insect physiology*, vol. 52, pp. 1054–1061.

Williams, NE 2011, 'Arthritis and Nature's Joints', in JK Campbell, O'RourkeM & MH Slater (eds), *Carving nature at its joints*, Cambridge: The MIT Press, pp. 199–230.

Williams, T & Carroll, S 2009, 'Genetic and Molecular Insights into the Development and Evolution of Sexual Dimorphism', *Nature Reviews Genetics*, vol. 10, no. 11, pp. 797–804.

Wilson, R, Barker, M & Brigandt, I 2007, 'When Traditional Essentialism Fails: Biological Natural Kinds', *Philosophical Topics*, vol. 35, no. 1-2, pp. 189–215.

Wilson, RA 1999, 'Realism, Essence, and Kind: Resuscitating Species Essentialism?', in RA Wilson (ed.), *Species: New interdisciplinary essays*, Mit Press, pp. 187–208.