

# Informationally-Connected Property Clusters

## 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 polymorphic natural kinds.

## 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.<sup>1</sup> This, for example, allows the HPC theorist to defend that biological species are natural kinds (Boyd 1999; 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,<sup>2</sup>

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.
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 Boyd 1988)

---

<sup>1</sup>“[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)

<sup>2</sup>*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, though, the HPC view can be adequately characterised by providing analogues of the two conditions that follow.

For an example of what an HPC looks like, consider the kind *eukaryotic cell*: according to the HPC account, this kind is constituted by a cluster of properties (that includes *Having mitochondria*, *Having an endomembrane system*, *Having a nucleus*, *Having chromosomes*, among many others) together with the mechanism that explains the fact that, e.g., *Having mitochondria* and *Having a nucleus* tend to co-occur. This mechanism will include the facts about individual cells that explain that they do not disintegrate in thin air, 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 a slightly different location. The mechanism also includes the facts about cell reproduction that explain that if these 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.

Kinds, and HPCs in particular, are, for many purposes, usefully thought of as sets of similar things. For many HPCs the “things” in question are kind *members*; but many HPCs do not have members, and some kinds are such that their members are not all similar to one another in the relevant sense—this latter case will figure prominently in the second half of the paper, where I discuss polymorphism.

The following terminology, thus, will be useful in what follows:

**Packet:** A *packet* of an HPC F is any member of a (possibly different) HPC, K, such that, typically, Ks are the smallest entities that have most of the properties in the F property cluster.

Packets are intended as such that *all* HPCs are sets of similar packets. For many kinds, packets are identical to members—that is, F and K are the same kind: for example, it is individual cells, i.e., individual members of the kind *eukaryotic cell*, that typically have most of the properties in the *eukaryotic cell* property cluster—properties such as the ones enumerated above.

For many other kinds, though, packets and members are different entities. Kinds of stuff, for example, do not have members, but they have packets. E.g., the packets of *water* will be the smallest kinds of entity composed of H<sub>2</sub>O molecules, with a boiling point of 100 °C, a freezing point of 0°C, etc.: water droplets.<sup>3</sup> One of the theses I will be advancing in what follows is that, in polymorphic kinds, packets do not coincide with members either.

In this paper I develop and clarify the notion, central in the HPC theory, of *co-occurrence*. I will argue that, while it is often assumed that co-occurrence means something in the vicinity of *spatio-temporal contiguity*, a more theoretically fruitful notion is what I will be calling *informational connectedness* [also *IC*

---

<sup>3</sup>There is some harmless indeterminacy in the size of packets of water, and of stuff in general, (how tiny must the droplet be?) which I think it would be unwise to regiment. In the characterisation of packet just presented, *typically* and *most* are indeterminate in analogously harmless ways.

henceforth]. Section 2, the main section of the paper, is dedicated to presenting and defending this information-theoretic unpacking of co-occurrence in the theory of HPCs: after introducing informationally-connected property clusters [also *ICPC* henceforth], I present three arguments in their favour: first, spatio-temporal-contiguity HPCs are a special case of ICPCs, so the expressive power of the HPC theory in its traditional understanding is a subset of the expressive power of the alternative offered here (subsection 2.1); second, it is a *strict* subset: there are HPCs which *have* to be described the IC way (subsection 2.2); 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 (subsection 2.3).

Section 3 showcases the theory of ICPCs by discussing its treatment of *polymorphism*, the phenomenon by which many natural kinds have members that fall under a small, stable number of different variants, dissimilar in relevant respects to one another.<sup>4</sup> The HPC account of kinds, and biological species in particular, has been accused, most forcefully by Ereshefsky & Matthen (2005), of being unable to account satisfactorily for the existence of polymorphism. After briefly discussing Boyd’s (1999) and Magnus’s (2011) reactions to this problem, I show how ICPCs can be polymorphic, 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 4 is the conclusion.

## 2 Co-Occurrence

Consider an individual *E. coli* bacterium, *Colin*, and the strain wherein Colin belongs. Assume that both Colin and its strain are HPCs.<sup>5</sup> Assume also that the cluster of properties that partially (and, together with a homeostatic mechanism, totally) individuates Colin (e.g., being Gram-negative, rod-shaped, motile, etc.) is the same cluster of properties that partially (and, together with a homeostatic mechanism, totally) individuates Colin’s strain.

Even given this identity among clusters, Colin is, obviously, not the same entity as its strain, just as I am not the same entity as the *Homo sapiens* species. The difference between Colin and its strain is the homeostatic mechanism that keeps properties in one and the same cluster co-occurring: processes such as intracellular homeostasis in Colin’s case and asexual reproduction (plus intracellular homeostasis) in the strain’s case.

---

<sup>4</sup>For brevity, in this paper I use “polymorphism” to refer to both polymorphism and polyphenism. Section 3 provides a more careful characterisation of the phenomena that I will be calling “polymorphism”.

<sup>5</sup>While it is by no means uncontroversial that individuals can be HPCs, nothing of substance will depend on this.

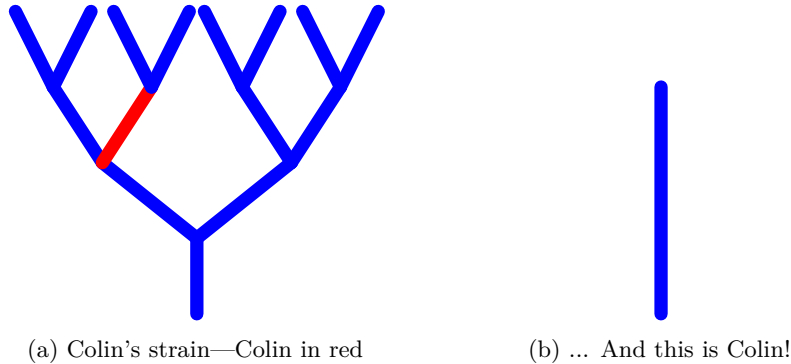


Figure 1: Two (dramatically idealised) spatio-temporal diagrams

This difference in homeostatic mechanisms results in a difference, among other things, in the pattern of instantiations of properties in the cluster: if we plot instantiations of properties in the cluster brought about by Colin's homeostatic mechanism in a space-time diagram we will get, at any given time, a high density of property instantiations wherever Colin is, and close to no instantiations elsewhere: the familiar four-dimensional space-time worm (fig. 1b). If, instead, we plot instantiations brought about by the strain's homeostatic mechanism (idealizing somewhat) we will get a binary branching tree, nodes corresponding to fissions—each segment between nodes a Colin-like worm (fig. 1a).

Suppose we are growing Colin's strain in a continuous-flow chemostat; the strain's spatio-temporal diagram, at any given time, will show a mix of age groups (where ages are distances to the immediately previous branching point), in varying proportions—see Lavric & Graham (2010) for details. The important thing to note is that such distribution of ages is as much a consequence of the strain's homeostatic mechanism as the clustering of Colin-like properties in the different individuals is: the process by which the strain's homeostatic mechanism ensures that Colin-like properties keep co-occurring in new bacteria also ensures that a certain mix of bacteria of different ages will be present at roughly the same time, and neighbouring times. For example, the very homeostatic mechanism that ensures that Colin-like bacteria keep appearing in the chemostat also explains that each of these bacteria has another bacterium, in a certain narrow range of ages and sizes, in its recent past—the bacterium from which it derives via fission.

To put the point in terms that will be useful in what follows: homeostatic mechanisms such as that of Colin's strain do not just raise the probability of finding Colin-like properties instantiated next to one another, at the same time—and this is the effect of homeostatic mechanisms that both HPC theorists and their critics have been focusing on. They also raise the probability of finding the same and other properties (the properties of younger or older bacteria of the same strain, in this case) instantiated *elsewhere* in space and time. Homeostatic mechanisms, that is, do not just bring about co-occurrence understood as *spatio-*

*temporal contiguity*, but also what we might call *informational connectedness*.

The main thesis of this paper is that *co-occurrence* in clause 2 of Boyd's characterisation of HPCs cited above should be understood as an informational connection. *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. (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.) The proposal I am making is that we view instantiations of properties in the cluster of an HPC as *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. In the example I have just presented, there is *informational connectedness* in Colin's strain: the presence of a bacterium of a certain age *a* at *(here, now)* increases the probability of the presence of a bacterium that is about to undergo fission at *(here, now - a)*.

For another example, consider the following particular aspect of the HPC *human*: 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). I propose to describe HPCs in a way sensitive to this kind of structural features: in the example, the state consisting of [the elbow being in such-and-such a position and the shoulder in thus-and-so a position] is a signal that carries information about the position of the wrist. We can encode this kind of informational connections by providing a set of pairs of probabilities, where each pair has, as first member, the unconditional probability of the wrist being at a certain location; and, as second member, its probability conditional on shoulder and elbow being at two certain other locations. An *informationally-connected property cluster* is a set of such pairs.

Although in both these examples locations are spatio-temporal regions, I intend *location* as a semi-technical term that might refer to, e.g., ranges of temperatures, or pH values, so that, for example, there is informational co-occurrence 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 tokened at  $x$

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

So as to be maximally explicit, I now introduce the notion of informationally-connected property cluster in a more formal fashion:

- *Locations* is a set of locations:  $\{l_1, \dots, l_n\}$ .
- *Cluster* is a set of properties:  $\{F_1, \dots, F_n\}$ .
- The notation  $F_i@l_j$  is introduced as shorthand for the sentence schema  $\lceil$ Property  $F_i$  is instantiated at location  $l_j$  $\rceil$ .
- *Atomic* is the set  $\{F_i@l_j: F_i \in \text{Cluster and } l_j \in \text{Locations}\}$ .
- *Complex* is defined as the closure of *Atomic* under conjunction.
- *Probabilities* is the set  $\{ \langle P(p), P(p|q) \rangle : p, q \in \text{Complex} \}$  where  $P(p)$  is the unconditional probability of  $p$  and  $P(p|q)$  is the probability of  $p$  conditional on  $q$ .

That is, *Probabilities* is the set of all ordered pairs, in which the first member is the unconditional probability of a certain state of affairs we would describe with the sentence “property  $P$  is instantiated at location  $l$ , and property  $P'$  is instantiated at location  $l'$ , and  $\dots$ ”; and the second member is the probability of that state of affairs conditional on another one, which we would describe with a similar sentence. Finally, given any two concrete sets *Locations* and *Clusters*,

**Informationally-Connected Property Cluster:** An *IC property cluster* is a subset of *Probabilities*.

In most interesting cases, there will be much more compact descriptions of an ICPC. For example, there will usually be a function taking the locations that appear in one of the members of any of the duples in an ICPC to the locations that appear in the other member—such was the case in the example of the wrist: its position is a function of the positions of shoulder and elbow. For another example, a large number of ordered pairs will usually be expressible as a function that takes instantiations of properties at locations to *distributions of probability* of instantiations of other properties at other locations. The characterisation just given, though, is the most general.

I claim that the most theoretically fruitful regimentation of “homeostatic property cluster” 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.

## 2.1 Contiguity HPCs are ICPCs

In the presentation, and the application, of the HPC account, it is often tacitly assumed that properties cluster around spatio-temporal locations<sup>6</sup>: 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.<sup>7</sup>

Indeed, contiguity HPCs are, for some purposes, a good approximation to the nature of many kinds (kinds of stuff, for example, or such non-polymorphic kinds as bacteria), 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 of the following kind:

$$\{\langle P(F@l), P(F@l | F'@l) \rangle : F, F' \in Cluster, l \in Locations\}$$

Where *Locations* is a certain set of 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

---

<sup>6</sup>This is probably the right place to prevent a possible misunderstanding: I have claimed that HPC theorists and their critics have assumed that properties in the cluster co-occur in the sense of being spatio-temporally contiguous to one another. But it is an ubiquitous (maybe a defining) trait of kinds that they are *scattered*: such kinds as *eukaryotic cell* are instantiated at many different places; wherever individual cells happen to be. How could HPC theorists be so deluded as to think that the properties of all of those samples are spatio-temporally contiguous?

Obviously, they are not: the claim is not that the HPC account tacitly assumes that *all* property instantiations in a cluster share spatio-temporal locations. The claim is, rather, that all property instantiations *in a packet* are spatio-temporally contiguous: *Having a nucleus* and *Having mitochondria*, whenever they are tokened in a certain individual packet (i.e., constituting an individual cell) are tokened at roughly the same place.

That this is the norm in HPCs I will be disputing in the sequel—for example, I will defend that packets in polymorphic species such as *domestic cat* do not correspond to individual organisms.

<sup>7</sup>I do not wish to commit myself to the claim—indeed, I do not believe—that Richard Boyd’s explicit, fully worked-out conception of homeostatic property clusters involves what I have called contiguity co-occurrence. 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 an explicit characterisation, it has leaned towards contiguity co-occurrence. His proposed solution to the problem of polymorphism—discussed in section 3.1—is a case in point.

I offer ICPCs as the kind of entities that Boyd *should be taken as referring to*.

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 eukaryotic-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.

## 2.2 HPCs Which Must be Described the IC Way

I present three families of examples:

**Process Kinds.** Many homeostatic mechanisms, notably the ones that individuate kinds of processes, only ensure informational co-occurrence: if, as it is sometimes claimed, some diseases are HPCs (Williams 2011), their homeostatic mechanisms will not generally ensure contiguity co-occurrence in the cluster. 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 being ten years apart, but they *are* informationally connected in the intended sense: the presence of an HIV infection at, say,  $\langle \textit{Sabrina}, 1981 \rangle$  increases the probability of the onset of AIDS at  $\langle \textit{Sabrina}, 1991 \rangle$ .

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.

**Social Kinds.** While in diseases and other kinds of processes contiguity co-occurrence is simply less central than informational co-occurrence, there are HPCs for which it is an open question whether contiguity co-occurrence so much as makes sense. Consider Boyd's suggestion that *feudal economy* is a natural kind (1999, p. 155). This amounts to saying that there are causal factors that explain that the institutions and practices typical of the feudal economy tend



to appear together. That is, the claim that *feudal economy* is a natural kind amounts to saying that the following is true:

1. If vassalage is present, manorialism is likely present.
2. If manorialism is present, serfdom is likely present.

... and the like. Now, there is no obvious spatio-temporal-contiguity version of such claims: for example, it is at least awkward to claim that manorialism is next to serfdom, which is nearby vassalage. On the other hand, there is a straightforward way to unpack these claims in the IC version: they amount to claiming that vassalage being present in a society increases the probability that manorialism is present too. Here, the *locations* in which such properties are tokened are societies: entities the spatio-temporal position of which, if they have any, is, for all 1 and 2 say, undecided.

**Some Kinds of Abstract Entities.** A final example is provided by what computer programmers call *declarative programming*. This is a programming paradigm whose languages have the following characteristics: their expressions are referentially transparent (that is, a variable can be substituted by its value without modifying the behaviour of the program); the order of execution of instructions is irrelevant; they describe *what* has to be done, not *how* to do it.

Declarative programming is, likely, an HPC: the properties I have just enumerated, and some others, are clustered together, but they are neither necessary nor sufficient for something to be an instance of declarative programming. Now, the *locations* these properties cluster around are computer programming languages; but computer languages are abstract entities, neither in space nor in time—and, I take it, nothing can be next to anything else that is neither in space nor in time.<sup>8</sup>

## 2.3 HPCs and the Information-Theoretic Toolbox

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

**Describing Inner Structure.** I wrote above that contiguity HPCs are a good approximation to the nature of some kinds. In fact, even for many of

---

<sup>8</sup>Of course, these facts about declarative computer languages are governed by causal processes inhering on (very concrete) computer programmers trying to go about their information-processing tasks: whenever a language is devised, for example, to specify the appearance of websites in terms which are independent from the browser on which these pages will be rendered, a language that has more or less the characteristics detailed above will make sense. There is nothing spookily non-causal about this. But the fact remains that such properties as, say, *irrelevance of instruction execution order* are nowhere to be found in space-time.

the HPCs for which a characterisation in terms of spatio-temporal contiguity is most natural and informative, there is a great deal of internal articulation and structure that gets lost without the use of IC property-clusters. Consider again the example of shoulder and elbow constraining the position of the wrist in a normal human musculoskeletal system. While encoding this kind of structural connection is straightforward in an IC property cluster, there is no way to do so by merely enumerating properties that tend to be clustered alongside one another. Contiguity HPCs do not distinguish between haphazard agglomerations of muscle and bone, and human musculo-skeletal systems; informationally-connected property clusters do.

**Quantity of Information and Criterial Properties.** Each pair of unconditional and conditional probabilities in an informationally-connected property cluster gives rise to a measure of the *quantity of information* that a (group of) property instantiation(s) carries about another (group of) property instantiation(s). If, for example, the pair is

$$\langle P(F@l), P(F@l | F'@l') \rangle$$

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

$$\log_2 \frac{P(F@l | F'@l')}{P(F@l)}$$

bits of information about the state consisting of  $F$  being instantiated at  $l$ . Any one signal might, and usually will, carry information about many different (groups of) property instantiation(s)—let us call each such group of instantiations  $PInst_i$ . The *Kullback-Leibler distance* (see Skyrms 2010, p. 36; the original discussion is Kullback & Leibler 1951) is the weighted average of all of these quantities of information:

$$I(F'@l') = \sum_i P(PInst_i | F'@l') \log_2 \frac{P(PInst_i | F'@l')}{P(PInst_i)}$$

Where  $I(F'@l')$  is the Kullback-Leibler distance of the signal consisting of property  $F'$  being instantiated at location  $l'$ . Different properties will give rise to different Kullback-Leibler distances. The properties with 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. They are the most criterial properties for the presence of the kind in question.

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

Finally, we can calculate the weighted average of Kullback-Leibler distances in a certain IC property cluster:

$$\sum_i P(F) I(F)$$

In kinds with higher average distance, more properties, and those most frequently instantiated, are more criterial: the cluster is more integrated, imperfect homeostasis is less frequent, and inductions, as a result, would be successful more frequently. The average distance could be, thus, suggested as a measure of the naturalness of the kind in question. These elucidations of criteriality and naturalness are only available under the IC regimentation of the HPC account.

In the following section I discuss in detail the case of polymorphic kinds. I propose the fact that such kinds can be satisfactorily characterised as ICPCs as a further argument in favour of the latter.

## 3 Polymorphism

Species polymorphism abounds. Two prominent examples are sexual dimorphism in mammals and caste polyphenism in ants (that appear in, e.g., queen and worker variants). According to Ereshefsky & Matthen (2005) this shows that species cannot be HPCs: the HPC account is based on the fact that properties in *one and the same cluster* keep co-occurring. If species are HPCs, then, this seems to lead to a view of species according to which every member is, in principle, very similar to every other member, and in which typologic variability is accidental, and due to failures of homeostasis.

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 possible, straightforward approach to polymorphic HPCs: 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).

There are at least two problems with this approach. The first 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, *heterostatic*, 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 heterostaticity, and this is unsatisfactory.

The second problem is that the suggestion 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.

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.<sup>9</sup>

Before going on to describe how appealing to informational connectedness helps with the problem of polymorphism, I will briefly discuss two other proposed solutions to the problem: Boyd’s own (1999) and (Magnus 2011).

### 3.1 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

---

<sup>9</sup>There is, though, one kind of phenomenon that fits the characterisation of polymorphism I have just given, which will not be treated the IC way. A clear example is the following: many species of butterflies present mimicry-based polymorphism and, e.g., females of *Papilio dardanus* differ in wing colouration, in some fourteen (stable, discrete) varieties (see Joron & Mallet 1998 for details); but no single homeostatic mechanism is sustaining an informational connection among these morphs. 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 IC solution, 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.

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: they all have very nearly the same properties; it’s only that many of them 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. Consider a female in the aquatic stage of the species Boyd is alluding to in the passage quoted above. It is supposedly a fact that this individual has the following property: if it were a male and in the first molt, it would be P. But, as Magnus (2011) points out, for many individuals in a polymorphic species it is very unclear that there are such counterfactual facts—for one, it might turn out to be metaphysically impossible for certain males to have been females.

According to a common opinion, it *will* be metaphysically impossible for males to be females when sex is fixed at conception, as it happens in many species. At least if origin is, as Kripke (1980) suggested, essential to individuals organisms.

### 3.2 Magnus: Cutting the Homeostatic Mechanism Loose

In identifying HPCs, according to Magnus, we should follow this process:

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 is, in effect, suggesting that we abandon the idea that property clusters partially individuate HPCs. HPCs should be individuated just by their homeostatic mechanisms. Magnus argues that proponents of the (traditional) HPC theory are under the spell of *similarity fetishism*, “[the idea] that a natural kind is primarily a set of similar things” (Magnus 2011, p. 863). In particular, he suggests that in ascertaining whether a certain organism is a member of a certain species we should be willing to go as far as and wherever the homeostatic mechanism that individuates the species takes us. It will take us to each morph, which is how the HPC account should accommodate polymorphism—by cutting the homeostatic mechanism loose. Similarity fetishism, as the name implies, is a vice, and the problem of polymorphism shows precisely in which sense it is vicious.

If similarity fetishism is a vice, Magnus’s remedy is something of a puritan overreaction: the solution cannot be to follow homeostatic mechanisms wherever they might wind up, 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.<sup>10</sup>

From there, the story is familiar: such mutants might have characteristics that impede gene flow with non-mutants, and sexual interactions between males and females of this mutant may end up in an episode of speciation. More explicitly, what I am suggesting is that Magnus’s alternative to similarity fetishism will, in some cases (in most, really), force us to the following process:

1. We look around and find clusters of properties—say, we identify one sexual dimorph of our target species.
2. We look for the mechanisms responsible for these clusters—as Magnus says, we can legitimately zero in on a mechanism such as the *ongoing sexual interactions between females and males*.
3. We identify natural kinds by the scope of those mechanisms—that is, whatever results from individuals with the original cluster of properties via sexual reproduction, should be counted as part of the species: this opens the door to counting 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 occasions, imply following the homeostatic mechanism of sexual reproduction across the species boundary. Cutting the homeostatic mechanism loose has, as Magnus rightly claims, the virtue of generating all of the polymorphs: but only at the price of generating also all of the mutants, and beyond, to newly evolved species. In the context of an effort to provide the HPC account with a process to generate plausible species extensions, this is an unaffordable price.

### 3.3 ICPCs and Polymorphism

Boyd should not have felt compelled to propose his conditional-property solution: in fact, there is a “problem of polymorphism” only under the wrong reading of

---

<sup>10</sup>The same point can be made, of course, about the homeostatic mechanism that is asexual reproduction.

*co-occurrence*. In the exchange between Ereshefsky and Matthen, and Boyd it is being assumed that co-occurrence means spatio-temporal contiguity; but, as I have argued, this need not (and sometimes cannot) be so: in many HPCs the relevant notion is that of informational co-occurrence.

And, 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 across a certain range of locations—also makes the density of instantiations of the properties in the *female* cluster stay above a certain value. For a very obvious example, the presence of a (say, male) cat here now, increases the probability (all the way to 1, really) of the presence of a female and a male cat 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.

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 mammal sexual dimorphism: the presence of males of a certain mammal here now raises the probability of the presence of females in the spatio-temporal vicinity.

What kinds of homeostatic mechanisms are sustaining these informational connections? 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 other types of cases balancing selection will play a role.

In any event, the following should be made clear: the fact that there is informational-connectedness among morphs does not mean that we should be able to *infer* the existence of a morph from the existence of another—we won't be able to draw such an inference, for example, if the causal mechanism that explains the informational connection among morphs is unknown to us.

In summary, the answer to the polymorphism problem was there all along, and it doesn't involve reference to dubious counterfactual-conditional properties. Recall the spatio-temporal diagram of Colin's strain in section 2. If we do a similar exercise, starting with a mammalian sexual dimorph, we will also end up with something like a branching tree. In this case, though, there will be two

kinds of branches, corresponding to the two sexual dimorphs. The homeostatic mechanism does not just explain the recurrence of properties of the dimorph we started with, but also the fact that these properties are informationally connected, in the manner explained in section 2, with the properties of the other dimorph. The final informationally-connected property cluster will be composed by the properties of both dimorphs, and the informational relations among them.

*Pace* Magnus, then, there are no reasons to shy away from similarity fetishism about *this* kind of cluster; at least no reasons having to do with species polymorphism.<sup>11</sup> 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. Whereas Magnus would have us individuate HPC solely by homeostatic mechanism, with the ensuing problems in drawing sensible species boundaries, we should keep the homeostatic mechanism very much in check by an IC property cluster.

### 3.4 Heterostasis

Ereshefsky and Matthen have claimed that polymorphism shows that the idea of homeostasis is intrinsically ill-suited to the analysis of biological species: it is as much part of the nature of species that, say, male individuals are similar to one another, as it is that male individuals and female individuals are comparatively dissimilar:

[I]n addition to Boyd’s “homeostatic” mechanisms we need to recognize “heterotic” mechanisms that produce variation, and “heterostatic” mechanisms that maintain it. (Ereshefsky & Matthen 2005, p. 10)

We can now see that the claim that Boydian homeostatic mechanisms are not heterostatic is true only under the spatio-temporal-contiguity version of the HPC theory. If HPCs are ICPCs, very many homeostatic mechanisms are heterostatic. This has nothing to do with polymorphism, either: consider again the HPC *eukaryotic cell*. The homeostatic mechanism that partly individuates it *is* homeostatic: it explains that all packets—that is, all individual eukaryotic cells—resemble one another in the relevant way; and it is also heterostatic: it explains that each individual packet has dissimilar parts and that, for example, mitochondria are nothing like endoplasmic reticula.

This kind of inter-packet homeostasis and intra-packet heterostasis is typical. For another example, the mechanism that maintains the HPC that is the human

---

<sup>11</sup>The resulting theory, though, is still an HPC theory, and does rely on the notion of similarity. For example, its predictions regarding species boundaries 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 HPC theory in this and other respects are quite independent of the issue of polymorphism, and should be assessed independently.



skeleton is homeostatic (all skeletons resemble one another) and heterostatic (in each individual skeleton, e.g., the mandible is nothing like the humerus).<sup>12</sup>

If species are thought of as ICPCs, polymorphism turns out to be intra-packet heterostasis of that very same kind: packets in a polymorphic species are groups of individual morphs connected among them by informational links. The species' sustaining mechanism is inter-packet homeostatic: informationally-connected groups of morphs keep recurring, similar to one another. It is also intra-packet heterostatic: it maintains the differences among morphs in any single packet.<sup>13</sup>

It is easy to show how an informationally-connected HPC 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 (say) 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:

- $\langle P(A@f(l)), P(A@f(l) | C@l) \rangle$
- $\langle P(D@l), P(D@l | C@l) \rangle$
- $\langle P(C@f(l)), P(C@f(l) | B@l) \rangle$
- $\langle P(A@l), P(A@l | B@l) \rangle$

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 HPC'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 HPC'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 homeostatic: it makes the packet consisting of [an A-plus-B morph and a C-plus-D morph, 200 to 300 m from one another] keep recurring. See Figure 2 for the result of a simulation of this ICPC at play.

<sup>12</sup>For *yet* another example, the social mechanisms that explain that all capital Rs are similar to one another also explain that the top half of a capital R (closed and with a curve part) is nothing like its bottom half (two disconnected straight segments).

<sup>13</sup>Throughout the paper I am taking for granted that, in polymorphic species, kind members are individual animals. A more radical, but also more parsimonious, option, would be to abandon this piece of pretheoretical wisdom, and identify members with the informationally-interconnected groups of organisms I am calling "packets". I will not explore this possibility here.

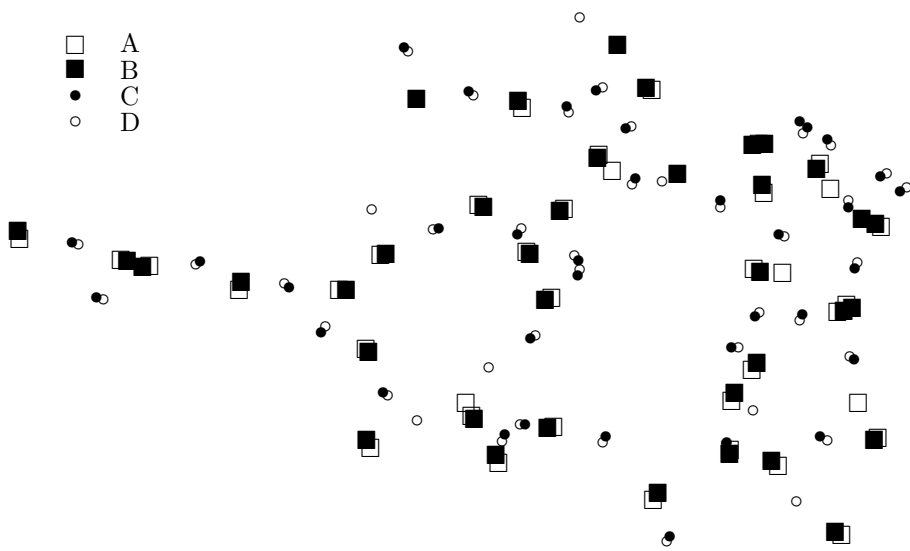


Figure 2: A toy model of a polymorphic kind with two “morphs”: black-square-plus-white-square and black-circle-plus-white-circle. Polymorphism as presented here is still compatible with imperfect homeostasis—note the occasional single square or circle.

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.*

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.

## 4 Conclusion

A homeostatic property cluster, once past the crudest level of idealization, very often has to be a collection of informational links among property instantiations. As I have tried to show, individuating HPCs by these kinds of structures has several theoretical merits: 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 facilitates the use of information-theoretic tools in the description of natural kinds.

On the other hand, this implies no fundamental modification of the HPC account. For example, HPCs 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.

## 5 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.
- 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 Flow 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.
- Joron, M & Mallet, JLB 1998, 'Diversity in Mimicry: Paradox or Paradigm?', *Trends in Ecology and Evolution*, vol. 13, no. 2, pp. 461–466.
- Kripke, S 1980, *Naming and Necessity*, Blackwell.
- Kullback, S & Leibler, RA 1951, 'On Information and Sufficiency', *Annals of Mathematical Statistics*, vol. 22, pp. 79–86.
- Lavric, V & Graham, DW 2010, 'Birth, growth and death as structuring operators in bacterial population dynamics', *Journal of Theoretical Biology*, vol. 264, no. 1, pp. 45–54, viewed <<http://www.sciencedirect.com/science/article/pii/S0022519310000329>>.
- 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.
- 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.
- Williams, NE 2011, 'Arthritis and Nature's Joints', in JK Campbell, M O'Rourke & 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.