

Transitional Gradation in the Mind: Rethinking Psychological Kindhood

Cameron Buckner

Forthcoming, British Journal for the Philosophy of Science

Abstract: I here critique the application of the traditional, similarity-based account of natural kinds to debates in psychology. A challenge to such accounts of kindhood—familiar from the study of biological species—is a metaphysical phenomenon that I call ‘transitional gradation’: the systematic progression of slightly modified transitional forms between related candidate kinds. Where such gradation proliferates, it renders the selection of similarity criteria for kinds arbitrary. Reflection on general features of learning—especially on the gradual revision of concepts throughout the acquisition of expertise—shows that even the strongest candidates for similarity-based kinds in psychology exhibit systematic transitional gradation. As a result, philosophers of psychology should abandon discussion of kindhood, or explore non-similarity based accounts.

1. Introduction: The Similarity Thesis

A fundamental idea about natural kinds is that their members are mutually similar, so much so that Quine even suggested ‘the notion of a kind and the notion of similarity or resemblance seem to be variants or adaptations of a single notion’ ([1994], p. 42). The similarity that kind members bear to one another is thought to account in part for kinds’ scientific import. For one, it will be rational to infer the properties of the entire kind from the properties of a sample, because the members of the sample and the members of the whole population will be similar in scientifically important respects; for another, there will be many true generalizations ‘out there waiting to be discovered’ about categories whose members are similar in many different ways. Moreover, sciences describing domains containing many distinct kinds will be good subjects for

taxonomy, for if two kinds are distinct, then the way in which members of one kind are similar will differ from the way in which members of another kind are similar, and members of different kinds can be distinguished by these characteristic similarities. Let us call this set of ideas the Similarity Thesis (hereafter ‘ST’); to summarize,

The Similarity Thesis: A natural kind is scientifically important in virtue of a distinctive similarity shared amongst all and only members of the kind.

The relevant notion of similarity, however, has proven notoriously difficult to characterize; Quine himself worried that it is of ‘dubious scientific standing’ and ‘logically repugnant’ ([1994], pp. 42-43). Specifically, members of kinds might be held similar in different ways, to different degrees, and at different levels of description, and it is not clear which types of similarity should be required for kindhood.¹

Notably, categories which grade into one another present special challenges to ST, for category divisions with transitional borderline cases frustrate our attempts to locate precise and distinctive metrics of similarity characterizing members of each kind. In the philosophy of biology, this problem has been acknowledged in the case of species. Though Putnam and Kripke offered tigers, beech trees, and *Homo sapiens* as paradigm examples of natural kinds (locating their essential similarity in shared genetic codes), ST has been heavily criticized in the species debate due to the fact that systematic heritable variation and gradual change—at phenotypic, genotypic, and epigenetic levels of description—are required for natural selection to occur. As a result, biologists have discovered many transitional forms between related species, and this systematic proliferation of borderline cases persistently embarrasses any attempt to identify

¹ See also Magnus ([2011]) for a critical discussion of the hold that ST has had on the theory of kinds in biology.

characteristic forms of similarity for each species. Synchronic cases of transitional gradation can be found most dramatically in the cases of species rings (Figure 1) and other forms of cryptic species complexes; but if we expand our interest diachronically to evolutionary time scales, every species will blur into its phylogenetic forebears via a variety of transitional forms.

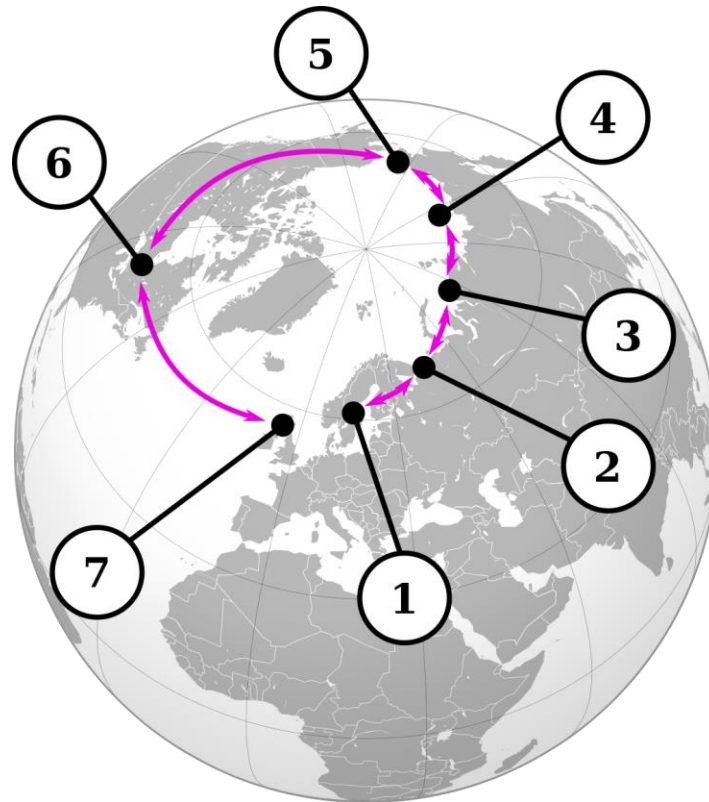


Figure 1. Common example of a species ring involving seagulls around the Arctic Circle. *Larus fuscus* (1) blends into its distinct Siberia population (2), which blends into *Larus heuglini* (3), *Larus argentatus birulai* (4), *Larus vegae* (5), *Larus smithsonianus* (6), and *Larus argentatus* (7). Each population can hybridize with its neighbors, but *Larus fuscus* and *Larus argentatus* are phenotypically distinct, genotypically distinct, and cannot successfully interbreed. (Image source: Wikimedia Commons, S. Solberg.)

To be clear, the difficulty is not just that a few hard cases exist—we could grant ST exception for even a large number of tricky borderline cases, if it worked for the less complex cases. The point is rather that the hard cases expose the inadequacy of ST even in the supposedly easy ones. Nomologically, there is nothing abnormal about cryptic species complexes, and all other species exhibit the same forms of intrapopulation variation to less dramatic degrees.² Furthermore, the most serious challenge to ST in systematics is not the mere existence of variation, but rather the fact that heritable variation and gradual change is a primary mechanism by which natural selection occurs ([Sober [1980]]). This is a resilient and subtle problem; though there are a variety of devices that could be used to bracket off or abstract away from this variation, its existence is required to explain, through speciation, how species members came to possess the properties that they do. As a result, all such devices must fail. In short, ST seeks to establish species as natural kinds to secure their explanatory legitimacy in biology, but any device that successfully abstracted away from the gradation that threatens species' kindhood would rob them of the very feature by which they participate in one of biology's central explanatory projects.

While controversy still surrounds this challenge to the kindhood of species, there is no doubt that it presents a serious complication, with many philosophers of biology having lost faith in the

² Of course, some mechanisms and pressures favor genetic and phenotypic homogeneity. Others, however, favor heterogeneity. The point is that there is no general law about species as such that secures homogeneity and for which ring species would have to be regarded as nomologically deviant.

ability of ST to arbitrate taxonomic disputes.³ By contrast, the theory of natural kinds in philosophy of psychology—especially variants of the homeostatic property cluster (HPC) view, which holds kindhood to be determined by an accommodation between a characteristic set of properties and underlying mechanisms that cause those properties to non-accidentally cluster in category members⁴—is currently enjoying a surge in popularity. One of the most prominent recent debates, centering around the work of Machery [2009], questions whether concepts are an HPC kind—on the assumption that any posit found not to be a natural kind should be eliminated from psychology. Instead, Machery recommends that psychologists focus attention on the genuine natural kinds in this area of study, which he takes to be the major subtypes of concepts: prototypes, exemplars, and theories.

I here urge caution, on the grounds that the subject matter of psychology, like that of biology, involves a proliferation of explanatorily-important transitional forms. This will be old news to some philosophers of psychology and especially philosophers of psychiatry, who have long argued that transitional forms between mental illnesses (such as anxiety and depression) stymie essentialist approaches to psychological kinds (Zachar [2000]; Haslam [2002]).⁵ My goal here is

³ It should be noted that several new approaches have recently been offered to the ‘species problem’ that are designed to rely less on distinctive similarities—such as, Ereshefsky & Matthen’s ([2005]) ‘population structure theory’.

⁴ In the interests of space, I here do not elaborate on the nature of accommodation; for more see Boyd ([1999]).

⁵ The clearest case so far is depression, where large, replicated empirical studies have failed to find any evidence of an underlying discontinuity between normal unhappiness and depressive disorders—see Ruscio and Ruscio ([2000], p. 2002).

thus not merely to demonstrate the existence of transitional forms in psychology—which should by now be obvious. Rather, my aim is to understand this challenge in its most general form (as applying to any similarity-based account of natural kinds, rather than narrowly against essentialist approaches) and to press it against some of the strongest candidates for such kinds in psychology (in particular, prototypes and exemplars). The argument, moreover, will generalize to any gradually-modified representational state, extending the concerns about transitional forms in psychology well beyond the commonly discussed examples in philosophy of psychiatry. In so doing, I problematize the eliminativist style of argument favored by authors like Machery—for if even the standard sub-types of concept (such as prototypes, ideal models, and exemplars) fail to count as natural kinds in the relevant sense, this assumption would lead to an ‘eliminative regress’ that would remove from consideration large swathes of psychology’s subject matter. Psychology must talk about something, so if the threat posed by transitional forms applies so broadly, this outcome would serve as a *reductio* of the use of similarity-based kindhood as a yardstick for legitimacy in psychology.

To make this case, I offer a general treatment of the metaphysical challenge presented by transitional forms (Section 2), and argue that this type of situation will also routinely appear in a core area of psychology, the study of concepts (Section 3). Finally, I explain why these transitional forms are explanatorily central to psychology, so no attempt to abstract away from them can succeed without serious costs (Section 4). My specific arguments here appeal to psychological theories about the structure of concepts—roughly, hypotheses about the ‘bodies of knowledge retrieved by default when categorizing, reasoning, drawing analogies, making inductions, and so on’ (Machery [2010]). My key premise is that learning is a process of gradual modification and selection that will—like random mutation and natural selection in evolution—

systematically introduce a series of intermediate forms between standard types of concept such as prototypes and ideal models. As a result, ST will face serious challenges when applied to taxonomic disputes (i.e. ‘lump or split?’) debates in the psychology of concepts.

A note before beginning: writing about the scientific categories psychologists deploy in the study of their subjects’ categorization procedures presents special terminological challenges. To pre-empt possible confusion, the central thesis of this article concerns the question as to whether the category CONCEPT, understood as a class of mental representations (a category whose extension includes individual concepts like SPOON, TREE, ELECTRON, MOTHER, etc.), forms one or a small number of similarity-based natural kinds in psychology. This concern should be strictly distinguished from worries about the natural kind status of any particular concept, like CHIMPANZEE, (a category whose extension includes individual chimpanzees like Washoe, Ai, Kanzi, Nim Chimpsky, etc.). We will discuss the natural kind status of some particular species concepts (such as CHIMPANZEE) in Section 2, but only as an analogy for the way that transitional gradation amongst conceptual structures can pose challenges to similarity-based approaches to psychological kinds. Just as transitional gradation amongst species members might pose a problem for the natural kind status of a particular species concept like CHIMPANZEE in biology, I claim, so will it pose problems for the natural kind status of the general category CONCEPT in psychology. With this all kept in mind, let us begin.

2. Transitional Gradation

Let ‘gradation’ between putative kinds name the metaphysical situation in which the extension of one category blurs into the extension of another category or categories without an obvious dividing line. The simplest case occurs when we have two candidate kinds, A and B, with A characterized by a set of property values P_a and B characterized by the disjoint set of

property values P_b , and in nature individuals can be arranged along a continuum from instances displaying mostly P_a , to instances displaying a mixture of P_a and P_b , to finally instances displaying mostly P_b . This simple scheme can be complicated in a variety of ways depending upon the subject matter. The properties involved can be binary (for example, ‘has a tail’), monotonic (‘height’), multivalent (‘color’), or complex (‘spotted’). While the continuum might be unidimensional, it will more typically be multi-dimensional (with variation in values for many different properties). Whatever complications are introduced, gradation so defined will come in degrees, depending upon how clustered instances are in multi-dimensional property space in roughly the ways suggested by the putative kind groupings. Gradation will furthermore be ‘transitional’ when there are systematic uni- or bi-directional processes of modification (such as evolution, development, or learning) that govern the gradation of individuals from one to another putative kind(s). The severity of transitional gradation as a threat to the (distinct) kindhood of A and B is determined by how numerous and important are the cases lying in the middle portion of the continuum, and how explanatorily central the transitional processes and instances are to the discipline(s) in which A and B are candidate kinds.

Rather than focusing on traditional essentialist notions of kindhood—which, due to their association with necessary and sufficient conditions for kind membership, have become an easy target in the life sciences—these issues may be more generally understood by focusing on the account of natural kinds more commonly invoked by philosophers of psychology, homeostatic property cluster (HPC) theory. According to Boyd, who offers the most worked-out version of such a view, natural kinds are ‘established through a sort of bicameral linguistic legislature in which we and the world jointly legislate’ ([1999], p. 89). Less metaphorically, the natural kind status of a category is determined by an accommodation between what Boyd calls a term’s

‘programmatically’ and ‘explanatory’ definitions. A programmatic definition specifies the causal or explanatory role that some body of theory, models, or other scientific practices expects will be played by members of a category, often taking the form of a cluster of properties or attributes. An explanatory definition then describes the underlying properties or structures common to members of that category that explain how the term’s programmatic definition could be at least approximately true (or, more generally, how inductive and explanatory practices making use of that programmatic definition could be at least approximately successful).

On this view, a term names a natural kind in some science when an accommodation can be reached, adequate to the needs and standards of that science, between that term’s programmatic and explanatory definitions; when there are in fact common causal powers or mechanisms that explain how its programmatic definition could be non-accidentally true of the members of its extension. I will hereafter refer to the causal powers or mechanisms found in a natural kind term’s explanatory definition as ‘underlying structures’, with two caveats to forestall misunderstandings: 1) the relevant structures may be extrinsic with respect to category members (such as predation risk) and 2) ‘underlying’ here is understood epistemically rather than metaphysically (i.e. HPC theory holds that we may successfully refer to kinds without knowing their explanatory definitions, not necessarily that explanatory structures always arise from ‘lower-level’ sciences or mereological parts).⁶

⁶ For example, it is often assumed that HPC theory requires that the causal mechanisms that secure a kind’s programmatic definition must be parts of kind members (in the way H₂O molecules compose samples of water) or from ‘lower level’ sciences (in the way that molecular physics is presumed to be a more fundamental science than chemistry). This assumption is likely a holdover from reductionist approaches to kindhood, however, as these are not the senses of

With this terminology in place, we can now distinguish two possible expressions for the Similarity Thesis within the HPC framework: first, members of a kind may be deemed similar because they share a cluster of characteristic or surface properties (i.e., the sort that would be found in a term's programmatic definition), and second, members may be deemed similar because they possess or are shaped by similar underlying structures (i.e., the sort that would be found in a term's explanatory definition). These two dimensions of similarity may be assessed independently.

Particular debates over kinds have too often been cast in absolute terms—either natural kinds or bust!—despite nearly all theorists conceding that kindhood comes in degrees. Perhaps more clarity may be achieved by schematically distinguishing different types of challenge to ST, with the concession that certain types or degrees of accommodation between a programmatic and an explanatory definition will better support a wider range of scientific activities than others. I propose a system of four classes of accommodation, with the latter classes presenting increasingly difficult challenges to ST (Figure 2). In the ideal, limiting case—a 'Class I' accommodation, which presents no challenge to ST—high surface similarity obtains and is

'underlying' required for Boydian accommodation. For example, evolutionary biologists have sometimes expressed surprise that many Batesian mimics only poorly resemble the noxious species they are supposedly mimicking. In a fascinating study, Chittka & Osorio ([2007]) show how quirks of generalization patterns in predator learning might actually make imperfect mimicry more adaptive. In other words, predator learning mechanisms might underlie the surprising morphological features of a mimic in the relevant sense (serving as the mechanism that causally secures and explains them), despite the fact that neither are those learning mechanisms a part of the mimic's body nor is psychology a 'lower' science than morphology.

explained by the operation of highly similar underlying structures. A more tenuous ‘Class II’ accommodation involves transitional gradation at the surface level of description, but which is explicable by modulation of highly similar structures at an underlying level of description. A rougher still ‘Class III’ accommodation involves transitional gradation in both surface properties and underlying structures. Finally, there is yet another challenge of the sort that troubles Craver, wherein differing levels of abstraction with which to describe underlying explanatory structures favor different kind boundaries at the surface level of description (a ‘Class IV’ accommodation). Since Craver ([2009]) has recently written extensively about this kind of challenge—his primary case study shows how more or less abstract descriptions of the hippocampus (pictorial images of specimens, informational flow diagrams, or computational models) recommend wider or narrower extensions for ‘memory’—I will not address it here. However, there is much to say about how it can interact with the other types of challenge in interesting and bewildering ways.⁷

⁷ Some authors have explored even weaker notions of kindhood based on even more extreme challenges to accommodation. For example, Haslam ([2002]) proposes ‘practical kinds’ and ‘fuzzy kinds’, which would both fall to the right of a Class IV accommodation because neither can be characterized by any distinctive form of similarity (surface or underlying). I do not discuss these notions here because I think it unlikely that the subtypes of concepts present this deep a challenge to accommodation.

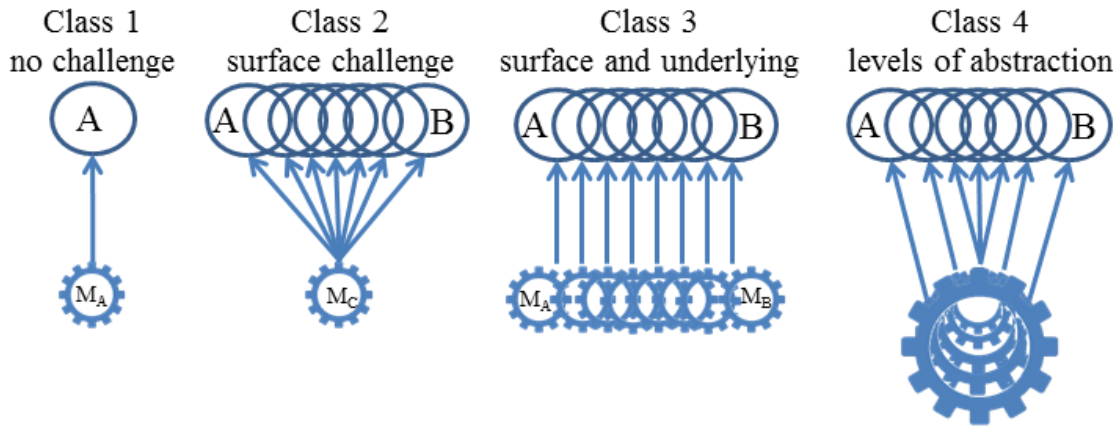


Figure 2. Possible relationships between a property cluster and underlying structures in terms of increasingly serious challenges to accommodation arising from transitional gradation between two putative kinds A and B. Circles denote property clusters and gears denote underlying structures (M_a denoting structures that explain A's characteristic surface properties, M_b denoting structures that explain B's characteristic surface properties, and M_c denoting structures the modulation of which can produce either A or B surface properties); overlapping shapes indicate degrees of similarity, and size of gears indicates degree of abstraction with which structures are described.

To be clear, the general problem posed by transitional gradation is not that nearby kinds must be considered mutually exclusive and that transitional cases between kinds might satisfy the criteria for both. In some cases, there may indeed be some important theoretical reason why kinds must be mutually exclusive; concerning species, for example, if too many individuals were members of more than one species, then they would join together those two putative species through interbreeding. However, this feature is particular to the species problem and emerges only because individuation criteria for species happen to involve the same process that introduces transitional gradation (namely, interbreeding). Rather, the deeper problem posed by

transitional gradation—the problem that generalizes beyond biology—is that where transitional gradation proliferates, nature does not provide us with any clear discontinuities that would recommend a particular metric of similarity for either kind (whether or not overlapping) as more or less natural than a host of other eligible alternatives. Metaphorically, we reach out to nature to feel for a joint and find only smooth continua. We may cut these continua at particular locations, but any particular cut will be arbitrary in the sense that, even with perfect empirical knowledge, the cut’s location would remain contestable just to the degree that transitional gradation proliferates. In cases of Class III challenge, where we find smooth continua at both surface and underlying levels of description, no natural similarity metric can be located, and applying similarity-based accounts of kindhood to settle borderline disputes regarding about cases will be otiose.

3. Transitional Gradation in Psychology: Prototypes and Ideal Models

While transitional gradation has received the most attention in philosophy of biology, in this section I review evidence that it will also complicate philosophy of psychology by exploring the recent debate as to whether concepts form a natural kind. In particular, I argue that the concepts debate will, like the debate over species, be characterized by (1) routine transitional gradation that is (2) explanatorily important. The problem is that learning, like random mutation and natural selection, is a process of gradual modification and selection, one in which suitable representations are incrementally revised in response to evidence and placed in positions of control over appropriate behaviors. As such, we should expect that transitional gradation in representations’ characteristic properties will pose an enduring challenge to the taxonomy of learned representations. Moreover, this variation will be explanatorily important for any project

that includes ontogeny in its explanatory purview. As such, as with species, ST is unlikely to resolve major disputes about the taxonomy of concepts.

To further narrow our attention, the present case study focuses on prototypes, a particular subtype of category representation. This topic is relevant both as a general test case for the application of the theory of kinds to an important posit in psychology and to recent claims of Machery ([2009], [2005]), who has argued that concepts are not a natural kind (on the HPC account). He holds that to explain human performance on categorization and inference tasks, we should eliminate ‘concept’ from psychological theorizing and instead focus directly on prototypes, exemplars, and theories. As mentioned above, however, this approach would lead to an eliminative regress if these subcategories are not themselves certified as HPC kinds. Thus, it is imperative for the plausibility of Machery’s position—and the viability of the HPC approach to ontological disputes in psychology more generally—that some standard types of category representation turn out to be HPC kinds.

In defense of claim (1), consider the work of Barsalou, which offers several *prima facie* threats to the conclusion that concepts (especially prototypes) are a natural kind. He has argued that in general prototype and exemplar theories cannot be empirically distinguished (because whatever a prototype theory can do at the time of learning, an exemplar theory can do at the time of inference—see Barsalou, [1990]) and that their structure is not stable (but rather constructed ‘on the fly’ in response to particular circumstances—see Barsalou [1987]). In his book, Machery rebuts these particular arguments; I will here focus on extracting a yet deeper challenge from Barsalou’s work comparing and contrasting prototypes with ideal models. Machery ([2009], pp. 117-118) considers the possibility that ideal models may be a fourth sub-kind of concept. While conceding that the data presented by Barsalou are suggestive, Machery left off the consideration

of ideal models there because he felt there had been insufficient empirical work on the subject. While this assessment was apt at the time Machery began his critique, ideal models have since been studied extensively by a thriving research program. This program, combined with general considerations about the nature of learning canvassed below, require us to take ideal models seriously now.

Let us then consider that evidence. Inspired in part by the Wittgensteinian idea that many concepts are characterized by ‘family resemblances’ rather than sets of necessary and sufficient conditions, prototypes are usually defined as being structured around weighted lists of features, none of which are essential for kind membership. Notably, Rosch and Mervis ([1975]) originally characterized prototypes themselves with a property cluster. Emphasizing that they offer a ‘description of structural principles’ rather than a ‘processing model’, they identified prototypes as analog category representations, exhibiting graded, similarity-based membership, consisting of weighted lists of features selected by frequencies of instantiation (relative to other categories), and for which no feature or subset of features was essential (with an individual’s category membership coming in degrees, as determined by the number and importance of features it possesses).

However, Barsalou ([1985]) found that subjects’ responses on categorization tasks thought to elicit prototypes were also frequently influenced by similarity to an ideal model. Ideal models are structured around ideal rather than typical features—those considered the ‘best’ for the category, given some interests or goals. For example, Barsalou suggests that the ideal ‘foods to eat on a diet’ are those with ‘zero calories’, and the ideal ‘things to take from one’s home in a fire’ are ‘highest possible value’ ([1985], p. 630). Such ideal values are rarely the most frequently instantiated in members of the category, and so these two sorts of category

representation would seem to be, focusing merely on their characteristic definitions, distinct kinds of representation generated in entirely distinct ways. If this were right, one would expect to find people's categorization judgments sensitive to typical features for some categorization tasks, and to ideal features for another, disjoint set of tasks; and thus (if we consider current evidence adequate), Machery's list of three kinds of category representation would need to be expanded to four.

Notably, however, the responses of Barsalou's subjects did not reflect two distinct subsets of graded category representations, one organized around typicality and another organized around ideality. Rather, the category representations investigated by Barsalou all seemed to be organized around a mixture of the two. Barsalou ([1985]) investigated the hypothesis that representations for goal-derived categories (such as 'restaurants to eat at') would be structured around ideality, whereas representations for taxonomic categories (such as 'mammal') would be structured around typicality (central tendency and frequency of instantiation). While ideality and typicality had the greatest influence on the structures of goal-derived and taxonomic categories respectively, both were significant predictors of category structure for all kinds of categories reviewed. Moreover, the degree to which typicality or ideality governed a category representation's structure appeared to be a function of, among other things, the types of discriminations made and the amount of the subject's goal-related experience (i.e. on structured discrimination tasks evaluable in terms of success or failure and for which feedback about success is readily available). This basic finding has since been confirmed with a variety of subjects on a variety of tasks, including tree expert's judgments on trees (Lynch, Coley, and Medin [2000]), U.S. bird experts and Itza Maya foragers on birds (Bailenson, Shum, Atran, Medin, and Coley [2002]), experienced fisherman on fish (Burnett, Medin, Ross, and Block

[2005]), trained subjects on artificial categories (Levering & Kurtz [2006]), and untrained subjects on role-governed categories (Rein, Goldwater, and Markman [2010]). Moreover, in cases where comparisons were possible, ideal features were not invariant but rather relative to the types of tasks on which the subjects had specialized (such as the contrast between bird experts and foragers found by Bailenson et al. [2002]).

If this story is correct, one might wonder why so many other studies on categorization since the 1970s did not reveal an influence of ideal features. Here, we have a ready and plausible explanation: these studies largely focused on the everyday taxonomic categories of university undergraduates. Few studies were performed on experienced subjects or on specialized categories, and studies on artificially-learned categories typically included only a minimal training phase. Thus, it is not surprising that the category representations revealed by these studies were structured mostly around frequency of instantiation, with the effects of goal-related experience having gone unnoticed.

These results put pressure on the thesis that prototypes, as defined in terms of typicality, form a natural kind. To connect back to our earlier discussion about ST, we might hope that ‘prototype’ would name a natural kind in virtue of particular prototypes sharing a similar, distinctive category structure—namely, a weighted list of features selected by validity or frequency of instantiation. The problem is that the category representations structured around typicality appear to grade smoothly into the category representations structured around ideality (through learning), with a significant number of category representations structured around both typicality and ideality. As with a species ring (recall Figure 1), the endpoints of this continuum are significantly dissimilar, but the drawing of any specific, similarity-based distinction between prototypes and ideal models appears arbitrary (in the sense indicated in Section 2). Thus, the

case of prototypes and ideal models demonstrates the presence of significant transitional gradation in psychology. Below, I argue that this transitional gradation presents at least a Class II challenge to accommodation—already a significant challenge to ST—and conclude by reviewing preliminary evidence that it actually rises to the level of the even more problematic Class III.

4. The Explanatory Import of Transitional Gradation in Psychology:

Ontogeny and Expertise

4.1 The importance of ontogeny

In this section, I relate the importance of the transitional gradation between prototypes and ideal models to a core project in cognitive science: the study of expertise. In short, to explain the transition from novice to expert we must appeal to the gradual transition in a subject's domain-related category representations, as a product of goal-related experience, from prototypes to ideal models. In this sense, the transitional gradation between prototypes and ideal models will be explanatorily important to psychology, just as transitional gradation in population members is to biology. Attempts to idealize away from this gradation will fail to support monolithic answers as to whether to lump or split sub-categories of concepts in response to challenges, for the cost of disregarding this variation would be too great to bear.

Let us consider the adequacy of some obvious responses to this transitional gradation problem facing the taxonomy of concepts. The simplest response would be to maintain prototypes and ideal models as distinct natural kinds by ignoring or explaining away the gradation between them as peripheral to our interests. But this solution does not seem very promising in the case of prototypes and ideal models, for at the very least, a theory of concepts ought to make predictions about our categorization decisions, and it cannot do so without

specifying the structure of category representations by saying what types of features are most important in deciding whether and to what degree new exemplars will be judged members of the target category. This is just what particular accounts of ‘typicality’ and ‘ideality’ specify, so this transitional gradation cannot simply be ignored.

Nevertheless, one might still hope that prototypes and ideal models could be regarded as distinct natural kinds if the borderline cases between them were somehow rare, accidental, or unsystematic. Crucially, I argue, the borderline cases cannot be so regarded for any project that includes the acquisition of expertise within its explanatory purview, for here this transitional gradation is both systematic and explanatorily important. Indeed, the possession of specialized category representations is regarded as one of the primary distinguishing marks of expertise (see Ericsson & Lehman [1996] for a review). In other words, the transitional gradation from frequency-based prototypes of novices to the distinctive category representations of experts will be, as transitional gradation amongst population members was for speciation, a primary explanans of the distinctive features of expert performance. A robust finding about expertise is that it is typically acquired only gradually through extensive, deliberate practice (ten years is a frequently cited estimate) on structured problem domains with readily available, accurate feedback. Such deliberate practice gradually produces an incremental specialization of the trainees’ conceptual schemes as they attempt to improve their performance on commonly practiced tasks. In short, any view that abstracted away from the gradation between prototypes and ideal models would render the acquisition of expert performance a mystery.

Let us consider the systematic nature of this gradation in more detail. A recent consensus in the ideal models literature is that central tendency, frequency of instantiation, and ideality are all determinants of graded category structure to one degree or another, depending upon several

variables such as category exemplar structure, presence of contrasting categories, and especially amount of goal-related experience.⁸ Why would greater experience lead to a focus on ideal rather than typical features? The most popular theory holds that repeatedly discriminating between competing options on tasks with feedback gradually biases our representations towards ideal rather than typical features, since such ideal values minimize prediction error in common discrimination tasks. The error reduction effect arises from the fact that when categories can be contrasted along common dimensions, constantly distinguishing between them produces ‘repelling forces’ between their category representations (Davis & Love [2010]), gradually rendering their representations more distinct. Idealizing, exaggerating, and caricaturing these differences will in some conditions support more accurate categorization decisions, because it can minimize the perceived similarity between a category representation and the exemplars of opposing categories.⁹ Moreover, the ‘direction’ in which the representations migrate will depend upon the specific discriminations most frequently made by the learner.

The defender of ST might here respond that even if it is granted that prototypes and ideal models cannot be regarded as distinct natural kinds, we can still lump them into a new superkind of category representation—which I will refer to as ‘C’. Such lumping would be defensible, the

⁸ It is worth noting that while both central tendency and frequency-based measures such as cue validity were often grouped together under ‘typicality’, these two measures routinely produce distinctly structured category representations and probably should never have been conflated.

⁹ Though theorists do not often go into this level of detail, this argument should be paired with a signal detection analysis of tasks reporting relative costs and rates of such as false positives to correct identifications. Caricaturing will typically minimize false positives, but perhaps at the cost of losing some correct identifications.

thought goes, if we can locate a set of sufficiently numerous and important properties P_c that cluster in instances of C to underwrite C 's kindhood without relying on typicality or ideality.

This lumping strategy, however, fails for the same reason as the splitting strategy just considered: a natural kind's specification (its 'programmatically definition', in Boydian terms) must include the set of properties possessed by category members that allow them to play their inductive and explanatory role in the science that studies them, which in this case must appeal to typicality and ideality. For example, we might attempt to characterize C as the set of category representations structured as weighted lists of features that are normative of category membership, without saying anything about how those features are selected. While this specification for C may be appropriate for certain 'high-level' explanatory interests (such as Weiskopf [2007]) it is problematic for Machery's purposes, because the psychological models on which Machery focuses are crucially concerned with predicting subjects' categorizations and inductions, and we cannot predict subjects' responses without knowing which features will show up in these lists. The next obvious idea would be to add a disjunctive property to this lumped specification of the form 'typical \vee ideal', which would fully capture these features. However, such disjunctive specifications are inimical to similarity, and to admit that they are required to capture the most explanatorily-important characteristics of concepts is precisely to abandon the idea that all concepts are mutually similar in the first (i.e. surface/programmatically) sense relevant to ST.

In the language of previous sections, we must conclude that the case of prototypes and ideal models presents at least a Class II challenge to accommodation—which is already a significant blow to ST. However, a more subtle form of similarity-based lumping (focusing on the term's 'explanatory definition') might still be acceptable on a Class II accommodation if we can locate

a structure underlying all of *Cs* instances, M_c , that causally explains the distribution of typical and ideal features in category representations—in other words, by conceding surface dissimilarity and focusing on similarity in underlying explanatory mechanisms.

We are not yet in a position to fully judge the latter license because the most likely source of these underlying structures would be neuroscience, and the neuroscience of category learning remains unsettled—though there is already some suggestive computational evidence. Specifically, the amount of a subject’s goal-related experience appears to be a variable that modulates shared underlying learning mechanisms, determining the degree to which these mechanisms produce representations structured around typicality or ideality. Several recent computational models of category learning based on the minimization of prediction error suggest that this bet may pay off. Many of these models operate on the idea that a single mechanism attempting to minimize category prediction error will emphasize typical features and central tendencies on common taxonomic categories with which subjects have smaller amounts of experience, but will emphasize ideal values along shared dimensions that allow subjects to distinguish options in commonly repeated discriminations (Voorspoels et al. [2011]; Voorspoels et al. [2013]).

While not strongly vindicating ST about concepts, a Class II accommodation along these lines might yet show how the case of concepts is less problematic than the case of species, where we also find transitional gradation in underlying explanatory structures (such as genetic sequences, epigenetic mechanisms, or environmental pressures). However, there remain two sources of pessimism—the first rather strong, the latter more speculative—suggesting that gradation between prototypes and ideal models must rise to the level of a Class III challenge to accommodation: evidence that experts develop distinct learning strategies throughout the

acquisition of their expertise, and the likelihood that the comparative study of concepts will expose transitional gradation in the neurological mechanisms underlying categorization. Should this case rise to the level of a Class III challenge to accommodation, then it will prove fatal to ST, for we will not be able to locate any distinctive similarity—surface or underlying—shared amongst all prototypes or ideal models. Before closing, I consider each in turn.

4.2 Mechanisms underlying expert concepts (in humans)

The study of expertise has found that experts not only develop specialized concepts, but also that these concepts allow them to iteratively bootstrap qualitatively distinct, specialized learning strategies, which in turn enable further gradual specialization of those concepts (see Ertmer & Newby [1996] for a review). Consider familiar examples from the study of chess expertise; whereas a chess novice will struggle to learn new strategies by tracking the placements and ordering of individual moves, grandmasters perceive games in terms of higher-order memory chunks encompassing entire board positions and multi-move sequences, allowing them to learn and remember more sophisticated strategies within the same working memory constraints. Across dozens of other subject domains, expert learning has similarly been shown to take advantage of more organized storage and retrieval of knowledge, increased perceptual and motor acuity, and especially distinctive forms of creative exploration and diagnostic engagement with training tasks. Some of these advantages are obvious, such as being able to use background theories to diagnose the causes of failure; others are subtler and broader-based, like substantially increased motivation, attention, and emotional resilience to training challenges. Moreover, experts can deploy all of these mechanisms more efficiently than the novice, for they are more sensitive to task demands, and more flexible in their use of strategies given their enhanced ability to deploy metacognitive and metatheoretic representations of tasks and interactions with one's

own abilities, background, strengths, and weaknesses. In short, the acquisition of expertise routinely includes ‘learning how to learn’ in qualitatively distinct ways, with the newly-attained concepts and strategies gradually changing the nature of learning many times during the acquisition of expertise. These iteratively developing learning mechanisms constitute significant transitional gradation at the level of the underlying mechanisms that explain the transition from prototypes to ideal models, and again the gradation itself is required to explain distinctive features of expertise.

To consider a yet more sophisticated ‘splitting’ objection, the defender of ST might argue that this evidence reveals that experts do not actually iteratively improve their novice concepts, but rather gradually lose their novice concepts and replace them with new and distinct expert concepts. This rebuttal does not, however, rest on a viable interpretation of the empirical evidence in this area. This point has already been somewhat illustrated through the discussion of the computational models that predict the gradual revision of novice prototypes into expert ideal models, but the defender of ST might argue that these models operate at a level of abstraction which obscures the emergence of distinct expert concepts. To further argue against this objection, I end the discussion of the expertise literature by showing how two other independent lines of research on strong candidates for qualitatively distinct expert concepts—configural rules and higher-order concepts—still illustrate the importance of transitional gradation in concept learning.

In the judgment and decision-making literature, an expert is said to categorize using a configural rule when the impact of one variable is highly-dependent upon the values of other cues in context (see Camerer & Johnson [1991] for a review). Configural rule theory is consistent with the idea that experts use graded category structures like a prototypes or ideal

models, but where the specific weightings of the cues is highly context-sensitive. The configural rules of experts have been found to differ significantly from the categorization techniques of novices, in that experts decide on the basis of fewer cues, use more context-sensitive weightings, focus more on causal cues, and use more ‘broken-leg’ cues (which are rare but highly-diagnostic of category membership, again echoing a theme of the ideal models literature). There is, however, no evidence that at any point in the acquisition of expertise learners suddenly switch from naïve prototypes to expert configural rules. Rather, research has shown that these distinctive features of expert categorization emerge gradually by ‘trying to fit increasingly sophisticated general rules to past cases’ (Camerer & Johnson, [1991], p. 208). In other words, they are iteratively built by applying and gradually modifying the conceptual structures of novices in response to evidence of success or failure on prior tasks. There is no way to make sense of these iterative revisions except by reference to the previous knowledge structures and their successes or failures on decision tasks. The proposed splitting interpretation is thus not viable here, as again we only find more transitional gradation.

On the other hand, it is obvious that experts in many domains do acquire some genuinely new concepts, for they routinely deploy specialized technical vocabulary. One of the most thoroughly studied domains featuring such vocabulary is medical diagnostics. Research on expert diagnosticians reveals an apparent challenge to the ubiquity of transitional gradation in psychology, for psychologists have discovered violations of the traditional idea that the hallmarks of expertise emerge monotonically as a linear function of expertise level (as originally suggested by the classic studies of chess expertise conducted by De Groot [1946/1978]). Boshuizen and Schmidt have found evidence that qualitatively distinct categorization strategies were favored by novice, intermediate, and expert physicians (see Rikers, Schmidt, and

Boshuizen [2002] for a review). In particular, novice medical students appear to categorize on the basis of a few surface cues emerging from ‘a limited understanding of biomedical knowledge and a lay type of clinical knowledge’ (Rikers et al. [2002], p28), intermediates on the basis of ‘extensive biomedical knowledge [involving specific anatomical or causal information]...acquired through individual study and lectures’ (Rikers et al. [2002], p28), and experts on the basis of specialized clinical concepts, with biomedical propositions appearing to play little explicit role in their reasoning. When asked to justify their diagnoses in speak-aloud protocols, these expert physicians favored technical terms unrecognizable to novices, such as ‘aorta-insufficiency, backward failure, cyanosis, endocarditis, hepatolienaomegaly, or hyperhidrosis’ (Rikers et al. [2002], p. 29).

While this finding of three qualitatively distinct categorization strategies at different levels of expertise might appear to challenge the ubiquity of transitional gradation in the study of expert concepts, a careful examination shows that it actually presumes and supports it. While many concepts that feature in expert reasoning are not at all possessed by novices, this is entirely consistent with the claim that experts retain many of the concepts they possessed as novices, which have been extensively revised and elaborated through gradual iterative revision just as indicated above. In fact, Boshuizen and Schmidt’s studies of these distinctive expert concepts show that they can only be properly understood in terms of their relations to the concepts of novice and intermediate learners. When probed, for example, these expert clinicians were able to articulate the reasoning behind these concepts, which generally took the form of a series of inferences involving clinical and biomedical concepts.¹⁰ In other words, the specialized expert

¹⁰ For example, note the mixture of anatomical information and clinical inferences from one physician who justified a diagnosis of endocarditis in the following terms: ‘The patient is a

concepts are really encapsulated clusters of concepts and inferences available to novices and intermediate learners, but which have gradually become tacit through extensive practice. Moreover, these encapsulated knowledge structures are deployed to issue diagnoses involving the very same categories as those of the novices and intermediates—such as, *heart disease, liver disease, diabetes, or stroke*.

This all bolsters the importance of transitional gradation, for the highly-specialized knowledge structures deployed by experts compose much larger conceptual structures used to place cases in these familiar categories in a much more sophisticated and context-sensitive manner. In the case of both configural rules and encapsulated technical concepts, a close examination of the data does not reveal discrete breaks in learning that could cleave expert from novice concepts in a principled way. Rather, it reveals only deeper interdependence between novice and expert concepts, and yet more complex transitional gradation.

4.3 Mechanisms underlying animal concepts

Finally (and more speculatively), if we take a comparative perspective on the study of concepts—if we think that conceptual abilities evolved and are possessed to greater or lesser degrees by a variety of non-human animals—it is likely that we will see transitional gradation in relevant underlying mechanisms derived from the transitional gradation through the phyletic record. Of course, the evolutionary emergence of ‘conceptual abilities’ remains a contentious question, and a number of researchers either hold that animals do not possess human-like concepts or that it is not yet clear how to empirically study the concepts they possess (Chater &

young man with a high fever who presents a septic syndrome. This suggests drug use. He shows signs of thromboemboli, due to an affected heart valve. The tachycardia fits with an associated aorta vitium’ (Schmidt & Boshuizen [1992], p. 275).

Heyes [1994]). Furthermore, there are systematic differences in the animal literature that render the particular consensus of the human concepts literature elusive: comparative psychology favors different taxonomic divisions (such as perceptual, associative, relational, and analogical concepts), distinct experimental paradigms (for example an emphasis on associative learning and same-different tasks), fewer mathematical models, and more learning principles (see Zentall et al. [2007] for a review). However, concepts are now studied in organisms separated from humans by as much evolutionary time as honeybees (Giurfa et al. [2001]; Chittka & Jensen [2011]), and unless we tie concepts tightly to a uniquely human trait like language, it is likely that such abilities emerged gradually over evolutionary time and will be exhibited to greater or lesser degrees by different species with different underlying neural mechanisms.

Many forms of variation have already been found in animal conceptual abilities, and correspondingly many forms of variation in underlying neural mechanisms that may explain these differences. Species should be expected to differ in the type and number of features they can associate with categories and how easily they can learn about them, as well as in relevant supporting capacities such as perceptual acuity, memory, and motivation. Some of these differences between species may be relatively sharp—such as an ability or inability to master higher-order analogical relations (Thompson & Oden [2000]; Penn, Holyoak, & Povinelli [2008]). Others appear transitional in the more problematic sense, like differences in the number, type, order, context-sensitivity, or degree of abstraction in relations that can be learned—differences that appear to emerge gradually in the phyletic record due to gradual changes in relative amounts of brain tissue devoted to the neocortex or hippocampus (Güntürkün

[1981]; Basil et al. [1996]).¹¹ Animals may also differ in their ability to structure categories by features regarded as typical (Jitsumori [1994]; Dépy, Fagot, & Vauclair [1997]) or essential (Philips, Shankar, and Santos [2010]). Finally, there is some evidence that animals themselves may demonstrate some of the expertise effects just mentioned (Helton [2008])—though perhaps not those that depend upon the more elaborate forms of metacognitive awareness and explicit representation of task demands, which probably do depend upon language (Carruthers [2008]). Some of these lines of comparative investigation are still preliminary, but there is already enough evidence that a defender of the Similarity Thesis should confront it.

At any rate, if transitional gradation in the mechanisms underlying conceptual abilities is found to proliferate in the study of expertise or in comparative psychology, then the case of concepts must rise at least to the level of a Class III challenge to accommodation. In this eventuality, ST will not help us settle borderline disputes about concepts, for nature will not provide us with clear discontinuities—at either the surface or underlying levels of description—that could help us identify non-arbitrary metrics of similarity characterizing subtypes of concepts. ST should here be entirely abandoned, and we must move to an alternative perspective concerning the theory of kinds' role in these disputes.

5. Conclusion

I have here discussed the general tension between the traditional idea that natural kinds are united by a distinctive form of similarity (ST) and the metaphysical phenomenon of transitional gradation. I argued that the problem of transitional gradation extends beyond the confines of biology, where it has most been acknowledged, to central areas of psychology where it has been

¹¹ There is some evidence that the relevant gradations found in the acquisition of human expertise may also be found in animals—see Helton ([2008]).

little acknowledged until now. While I have focused on the study of concepts, the sorts of gradual adaptation and iterative bootstrapping found in concept learning are likely to be organizing principles in (at least) all of the life sciences, and these principles introduce transitional gradation fundamentally at odds with the appeal to similarity with which the search for kinds began. Moreover, such transitional gradation is likely to be found in any gradually modified representational state, which encompasses a large swathe of psychology's subject matter. Though some theorists of kinds reject this appeal to similarity entirely (Magnus [2011]), such a move requires substantial conceptual redeployment, and those hoping to use the theory of kinds to settle taxonomic disputes in the life sciences should plan accordingly.

As a final call for increased caution, I suggest that the problem with ideal models is probably just the tip of the iceberg in terms of our category representations' transitional plasticity. The same appeals to the adaptability of expertise could as well recommend transitions to yet other types of feature depending upon what expertise in that domain requires the learner to master. For example, other studies have revealed that experts' categories can also gravitate towards deep causal structure when such structure best predicts category membership (Rottman, Gentner, & Goldwater [2012]). Moreover, increasingly popular Bayesian models of category learning also highlight the importance of intermediate structures in category learning by construing categorization as probability density estimation and taking learners to iteratively develop conceptual clusters by continually updating probability estimations as new stimuli are encountered (Griffiths et al. [2007]).¹² This particular Bayesian model also suggests transitional

¹² Of course, one might object that Bayesian approaches cannot provide a model of underlying mechanisms of categorization, given that pure Bayesian inference is not computationally

gradation between prototypes and exemplars, given its prediction that subjects will display prototype-like judgments early in training and exemplar-like judgments later in training—explaining this influential but puzzling behavioral finding from Smith & Minda ([1998]). These new models all enable a variety of predictions about gradual changes found in the trajectory of concept learning that were beyond the purview of previous theories, but which now must be taken seriously. The general problem—good for organisms with brains, bad for the theory of kinds—is that, as with mutation and gene exchange in natural selection, learning endows us with a powerful form of iterative, open-ended adaptability, and the structure of our category representations will ultimately demonstrate as much diversity as the types of environmental relationships that we are able to learn to track.

Before closing, it is worth noting that none of the arguments here establish that categories vulnerable to gradation are wholly arbitrary or useless, or that they should be eliminated from psychology. We should, as Machery urges, stop thinking that concepts are all alike, but if we reject the assumption that only similarity-based kinds are valid subjects for science, then concept eliminativism no longer follows. As some have suggested, ‘concept’ may instead name a ‘practical kind’ whose use is determined pragmatically through experimental or clinical practice (Haslam [2002]), or concepts may yet be amenable to pluralist or promiscuous approaches that allow context-sensitive category membership conditions and/or cross-classification by multiple incommensurable taxonomies (Weiskopf [2007]; Dupré [1996]; Craver [2009]; Rice [2014]). Concepts may be united not by synchronic similarities but rather by historical relations or transitional processes (Millikan [1999])—so long as we abandon the hope that these relations tractable; but such models typically rely on more plausible approximate inference algorithms that are biologically feasible.

will secure significant inductive generalizations holding for all concepts. The arguments here simply establish that membership in any such kinds cannot be univocally fixed by non-arbitrary metrics of similarity, and that attempts to settle their borderlines by appeal to such natural metrics is likely to end in frustration. The more general moral is that psychology should move instead towards a broader, model-based perspective that aims to predict and explain the full range of variation in conceptual abilities as functions of variables like amount and type of goal-related experience, informational structure of the domain, context, and general psychological resources like memory, motivation, and attention.

The University of Houston
513 Agnes Arnold Hall
Houston, TX 77204-3004
cjbuckner@uh.edu

Acknowledgments

I am grateful to Colin Allen, Felipe de Brigard, Carrie Figdor, Edouard Machery, the attendees of the 2014 Southern Society for Philosophy and Psychology, and two anonymous reviewers for discussion and feedback on earlier drafts of this paper. This work was supported in part by a fellowship from the Alexander von Humboldt Foundation.

References

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., and Coley, J. D. [2002]: 'A Bird's Eye View: Biological Categorization and Reasoning within and across Cultures', *Cognition*, 84(1), 1-53.

- Barsalou, L. [1985]: 'Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded structure in Categories', *Journal of Experimental Psychology: Learning*, **11**(4), pp. 629-654.
- Barsalou, L. [1987]: 'The Instability of Graded Structure: Implications for the Nature of Concepts', In U. Neisser (ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, Cambridge: Cambridge University Press, pp. 101-140.
- Barsalou, L. [1990]: 'On the Indistinguishability of Exemplar Memory and Abstraction in Category Representation', in T. Srull and R. Wyer (eds.), *Advances in Social Cognition*, Hillsdale: Lawrence Erlbaum Associates, pp. 61-86.
- Basil, J. A., Kamil, A. C., Balda, R., and Fite, K. V. [1996]: 'Differences in Hippocampal Volume among Food Storing Corvids', *Brain, Behavior and Evolution*, **47**(3), pp. 156-164.
- Boshuizen, H., and Schmidt, H. G. [1992]: 'On the Role of Biomedical Knowledge in Clinical Reasoning by Experts, Intermediates and Novices', *Cognitive Science*, **16**(2), pp. 153-184.
- Burnett, R. C., Medin, D. L., Ross, N. O., and Blok, S. V. [2005]: Ideal Is Typical. *Canadian Journal of Experimental Psychology*, **28**(1), pp. 41-50.
- Camerer, C. F., and Johnson, E. J. [1991]: 'The Process-Performance Paradox in Expert Judgment: How Can Experts Know So Much and Predict So Badly?', In K. A. Ericsson and J. Smith (eds.), *Toward a General Theory of Expertise: Prospects and Limits*, Cambridge: Cambridge University Press, pp. 195-217.
- Carruthers, P. [2008]: 'Meta-Cognition in Animals: A Skeptical Look', *Mind and Language*, **23**(1), pp. 58-89.

- Chittka, L., and Jensen, K. [2011]: ‘Animal Cognition: Concepts from Apes to Bees’, *Current Biology*, **21**(3), pp. R116-R119.
- Chittka, L., and Osorio, D. [2007]: ‘Cognitive Dimensions of Predator Responses to Imperfect Mimicry’, *PLoS Biology*, **5**(12), E339.
- Craver, C. F. [2009]: ‘Mechanisms and Natural Kinds’, *Philosophical Psychology*, **22**(5), pp. 575-594.
- Davis, T., and Love, B. C. [2010]: ‘Memory for Category Information Is Idealized through Contrast with Competing Options’, *Psychological Science*, **21**(2), pp. 234-242.
- De Groot, A. D. [1946/1978]: *Het Denken Van Den Schaker [Thinking Processes in Chess Players]*. Amsterdam: North Holland.
- Dépy, D., Fagot, J., and Vauclair, J. [1997]: ‘Categorization of Three-Dimensional Stimuli by Humans and Baboons: Search For Prototype Effects’, *Behaviour Processes*, **39**, pp. 299–306.
- Dupré, J. [1996]: ‘Promiscuous Realism: Reply To Wilson’, *The British Journal for the Philosophy of Science*, **47**(3), pp. 441-444.
- Ereshefsky, M., and Matthen, M. [2005]: ‘Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory’, *Philosophy of Science*, **72**(1), pp. 1-21.
- Ericsson, K. A., and Lehmann, A. C. [1996]: ‘Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints’, *Annual Review of Psychology*, **47**(1), pp. 273-305.
- Ertmer, P. A., and Newby, T. J. [1996]: ‘The Expert Learner: Strategic, Self-Regulated, and Reflective’, *Instructional Science*, **24**(1), pp. 1-24.

- Giurfa, M., Zhang, S., Jenett, A., Menzel, R., and Srinivasan, M. V. [2001]: ‘The Concepts of ‘Sameness’ and ‘Difference’ in an Insect’, *Nature*, **410**(6831), pp. 930-933.
- Griffiths, T. L., Sanborn, A. N., Canini, K. R., and Navarro, D. J. [2008]: ‘Categorization as Nonparametric Bayesian Density Estimation’, *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, pp. 303-328.
- Güntürkün, O. [1987]: ‘Brain Differences Determine Different Limits of Intelligence’, *Behavioral and Brain Sciences*, **10**(4), pp. 689-689.
- Haslam, N. [2002]: ‘Kinds of Kinds: A Conceptual Taxonomy of Psychiatric Categories’, *Philosophy, Psychiatry, and Psychology*, **9**(3), pp. 203-217.
- Helton, W. S. [2008]: ‘Expertise Acquisition as Sustained Learning in Humans and Other Animals: Commonalities across Species’, *Animal Cognition*, **11**(1), pp. 99-107.
- Hull, David L. [1965]: ‘The Effect of Essentialism on Taxonomy—Two Thousand Years of Stasis (I)’, *The British Journal for the Philosophy of Science*, **15**(60), pp. 314-326.
- Jitsumori, M. [1994]: ‘Discrimination of Artificial Polymorphous Categories by Rhesus Monkeys (*Macaca Mulatta*)’, *Quarterly Journal of Experimental Psychology*, **47**, pp. 371–386.
- Kripke, S. A. [1980]: *Naming and Necessity*, Cambridge: Harvard University Press.
- Levering, K., and Kurtz, K. J. [2006]: ‘The Influence of Learning to Distinguish Categories on Graded Structure’, In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pp. 26-29.
- Lynch, E., Coley, J., and Medin, D. [2000]: ‘Tall Is Typical: Central Tendency, Ideal Dimensions, and Graded Category Structure among Tree Experts and Novices’, *Memory and Cognition*, **28**(1), 41-50.

- Machery, Edouard [2009]: *Doing without Concepts*, Oxford: Oxford University Press.
- Machery, Edouard [2010]: 'Précis of: Doing Without Concepts', *Behavioral and Brain Sciences*, **33**, pp. 195-244.
- Machery, Edouard. [2005]: 'Concepts Are Not a Natural Kind', *Philosophy of Science*, 72(3), pp. 444-467.
- Magnus, P. D. [2011]: 'Drakes, Seadevils, and Similarity Fetishism', *Biology and Philosophy*, **26**(6), pp. 857-870.
- Millikan, R. G. [1999]: 'Historical Kinds and the "Special Sciences"', *Philosophical Studies*, **95**(1), pp. 45-65.
- Quine, W. [1994]: 'Natural Kinds', in D. Stalker (ed.) *The New Riddle of Induction*, La Salle, Illinois: Open Court, pp. 42-56.
- Penn, D. C., Holyoak, K. J., and Povinelli, D. J. [2008]: 'Darwin's Mistake: Explaining the Discontinuity between Human and Nonhuman Minds', *Behavioral and Brain Sciences*, **31**(02), pp. 109-130.
- Phillips, W., Shankar, M., and Santos, L. R. [2010]: 'Essentialism in the Absence of Language? Evidence From Rhesus Monkeys (*Macaca Mulatta*)', *Developmental Science*, **13**(4), pp. F1-F7.
- Rein, J. R., Goldwater, M. B., and Markman, A. B. [2010]: 'What Is Typical about the Typicality Effect in Category-Based Induction?', *Memory and Cognition*, **38**(3), pp. 377-388.
- Rice, C. [2014]: 'Concepts as Pluralistic Hybrids', *Philosophy and Phenomenological Research*.
Online First.

- Rikers, R. M., Schmidt, H. G., and Boshuizen, H. P. [2002]: 'On the Constraints of Encapsulated Knowledge: Clinical Case Representations by Medical Experts and Subexperts', *Cognition and Instruction*, **20**(1), pp. 27-45.
- Rosch, E., and Mervis, C. [1975]: 'Family Resemblances: Studies in the Internal Structure of Categories', *Cognitive Psychology* **7**(4), pp. 473-605.
- Rottman, B. M., Gentner, D., and Goldwater, M. B. [2012]: 'Causal Systems Categories: Differences in Novice and Expert Categorization of Causal Phenomena', *Cognitive Science*, **36**(5), pp. 919-932.
- Ruscio, J., and Ruscio, A. M. [2000]: 'Informing the Continuity Controversy: A Taxometric Analysis of Depression', *Journal of Abnormal Psychology*, **109**(3), pp. 473-487.
- Ruscio, A. M., and Ruscio, J. [2002]: 'The Latent Structure of Analogue Depression: Should the Beck Depression Inventory Be Used to Classify Groups?', *Psychological Assessment*, **14**(2), pp. 135-145.
- Schmidt, H. G., and Boshuizen, H. P. [1992]: 'Encapsulation of Biomedical Knowledge', In *Advanced Models of Cognition for Medical Training and Practice*, Berlin: Springer, pp. 265-282.
- Smith, J. D., and Minda, J. P. [1998]: 'Prototypes in the Mist: The Early Epochs of Category Learning', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, pp. 1411-1436.
- Sober, Elliott. [1980]: 'Evolution, Population Thinking, and Essentialism', *Philosophy of Science* **47**, pp. 350-383.

- Thompson, R. K., and Oden, D. L. [2000]: ‘Categorical Perception and Conceptual Judgments by Nonhuman Primates: The Paleological Monkey and the Analogical Ape’, *Cognitive Science*, **24**(3), pp. 363-396.
- Voorspoels, W., Vanpaemel, W., and Storms, G. [2011]: ‘A Formal Ideal-Based Account of Typicality’, *Psychonomic Bulletin and Review*, **18**(5), pp. 1006-1014.
- Voorspoels, W., Storms, G., and Vanpaemel, W. [2013]: ‘Idealness and Similarity in Goal-Derived Categories: A Computational Examination’, *Memory and Cognition*, **41**(2), pp. 312-327.
- Weiskopf, D. [2009]: ‘The Plurality of Concepts’, *Synthese*, **169**(1), pp. 145-173.
- Zachar, P. [2000]: ‘Psychiatric Disorders Are Not Natural Kinds’, *Philosophy, Psychiatry, and Psychology*, **7**(3), pp. 167-182.
- Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K., and Rattermann, M. J. [2008]: ‘Concept Learning In Animals’, *Comparative Cognition and Behavior Reviews*, **3**, pp. 13-45.