NOTE TO USERS

FLEXIBILITY IN CATEGORIZATION AS A FUNCTION OF VARIABILITY OF REPRESENTATION TYPE: STUDIES IN FEATURE WEIGHTING AND CATEGORICAL BIASING

By
SAMUEL D. HANNAH, BA

A Dissertation
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

CONCEPTUAL VARIABILITY AND CATEGORIZATION FLEXIBILITY

DOCTOR OF PHILOSOPHY (2004) McMaster University
(Psychology) Hamilton, Ontario

TITLE: Flexibility in Categorization as a Function of Variability of Representation Type: Studies in Feature Weighting and Categorical Biasing

AUTHOR: Samuel D. Hannah, BA (Simon Fraser University)

SUPERVISOR: Professor L. R. Brooks

NUMBER OF PAGES: xiii, 180

Abstract

For nearly the last fifty years, most research on the acquisition and application of conceptual knowledge has focused on structural relations and abstract descriptions as the composing a concept. In this thesis I show that conceptual content is much richer than this, and includes instantiated forms of knowledge as well as abstract descriptions. This instantiated knowledge, both at the featural and procedural level, is necessary to explain error patterns in probe and biasing studies. Furthermore, this variety in content is shown to be systematically linked to different decision patterns, providing an explanation for the flexibility of categorization. Much critical learning done in concept formation, therefore, involves learning optimal feature descriptions and task-appropriate procedures. These ideas are discussed in terms of embodied cognition.

Acknowledgements

This work would have been impossible without a number of people. First among them is my supervisor, Lee Brooks. Lee taught me more than just how to do research in cognitive psychology. He showed me that creative thinking and rigourous thinking are not mutually exclusive, that generosity to peers, juniors, senior, dogs, cats is a necessity of life, and how to write and speak about my research (this latter project is still a work in progress). After Lee, there is no obvious order to those I must thank, so I will thank people in a randomized order. Judy Shedden must be thanked for her enthusiasm, tremendous kindness and sharp mind. And her parties. Geoff Norman must be thanked for his encouragement, his thick skin at lab meeting, his constant humor and his statistical élan. Sue Becker must be thanked for her intelligent analysis of various rambling PRs, delivered with gentility and thoughtfulness. Bruce Milliken must be thanked for helping me learn to balance caution and courage in research, and for his warmth and consideration. My labmates, past and present, must be thanked: Kevin, Vicki, Chan, Aimee, Nicole and Tav. Vicki Leblanc, whose previous work opened the door for this, and who must be thanked for putting up with me as a officemate for an entire year without ever failing to be supportive and friendly, even during her last stressful year. And for still putting up with me. Kevin Eva, for setting absurdly high standards I enjoyed not living up to, as well as his grace and generosity. Chan Kulatunga-Moruzi must be thanked for her laughter and for the example of determination she set in the face of many obstacles. Aimee Skye for her energy—which I have fed off constantly—her hospitality and her delicate nature (one of those three is a joke) Nikki Woods, for bringing a lot of fun (along with Aimee) into the lab, and encouraging me to improve my wardrobe. And for updating me on the latest in Maury Pauvich. Tavinder Ark must be thanked for her stories and adventures. And there are many, many friends in this department who have kept my spirits bouyed and my mind engaged, some of whom I will no doubt leave out by absent-mindedness: my beer-brewing partners Alex Ophir and Rick LeGrand, Maria Jesus Funes Molina, Fil Cortese, Jason Tangen and Melanie McKenzie, Nicole Anderson, Christine Tsang, my evil nemesis Sandra Hessels, Jason and Launa Leboe, Lynne and Al Honey, Mark Honey and Laura Theall-Honey, Nicole Conrad, Matt Crump, Chris Taylor, Eric Richards, Zahra Hussain, Lisa Betts, Carl Gaspar, Ale Friere, Ellen MacLellen, Phil Cooper, Kamini Persaud, Andrew Clark, Lisa DeBruine, Pat Barclay, Damian Jankowicz, Carrie Sniderman, Elliot Beaton, Eric Bressler, Kelly Stiver, Julie Desjardins, Vicki Armstrong, Mayu Nishimura, Melisssa

Dominguez, Sandy Martin-Chang, Curt Nordgaard, Vito Scavetta, Marta Sokolowska, and Jessica Phillips-Silver (last only on paper, Jessica). Of course, my family must be thanked for all their love and support: big brother Victor, Mary-Ellen, my not-so-little nephew Rio, and brother-in-law Wayne. For reasons of space I must forgo thanking all those people on the West Coast—or Best Coast—who have kept me going and kept believing in me.

Table of Contents

List of Figures

List of Figures, cont'd

List of Tables

List of Tables, cont'd

Preface

This thesis includes two papers submitted for publication to the Journal of Experimental Psychology: Learning, Memory and Cognition. Both papers were submitted in July, 2004, and both were written by Lee Brooks and myself. I was first author on both papers. As required by the regulations of the university when submitting a thesis that includes work with authors in addition to myself, I will now detail as exactly as possible my contributions to these works, and when the work reported in the papers was conducted.

The research for the first paper began in November of 2001 and concluded in February of 2002, with analysis continuing through the summer of 2003. The research for the second paper began in late November of 2000 with the creation of the stimuli, and experimentation continued through to September of 2003.

In terms of experimental design, I developed the stimuli and test procedures used in both papers. These were based on stimuli used previously in the Brooks lab, but simplified, and amended in some cases to allow for a superordinate structure. The choice of developing transfer materials by skewing training features so as to maintain configural relations, and thus a moderate degree of overall similarity, was my choice. The training procedure was modified from procedures current in the Brooks lab, amended to introduce a role for causal stories and explicit feature learning. The direct instruction regarding diagnostic features was my idea.

I developed the analyses used in both papers, with some valuable advice from John Vokey regarding the analysis of multiway frequency data and log-linear regression. My contribution in this area includes not only the choice of inferential tests, but the selection of dependent variables, and the segregation of analysis by listing and counting strategies. The decision to implement a strategy report at the end of test, although used in prior work in the Brooks lab, was also mine, as was the decision to implement a yoked control condition in the first paper. This was also a technique previously used in other research coming out of the Brooks lab.

My contribution to the conceptual development of the two papers is also considerable. The notion of a feature-goodness heuristic was my development, as was the connection between the feature-goodness heuristic and Whittlesea's discrepancy-attribution hypothesis. The rival notion of the accessibility of alternatives as a mediating factor in biasing, and the elaboration of the forms that argument, were

also part of my contribution. In the first paper, I also developed the rival explanations to the feature-goodness hypothesis regarding the effect of an instantiated features.

In the second paper, the notion of an attentional routine as a form of instantiated procedural knowledge was my development, as was the connection between instantiated procedural knowledge and both Hintzman's (1986) MINERVA II model, and the work of Jacoby, Lindsay and Hessels (2003). I also introduced the notion that conceptual structure is dependent on the level of feature description, and that the application and formation of conceptual knowledge are guided by a principle of minimal effort. I also worked out the relation between the ideas contained in the papers and those of a number of other researchers. For example, I developed the connection between categorical biasing and the flexibility found in similarity judgments, the biasing of attentional routines and inattentional blindness, and the relation between the ideas expressed in the papers and the embodied cognition perspective. I also developed the relation between this work and that of Shephard, Hovland and Jenkins (1961), Tversky (1974), Nosofsky (1986), Ashby et al. (1999), Alfonso-Reese et al. (2001), Yamauchi and Markkman (2001) and Markman and Maddox (2003). I made a substantial contribution to the notion of instantiated features as sufficient features, and the importance of the degree of association between instantiated features and categorical identity versus that of informational features and categorical identity.

Chapter 1

Biasing Categorization Decisions: A Problem and a Beginning

LeBlanc, Norman, and Brooks (2001) showed that medical students and medical residents could have their diagnoses pushed between a correct and a plausible alternative diagnosis simply by having them first evaluate the plausibility of a tentative diagnosis (either the correct or the alternative). This biasing effect was substantial, producing among doctors in residency shifts of roughly 40 percentage points in the likelihood of their classifying a person as having either the correct or the alternative disease. While these results have important implications for issues of medical expertise and medical error, the evidence of such categorical plasticity also raises some intriguing theoretical questions with regard to concept formation and categorization.

Outside of LeBlanc et al.s applied work, such categorical biasing effects have not been described in any categorization research. This absence is particularly conspicuous when put alongside the substantial effects readily elicited under real-world conditions. Categorical biasing effects have not been found in categorization research so far because of the narrowness of the underlying assumptions regarding the nature of categorization and concept formation. Throughout much of the last 50 years of experimental research into concept formation and categorization there has been a tendency to approach the topics as primarily involving the abstraction of categorical structure, that is, the abstraction of relations among features across categories in some domain of knowledge. It is the contention of this thesis that the abstractionist and structural assumptions that have dominated research into concepts and categorization are inadequate, separately and jointly, to explain how people learn and use concepts in everyday life, at least with regard to physical categories.

An emphasis on learning the structure of categories has lead to a lot of research about how relations among features are learned and represented (e.g., Love & Markman, 2003), or how the distribution of features in a domain of knowledge influences categorization decisions (e.g., Alfonso-Reese, Ashby, &

Brainard, 2002). Until recently, however, little emphasis has been given to the role of feature appearance and other forms of concrete knowledge, such as knowledge of motor routines performed on specific items, emotional valencies elicited by specific members, or attention patterns. The abstractionist assumption is particularly prominent when considering how theories treat the representation of features of a concept, as most accounts make little or no distinction between a semantic representation and a perceptual representation of the same feature. This emphasis on the abstraction of higher-order relations mitigates against the discovery of phenomena such as the biasing effect in two ways: it encourages researchers to focus on the role of relatively stable information, and it discourages examination of the variety of information that can be used to make a decision.

This research has several aims that will be pursued primarily by using an analog of LeBlanc et al.'s (2001) biasing effect, substituting simpler, artificial categories that are highly controllable for medical categories, and substituting university undergraduates for medical personnel. The use of research-created categories allows for a rigorous exploration of the phenomena. The use of nonspecialist undergraduates, besides being highly convenient, allows us to rule out the possibility that the biasing effects are dependent upon specialized knowledge, or processes attendant on the acquisition of expertise.

My first aim is to provide supporting evidence for the contention that feature knowledge takes at least two forms, namely specific feature representations and abstract feature representations, and both forms are critically involved in categorization processes. This inclusion of specific, non-relational knowledge is extended to the learning of specific attentional patterns. The possession of a rich variety of information in a variety of forms (such as multiple descriptions of the same feature) is critical to understanding the conceptual flexibility pointed at by the categorical biasing effect. Furthermore this demonstrates that learning is not confined to the learning of relations among features, but also involves learning *how* to describe the features themselves. I will argue in both the first set of studies and in the General Discussion that this is important because it means that the patterns of relations learned depend, in part, on the level of feature description. The categorical structure of a domain can, to some extent, be modified by changing the nature of feature descriptions.

The second aim is to show that to understand categorization issues of processing must be given more attention than they are now. Just as issues of categorical structure cannot be understood independently of issues of feature description, so also issues of conceptual content cannot be neatly separated from issues of conceptual processes. There is a richer array of processes applicable to categorization than has generally been acknowledged. A demonstration that specific feature representations influence categorization and that features can be represented at different levels of abstraction implies that learning how to represent features is an important process to concept formation. This thesis will also show that different levels of feature representations support different types of processing, including categorization heuristics similar to those proposed recently by Whittlesea and Leboe (2000). Finally, this work is positioned with regard to classic and contemporary approaches, especially a newly emerging approach called *embodied cognition.*

To understand the theoretical importance of stimulus specificity as an influence in categorization, we need to examine how the field has developed over the last few decades. This will allow me to show that the central concerns have remained focused on the abstraction of category structure as central to categorization and concept formation. After sketching a brief history of categorization research, I will show how the view that there are at least two levels of abstraction in how features are represented, including a highly specific level, is significantly different from the majority view of categorization, and also how it is consistent with some earlier findings and some specific theoretical accounts that have arisen recently. This will then allow me to show that the phenomenon of biasing is problematic for many established accounts, and that others make predictions about what controls biasing that diverge from those predictions based on assuming people use heuristics that weight features in terms of their appearance.

Chapter 2

A Historical Overview and Implications of Current Findings

2.1 A History of Structural Abstractionism

To frame this discussion of how the ideas presented in this thesis relate to the field, I will use a highly influential paper, that of Shepard, Hovland, and Jenkins (1961). Not only does this paper continue to be cited (e.g., Love & Markman, 2003; Nosofsky, 1988), but its view of categorization and concept learning has shaped the field such that even papers that do not cite it bear its imprint. Shepard et al. begin their paper by making a critical comparison. They compared the learning of artificial categories marked by overlapping, or shared, features with the learning of artificial categories that are free of feature overlap. They found that categories with overlapping features are more difficult to learn than categories without overlapping features , and that therefore how features are distributed across categories is critical to the learning of categories. They then found that the complexity of such structures is mirrored in the complexity of rules generated by persons learning the categories, suggesting that verbal rules are closely tied to concept learning. Lastly, they found that people made many fewer errors than would be predicted simply by considering the confusions arising from overlapping features.

This last point suggested to Shepard et al. (1961), and to many researchers that followed, that the learning of categories and the formation of concepts is supported by much more than simply mapping individual features to category labels. Instead, Shepard et al. argued that concept learning involved the abstraction of the dimensional structure of categories via selective attention, and guided by explicit hypotheses expressible in the form of verbal rules.

This reinforced earlier findings by Bruner, Goodnow, and Austin (1956) that categories defined by conjunctive rules were more difficult to learn than categories defined by disjunctive rules, again pointing to the centrality of both structure and verbal rules in controlling concept formation and category

learning. Bruner et al.'s use of information theory lead them to focus on the statistical properties of features as being critical to understanding how features are processed. It was not the concrete details of the features that mattered, but simply their distribution across categories. Thus, while the work of Shepard et al. (1961) led to a view of concept learning as abstraction with regard to category structure, the work of Bruner et al. led to a view of concept learning as abstraction with regard to features. The stimuli both sets of researchers tended to use involved cards depicting one or more simple coloured geometric shapes, such as squares, circles, stars and so on. These simple entities had few internal features, much like the stimuli used by mathematical modelers of categorization today (Alfonso-Reese et al., 2002; Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Maddox, 2001; Maddox & Bogdanov, 2000; Nosofsky & Palmeri, 1997; Nosofsky, Palmeri, & McKinley, 1994). Feature overlap occurs completely or not at all, such that shared features are identical to one another, and not merely similar in some respects. Many of the studies discussed by Bruner et al., for example, used stimuli, similar to those used in the Wasson Card Selection Task. An exemplar in one category may be a card with a large red triangle, another example of the same category may be a small blue rectangle, while an examplar of a second category may be a small red rectangle. Colour, size and shape may vary within and across categories, but the specific values would be the same: the rectangle shown in one example could be the identical size and proportions as that shown in the example of another category. These stimuli reflect the assumption that what matters with regard to feature processing is the distribution of feature occurrence, not how similar a feature is to previously seen features.

The all-or-none assumptions regarding feature representations embodied in a distribution-based analysis of feature processing are also prevalent in Rosch and colleagues' early investigations into common semantic categories (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johson, & Boyes-Bream, 1976). Rosch and Mervis' study cashed Wittgenstein's (1953/1994) observations regarding categorical structure. Wittgenstein argued that at least some important and common categories had no necessary feature, but instead possessed a more complex and diffuse structure involving a set of commonly occurring features, no one of which is shared by all members of a category. Wittgenstein used the example of the sharing of facial features among family members to illustrate the idea, because all members more or less resemble each other, and any two members usually share at least one feature although there is usually no feature shared by every family member. Rosch and Mervis established that common semantic categories possess just such a "family-resemblance structure" by having students write out lists of the features associated with different categories. Not only was there no necessary feature defining these common semantic categories, but there were no sufficient features either. Features were not uniquely associated with category identity, often overlapping across several categories. They calculated how often a given feature was listed across members of a category, relative to how often the same feature appeared in the feature lists of other categories (or, "cue validity"). Items that were rated as highly typical of their category had many features that were common across members of that category but uncommon among members of rival categories. Note again the role that featural overlap

plays, as well as the focus on categorical structure. Rosch et al.'s investigation into the taxonomic organization of everyday concepts also relied on word lists, and used such measures as how many features were listed for the first time at a given level of abstraction. Again, what matters is whether a word or phrase, or a synonym, appears or doesn't appear on a word list, with synonyms treated as fundamentally identical variants of one another. What does not matter is the quality of a feature, only its presence or absence.

In Rosch and Mervis's (1975) work, the importance, or weight, of a feature in influencing a categorization is a direct function of that feature's cue validity. This statistical treatment of feature weighting has largely gone unchallenged, and subsequent research into feature weighting has been largely nonexistent until recent years. Current treatments have examined the role of causal beliefs in shaping feature influence (Ahn, 1998; Rehder & Hastie, 2001; Sloman, Love, & Ahn, 1998). In these accounts, it is the extent to which a feature is believed to play a central role in some causal process that influences its influence in categorization decisions. Despite this nonstatistical approach, features are again handled as members of short verbal descriptions or word lists.

Defining physical categories by word lists is a valid approach if we assume that their mental representations are language-like descriptions of objects. This "mentalese" assumption (Fodor, 1985) of parity between mental representations of physical categories and verbal descriptions of the same runs deep in treatments of categorization and concept formation. One of the first models of conceptual organization, Collins and Quillian's (1969) spreading activation model, consisted of a network of interconnected nodes, with each node consisting of an abstract description of some feature or set of features. Many later models continued this tradition. For example, Anderson's ACT* and ACT models (Anderson, 1991; Anderson & Ross, 1980) of memory and concept organization, and his rational model (Anderson, 1992; Anderson & Fincham, 1996) of concept formation and categorization explicitly assume that concepts are propositional in nature. Thus his ACT* model treats categories as descriptions linked by semantic relations ("isa", "causes", etc.), while his rational model allows categories to emerge by applying Bayesian statistics to sequences of exemplars encoded as verbal descriptions.

The abstractionist assumptions discussed above were an explicit part of the program initiated by Posner and Keele (1968) in their study of prototype formation. Posner and Keele presented participants with dot patterns that were derived from one of two prototypes, but did not present the prototypes. At test, participants were faster and more accurate on classifying the prototypes than they were at classifying novel distortions, even though those had never been presented in training. Participants were as quick and accurate to classify the prototypes as to classify old exemplars, and many participants were consistently quicker. Novel derived variants were more accurately and more quickly classified as they converged upon the prototype patterns. Posner and Keele interpreted their results as evidence that we automatically abstract prototypic representations from the specific exemplars that we encounter, and this mechanism of automatic prototype generation was the basis of concept formation. Posner and Keele's work suggested to many that even our perceptual experience is guided by an informational

summary of our past. That concepts are organized around representations abstracted across encountered exemplars—and, thus, correspond exactly to no specific exemplar—is still a common explanation of categorization and concept formation (e.g., Edelman, 1998; Hampton, 1995).

An alternative to the prototype view is that of the exemplar or instance view. Instance accounts rely on the idea that we can account for all of the effects that prototype models focus on simply by assuming that people preserve a memory of the specific items they encounter in some classification context, and then retrieve those instances when we encounter something similar in a similar context(e.g. Brooks, 1978; Hintzman, 1986; Kruschke, 2001; Medin & Schaffer, 1978; Nosofsky et al., 1994; Regehr & Brooks, 1993). Hintzman, for example, developed a model, MINERVA II, that produces a transient prototype by retrieval of multiple instances associated with the current cues, or, a "chorus of instances". The prototype exists only as long as the underlying instances are activated in some context. Transient prototypes could thus be generated on the fly (Kahneman & Miller, 1986) without any dedicated abstraction mechanism involved. Brooks (1978, 1987) described this approach using terms such as "nonanalytic", "noncomputational", and "decentralized". Yet even these approaches operationalized features as abstract representations.

Most research supporting instance/exemplar accounts employ categories defined by a set of binary-valued dimensions, with membership usually following some well-defined rule. This commonly takes the form of a feature-counting rule, such as, "i is a member of J if it has at least 2 of the following 3 features ..." Such rules reflect the family-resemblance structures characterized by featural overlap as described by Rosch and Mervis (1975). The binary nature of the feature dimensions means that any exemplar can be represented as a string of ones and zeroes. This description of features only holds if it is assumed either that all features take on a single value, or that variance in the manifestation of the features is irrelevant. If what matters, however, is not that cats have paws, as do dogs and monkeys, but that they usually have a particular type of paw, then our notion of cat depends in part on a specific manifestation of paws, not merely a generic feature value. Variations in the manifestations of features are then critical. Ironically, therefore, a theoretical approach that seeks to assert the importance of specific experience treats concrete exemplars as a collection of abstract features, and thus builds its approach upon a bed of abstraction.

2.2 The Importance of Representational Specificity

Recent work by Brooks and Hannah (2000; 2004) has found evidence that under some conditions a single feature can drive a categorization decision even though it is surrounded by more numerous competing features. It is worth considering this study in some detail. Participants were trained to recognize members of two species of imaginary animals. The imaginary conspecifics looked very similar to one another and very different from their rivals. Overlap between categories arose, but only at an abstract level. For example, although most "bleebs" had rounded heads and most "ramuses" had

angular heads, one bleeb had an angular head and one ramus had a rounded head. However, these divergent heads looked very different from their rivals. Angular ramus heads were diamond shaped, while the angular bleeb head was rhomboid. This pattern of abstract overlap only is characteristic of many basic-level categories at least. Dogs and cats can both be said to have paws, but their paws look very different.

At test, people are presented with items containing three features from one category and a single feature from the other. These creatures were created so that no single feature at an informational level is sufficient for categorization. If people were abstracting the distributional structure of the domain, then we might expect them to classify most items according to the species with a majority of features present. If all the features of test items are perceptually novel, yet embody the same relevant dimensions learned in training, then this is just what happens. This tells us that the training is sufficient for people to grasp at least the outlines of the informational structure of the domain. That is, the interfering effect of featural overlap that Shepard et al. (1961) found holds only when the identical feature appears in the exemplars of two categories. If the overlapping features are merely similar, sharing some, but not all properties (e.g., exemplars in two categories have 2 legs, but the height, shape, etc. of the legs differ), then this interference largely disappears, even though the features still overlap on some critical dimension. However, if a single feature seen in training is included in the test items, and drawn from the rival species, then most people will classify the item based on that one familiar feature, and the interference of Shepard et al. returns.

This demonstrates, first, that the informational content of features is not necessarily equivalent to the specific appearance of that content: *informational features* are not identical to *instantiated features*. Features may be represented in terms of some general property embodied by the actual feature (e.g., number or length) or they may be represented by the specific embodiment encountered. Second, it demonstrates that people can be more sensitive to the specific featural manifestations in categorization decisions than to the abstract, informational content of the features. Such sensitivity is masked, however, if all features at test are equally familiar or equally unfamiliar in their manifestations as they are in most experiments in categorization since Shepard et al. (1961) published their work. That features can be described as having at least two levels of representation is important, for it means that in addition to learning about the structure of the categories, learning about how best to represent features may also be an important factor in concept formation.

The idea that feature appearance is an important target of category learning finds support in the work of Schyns and colleagues (Schyns, Goldstone, & Thibaut, 1998; Schyns & Murphy, 1994; Schyns & Rodet, 1997), who have argued that features can be seen as categories at a small scale. Using artificial materials with blob-like internal features that can be parsed in different ways, they have shown that differences in learning history can give rise to differences in how new items are parsed into features. In other words, the features that are seen in an item are in part a construction on the part of the percipient, a construction shaped by prior experience of specific acts of parsing. This points to the preservation

not only of specific contents, but also to the preservation of specific mental operations, in this case, of feature-parsing operations. In this case, it would be inaccurate to describe concepts as solely involving abstract representations not only because some conceptual content is highly specific and concrete, but also because some conceptual content involves nonpropositional procedural knowledge.

This is consistent with Brooks' (1987, 1990) view of instance theory as involving the preservation of prior instances of *processing* rather than the preservation of *exemplars*, and will be an important foundation for the tentative framework sketched out at the end of this thesis. It is an insight that provides important linkages between categorization and concept formation with a theoretical approach in memory that is often called transfer-appropriate processing (TAP) (Morris, Bransford, & Franks, 1977; Roediger, Weldon, & Challis, 1989). The TAP account is based on the idea that memory preserves a record of all mental operations performed within some context, and that memory usage, and cognition in general, can be seen as a skilled activity (Jacoby, Baker, & Brooks, 1989; Kolers & Roediger, 1984; Morris et al., 1977; Whittlesea, Brooks, & Westcott, 1994).

Given the sensitivity to specific manifestations people showed in Brooks and Hannah's (2000; 2004) studies, it is possible that learning about the specific characteristic features may even be more important than learning about the higher-order relations defining a domain. In many ways, simple features are more informative than has been realized, at least for purposes of classification of physical objects. For most things in the world, the features of category mates resemble one another much more strongly than do members of different categories. That makes the specific appearance of features highly informative. Both dogs and cats may have paws, and thus be said to share a feature if we represent the feature at a purely semantic or generic way. However, the paws of cats and dogs are distinctive, and their distinctive appearances are closely correlated with categorical identity. As a result, we need see only the paw of a cat feeling around a corner to know from that alone that there is a cat around the corner. As a feature of a semantic category, "paw" is not sufficient for classification. As a feature of a physical category, a physical paw is.

The use of perfect overlap in categorization studies from the time of Shepard et al. (1961) on, or earlier, leaves no room to learn about systematic variance in feature manifestations, and so forces people to learn about higher-order properties because this is the only area of systematic variance. The reliance on higher-order information found in many concept learning studies then may be an artifact arising from researchers eliminating sources of information that are normally used in much everyday concept learning.

Brooks and Hannah's (2000; 2004) work implies that records of prior specific experiences of feature processing are retrieved to guide current categorization decisions. This idea is central to an ambitious account put forward recently by Barsalou (Barsalou, 1999; Goldstone & Barsalou, 1998; Solomon & Barsalou, 2001). Barsalou's program is aimed at grounding abstract thought and symbolization in prior perceptual experiences. According to Barsalou, when reasoning about a problem or when making a decision about things that are not immediately visible, we construct a mental simulation

by drawing on past perceptual and motor experiences. Barsalou's argument suggests that symbols and schemata are grounded in the experience of specific items encountered in specific contexts, rather than standing apart from such concrete knowledge. For such "perceptual symbols" to exist and to be manipulated freely, features of items would have to be encoded as individual memory records to be retrieved separately from the whole item in which they were embedded. Brooks and Hannah's (2000; 2004) work on the effect of individual, perceptually familiar features provides empirical support that features are encoded as records in their own right.

2.3 Bringing in Processing

Shepard et al.'s (1961) work emphasized the importance of concept learning via abstraction of relevant dimensions and higher-order relations. Shepard et al. took this as implicating a process involving selective attention, reminiscent of Kruschke's recent arguments (Kruschke, 2001; Kruschke & Blair, 2000) about the relevance of Mackintosh's (1997) attention-based account of blocking to concept formation. Many contemporary knowledge-based accounts of concept learning share this emphasis on the modulation of attention as an important mechanism in directing concept learning, with background knowledge guiding attention. (e.g., Ahn, 1998; Heit & Bott, 2000; Murphy & Medin, 1985; Wisniewski & Medin, 1994). Brooks and Hannah (2000, 2004) suggested that their results imply that the function of feature-lists and other weak rules produced by participants in categorization research is to direct attention to important features. Excepting a handful of researchers such as Kruschke and colleagues, how attention functions and interacts within a concept-learning task is ignored. Attention is commonly invoked as a placeholder process, the invocation of which allows categorization researchers to focus their efforts on other aspects of the learning situation.

Processing details are equally sketchy in most categorization accounts when we consider issues of applying conceptual knowledge. Although instance/exemplar theories often invoke the notion of episodic memory as an explanatory construct, the details of episodic memory are largely ignored. The exemplar-based formulations of this approach focus on concepts as collections of objects previously encountered; objects can be readily represented as descriptions, leaving this approach very compatible with treating concepts as abstract descriptions (e.g., Yamauchi & Markman, 2000). Brooks' instance theory has been influenced by TAP-based accounts of memory that argue for a preservation of records of mental processing, not of objects encountered (Jacoby & Brooks, 1984; Brooks, 1990). An implication of Brooks' approach is that concepts should include such procedural knowledge such as feature-parsing routines, motor routines, and so on.

As an illustration of the tendency to dismiss processing issues in favour of structural and representational concerns, consider a relatively recent entry in the ongoing debate between exemplar and prototype accounts. Smith and Minda (1998) have added an intriguing set of findings to this debate, showing that when exemplar and prototype models are compared over time, early stages are fit well

by a prototype model, but an exemplar model fits later stages best. As interesting as these findings are, *why* this time course arises is not clear. It could be because there are stored representations of prototypes and stored representations of exemplars, but the selection of conceptual elements in the early stages of learning is governed by rules, and favours the more abstract prototype representations. Then as more experience is gained with the material and responding becomes automatised and influenced by raw similarity, exemplar representations win out. Alternatively, it could have nothing do with separate classes of stored representations and their accessibility, but that encoding of items changes over the course of learning, which changes the distinctiveness of the preserved traces, which changes the nature of what is retrieved. In early learning, a new exemplar is weakly similar to many different exemplar representations, and so many are retrieved, forming something like Hintzman's (1986) on-line prototype. Later in the learning, however, items are elaborately encoded, so that a new item may be strongly similar to only one or two exemplar representations, and only these are retrieved, producing exemplar-driven behaviour. Thinking only about what type of representation is apparently active leaves us no way to choose between these or other alternatives, and this limits the depth we can bring to the understanding of these issues.

2.3.1 Feature Learning and Processing

Concept learning as described by Shepard et al. (1961) is the process of abstracting knowledge of category structure from exemplars, influenced by the complexity of the domain's structure, and guided by rule statements. Features played little role in their account, except as possible sources of response confusion due to (exact) feature overlap. A recent illustration of this view is presented by a recent paper by Alfonso-Reese et al. (2002), in which they argue that the complexity of some learning domain's covariance structure determines the complexity of learning in a categorization task. Processing in most contemporary accounts is limited to some type of abstraction process in learning, either guided by or embodied by explicit structural hypotheses in the form of verbal rules, and augmented by some type of memory retrieval. The latter is deemed to be involved in concept use, and some form of similarity comparison is taken as central to this memory-based process. However, the perceptual overlap effect (Brooks & Hannah, 2000; 2004) suggests that in addition to such processes, feature encoding and representation processes are also critical to categorization and concept learning.

Recent authors such as Wisniewski and Medin (1994) and Schyn's and colleagues (Schyns et al., 1998; Schyns & Murphy, 1994; Schyns & Rodet, 1997) have discussed the importance of features, but no one has postulated that there are at least two different levels of abstraction of feature representation, nor explored the processing implications that follow upon variability in feature representation. The existence of different levels of feature representations itself means that learning how best to represent features, and deciding which feature representation to use, are processes critical to categorization and concept learning and use. However, even more is implied. Different types of feature representations can support different types of categorization processes.

To take an example I will revisit in the introduction to the first paper, the computation of similarity is not a single process. There are at least two different ways of performing similarity comparisons: a set-theoretic process, and a metric process. A set-theoretic approach to similarity determines the similarity between two things by counting how many features of one item are shared by another (Tversky, 1977), while a metric approach to similarity is based on the distance between some items in a "similarity space" (Shepard, 1987). A feature-counting approach embodied in a set-theoretic computation of similarity is reliant on informational features. A metric approach to similarity, with its emphasis on the degree that two items resemble each other, is consistent with a reliance on instantiated feature representations. Just as different types of feature representations are compatible with different types of similarity computations, they are compatible with different types of other processes.

2.3.2 Biasing and Categorization

The existence of categorical biasing implies a high degree of plasticity in how categorization decisions are made, and this is not easily reconciled with standard approaches to categorization. Take, for example, one of the most prominent implementations of a metric similarity approach, that of Nosofsky (1988, 1986, 1997). At the heart of most of his approach is an exhaustive comparison of the total similarity of exemplar i to all members of category J, relative to the summed similarity of i to all members of all categories. Similarity is simply the squared distance between two exemplars in some psychological space, weighted by attention. As attentional weights are set to one or less than one, the effect of attending to shared dimensions is to reduce the perceived distance between an item and other exemplars of some category, enhancing similarity. So how does suggesting category J before showing an exemplar increase the probability of concluding that i is a member of J? The similarity comparison is exhaustive, and therefore it should make all possible relevant comparisons regardless of what is suggested. A suggestion could possibly influence the perceived similarity between i and members of J by drawing attention to the suggestion relevant dimensions of i, and so change the weighting of the distance between i and members of J. In my experiments, the categories are imaginary animals, categorized by physical features: tails, torso shape number of legs and so on. Importantly, all my materials—in most experiments—share the same relevant dimensions. They all have torsos of some shape, some number of legs and so on. Attending to the relevant dimensions, therefore, should enhance the perceived similarity between the current item and *all* relevant categories, including the rivals to the suggested category. Without some way of biasing the extent of the similarity comparisons or limiting the effects of attention only to the distance between the current item and the exemplars of the suggested category, suggestions cannot have any effect on categorization decisions in a model like Nosofsky's.

Tversky's (1977) treatment of similarity appears more conducive to explain biasing. The data supporting Tversky's set-theoretic approach exhibit some of the flexibility indicated by LeBlanc et al.'s categorical biasing effect. Tversky found that similarity ratings for pairs of items were asymmetrical,

and depended on which item in a comparison pair was chosen as the reference item. For example, he found that people rated China as being more similar to Korea than Korea was similar to China. Tversky noted this was consistent with the idea that people were comparing the number of features of a comparison item to the number of features defining a reference item, and basing their similarity estimate on the proportion of the standard items shared by the comparison item. Because people knew little about Korea compared to China, the set of features defining Korea would be much smaller than the set for China, and so the proportion of overlapping features differs depending on which is the reference item, even though the actual number of overlapping features is identical in either case. For example, if people know twice as many facts about China as they do about Korea, and everything believed to be true of Korea is believed true of China, then when comparing China to Korea there would be 100 per cent overlap in features (all of Korea's features are shared by China). However, when making the reverse comparison there would only be a 50 per cent overlap when comparing Korea to China because of the additional features known about China (only half of China's features are shared by Korea). This leads people to rate China as more similar to Korea than Korea is to China.

Here the situation is the reverse from what we described for Nosofsky (1988, 1986, 1997). There, search should be unaffected by a suggestion, but the similarity computation could be affected by suggesting a category. In Tversky's case, similarity arises from a simple proportion calculation that hinges on how many features are counted, and there is no a priori reason that the counting scheme should be influenced by suggesting a category. However, suggesting a particular disease could influence what diseases are compared by influencing what diseases come to mind. To use the terminology of memory researchers (Tulving & Pearlstone, 1966), a category suggestion could bias a set-theoretic type of categorization decision process by influencing the *accessibility* of alternatives [1]. That is, people will conclude for disease X if the patient's symptoms are more similar to the symptoms of disease X as compared to the alternatives to disease X. A suggestion could influence what alternatives to X were retrieved and therefore accessible for a set-theoretic decision-process to act on. A suggestion would not bias the decision-making process per se, but would bias the input to that process.

Any non-similarity, rule-based decision process could be biased in the same manner. By "rule", I mean a potentially verbal, algorithmic decision process, in which the process for combining features to reach a decision is explicitly given, and therefore should reach the same conclusion given the same inputs. For example, many studies using artificial categories use categories created around what Shepard et al. (1961) called a "Type IV" structure. Exemplars for such structures can usually be perfectly categorized using some type of counting rule that involves all dimensions, such as "It is a J if it has 3 of the following 5 features." Because features, if correctly recognized, should always be counted the same way, these rules, given the same inputs, should lead to the same conclusions every time they are applied. However, a suggestion could influence either attention to features—by influencing what

[1] Tversky and Kahneman's term "availability" conveys a similar meaning as "acccessibility" as used here, but it is avoided because of possible confusion with the use of "availability" as defined among memory researchers.

categories are thought of, as for set-theoretic similarity-based procedures—or they may influence the interpretation of features.

Both types of decision procedures could be biased by a suggestion manipulating either the attention to features or the interpretation of features (which could manifest as a failure to report a feature if it was incorrectly interpreted as background variation), and there is some support in the LeBlanc et al.'s (2001) data. They found that the effect of a suggestion not only influenced the probability of concluding for the correct or an alternative hypothesis, but also influenced feature reporting with suggestions increasing the number of suggestion-consistent features reported and decreasing the number of suggestion-inconsistent features. This is consistent with the notion that the suggestion affects attention to features or interpretation of features, biasing input to some similarity comparison mechanism.

However, the impact of the suggestions upon feature reporting was much smaller than it was upon diagnosis. For example, among doctors in residency, the suggestions shifted the probability of concluding for the correct diagnosis by 45 percentage points, and shifted the probability of concluding for the alternative diagnosis by almost 40 percentage points. However, the frequency of reporting for the correct features shifted only nine percentage points because of the suggestions, while the frequency of reporting features consistent with the alternative diagnosis shifted by 15 percentage points due to the suggestion. Second, the diseases involved were ones well known even to medical students (e.g., lupus, stomach cancer, Cushing's disease), and the photographs of the patients that participants were asked to diagnose were taken from medical textbooks, and contained highly predictive features, which in a second study (Brooks, LeBlanc, & Norman, 2000) were rated by expert diagnosticians[2] as obvious. Medical education stresses the importance of such features, and students are encouraged to look for such signs. Lastly, medical students and doctors are trained to engage in counterfactual thinking in the form of generating alternative diagnoses—or, "differential diagnoses"—even to self-generated diagnosis, which should encourage the generation of alternative hypotheses, which should offset to some extent any narrowing of the accessibility of alternatives.

2.3.3 Biasing and Discrepancy Attribution: A Feature-Goodness Heuristic

An alternative to the assumption that suggesting a diagnosis may bias attention to features or feature search by restricting the accessibility of alternatives lies in the discrepancy-attribution (DA) hypothesis of Whittlesea and colleagues (Whittlesea, 1997; Whittlesea & Williams, 2001a, 2001b). In this hypothesis, Whittlesea postulates that unexpected deviations from locally developed norms[3] for relative fluency, or coherency, of processing elicit attempts to explain discrepancies in processing. When an item is suddenly processed more fluently or less fluently than expected, people seek an explanation for this deviation, and look to such salient cues as the demands of task and their beliefs about responding within a task. In a recognition task, for example, if a test word is more fluently processed

[2]General internists with a minimum of 10 years of clinical practice
[3]That is, norms established within a specific situation

than previously encountered words, then people will tend to attribute this greater fluency to having seen the item in study, and judge the item as old. If the task is judging the density of visual noise, and one display is processed more coherently than previous displays, then people will tend to attribute this to the noise display being less dense.

Because the resulting experience is an outcome not only of fluency but also of the explanations for fluency differences, Whittlesea's DA hypothesis implies that cognitive processes and decision-making are based on heuristics that are inherently susceptible to biasing by secondary information. The DA hypothesis not only allows for biasing in almost any decision or judgment, it expects it. Change what cues in the environment are salient, or change the nature of the task, or people's beliefs about their responding within that task, and the nature of the decisions and experience changes. For example, Whittlesea, Jacoby, and Girard (1990) gave people a recognition task after a brief study session, embedding the test words in visual noise displays of variable density. For the novel test words, most of the variance in processing fluency came from being embedded in different densities of visual noise. As expected, people tended to wrongly attribute the new words embedded in low density masks—which were easier to process than the other novel words—to their being shown at study. The authors then reversed the task with another group of subjects, giving them new and old words embedded in masks of constant density, and having their participants judge the density of the noise displays. As the authors predicted, the visual noise displays containing old words tended to be judged as less dense than the noise containing new words, suggesting that the greater fluency arising from processing the old worlds was attributed to the noise being less dense.

Whittlesea and Leboe (2000) showed that people tended to rely on the perceptual similarity of novel features to old features to categorize items. They linked such a "resemblance heuristic" to the fluency or coherency of processing. Such a heuristic could be used in different ways. For example, features that are more fluently processed than surrounding features could be interpreted as being more reliable or more significant features than more difficult or less coherently processed features. That is, people may use fluency to estimate *feature goodness*, using differences in the subjective goodness of features to resolve conflicts among attended features. Or it could be that the extensiveness of search is influenced by discrepancies in processing coherency among features. Extensive search may only occur when features are equivalent in the ease with which they are processed; when sharp fluency differences exist, attention may be directed only to the more fluently processed features. This *feature sufficiency* heuristic is similar to the argument that Johnston, Hawley, Plewe, Elliott, and DeWitt (1990) made for fluency differences as guiding attention within a novelty detection task, although in this task they suggested that attention would be directed to the least fluently processed targets.

Using the DA hypothesis, it becomes relatively straightforward to explain how both the findings of Brooks and Hannah (2000, 2004) and categorical biasing effects could involve a feature-goodness heuristic. Highly familiar features would recruit prior records of feature processing more than unfamiliar features, and these recruited traces could aid processing the familiar feature. This would lead to the

familiar features being more fluently processed than less familiar features, leading to familiar features being deemed more important or more significant than less familiar surrounding features. Similarly, a suggestion could recruit prior episodes of feature processing consistent with the suggestion, making the processing of suggestion-consistent features present in the stimulus more coherent relative to other features. Counting rules, or other such strong rules, would circumvent or supplant any reliance on fluency differences. Alternatively, we could say that people reliant on informational features are by definition attending only to the informational content of feature representations, and ignoring other content such as processing fluency.

Different types of representations would then not only be linked to different types of rule statements, but also tied to different types of decision processes. Informational features would necessarily be the only type of features amenable to computational types of processes, while instantiated features would be usable in noncomputational heuristic decision processes[4].

2.4 Outline of Thesis

Although the focus of this thesis is to explore issues of feature representation and decision processes using an analog of LeBlanc et al.'s (2001) categorical biasing effect, I will not start with biasing. Instead, I will start with a submitted paper that lays some of the groundwork for the biasing research by addressing how a familiar-looking feature influences decision-making differently than does an unfamiliar-looking feature. This work will show that a reliance on instantiated features (feature appearance) leads to errors involving a discounting of conflicting information, and a reliance on informational features (such as feature labels) tends to produce errors involving a neglect of conflicting information. This is consistent with the idea that people use the specific appearance of features in conjunction with a feature-goodness heuristic, evaluating features for significance based on the similarity to specific previously encountered features. This work also shows that a reliance on different feature representations is connected to different types of rule statements, with a reliance on instantiated features yielding feature-list strategies, and a reliance on informational features yielding feature-counting strategies.

The second paper, also recently submitted, uses this to explore the categorical biasing effect. First, a categorical biasing effect is produced with the same materials used in the first paper, and then it is shown that changing the familiarity of test features modulates the biasing effect for those giving feature-list strategies, but not for those giving counting strategies. This result is also consistent with the use of a feature-goodness heuristic by people reliant on instantiated features. This second paper goes on to show that the biasing effect can be increased by having the overlap occur among members of different superordinate classes, and that this *superordinate biasing effect* is linked to a spatial segregation of features, rather than a cognitive segregation of categories. An influence of the cognitive context would

[4]This leaves open the possibility that informational features can also be used in heuristic decision procedures

suggest that the accessibility of alternatives was a critical factor in producing a categorical biasing effect, while the influence of spatial segregation is interpreted as involving the activation of stimulus-specific procedural knowledge in the form of attention routines. Finally, the last study in the second paper shows that the biasing effect is not influenced by the number of categories, which is a plausible factor if the accessibility of alternatives is important in mediating biasing effects.

The last empirical section in this thesis is an unsubmitted study that shows that the categorical biasing effect is highly robust. I first show that it survives transformations of the test stimuli intended to further reduce their overall familiarity of the test items while having minimal effect on the individual features. Then I show it also survives the introduction of perceptual overlap into the training items, showing that mere awareness of featural ambiguity in training does not diminish susceptibility to biasing suggestions at test.

In the general discussion, I try to bring these results together, and argue that the results and conclusions are consistent with a view of concepts as highly decentralized collections of distributed experiences and responses (e.g. Brooks, 1987; Barsalou, 2000). Furthermore, this distributed view of conceptual structure and the results of the research presented here are highly compatible with a memory account known as *transfer-appropriate processing* (TAP). TAP's perspective on memory formation and use implies a high degree of variability in how the acquisition and use of knowledge, just as I am arguing that the formation and usage of conceptual knowledge possess a high degree of variability, more so than has been recognized. Most importantly, I will try to show in the general discussion that such plasticity is not compatible with many standard approaches to studies in concepts and categorization, but is highly compatible with the embodied cognition perspective. I will try to show how this work both provides supporting evidence for embodied cognition approaches, and at least slightly supplements their theoretic equipment with some ideas that are new, at least to the embodied cognition account.

Chapter 3

The Effect of Instantiated Features on Decision-making

Running head: FAMILIARITY AS A DETERMINER OF FEATURE IMPORTANCE

Feature appearance as a determiner of feature importance in classification

Samuel D. Hannah and Lee R. Brooks
McMaster University

Draft July 13, 2004
Submitted to the Journal of Experimental Psychology: Learning, Memory and Cognition
July 14, 2004

Please address correspondence to:
Sam Hannah
Department of Psychology
McMaster University
Hamilton, Ontario
Canada, L8S 4K1
e-mail: hannahsd@mcmaster.ca
phone: (905) 525-9140, ext 24824
fax: (905) 525-9140

Abstract

Both humans and birds have two legs, but legs that look human never appear on birds. This paper explores this normal constraint of feature appearance and informational description being differently associated with categories. Our training items were standard family resemblance categories, but the feature instantiations were strongly associated with category. Test items pitted feature instantiations against their informational value. All but one feature in a test item were novel instantiations of informational features associated with one category. The remaining feature was a reinstantiation of a feature that in training had appeared, in its perceptual form, only in the opposite category. Although both informational and instantiated representations were important for categorization, some participants relied on the verbal feature representations while others relied more heavily on feature appearance. Most of this reliance on feature appearance involved a heavier weighting of familiar instantiations rather than failing to notice or appropriately encode novel features.

Feature appearance as a determiner of feature importance in classification

The role of variability in the appearance of a feature has largely been ignored in research on categorization until recently. Yamauchi and Markman (2000) showed that variability of feature manifestation on irrelevant dimensions impedes category learning, and Markman and Maddox (2003) extended that finding to variability along relevant dimensions. Markman and Maddox's data contain an interesting finding: perceptual variability was only a handicap for category learning when values associated with one category appeared in exemplars of the second category. This exact, or perceptual, overlap seemed to have an interfering effect on learning. Such findings of interference from feature overlap were noted by Shepard, Hovland and Jenkins (1961), and formed an important part of their argument for the centrality of relational abstraction in category learning.

However, Brooks and Hannah (2000; 2004) argued for a different relation between the surface appearance and an underlying abstract (informational) representation of a feature. In that work, the training items were family resemblance categories, but the feature instantiations were strongly correlated with category (Figure 1). This was intended to reflect a very common relation in the world. For example, both humans and birds have two legs, but despite variability in the appearance of different human legs, legs that look human never appear on birds. The more abstract representation of two legs might be important for some tasks, such as generalization to markedly novel feature appearances, understanding human locomotion and seeing relations between humans and other animals. However, the often-perfect association between a particular feature

instantiation and a category makes that instantiation an important clue for category identification.

To demonstrate the influence of particular instantiations, some of the test items pitted feature instantiations against their informational (abstract) values. All except one feature of each of these test items were novel instantiations of informational features differentially associated with one category, but the appearance of the remaining feature had previously occurred only in the opposite category – a perceptual lure feature (as with B in Figure 1). For these test items, people based their categorization on that single familiar lure feature, despite the presence of a greater number of rival features. If the lure feature also was given a novel instantiation (retaining an informational value characteristic of the competing category, as with A in Figure 1), then people tended to classify into the category that has the most features present in the item. Overlap at a more abstract level is still present in these all-novel-instantiation items, but the interference was sharply diminished.

Brooks and Hannah (2000; 2004) argued that such interference from perceptually familiar features, the perceptual overlap effect, could only arise if people used specific feature appearance when making categorization decisions. However, performance on the all-novel stimuli implied that people also relied on more abstract feature representations. This lead to our argument that there are two levels at which features are represented: as instantiated features, and as informational features.

Yamauchi and Markman (2000) make the crucial point that perceptual variability and similarity have different effects depending on task. They found, for example, that perceptual variability has little effect on performance in an inference task, but a large

effect in a categorization task. However, their paper also seems to imply that surface variability is problematic, preventing people from uncovering the structural relations defining some domain, and this seems further reinforced in Markman and Maddox's (2003) follow up. Their findings of impairment resulting from surface variability occurred in conditions in which the variance within categories approaches that of the variance between categories. In the real world this is likely to occur only when categories are very finely grained, such as when discriminating among different varieties of Monarch butterfly, or recognizing different individual faces. For a wide range of ordinary categorization decisions, the surface variability is systematically related to category identity. Because the perceptual instantiations for a wide range of physical categories are more strongly associated with categorical identity than are the informational features, instantiated features are usually a more reliable guide for categorizing than are informational features. The structural knowledge emphasized by Markman and colleagues is clearly crucial for many tasks, but the particular instantiations also have a critical role.

Feature representations, relational knowledge and rules

Although perceptual overlap between categories is very uncommon for the majority of the things we must use or recognize, researchers using artificial categories commonly employ it. This reliance on perceptual overlap strikes us as likely due to the belief that single features are not sufficient to support a categorization decision on their own. Arguments against the sufficiency of individual features to support classification, however, pertain to informational features, not instantiated features. If we were to treat the paw of a cat at an informational level—as the verbal label paw—then we have no

basis to decide whether an instance of paw signifies cat or dog or monkey, as all of them possess features to which the label paw applies. On the other hand, if we think of a cat's paw in terms of a particular manifestation, then such overlap disappears, and the feature is sufficient to classify an item as a cat. One need see only the paw of a cat around the corner to know that there is cat around the corner.

Brooks and Hannah (2000; 2004) have argued that representation of both informational and instantiated features are important in understanding the role of spoken rules. The rules volunteered about everyday categories and offered in formal instruction (e.g. medicine) are usually just feature lists with little indication of weighting and no formal decision procedure. As such, they are insufficient to account for the competence of expert classification. Their value, however, may lie in their ability to direct attention to, and thereby promote the learning of, particular instantiations of the features that are characteristic of the category. Brooks and Hannah's evidence implies that such feature-list rules did act to produce a reliance on the specific instantiations that appeared in training items and that were named in the rules, rather than solely a reliance on any feature that could be named by a term in the rule.

While most of Brooks and Hannah's (2000; 2004) participants produce feature-list rules, a few produced rules containing a specific decision-making procedure based on the number of features present (counting rules). These participants treated the combination of features as important, as they at least implicitly acknowledged that no single feature is sufficient to make a decision. For these participants, what matters is whether a feature is present or absent in a set of features, not how the present features are manifested. These participants, therefore, seemed to be reliant on informational

representations alone, an interpretation supported by post-hoc evidence that they show less of a perceptual overlap effect. However, when induction is used as a training method, very few such participants emerge, making formal analysis very difficult. We have found that such counting rules become much more common when the instructions include the names of the features that are relevant for each category, a condition we will use in this paper. Such instructions would be similar to a component of medical education, as well as many other educational programs.

Objectives of this paper

The purpose of this paper is twofold. First, we want to formally demonstrate the importance of the distinction between informational and instantiated features by contrasting the types of responses made by persons whose reports suggest reliance on only the informational content of features (e.g. those who give counting rules) with those whose reports suggest they are also sensitive to the particular instantiations of features. Therefore, our first major objective was to see if participants whose reports suggested a different degree of reliance on informational and instantiated features in fact show different patterns of decision-making.

Our second concern is to explore the basis for the interference arising from overlapping instantiated features. We entertain three possibilities: (1) A familiar instantiation may be more strongly weighted than less familiar instantiations of an informational feature. That is, people may evaluate the significance or importance of features based on their ease of recognition. In subsequent discussion, we will term this a feature-goodness heuristic.

(2) A familiar instantiation may disrupt the processing of rival features. Such disruption of attention could happen several ways. For example, less familiar features may be neglected because the more familiar feature is interpreted first, producing a top-down inhibition of processing inconsistent feature information (Johnston & Hawley, 1994; McClelland & Rumelhart, 1981). Even if we reject an inhibitory account, a perceptually familiar feature may dominate processing resources, biasing attention away from the less-familiar features. Alternatively, the extensiveness of search may in part be calibrated by the familiarity of the first processed features. In the medical literature, the phenomenon known as satisfaction of search suggests that feature search is highly vulnerable to surrounding information (Berbaum, Franken Jr., Dorfman, Rooholamini, Kathol, et al, 1990, Berbaum, Franken Jr., Dorfman, Rooholamini, Coffman, et al, 1991; Knottnerus, 1995; Nodine, 1992). Berbaum et al (1990, 1991) for example, showed that the detection sensitivity of radiologists at detection lesions on X-rays decreased as the number of lesions increased.

(3) A familiar instantiation may guide the interpretation of other features. A familiar feature that is interpreted first could set up an interpretive framework that guides the interpretation of the subsequently processed features. Although we may think of most features for common objects as obvious, and thus immune to any interpretive plasticity, some evidence suggests that feature processing may be quite plastic. For example, Wisniewski and Medin (1994) found that the interpretation of features present in children's drawings could be manipulated by manipulating the cover story accompanying the drawings. Brooks, LeBlanc and Norman (2000) found that even medical experts working with pictures of patients suffering from well-known diseases missed critical

features that they later called obvious. While this latter finding could reflect an imperfect search process, it would also arise if the 'missed' information were reinterpreted as normal background variation.

In the first experiment of this paper, participants will be asked for each test item to identify the novel features that are relevant to the "correct" category – that is, to the category with a majority of the informational features, consistent with the rule we followed in creating the categories. This is as if the participants in the experiment shown in Figure 1 had responded to the following questions for test item B: "what body shape," "what head shape," what body markings?" From their answers we would know whether or not they had misidentified the features as being consistent with a ramus (alternative 3 above). If they identified the features consistent with the bleeb category, yet subsequently identified the item as a ramus, we would know that the ramus categorization was not because they had failed to see the evidence consistent with calling it a bleeb (alternative 2 above). We would conclude that they had weighted the familiar instantiation more heavily than the more numerous novel instantiations (alternative 1 above).

Experiment 1

To explore the alternative interpretations of the effect of familiar instantiations, we use an iterative procedure in which participants always make two classifications of each test item. In between these immediately successive classifications, they will describe the appearance of critical features, as pointed out by the experimenter. Each test item will have a lure feature, that is, a feature consistent with a competing category. For one group of people this lure feature will perceptually match a feature seen in a different category in training (perceptual overlap). For another group, the lure feature will be a

perceptually novel instantiation of the values characteristic of another category (informational overlap).

What is of chief interest is what happens after participants on the first classification round give a category not supported by the most numerous informational features (an error if scored by an counting rule), especially when that response involves classifying the item as a member of the lure, or overlap, category. If the original response was due to heavier weighting of the familiar instantiation, pointing out the less familiar alternatives should have little effect, and they should persist with the original classification (persistence response). If the original response was due to failure to process alternative features, then when participants are forced to attend to the more numerous features consistent with the rule used to create them, they should revise their original answer (revision response). Lastly, if their original response entailed a reinterpretation of one of more of the rule-consistent features, then they should describe one or more of the rule-consistent features in terms consistent with the overlapping category (reinterpretation response). Responses involving a classification in terms of neither the correct nor overlapping category, in either the first or second pass, are confusion responses, and likely to arise from random memory slips. Although all of these responses would be treated as errors if scored by an additive (counting) rule, it must be emphasized that that counting rules may often not work in the real world. Rarely, for example, are medical diagnoses simply a matter of counting the number of supporting features. Reliance on feature quality, such as the familiarity of feature appearance, is often likely to be a reasonable strategy, and possibly an optimal one.

We have two main contrasts: response patterns between the perceptual overlap (PO: perceptually familiar lures) and the novel (informational) overlap groups (NvO: novel lures), and between participants reporting the use of different strategies, with the latter being the more critical comparison. We expect participants reporting counting strategies (counting group) will show a relative insensitivity to featural similarity, while those reporting just lists of features (listing group) should show sensitivity to instantiations of features.

Within the counting group, there should be no differences between participants responding to different lure types, because these differ only in perceptual familiarity. In reality, because some people reporting a counting strategy may not in fact use it until partway into the testing situation, or may not use it on every trial, some differences may exist, but these are unlikely to be significant. For listing participants, there should be differences across lure groups, as in both groups, there is sensitivity to perceptual familiarity, but familiarity is working in opposite directions. For the PO group the lure feature is the most familiar, while for the NvO group it is the other features that are the most familiar. Overall, participants receiving PO test items should make more lure-based errors than NvO participants, replicating previous findings of Brooks and Hannah.

This experiment deviates in two more ways from standard categorization paradigms: four categories are used instead of two, and features are taught explicitly rather than by induction. More than two categories is necessary if the differential activation of alternatives is to be a real contributor to feature processing. With only two categories, the options may come to form cognitive complements of one another such that suggesting one category activates representations of both. Explicit teaching is necessary

to ensure that participants have available an adequate descriptive and parsing scheme that applies to the test items, and so don't follow the overlap feature because nothing else is recognizable (as well as being more analogous to everyday instruction). Furthermore, it produces a much higher rate of counting strategies, which is critical for our comparison of strategies. Finally, it has the advantage of being closer to the methods of instruction used in many applied areas, such as medicine.

Methods

Participants

A total of 83 people participated in the experiment; two participants were replaced for not following directions and one for failing to meet the learning criterion. A total of 80 participants, therefore, supplied data that were analyzed, with 40 participants run in each condition. Participants were run in cohorts ranging from two to six participants per session. All participants were McMaster University students enrolled in a first-year psychology course, spoke English as their first language, and received course credit for participating.

Stimuli and apparatus

Stimuli consisted of line drawings of imaginary animals on overhead transparencies. Stimuli were displayed using an overhead projector, measured approximately 30 cm X 55 cm on the screen. The drawings consist of four "species" of imaginary animals created around a family resemblance structure based on three features: tail type, torso shape and number of feet/legs.

Each training category consisted of a prototype animal, with all the relevant features, and three exemplars that differ from the prototype by one relevant feature (one-

away exemplars). Category membership, therefore, is defined by a two-out-of-three-features family-resemblance rule. Examples of the training and test items are shown in panels A and B, respectively, of Figure 2. For all one-away exemplars, the value on the deviant feature matched the rule-consistent value for that feature for one of the other three categories, but at an informational level only. For bleebs, for example, the tail-deviant exemplar has a novel manifestation of a curly tail, which is the rule-consistent value for prin tails.

The 24 test items consisted only of one-away items, and were generated by skewing the features of each training items approximately 20° clockwise or counterclockwise, producing two skewed versions of each feature, both relevant and irrelevant. These skewed features were reassembled to yield two skewed test items for each training item. All features for any item were skewed in the same direction. This skewing was intended to make the items appear moderately rather than bizarrely unfamiliar. Our overall intention was to create a test set in which participants might be tempted to use either perceptual or informational features, depending on their availability. For this purpose, we thought it wise to make the items seem unfamiliar but not so unfamiliar that no recourse to the remaining perceptual information would be made.

For one set of 24 test stimuli, the perceptual overlap (PO) stimuli, the informational overlap feature was replaced with the unskewed feature found in the overlapping category. The informational overlap found in training, therefore, became a perceptual overlap at test. For another set of 24 test items, the novel overlap (NvO) stimuli, the old informational overlap feature was replaced with a novel feature that also overlapped informationally. At test, participants saw either the PO or the NvO stimuli.

For PO participants, the overlap feature was the most familiar feature in the item. For NvO participants the overlap feature was the least familiar feature, given that the other features were skewed versions of old instantiations.

Procedure

Training. Participants were told at the start of training that their task was to learn a set of four species of imaginary animals, and to apply that knowledge later to classifying new exemplars into one of the four categories. Participants then saw eight presentations of each item, in a variety of presentation formats, and with a variety of study tasks, specified below. Performance on the final presentation of individual training items and a presentation of the same items in a different order after test was used to assess learning. Only participants who achieved a minimum of 70% accuracy on both assessment rounds were accepted as having learned and retained the material (only one participant failed this criterion). This represents a performance level that is closer to perfect than it is to chance (criterion = chance + .6[1-chance]).

Training began with the experimenter pointing out the consistent and inconsistent features, both to facilitate learning and to ensure the participants would have a descriptive vocabulary for those features that transferred to the test materials. Neither the existence of informational overlap nor that of the two-out-of-three rule was pointed out. In addition to the initial pass through the training set, in which characteristic features are pointed out and named and inconsistent features acknowledged, participants received an additional seven presentations of the test items. In successive passes in training, the participants silently identified the characteristic features of training items (two times, with feedback on the features from the experimenter), silently identified the category of

the training items (one time, with feedback on the category from the experimenter) and overtly identified training items (two times, with feedback on the category from the experimenter). On two display rounds, participants were told to study items in any way they found effective. Training was thus extensive, taking about 40 minutes, and provided substantial support for learning the informational structure of the categories and the informational descriptors for individual features.

Test. Testing began with the experimenter reminding participants of their task, and providing a hint about the family-resemblance nature of the items: "...as with the items seen in training, there is no single feature common to all." The experimenter presented the 24 test items individually in a randomly generated order that was held constant across all participants. Each item was presented two immediately successive times. For the first presentation, the test item was displayed for eight seconds, participants wrote down the identity of the item, and then either listed the features they based their decision on or indicated that they were basing their judgment on an overall impression. Second, the experimenter indicated the two features consistent with the rule, and asked participants to describe them—e.g., "Please describe the torso and feet." Upon completing this step, participants repeated the first step by identifying the item, and justifying their decision. The item was kept on display until all participants indicated they were done the third step. This third step took approximately five to ten seconds.

Analysis

Main analyses. We first scored test responses according to whether the initial identification was consistent with the two-out-of-three rule. Errors according to this rule were further broken down into four categories: persistence responses, revision responses,

overlap reinterpretation and confusion responses, as defined in the introduction to this experiment. Analyses were based both on error frequencies across all error categories, and across overlap errors only (errors excluding confusion responses). Differences in error patterns between groups were tested using the maximum-likelihood χ^2 ($\underline{L^2}$), which is recommended under conditions of partitioning (Vokey, 2003). The differences between test groups on measures of overall accuracy, the rate of overlap responses (classifying items as a member of the overlap, or lure, category) and performance on the two assessment rounds were also examined.

Subgroup analyses. Participants were asked at the end of the experiment to describe how they made their decisions regarding the identity of the test, and from this response they were assigned to one of three strategy groups: feature counting, feature list, single-feature rule. The feature-counting subgroup were those participants within each lure group who stated that their strategy involved classifying items based on the number of features present (e.g., "I based my decision on which species had the most features present", or "Every item had two features from one species, so I just looked for two consistent features"). Feature-list participants simply listed several features as important, without mentioning how these features were combined or conflicts resolved (e.g., "I based my decisions by which of the three relevant features were present," or "I mainly relied on torsos and tails. Sometimes I looked at feet, but not much.") The single-feature participants stated that they exclusively used a single, specific feature (e.g., torso) to make distinctions. We initially compared those reporting a counting strategy with those reporting only a feature-list without any specific decision rule. Those few (six in each lure type condition) who report relying only on a single feature were ignored for these

further analyses because it is not clear whether these represent an abbreviated counting rule or an abbreviated feature list. After this overall comparison, these strategy groups were partitioned by lure-type condition, and comparisons made within each strategy group.

Results

Training and overall performance

A 2 X 2 mixed-design ANOVA, with lure group (PO, NvO) as a between-subjects factor and assessment round (end of training, after test) as a within-subject factor shows no main effect of groups on identification of training items, $F(1,78) < 1.0$, $MSE = 0.06$, $p > .8$. As inspection of Table 1 reveals, performances on training items are nearly identical for both test groups.

Analysis of the overall accuracy using an independent t-test finds that lure groups differed significantly in mean accuracy on test items, $t(65) = -3.13$, $p < .005$, df corrected for heteroscedasticity. As can be seen in Table 1, the participants classifying stimuli containing a PO feature were more disrupted than were participants classifying stimuli containing an NvO feature.

There was also a significant difference in the willingness of participants to assign test items to the overlap category, $t(55) = 4.05$, $p < .0001$, df corrected for heteroscedasticity. The participants classifying items with perceptually familiar lure features were much more likely to classify items as members of that overlapping category. Importantly, it is not simply that the NvO participants make fewer errors, and thus fewer overlap responses. NvO participants also made fewer relative overlap responses than do PO participants (66.2% of total group-wise errors vs. 85.7%,

respectively; normal approximation to the binomial: $z = -4.88$, $p < .0001$). This replicates the perceptual overlap effect (Brooks & Hannah, 2000; 2004) despite the substantial changes in training that emphasized informational features.

Error patterns

Main analysis. The two lure groups show a significant difference in the patterns of responses across error categories, $L^2 (3) = 13.48$, $p < .005$. As can be seen in Table 2, the PO participants made a greater number of persistence responses than did NvO participants, and these constituted a larger proportion of total errors than is true of the NvO participants (55% vs. 36%). Although numerically they also made more revision responses than did NvO participants, these represented an equivalent proportion of total error (26% vs. 25%).

However, it could be that the significant L-square was solely due to an increase in lure-based responses in the PO group, making confusion responses proportionally less frequent. This alone could produce significance. A set of four post-hoc item-based comparisons examined differences between the PO and NvO conditions in the mean number of participants making errors in each error category. Paired t-tests were used, with a Bonferroni correction applied to alpha, such that $\alpha = .0018$ (with a total of eight cells, there are a total of 28 possible comparisons; $\alpha = .05/28 = .0018$)

We found a reliable difference in the mean number of participants making persistence responses across items, $t (23) = 4.32$, $p < .0005$. In the PO condition, an average of 4.2 participants (10%) per item made persistence responses, but only 1.2 participants (3%) per item made persistence responses when the perceptually overlapping feature was replaced with a novel feature that overlapped only at a descriptive, or

informational level. Similarly, for the PO group there was an increase in the mean number of participants making revision responses, from a mean of 0.75 participants (2%) of NvO subjects making revision responses per item, to a mean of 2.5 participants (6%) of PO subjects, t (23) =3.62, $p < .0018$.

No other difference was significant, and therefore the difference in the pattern of errors in the two groups seems due mainly to a greater tendency of PO stimuli to elicit persistence and revision responses in participants. Persistence responses would seem to be more important of the two: Persistence responses accounted for nearly 55% of total errors for the PO group, but only for 36% of total errors for the NvO group, while revision responses were proportionally constant (26% vs. 25% of total errors).

To seek converging evidence of this, we re-analyzed errors as a proportion only of overlap errors, eliminating the conflation of overall error with overlap-specific error. This analysis, unfortunately, failed to find significant evidence differences in error pattern across lure groups, $L^2(2) = 1.48$, $p > .45$. The patterns do remain largely the same. Persistence responses were still proportionally larger for the PO group than for the NvO grop (64% and 55%, respectively). Now, however, revision responses occurred relatively more often for NvO participants than for PO participants (37% and 31%, respectively).

Subgroup analyses. Results for participants in all strategy subgroups and both lure groups are summarized in Table 3. Comparison of participants between strategies (counting, listing), regardless of lure type, finds a substantial difference between those who describe counting features and those who merely report a list of features that they used, $L^2(3) = 15.18$, $p < .0025$ (counting: $N = 42$; listing: $N = 26$). Differences between

strategy groups appeared restricted to persistence and revision responses, with counting participants making a smaller proportion of persistence errors than listing participants (28% vs. 53%, respectively), but making a greater proportion of revision responses (45% vs. 21%, respectively).

When we re-analyzed errors as to consider only overlap responses, we found that the differences between strategy groups remains significant, $\underline{L}^2(2) = 15.18$, $\underline{p} < .001$. Furthermore, differences in the error pattern became exaggerated, (persistence responses: 36% vs. 68 %, counting versus listing; revision responses: 58% vs. 27%, counting vs. listing). Those reliant on instantiated features, regardless of lure type, overwhelmingly made errors involving the discounting of rule-consistent features, whereas those reliant on informational aspects of features made mainly attention-related errors.

When we examined performance within strategies, we found little difference in the relative frequency of response types, as expected from people applying the same strategies and reliant on the same sources of information. For counting participants, $\underline{L}^2(3) = 0.28$, $\underline{p} > .95$ ($\underline{n}$ = 19, PO; $\underline{n}$ = 23, NvO). As can be seen in Table 3, even the absolute numbers of responses in each category were very similar, although, this may reflect a ceiling effect; there is, however, a trend for the PO participants to make more errors overall. Regardless of the familiarity of overlapping features among the test items, counting participants most commonly erred (deviated from a counting rule) by making an initial categorization based on the overlap features, which they changed when the two rule-consistent features were pointed out (revision).

However, listing participants (listing features without giving a decision rule) were reliably different in the overall pattern of responses, $\underline{L}^2(3) = 12.26$, $\underline{p} < .01$ ($\underline{n}$ = 15, PO; $\underline{n}$

= 11, NvO). When confusion responses were dropped, however, the relative frequency across the three response categories changed only slightly, but significance disappeared, $L^2(2) = 1.66$, $p > .40$. The significance difference in the patterns across all errors may be due to the PO group simply making more lure responses of all types while confusions remained constant, but there is a trend for the PO group to make even more persistence responses than the NvO group (60% vs. 33%). Both groups did make similar pattern of responses, tending to persevere when following a lure feature and discount more numerous rival features after acknowledging them.

Discussion

This experiment has extended the perceptual overlap effect to conditions that include a larger number of categories and questions than before, conditions that might have increased emphasis on informational features. Participants classifying items containing a single perceptually overlapping lure feature were more likely to classify the item on the basis of that lure feature alone than were people classifying items containing a perceptually novel lure feature that conveyed the same information. This was despite the training and overall test accuracy indicating that people in both lure-type conditions had abstracted the informational structure. The lowest accuracy rate among our four Strategy X Lure conditions was 72% (listing-PO condition).

This experiment also met our two major objectives. Our first major objective was to see if participants whose reports suggested a different degree of reliance on informational and instantiated features in fact made responses consistent with this interpretation. Those giving counting strategies, we argued, are more reliant on informational features than are those participants giving listing strategies, who in turn are

heavily reliant on instantiated features. These two groups made distinctly different patterns of errors that in fact showed the predicted differential reliance on informational or instantiated features.

Our second major objective was to get evidence that would elucidate the basis of the effect of instantiated features. We believe the pattern of errors suggests that a reliance on instantiated features is related to the evaluation of features in terms of their soundness as evidence. This prevalence of persistence errors on the part of listing participants occurred despite the fact that when asked to describe them they did so in a manner consistent with the training descriptions, indicating that they were aware of the informational content of the features that they were discounting. The low rate of reinterpretation responses not only allows us to rule out reinterpretation of ambiguous features as a factor driving the PO effect, but also provides evidence that our items were not noticeably ambiguous. Uncertainty about the identity of features was not a contributor to the PO effect elicited here.

Instead, our listing participants seemed to be following a rule that says: "If something looks weird, disregard it, irrespective of its informational content." However, the fact that NvO participants are accepting the very features that PO participants reject, implies that "weird" is a relative notion, and the rule might be better stated as, "Trust the best-looking features." We take this to mean that the interfering effect of perceptual overlap arises, at least in part, via the evaluations of the reliability, or 'goodness', of features. That is, perceptual familiarity affects the application of a feature-goodness heuristic, in which the ease with which features are recognized is used to gauge the reliability of features.

Feature-goodness heuristic

Our argument that the persistence responses by listing participants reflect a heuristic rests on several observations. First is the nature of the persistence responses themselves. The fact that people maintain their original decision after acknowledging rival evidence requires some deliberation on their part, suggesting that they are deliberately weighting that one feature more than its rivals. This point is strengthened by the overall level of performance on test items, which suggests that listing participants generally have a good command of the domain. Second is the observation that these persistence responses vary systematically with strategy type, becoming much less common for participants using a counting rule. When an algorithmic alternative is explicitly stated, it seems to supplant the more inferential reliance on feature familiarity, which is what we would expect if the persistence responses reflected a heuristic. Third, listing NvO participants did not show a different pattern of error responses from that of listing PO participants, but did in the overall rate of perceptual overlap errors. This is not surprising if both are using the same heuristic to evaluate features, but responding to items that differ in terms of which items are the most familiar. Last, our evidence for a heuristic based on the similarity of features to those previously encountered receives further support by Whittlesea and Leboe's (2000) evidence of a feature resemblance heuristic in the classification of nonsense letter strings. Thus in two independent studies using different materials and procedures, researchers have uncovered evidence that people use the perceptual similarity of features to make classification decisions.

Listing NvO participants made a substantial proportion of persistence responses, and this may seem odd, given that their lure features are perceptually novel. However,

we believe we can readily explain this in terms of a heuristic involving the ease of feature recognition or naming, rather than feature familiarity per se. A novel feature may still be easy to label even though it is entirely unfamiliar in some context. We will give a more detailed interpretation in the General Discussion.

Familiar features and the disruption of processing

There is also evidence that familiar features encouraged the neglect of rival features, suggesting that familiar features may disrupt attention, either by affecting search or by affecting processing of the feature itself. Even among counting participants, a perceptually familiar lure produced more revision errors than did a perceptually novel lure, suggesting that the application of such rules was not completely independent of familiar features, despite our evidence that such persons are more reliant on informational rather than instantiated features. It may be that a counting strategy allows people to override the effect of instantiated features, but not be completely unaffected by them. In collaboration with Dr. Judy Shedden, we are in the process of conducting reaction-time and ERP experiments on similar materials to examine this possibility.

Alternatively, decision-making and search processes may be guided by separate representations, under some circumstances such that decision-making can draw upon informational representations while search procedures draw upon instantiated representations. In the introduction to Experiment 1, we laid out three different ways in which instantiated features could affect search/attention: by inhibition of processing of alternate features, by dominating processing resources, or by the calibration of search extent by the familiarity of encountered features.

However, it is possible that nothing interesting is going on with regard to the revision responses. Even though our counting participants were more likely to break with their own rule more often in the presence of a perceptually familiar lure than in the presence of a perceptually novel lure, we cannot be sure when in the testing phase this rule was adopted, nor should we expect it necessarily to be applied across all items. Furthermore, the failure to execute his or her own rule could arise from fatigue, distraction, or boredom. Under cases of inattention to the task, a familiar feature may exclusively dominate decision making because of its higher salience. These rather uninteresting sources of error become more plausible when we consider the eight-second display time we used to get performance off of the ceiling. Preliminary evidence from our reaction-time data, however, suggests that, with two categories to decide among, the application of a two-out-of-three-feature rule takes approximately three seconds on average (Hannah, Brooks & Shedden, 2004).

Experiment 2

We have argued that the listing participants in Experiment 1 are reliant on instantiated features. However, we have no direct proof of this given that the instantiated features were not directly described in the rules given by the participants. To rule out the possibility that something in the nature of the rules generated by the participants, or in the consistency of application of their own rules is responsible for the differences in the pattern of errors observed, we ran Experiment 2. By giving the rule/strategy statements generated by subjects in the counting and listing subgroups of Experiment 1 to new participants who received no perceptual training, we can see if prior experience with the specific manifestations of features is necessary to produce the patterns of errors observed.

If counting participants are relying primarily on the informational aspect of feature representations, then this should communicate well, and we should see no difference in the error patterns of the counting rule generators and their yoked partners. If, as we have been arguing, the listing participants are relying primarily on instantiated feature representations and such prior experience is generally not communicated in their rules, then we should see a difference between listing rule generators and their yoked partners.

In examining the strategy statements of the entire group of PO participants, we found several that were clearly so incomplete they were not useful as rules. Among the counting participants, we found one person out of 19 whose statement was too inarticulate to be useful ("Mostly the features, i.e., tail & body shape together, or # of legs and tail together, or body shape & legs together."). Four out of 15 listing participants produced statements that were too incomplete to use as rules. For example, "I tried to remember which characteristics applied to which species. As well, it was very helpful when asked to describe certain features. Also, practice helped!!!", or "I tried to remember what was similar about them even when they looked slightly different. Even as they evolved I saw certain similarities w/the feet & torsos." Because a greater percentage of listing rules were inarticulate descriptions, by weeding out such statements we actually make the yoked groups more similar to one another than the generator groups were. This biases against finding a difference in our results.

Method

Participants

A total of 24 people participated, all McMaster undergraduate students who received course credit in a first-year psychology course for doing so. We dropped four

people for not following directions; therefore, 20 participants supplied data for this experiment, ten in each condition.

Stimuli and apparatus

Ten rule/strategy statements were randomly chosen from those generated by each of the Counting subgroup and the Listing subgroup, subject to the constraint that the rule be coherent and usable. The researcher intuitively decided whether a rule met this constraint. These statements were attached to the participants' response sheets, one rule/strategy statement per participant. Feature list rules simply listed what features to look for, e.g., "Just identify what type of body, then legs and how many were there, then if they have a tail and what kind was it (e.g., bushy or curly)." Feature-count rules provided a process for combining features, e.g., "Mostly, link two features of the animal to one of the species. All species have two common features." As none of the rules gave complete descriptions of specific feature values and linked them to species, we trained yoked participants to learn the verbal labels associated with each feature. Participants were given tables providing verbal descriptions of the values of the relevant features for each category. At test, the same line drawings used in the previous experiment's test section were used. Both the tables of descriptions and the line drawings were shown on an overhead projector.

Procedure

In training, participants were asked to memorize the association between the categories and the features so that they could apply the rules they were given to classifying test items into one of the species. Training took place in three stages, the first consisting of the complete feature table being displayed for five minutes while

participants studied it. In the second stage, a partially empty table was displayed for two minutes while participants filled in the missing features for each category; other features were simply blocked out. At the end of two minutes, the complete table was presented, and participants corrected their answers. Participants were then given two minutes to fill in an empty table. At the end of this time, the full table was displayed, and participants corrected their answers. Participants were required to make no more than three errors, with no more than one error being made in any category to qualify as having learned the categories. The test procedure was identical to that used in the first experiment. Participants were asked at the end of the test section what strategy they employed.

Results and Discussion

No significant difference existed between the two yoked groups in filling in the table at the end of training, $t(18) = -1.26$, $p > .20$ (listing yokes accuracy = 91.7%, counting yokes accuracy = 96.7%). On the test items, both sets of rule generators tended to be more accurate than their yoked partners. However, no significant differences were found, although the difference between counting generators and counting yokes approached significance, $t(9) = 1.86$, $p < .1$. For listing generators versus listing yokes participants, $t(9) = 1.15$, $p > .25$.

Performance across all error categories is summarized in Table 4. Listing yokes differed from listing generators in the error patterns across all error categories, $L^2(3) = 7.83$, $p < .05$, while there was no difference in overall error pattern between counting yokes and counting generators, $L^2(3) = 2.75$, $p > .4$. We were particularly interested, however, in potential differences among persistence and revision errors, and so we partitioned the data, segregating persistence and revision errors from reinterpretation and

confusion errors. Separate analyses were conducted on each sub-table (indicated in the table by boldface and regular curly brackets).

Of these subanalyses, the only significant difference was between listing yokes and listing generators for persistence and revision errors, $\underline{L}^2(1) = 4.25$, $\underline{p} < .05$. Considering only the distribution across these two response categories, listing yokes made fewer persistence responses than the listing generators (47% vs. 68%, respectively) and more revision responses (53% vs. 32%, respectively) than their generating counterparts. The listing yokes looked more like the counting generators than like the listing generators. It should be noted that to the extent that the counting yokes did differ from their matched counting generators, it was by displaying an extreme version of the feature-count behavior. They show a nominally greater tendency to make more revision errors than their matched counterparts or either of the listing groups.

The listing yokes did not behave like their generator counterparts, while counting yokes did. Listing yokes looked more similar to the counting groups from both experiments than to their own generators. Therefore, the rules originating with the counting generators more accurately captured the sources of information that they relied on, but something critical to the decision-making of the listing generators was left out of their statements. We suggest that this something is knowledge of the appearance of the training features, and the similarity between training and test features. Yoked participants still made a relatively high rate of revision errors, again suggesting that such revisions are largely independent of previous perceptual experience with features. However, this observed difference between listing yokes and generating participants is not due to the inadequacy of the rules. We failed to find any reliable difference in overall

test accuracy, and the observed differences were small. The rules were therefore conveying useful information.

Furthermore, these results rule out the possibility that the differences in error pattern between strategy groups points to a difference in the level of knowledge of the category. Both yoked conditions produced error patterns more like those of the counting participants of Experiment 1 than like those of the listing participants, but performed worse than, or no better than, the listing participants. We should point out that considering the results of the NvO Feature-list participants and the PO counting participants Experiment 1 also weakens such an argument. These two groups perform equally well, yet the NvO Feature-list participants mainly made persistence errors, while the PO Counting participants mostly made revision errors.

When asked their strategies at the end of test, 70% of listing yoked participants reported turning the feature list rule into a feature-counting rule, while 80% of counting yoked participants reported using a feature-count rule. Eight of the ten listing yokes reported that they found the rules usable, and nine of the ten counting yokes found their rules usable. That so many listing yokes reported counting strategies, while almost half of the original PO group in Experiment 1 did not, suggests that such strategies are highly salient to people when there is no perceptual familiarity to compete with such a strategy.

General Discussion

Consistent with Brooks & Hannah (2000; 2004) and Hannah & Brooks (2004), we have found additional evidence that both informational and instantiated features are necessary to account for categorization responses, and replicated the perceptual overlap effect. We showed that people reliant on specific feature instantiations made a different

pattern of error responses than did those reliant on informational representations of features. This fulfilled the first objective of this paper, by showing that participants whose reports suggested a differential reliance on informational and instantiated features showed different decision-making patterns.

Moreover, we have explicated how the effect of instantiated features operates, fulfilling the second objective of this paper. Before the final categorization of each test item, we forced people to name the relevant informational features of that item. These are the features that had been used to provide feedback and that the participants had been required to name in the initial instruction. On this final categorization, therefore, any tendency to persist in responding based on the instantiated features could not be attributed to a failure to notice or correctly interpret the more numerous informational features. If participants, however, reverse an initial categorization after identifying the relevant informational features, then this would suggest the initial error was due to a failure to notice or correctly interpret the informational features.

In Experiment 1, participants who merely listed the features as a statement of their categorization strategy tended to persist in categorizations consistent with the instantiated feature even after having named the relevant informational features present in a test item. That is, they were actively discounting the less-familiar informationally consistent features in favor of the more recognizable feature. In contrast, the participants who reported counting the number of relevant features to determine a categorization showed little influence of the instantiated features. They made fewer errors, as defined by the informational features, and little tendency to persist in a categorization consistent with

the instantiated features once they had specifically named the instantiated features present in the item.

The yoked control participants in Experiment 2 confirmed that the tendency to generate categorizations based on specific instantiations of features is not due to anything explicitly expressed in the rule. We infer that the knowledge relied on by the feature-listing participants in Experiment 1 that was not expressed in their rule was familiarity with the feature instantiations used in training.

Our ability to elicit a perceptual overlap effect at all under these conditions is evidence for the robustness of the participants' reliance on instantiated features. Training provided more substantial support for learning the informational structure of the categories and the informational descriptors of the features than had been provided in prior experiments by Brooks and Hannah (2004), or other experiments involving perceptual overlap (e.g., Markman & Maddox, 2003; Shepard, Hovland & Jenkins, 1961; Yamauchi & Markman, 2000). The experimenter taught participants from the outset what features were relevant, and provided them with labels—that is, informational representations—that transferred perfectly to test. Before starting the test phase, the experimenter explicitly told participants that there was no single feature common to all members of a category. Nonetheless, over 40 percent of our participants clung to a feature-goodness heuristic. This suggests that the reliance on specific instantiated features is quite ingrained, a finding that runs counter to assertions that verbal, rule-based representations represent a default (Ashby, Alfonso-Reese, Turken and Waldron; 1998).

Scope and sufficiency of feature representations

An informational level of description allows for generality by applying to many different surface forms. This supports transfer to situations that are different on the surface and supports the discovery of higher-order principles and relations. The structural mapping that Genter and Markman pointed to as critical for the development of high-level knowledge (Genter, 2003; Gentner & Markman, 1997; Gentner & Medina, 1998; Markman, 1996; Markman & Genter, 1993, 1997) is more likely to be successful when the features of the categories in a domain are described concisely enough to reveal the critical commonalities across categories or situations. Informational features can support transfer in a way that instantiated features cannot, freeing us from the clutches of the stimulus, and allowing us to discover higher-order relations and deep analogies.

Different feature appearances, however, are often meaningfully different in ways that are missed by abstract representations. Because of the strong association that a particular feature manifestation has with a particular category, a single instantiated feature or a feature very similar to it can be sufficient to identify an item as a member of a category. When trying to recognize some animal curled up under a table, we are helped in identifying the lazy creature by knowledge of how a cat's paw looks. The tight association between a category and the specific instantiations of features means that attending to only a small subset of features is likely to be a reliable and highly frugal way of representing whole categories (see Goldstein and Gigerenzer, 2002, for a cogent argument for heuristics as optimally "fast and frugal" decision-making processes).

In many everyday categories, sets of instantiated features are closer to being definitional than are informational features. The appearance terms used in medical rules,

for example, are approximations that are useful for instruction, communication and monitoring. In service of these functions, the number and specificity of terms are reduced well below that which language could afford. Consequently, it is easy to generate feature manifestations that are consistent at a general language level with individual terms that would not be accepted by an experienced practitioner as good evidence for that disease. The terms in these rules need to be grounded (coordinated with perception) on both a general language level, to serve the needs of beginners, and at a concept-specific level, to deal with the complexities of the world. The multiple instantiations that are associated with a particular term in a category rule constitute the concept-specific grounding for that term (see Solomon & Barsalou, 2001, for similar evidence for "local" and "global" groundings of words, and Brooks and Hannah, 2004, for an extension of the current argument).

Obviously, however, the advantage in identification that results from relying on a large number of instantiated features (e.g. the various examples of human legs previously experienced) implies a cost in ease of communication and monitoring. If each of the instantiated features is designated by a separate term in a rule, then the rule for a complex category is likely to become very long. Avoiding this cost is an advantage of grouping the potentially large number of manifestations under a single informational term. The verbal term two legs for the category human is a label for a list of previously experienced manifestations of that structure. The term is grounded in the manifestations in that it is not applied unless there is a match to some one of them, but it can still function in a simple manner for communications or comparisons with other concepts. As part of an identification rule, two legs is an invitation to learn the appearances of two legs in the

context of humans, rather than naming a general language criterion for membership. That is, the term is naming a focus of attention and attendant learning. As part of a general comparison between humans and other mammals, the term can function adequately at a general language level, separately from any particular instantiation. Which manifestations are grouped under a single informational term obviously is strongly affected by the learner's model of the concept.

Some consequences of variability of levels of feature representation

Feature-dependent structural knowledge. The view that perceptual variability is primarily interference, which seems to be implied in Yamauchi and Markman (2000), treats category learning as essentially the abstraction of structural or statistical relations. This is a long established view, explicitly exemplified in classic works such as Shepard, Hovland and Jenkins' paper (1961), and in contemporary works, such as the recent paper by Alfonso-Reese, Ashby and Brainard (2002). These latter authors conclude that the complexity of the covariance structure of a domain determines the difficulty of category learning for that domain.

Although the covariance structure of a domain (the pattern of feature overlap) may be what makes a categorization task difficult, a change in feature description can be what makes it easy. If we change the specificity of feature description, we also change its pattern of overlap across rival categories. There is not one single covariance structure or structural description that applies to a domain, but a family of such structural descriptions that vary according to how features are described. If we encode a cat's paw as a simple verbal label, paw, that feature will be more weakly associated with cats than if we encode the paw in a much more detailed way. By finding a way of describing features such that

there is no overlap of representations, we simplify the covariance structure of a domain and in turn simplify learning. This structural flexibility suggests that the variability found in similarity judgments (e.g., Medin, Goldstone & Gentner, 1993; Tversky, 1977; Tversky & Gati, 1978) should also be found in classification decisions. Of course, if no systematic variance in the manifestations of features occurs in an experiment, so that no alternate feature descriptions can be generated, then the only learning permitted is the learning of structural relations.

Feature learning and feature lists. By grounding each informational feature in a set of specific instantiations, we can describe many real-world categories with a very small number of informational features. If I know what a cat's paw or its face or tail looks like, then almost certainly any thing having one of these features is a cat, no matter how obscured the rest of the features are. Because of the different grounding of the terms in different concepts, we do not have a complex structure for identification in the sense of Alfonso-Reese, Ashby and Brainard (2002). Such a strategy would lead to classification rules that take the form of short feature lists. As Brooks and Hannah (2000, 2004) discuss, this is exactly the kind of classification rule common in medicine, and probably in most ordinary physical-object categories. Each feature in such rules would often be close to sufficient, and the presence of at least one could be treated as necessary, pending further investigation, even if we could generate logically possible counterfactual members that had none of the listed features. Feature lists grounded in category specific instantiations converge upon classical descriptions of rules as necessary and sufficient features, and diverge from the fuzzy rules of family-resemblance descriptions of conceptual structure (Rosch & Mervis, 1975).

Importantly, such feature lists are likely to be the only kind of "rules" that can work in the real world, where the materials are ill defined and highly variable, unlike those used in our experiments and in most experiments involving artificial categories. Certainly, it is more generally viable than the counting rules that were adopted in the current experiments, rules that were useful only because of the structure of our stimuli. In the context of natural categories, how many features does it take to make a cat? Given that a person suffering a heart attack can manifest anywhere from six to zero signs, an emergency-ward doctor who tried to make diagnoses by counting features would be taken to be a rank beginner.

The most critical learning work may involve learning the optimal feature descriptions for some domain, not the structural relations. What is "optimal", of course, depends on what the learner expects to do with the categories defining the domain. If it is not recognition among similar items, but grasping a higher-order relation that comes dressed in very different surface clothing, then the learner may want to seek out very general features that support coherence across wide variance. Like Schyns and colleagues (Schyns, Goldstone, & Thibaut, 1998; Schyns & Murhpy, 1994; Schyns & Rodet, 1997), we are arguing that encoding variability (Martin, 1968) is critical to understanding how feature knowledge is used in categorization. However, we are arguing that encoding variability is reflects not just of past history, but also of how we expect to use our feature knowledge.

Task sensitivity. Yamauchi and Markman's (2000) work shows that different tasks, such as inference and categorization, require different kinds of feature representations. This is an important point to which we are sympathetic, and we believe

our work extends their argument. Even within a categorization task, different levels of feature representation support different aspects of the same overall job[1]. Informational features allow wide generalization across situations, etc., providing the kind of scope necessary for communication and abstract reasoning. Instantiated features are more strongly associated with their category, providing the kind of discrimination necessary for rapid identification and accurate identification under limited or restricted viewing conditions.

Stimulus generalization. Shepard et al. (1961) concluded that primary stimulus generalization could not account for the high level of performance on their tasks, and thus some abstraction process involving selective attention must be involved, guided by explicit hypothesis testing embodied in verbal rules. However, their materials had the extensive perceptual overlap that has become characteristic of artificial categories, yet is not characteristic of many real-world categories. While we do not doubt that people can learn properties sufficient to define the structure of a domain at an abstract level, it is not clear that they ordinarily do when feature instantiations vary systematically between categories. Stimulus generalization around known instantiations may not describe well what went on in Shepard et al.'s experiment, but it may well describe what goes on in many real-world category-learning situations. At least, researchers may have underestimated its importance in the learning of many real-world categories. The abstract features offered by people for real-world categories may more often represent their causal or interpretive models of the concept than terms designed to provide a sufficient rule for classification.

Our findings and existing models and frameworks

An alternate interpretation of our results is that existing models could capture the effects of familiar-looking features by representing the features at a more detailed level than is often done. However, such a change would not be enough. By changing feature representations in models from informational features to instantiated features, effects involving instantiated features could be captured, but those involving informational features—such as good transfer to perceptually novel items, or the use of strategies involving informational features—would be lost. Thus, at a minimum, models must be changed to allow for both levels of feature representation. Along with Damian Jancowicz, we are currently experimenting with a simple two-layer heteroassociative neural network in which instantiated and informational features are represented as competitive inputs. Although this work is still highly preliminary, the model has produced the perceptual overlap effect, as well as some other interesting results suggesting a link between perceptual overlap with cue interaction (or, simultaneous blocking) effects (e.g, Kruschke, 2001; Tangen & Allan, 2003).

Our data suggest as well that the coordination between these two types of features has to be subject to strategy. Clearly, our listing participants are placing much more emphasis on instantiated features than are the counting participants even though they had been through the same training procedure. Further, the pattern of errors made by our listing NvO participants is quite distinct from the counting group. If they are reliant on informational features only, then they are doing something quite different with them than are the members of either counting group. Additional problems are raised when we consider that the familiarity of lure features modulates the size of categorical biasing

effects for people using feature-list strategies, but not for people using counting strategies (Hannah & Brooks, 2004).

SCAPE. It is for this reason that we suspect that an approach like Whittlesea's SCAPE account can be of use (Whittlesea, 1997; Whittlesea & Leboe, 2000; Whittlesea & Williams, 2001a, 2001b). Whittlesea argues that in addition to processing the content of the stimuli we encounter, the relative fluency of processing itself can provide information related to the task. Differences in the fluency of processing of items are often correlated with task-relevant properties, and so deviations from context-specific expectations regarding the fluency of processing are informative. Inferences about such fluency differences are often accurate within the task context.

For example, in a categorization task, feature manifestations that are shared among many category members but not with members of other categories are likely to be especially fluently processed. Relative fluency, then, could be the way properties such as the strength of association of features are actually judged. One way of reading our results is that Counting participants are reliant on identifying the feature labels alone (the content of the stimulus) while Listing participants are weighting the processing of feature labels by the ease with which the features are processed (relative fluency). Not only does this give a fluency advantage for familiar features, but also for features that are good matches to general labels. Feature manifestations that are novel within the specific context of the task can still retrieve large pools of prior instances of feature processing due to processing from outside current context if they are especially good matches to many experiences of that label outside of the current context. For example, in our experiment, a semi-circular torso in training always meant a torso of a certain size, with

the convex surface upwards; the same torso reversed in orientation is still similar to many instances of "semi-circle", even though novel to the current context. Independent of the details of the application, judgment of the processing of features, as advocated by SCAPE, is likely to be a useful resource in the interpretation of our experiments

COVIS. Our argument for two levels of feature representation is reminiscent of Ashby, Alfonso-Reese, Turken and Waldron's (1998) dual-system theory of categorization. However, their central distinction is importantly different than ours. Their chief distinction is between verbal, semantic processing versus implicit, perceptual processing, with the verbal process being the default. Our distinction is between representations of specific features and representations of generic features. We have used the term 'perceptual familiarity' throughout our paper because it was perceptual features that were manipulated in these experiments. However, in principle, representation specificity could just as easily apply to verbal materials. In fact, along with some colleagues in medical cognition, we have begun just such a series of experiments (Dore, Weaver, Norman, Brooks & Hannah, 2004). Undergraduates were trained to diagnose imaginary psychiatric conditions, and given test cases containing a mixture of semantic features. Some features were described using a familiar wording and others described using an unfamiliar wording. Early results show participants favoring the diagnoses linked to the features cast in the familiar wording over those cast in an unfamiliar wording.

Conclusion

We have extended Brooks and Hannah's (2000, 2004) finding that there seem to be at least two levels of feature representation, and shown that relying on one or another

involves different decision-making processes, as indicated by different patterns of errors. In many ordinarily encountered categories, instantiated features are more strongly associated with category identity than are informational feature representations. This makes it reasonable to adopt a feature-goodness heuristic, in which ease of feature recognition is used to evaluate the significance of features.

References

Alfonso-Reese, L., Ashby, F. G., & Brainard, D. H., (2002). What makes a categorization task difficult? Perception and Psychophysics, 64, 570-583.

Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., & Waldron, E.M. (1998). A neuropsychological theory of multiple systems in category learning. Psychological Review, 105, 442-481.

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Rooholamini, S. A., Coffman, C. E., Cornell, S. H., Cragg, A. H., Galvin, J. R., Honda, H., Kao, S. C.S., Kimball, D. A., Ryals, T. J., Sickels, W. J., & Smith, T. P. (1991). Time course of satisfaction of search. Investigative Radiology, 26, 640-648.

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., Behlke, F. M., Sato, Y., Lu, C. H., El-Khoury, G. Y., F., Fred W., & Montgomery, W. J. (1990). Satisfaction of search in diagnostic radiology. Investigative Radiology, 25, 133-140.

Brooks, L. R., & Hannah, S. D. (2000). Relation between perceptual and informational learning of family resemblance structures. Paper presented at the 41st Annual Meeting of the Psychonomics Society, New Orleans, LA.

Brooks, L. R. & Hannah, S. D. (2004). Feature lists and rules: The case for two levels of feature representation. Manuscript submitted for publication.

Brooks, L. R., LeBlanc, V.R. & Norman, G., R. (2000). On the difficulty of noticing obvious features in patient appearance. Psychological Science, 11, 112-117.

Dore, K., Weaver, B., Norman, G. R., Brooks, L.R., & Hannah, S.D. (2004). [The effect of instantiated wording on diagnosis of pseudo-psychiatric cases]. Unpublished raw data.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), Advances in the study of language and thought (pp. 195-235). Cambridge, MA: MIT Press.

Gentner, D. & Markman, A. B. (1997). Structure mapping in analogy and similarity. American Psychologist, 52, 45-56.

Gentner, D. & Medina, J. (1998). Similarity and the development of rules. Cognition, 65, 1-42.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. Psychological Review, 109, 75-90.

Hannah, S. D., & Brooks, L.R. (2004). The role of instantiated knowledge in producing categorical biasing. Manuscript submitted for publication.

Hannah, S.D., Brooks, L.R. & Shedden, J. (2004). [Strong rules override, but do not displace, sensitivity to feature appearance]. Unpublished raw data.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 110, 306-340.

Johnston, W. A., & Hawley, K. J. (1994). Perceptual inhibition of expected inputs: The key that opens closed minds. Psychonomic Bulletin & Review, 1, 56-72.

Knottnerus, J. A. (1995). Diagnostic prediction rules: Principles, requirements, and pitfalls. Medical Decision Making, 22, 341-163.

Kruschke, J. (2001). Toward a unified model of attention in associative learning. Journal of Mathematical Psychology, 45, 812-863.

Markman, A. B. (1996). Structural alignment in similarity and difference judgments. Psychonomic Bulletin and Review, 3, 227-230.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. Cognitive Psychology, 25, 431-467.

Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. Psychological Science, 8, 363-367.

Markman, A.B., & Maddox, W.T. (2003). Classification of exemplars with single- and multiple-feature manifestations: the effects of relevant dimension variation and category structure. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 107-117.

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. Psychological Review, 75, 421-441.

McClelland, J. L., & Rumelhart, D.E. (1981). An interactive model of context effects in letter perception. Part 1. An account of basic findings. Psychological Review, 88, 375-407.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. Psychological Review, 100, 254-278.

Nodine, C.F., Krupinski, E.A., Kundel, H.L.,Toto, L., & Herman, G.T. (1992). Correspondence: Satisfaction of Search. Investigative Radiology, 27, 571-573.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Schaffner, K. F. (2000). Medical informatics and the concept of disease. Theoretical Medicine and Bioethics, 21, 85-101.

Schyns, P. G., Goldstone, R. L., & Thibaut, J-P. (1998). The development of features in object concepts. Behavioral and Brain Sciences, 21, 1-54.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D.L. Medin (Ed.), The psychology of learning and motivation: Advances in research and theory, vol. 31, (305-349). San Diego, CA: Academic Press.

Schyns, P. G., and Rodet, L. (1997). Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory and Cognition, 23, 681-696.

Shepard, R. N., Hovland, C. I. & Jenkins H. M. (1961). Learning and memorization of classification. Psychological Monographs, 75 (13, Whole No. 517).

Solomon, K. O. & Barsalou, L. W. (2001). Representing properties locally. Cognitive Psychology, 43, 129-169.

Tangen, J. M. & Allan, L. G. (2003). The relative effects of cue interaction. Quarterly Journal of Experimental Psychology, 56B, 279-300.

Tversky. A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky. A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B.B. Lloyd (Eds.), Cognition and categorization (pp. 79-98). Hillsdale, NJ: Lawrence Erlbaum.

Vokey, J. R. (2003). Multiway frequency analysis for experimental psychologists. Canadian Journal of Experimental Psychology, 57, 257-264.

Whittlesea, B. W. A. (1997). Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In D. L.

Medin (Ed.), The psychology of learning and motivation: Advances in research and theory: Vol. 37. (pp. 211-264). San Diego: Academic Press.

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. Journal of Experimental Psychology: General, 129, 84 – 106.

Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. Journal of Experimental Psychology: Learning, Memory and Cognition, 27, 3-13.

Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. Journal of Experimental Psychology: Learning, Memory and Cognition, 27, 14-33.

Wisniewski, E.J. & Medin, D. L. (1994). On the interaction of theory and data in concept learning. Cognitive Science, 18, 221-281.

Yamauchi, T., & Markman, Arthur B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. Memory and Cognition, 28, 64-78.

Author note

Samuel D. Hannah and Lee R. Brooks, Department of Psychology, McMaster University, Hamilton, Ontario.

Funding for the first author has been provided by the Ontario Graduate Scholarship Fund, and for the second author by the National Science And Engineering Research Council. We would like to thank Seth Chin-Parker, Kevin Eva, Alan Neville, Geoff Norman, Brian Ross, Aimee Skye and Nikki Woods for useful suggestions and criticism. We would also like to thank Evan Heit and three anonymous reviewers for useful feedback on an earlier version of this paper.

The research reported here forms part of the first author's Ph.D. thesis.

Correspondence concerning the paper can be addressed to Sam Hannah, Department of Psychology, McMaster University, Hamilton, Ontario, Canada, L8S 4K1. The author can be reached via e-mail at hannahsd@mcmaster.ca.

Footnotes

[1] We would like to thank Brian Ross for pointing out this relation between our work and that of Yamauchi and Markman.

Table 1

Overall Performance for Perceptual Overlap and Novel Overlap Groups, Experiment 1 (Standard Deviations in Parentheses).

	Assessment round		Test items	
Lure group	After training	After test	Accuracy	Overlap errors
Perceptual Overlap (N = 40)	98.6% (2.6)	96.9% (5.8)	81.0% (18.8)	16.7% (16.1)
Novel Overlap (N = 40)	97.85 (6.3)	98.1% (4.9)	92.0% (11.6)	5.3% (1.2)

Note. Assessment rounds entailed identifying training items.

Table 2

Errors By Type For Perceptual Overlap And Novel Overlap Groups, Experiment 1

(Errors as Percentages of All Errors) {Errors as Percentages of Overlap Errors Only}.

Lure type	Error types			
	Persistence responses	Revision responses	Reinterpretation responses	Confusion responses
Perceptual Overlap	100 (54.9%), {64.1%}	48 (26.4%), {30.8%}	8 (4.4%), {5.1%}	26 (14.3%)
Novel Overlap	28 (36.4%), {54.9%}	19 (24.7%), {37.3%}	4 (5.2%), {7.8%}	26 (33.8%)

Table 3

Errors by Type for Strategy and Lure Groups, Experiment 1 (Errors as Percentages of All Errors) {Errors as Percentages of Overlap Errors Only}.

		Error types			
Strategy	Lure Group	Persistence	Revision	Reinterpret.	Confusion
Feature Counting	Perceptual Overlap (n=19)	14 (29.2%), {36.8%}	22 (45.8%), {57.9%}	2 (4.2%), {5.3%}	10 (20.8%)
	Novel Overlap (n = 23)	4 (25.0%), {33.3%}	7 (43.8%), {58.3%}	1 (6.3%), {8.3%}	4 (25.0%)
	Across Lure Groups	18 (28.1%), {36.0%}	29 (45.3%), {58.0%}	3 (4.7%), {6.0%}	14 (21.9%)
Feature Listing	Perceptual Overlap (n=15)	62 (60.2%), {70.5%}	21 (20.4%), {23.9%}	5 (4.9%), {8.3%}	15 (14.6%)
	Novel Overlap (n=11)	12 (33.3%), {57.1%}	8 (22.2%), {38.1%}	1 (2.8%), {4.8%}	15 (41.7%)
	Across Lure Groups	74 (53.2%), {67.9%}	29 (20.9%), 26.6%}	6 (4.3%), {5.5%}	30 (21.6%)

Table 4

Mean accuracy and error rates by response type by strategy and yoke condition, Experiments 2. (Errors as Percentages of All Errors) **{Errors as Percentages of Persistence and Revision Responses}** {Errors as Percentages of Reinterpretation and Confusion Responses}.

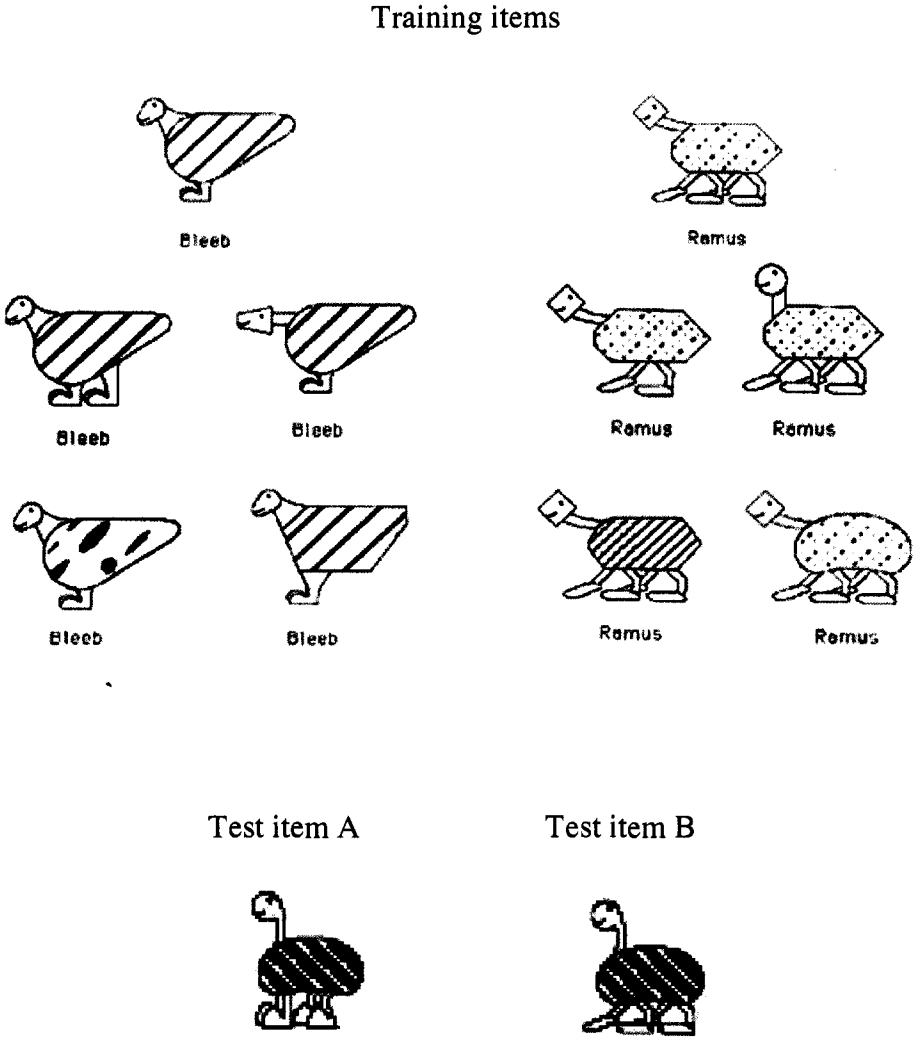
Groups (N = 10)	Accuracy	Error types			
		Persistence	Revision	Reinterpret.	Confusion
Counting Generators	85.8%	11 (32.4%), **{45.8%}**	13 (38.2%), **{54.2%}**	1 (2.9%), {10.0%}	9 (26.5%) {90.0%}
Counting Yokes	72.1%	19 (28.45%), **{35.2%}**	35 (52.2%), **{64.8%}**	3 (4.5%), {23.1%}	10 (14.9%) {76.9%}
Listing Generators	79.2%	28 (56.0%), **{68.3%}**	13 (26.0%), **{31.7%}**	3 (6.0%), {33.3%}	6 (12.0%) {66.7%}
Listing Yokes	67.1%	25 (31.6%), **{47.2%}**	28 (35.4%), **{52.8%}**	9 (11.4%), {34.6%}	17 (21.5%) {65.3%}

Figure captions

Figure 1. Items from one experiment in Brooks & Hannah (2000; 2004). Test item A tended to be called a bleeb, consistent with a majority of the characteristic features for bleebs: two legs, rounded body, rounded head, stripes. Test item B tended to be called a ramus, consistent with the one perceptually familiar feature: four legs. Note that both test items have the same informational description: four legs, rounded body, rounded head, stripes. They differ only in that for B the instantiation of the four legs had previously been seen on ramus training items.

Figure 2. In panel A are the definitions and prototypes (upper row) and examples of 1-away exemplars (lower row) for the four imaginary animal species used in training in Experiment 1. Panel B shows examples of a test item for the Perceptual Overlap group (PO, left) and the Novel Overlap group (NvO, right). For both groups the test item is a skewed version of the bleeb 1-away depicted at the far right in A. The tail in the PO item is identical to that seen in another category, while the NvO feature is a perceptually novel instantiation of the label 'curly'. Arrows highlight the pattern of feature overlap across training and test examples.

Figure 1


Training items
Bleeb
Ramus
Bleeb
Bleeb
Ramus
Ramus
Bleeb
Bleeb
Ramus
Ramus
Test item A
Test item B

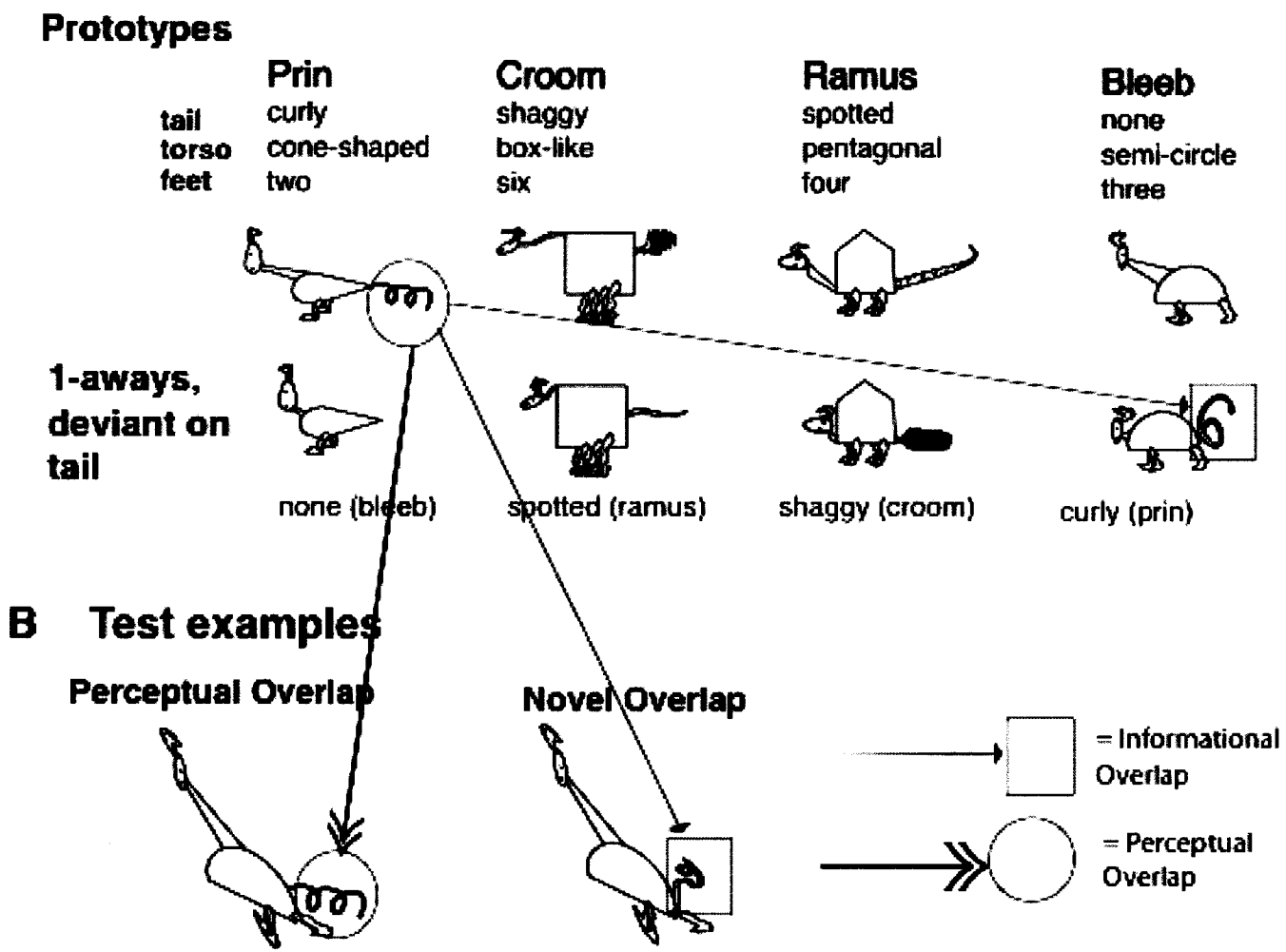

A Training exemplars
Prototypes
Prin
Croom
Ramus
Bleeb
tail
torso
feet
curly
cone-shaped
two
shaggy
box-like
six
spotted
pentagonal
four
none
semi-circle
three
1-aways, deviant on tail
none (bleeb)
spotted (ramus)
shaggy (croom)
curly (prin)
B Test examples
Perceptual Overlap
Novel Overlap
= Informational Overlap
= Perceptual Overlap

Figure 2

Chapter 4

Categorical Biasing Effect: Not Just for Doctors Anymore

Running head: INSTANTIATED FEATURES AND CATEGORICAL BIASING

The role of instantiated knowledge in producing categorical biasing

Samuel. Hannah and Lee R. Brooks
McMaster University

Draft of July 13, 2004

Submitted to the Journal of Experimental Psychology: Learning, Memory, and Cognition

July 14, 2004

Corresponding Author:

Sam Hannah
Department of Psychology
McMaster University
Hamilton, Ontario
Canada L8S 4K1
e-mail: hannahsd@mcmaster.ca
fax: (905) 529-6225

Abstract

Before categorizing novel exemplars, participants first evaluated the likelihood that the item was either a member of the correct category or a plausible alternative category. This was sufficient to bias categorization toward the suggested category. In a series of experiments, we show that several factors related to the accessibility of alternative categories have no effect, but knowledge of feature appearance and feature arrangement did affect the susceptibility to such biasing suggestions. We argue that the influence of such stimulus-specific knowledge is linked to the distinction between informational and instantiated features, and to the use of a feature-goodness heuristic.

The role of instantiated knowledge in producing categorical biasing

LeBlanc, Norman and Brooks (2001) found that medical students and medical residents could be biased towards a correct diagnosis or a plausible, alternative diagnosis simply by having them evaluate the plausibility of that diagnosis. This biasing effect was large; providing an alternative diagnosis to evaluate shifted the probability of concluding for the alternative by 20 to 70 percentage points. The suggestion also affected the reporting of features by the participants, increasing the likelihood of reporting features consistent with the suggested diagnosis and decreasing the likelihood of reporting features inconsistent with the diagnosis. This categorical biasing effect arose despite the diseases being fairly well known even to medical students (e.g., stomach cancer, lupus, Cushing's disease), and despite the photographs being taken from medical textbooks, and thus presumably representative of the disorders.

Biasing such as this has been used in the psychological literature as a marker for the contribution of top-down processes in object and category identification. However, recent work by Brooks and Hannah (2000; 2004) and Hannah and Brooks (2004) suggests that different levels of feature representation may be critical in accounting for when such biasing effects occur. The absence of such effects in research using novel categories may be because such research rarely, if ever, manipulates the level of specificity of feature representation. We will argue that allowing participants to vary the specificity with which they represent features is critical not only for accounting for categorical biasing effects, but also other forms of plasticity in categorization decisions.

Representational variability: Instantiated and informational features

Variation in the level of feature representation can be illustrated by considering the stimuli used by Brooks and Hannah (2000; 2004), depicted in Figure 1. At the top are exemplars for two species of imaginary animals (bleeb and ramus) that Brooks and Hannah's participants learned to classify. As with many real-world categories, no specific manifestation of any feature occurs in both categories, but more general feature properties do. That is, there is no perceptual overlap, but there is informational overlap. For example, four of the five bleebs have rounded heads, and so does one ramus. However, the rounded heads of the bleebs look very different from the rounded head of the ramus. Someone in Brooks and Hannah's experiments who notices the overlap of "rounded head" can respond in at least two ways. They may represent the bleeb feature in a specific manner: "head rounded like that [insert approximate perceptual image here]." Alternatively, they could maintain the abstract representation but seek higher-order relations to resolve conflicts, generating a rule such as, "bleebs have two or more of rounded head, rounded body and stripes." In most previous research using artificial categories, all features occur in the same form in both categories; that is, there is perceptual overlap between categories for all features. With such materials, the only recourse to differentiate categories is to rely on abstract feature representations and higher-order relations to resolve conflicts.

Brooks and Hannah (2000; 2004) used these materials to show that people do make use of both specific feature appearances (instantiated features) and more abstract feature representations (informational features) when making categorization decisions. Even though Test Item A in Figure 1, for example, is novel and contains a feature that is

informationally consistent with ramus (the four legs), participants will classify it as a bleeb with over 80% accuracy. Thus, people can use something other than mere perceptual similarity to guide their classifications: they can rely on the three informational features that indicate bleeb. However, behavior changes drastically when we replace the overlap feature (or lure feature) used in Test Item A with the legs that had been seen in ramus training items, as depicted in Test Item B—changing the informational overlap to a perceptual overlap. Now people seem to place much less reliance on the various informational features they used when classifying Test Item A, and 60% call Test Item B a ramus. Under some conditions, then, instantiated features will not only be used, but will override more numerous informational features.

Reliance on instantiated features is a reasonable strategy to use for many ordinary physical categories given that feature manifestations often are strongly associated with categorical identity. Both humans and birds can usefully be described as having two legs, but human looking legs never occur on birds. Because feature appearance is normally a reliable guide to categorization, reliance on it allows classification based on only one or two features. This allows rapid categorization, and categorization under limited viewing conditions or with degraded information. If you see something that looks very much like the tail of a golden retriever go past your fence, the inference that there is a golden retriever on the other side of the fence is very likely to be correct.

Some consequences of representational variability

Hannah and Brooks (2004) replicated and extended Brooks and Hannah's (2000; 2004) finding, providing evidence that people who relied on different types of feature representations produced different kinds of strategy or rule statements and used different

decision-making processes. People reliant on informational features produced counting rules in which features were equally weighted, and seemed to make decisions by following such rules. People reliant on instantiated features produced simple lists of features when asked how they made decisions on test items, and seemed to weight features by their similarity to features seen in training, or ease of recognition. We called this latter strategy a feature-goodness heuristic.

Variation in representation is systematically linked, therefore, to flexibility in decision-making procedures. The existence of a feature-goodness heuristic implies that categorization decisions can be manipulated by changing the ease of feature recognition, and that some categorization processes may thus be highly susceptible to biasing. Even the structure of a categorization domain is variable, and changes as the level of specificity of feature representation changes (Hannah and Brooks, 2004). The more general a feature representation is, the more widely across the domain it is distributed. Because the distribution of features is the categorical structure of a domain, categorical structure is itself plastic, and dependent on level of feature representation.

All of the above suggests that categorization is a highly flexible process, and this flexibility emerges in part from variability in the level of feature representations. Categorization should display a similar degree of variability as found for similarity judgments (e.g., Medin, Goldstone & Gentner, 1993; Tversky, 1977; Tversky & Gati, 1978). especially when instantiated features—and their attendant heuristic classification process—are heavily relied on. The categorical biasing effect readily elicited in the medical literature may, therefore, reflect variability that can be traced directly to the nature of the knowledge relied upon to make the decision.

Alternative factors producing categorical biasing:

There are at least three other plausible sources for categorical biasing effects: degraded or ambiguous stimuli, vague classification criteria, and low accessibility of alternative categories. First, it seems plausible that categorical biasing effects would be very difficult to produce in the categorization of clear presentations of everyday objects. Prefacing the display of a pen by asking whether it could be a rocket seems intuitively unlikely to encourage people to call a pen a rocket, even though both have a narrow cylindrical structure. However, comparable effects with everyday pictures are readily produced with highly degraded stimuli. Bruner and Potter (1964), for example, demonstrated that participants who were shown pictures of objects starting from very blurred focus had more difficulty correctly identifying objects than participants who were shown the same objects starting from medium blur or light blur. Bruner and Potter suggested that their participants developed hypotheses on the basis of the degraded information and were able to maintain those hypotheses because of the ambiguous nature of the stimuli. Similar results using different types of stimulus degradation were reported by Jacoby, Baker and Brooks (1989), and by Snodgrass and Hirshman (1991).

Second, the contribution of vague criteria to categorical biasing effects is well illustrated in studies by Wisniewski and Medin (1994). They had people classify children's drawings after either giving them a cover story that the drawings came from urban versus rural children, or creative versus noncreative children, or no cover story at all. The purpose of this study was to show that subjects who were given a prior theory were much more likely to note abstract features, or features as exemplars of a higher-order concept, than those who had no prior theory, who were more likely to identify

concrete features. The authors concluded, "Our findings suggest that theory- or knowledge-driven processes interact and are tightly coupled with data-driven processes in determining how and which abstract features are specialized. Knowledge-driven processes influence data-driven processes and vice versa" (p.265).

In common with many situations investigated in social cognition, Wisniewski and Medin's (1994) participants were not told which features were relevant to the categories; instead they devised relationships between theory and features as the task unfolded, based on presumably common sense relationships (e.g. creative children would make more detailed drawings). This degree of uncertainty, while certainly representative of many social judgment situations (e.g. judgments of honesty), is quite unlike the learned categorization tasks on which we will focus. Taking medical diagnosis as one example, students spend long hours learning the relationship between clinical and laboratory findings and diagnoses. In contrast to the Wisniewski and Medin (1994) study, we would not expect to see subjects invoking different features or reinterpreting features. Most of the biasing found by Wisniewski and Medin could plausibly be said to depend on the presence of categories with indefinite criteria as well as noticeably ambiguous features.

A third type of factor that may contribute to categorical biasing effects for both medical and everyday categories is the number and accessibility of alternatives. A suggested diagnosis or category may be difficult to resist to the extent that alternatives to the suggestion are more difficult to generate. At the minimum, this means that there have to be a reasonably large number of alternatives. If there are only two possible categories, as is the case with many laboratory studies of categorization, both alternatives are likely to be available on all trials. Not only are there many disease categories, but also each

disease is a name for a set of features that are linked by a common causal mechanism. Different causal mechanisms may form distinct semantic contexts around features, making categories and their associated features more or less available. Suggesting a disease may instantiate a semantic context or schema, aiding the processing of features and the retrieval of information consistent with that disease. Throughout the 1970s and 1980s research in both scene processing and memory retrieval suggested that the processing and retrieval of elements inconsistent with an established context or schema was impaired compared to information consistent with a context (e.g., Biederman, Mezzanotte & Rabinowitz, 1982; Godden & Baddely, 1975).

The aims of this paper

Ambiguous stimuli, vague classification criteria, and low accessibility of alternatives undoubtedly contribute to the production of categorical biasing. However, the experiments in this paper are designed to produce a categorical biasing effect in a situation that cannot be linked to these factors. The materials will consist of a small number of categories with well-known features, the features will be individually unambiguous and the decision conditions unhurried. Instead, we will emphasize the importance of the distinction between instantiated features and informational features – a distinction we believe to be important for a wide variety of phenomena in categorization.

We will show that reliance on instantiated features produces a different pattern of biasing effects than does reliance on informational features alone. Such a reliance on instantiated features will prove to result in an interaction between biasing suggestions and the perceptual familiarity of lure features in an item. Reliance on either instantiated or informational features, however, leaves people susceptible to biasing by activating

concept-specific attentional patterns learned in the course of attending to the relevant features of training items (attentional routines). Findings supporting the contribution of learned attentional routines to biasing emerged unexpectedly in the course of this work and will be described later in this paper.

We argue then that biasing arises because concept representations include concept-specific feature knowledge (instantiated features and attentional routines), as well as informational features. If the stimulus contains a feature whose perceptual manifestation normally occurs in another category, then suggest that other category is likely to be substantially more seductive for those reliant on instantiated features. Furthermore, we will argue that such information must be applied using a flexible classification strategy such as been argued for in reasoning and decision-making (Goldstein & Gigerenzer, 2002; Tversky & Kahneman, 1983). This combination of instantiated features and a feature-goodness heuristic allows us to produce categorical biasing under conditions that eliminate the well-known explanations just reviewed. Independent of such feature-based processing, suggesting a category activates other processing relevant to that category, such as the programming of visual attention or feature-search routines. This biases the search for features, and thereby biases the categorization decision. It should be kept in mind that in LeBlanc, Norman and Brooks' (2001) results, considering a tentative diagnosis affected not only the final diagnosis, but also the reporting of features.

Experiment 1: Initial Demonstration of Categorical Biasing

We designed this experiment to capture the role of stimulus-specific feature manifestations, for which we have just argued. As with the stimuli used in Brooks and

Hannah (2000, 2004), all of the training items except for prototypes are characterized by informational but not perceptual overlap between categories (Figure 2). To allow for an influence of perceptual familiarity, all but one of the features in the test stimuli were novel. The informational overlap feature in each test item was changed into a perceptual overlap (PO) feature (Figure 3). The most familiar feature among the test items is, therefore, a lure feature.

To capture factors related to the accessibility of alternative categories, we embedded feature instruction within causal stories regarding the evolutionary adaptiveness of each characteristic feature. These stories were intended to create distinct semantic contexts for each imaginary animal, minimizing access to alternatives inconsistent with an activated context. Additionally, four categories instead of the more usual two were used. Robinson and Hastie (1985) found that people when have more than three mutually exclusive alternatives to consider, evidence for one alternative is no longer taken as evidence against the others, as if they are no longer able to keep all alternatives accessible. Both these aspects of the design will be discussed in greater detail in later experiments in which their effects are separately evaluated.

The characteristic features associated with a category were explicitly taught to participants, as is done in formal instruction situations such as medical education. This necessitated acknowledging the existence of the rule-inconsistent feature appearances, but the experimenter did not point out that these deviant features actually overlapped with another category. It is very possible that the extensive instruction combined with explicit acknowledgment of deviant features could facilitate the abstraction of the statistical structures governing the categories. If people are customarily oriented towards acquiring

abstract knowledge, then the procedures employed in Experiment 1 should work strongly against finding any categorical biasing effect. The emergence of a reliance on perceptually mediated classification strategies under such conditions, therefore, would suggest that such strategies are deeply ingrained.

Method

Participants

We set an a priori target of 20 participants for each between-subject condition (biased group and unbiased control). Participants for all experiments were McMaster University students enrolled in a first-year or second-year psychology course and who spoke English as their first language. All participants received course credit for participating. Participants were run in cohorts ranging from two to eight participants per session. Participants who failed either to follow directions or to meet learning criterion were replaced until the limit of 20 participants per group was reached. In the unbiased control group, one person failed to meet the learning criterion, resulting in a total of 41 participants being run.

Stimuli.

Stimuli consisted of line drawings of imaginary animals presented on an overhead projector. The drawings consist of exemplars of four species of imaginary animals, called bleeb, ramus, croom, and prin. Each category was created around a family-resemblance structure based on three features: tail type, torso shape and number of feet. The training set for each category is composed of one animal with all the relevant features (prototype) and three exemplars that differ from the prototype by a single relevant feature (one-away exemplars). Category membership was defined by a two-out-of-three features rule. All

members of a category, therefore, have at least two features characteristic of that category. Examples of the training stimuli are shown in Figure 2.

For all one-away exemplars, the informational value of the deviant feature is identical to the rule-consistent value for that feature for one of the other three categories, but has a unique perceptual manifestation (informational overlap). For bleebs, for example, the torso-deviant exemplar has a rectangular or box-like torso, which is the rule-consistent value for croom torsos, but the bleeb box-like torso is not perceptually identical to the croom box-like torso. We will refer to the rule-inconsistent feature in a one-away item as the overlap feature or the lure feature. The category from which the overlap feature was "borrowed" will be referred to as the overlap category, and the category corresponding to that indicated by the two-out-of-three rule we call the correct category[1].

The 24 test items, examples of which are shown in Figure 3, consist only of one-away items, and were generated by skewing the features of each training items approximately 20° clockwise or counterclockwise, producing two skewed versions of all features. These skewed features were reassembled to yield two test items for each training item. Most importantly, the original overlap feature was replaced with its unskewed perceptual equivalent. The informational overlap found in training, therefore, became a perceptual overlap at test. This skewing was intended to make the items appear moderately rather than bizarrely unfamiliar. Our overall intention was to create a test set in which participants might be tempted to use either perceptual or informational features, depending on their availability. For this purpose, we thought it wise to make the items

seem unfamiliar but not so unfamiliar that no recourse to the remaining perceptual information would be made.

Procedure

Training procedure. The experimenter told participants at the start of training that their task was to learn a set of four species of imaginary animals, and to apply that knowledge later to classifying new exemplars into one of the four categories. Participants then saw eight presentations of each training item, spread over three blocks. Items were presented three times as quartets (one item from each of the four categories) in the first block, three times as pairs in the second block and twice as single items in the final block.

We used performance on this final presentation of the individual training items to assess whether participants had learned the training set. Identification was also assessed at the end of test using a different ordering of the same items to ensure that learning was sufficiently robust for relevant knowledge to be available throughout test. Only participants whose performance on both rounds exceeded a learning criterion had their test data included in subsequent analyses. We set the learning criterion to a minimum of 70% accuracy on both assessment rounds. This 70% level reflected a level of performance closer to perfect than to chance (criterion = chance + 0.6[1.0 – chance]; chance with four categories = 0.25). This formula for determining the learning criterion was used for all experiments reported in this paper.

Training began with participants receiving direct instruction with regard to the diagnostic features, and their descriptive terms. The experimenter pointed out both the correct and overlap features, named the correct features and gave an explanation

regarding the adaptive functions of both features. Although the overlap feature was pointed out, its overlapping nature was not, nor was the two-out-of-three membership rule pointed out. At different points in the experiment, participants were required to (a) silently identify the consistent features of each displayed exemplar, followed immediately with feedback to the whole cohort from the experimenter, (b) silently categorize exemplars, with feedback to the cohort, (c) write down the classification of exemplars, with feedback, (d) allowed to study items as they wish (free study).

Test procedure. Before presenting each test item, the experimenter asked biased participants to consider the likelihood that the item presented was a member of a given category; e.g., "How likely is it that [trial] number ten is a ramus." After viewing the item, participants rated the likelihood that the suggested category was the correct category for the item. They assigned a 0% likelihood if they had no doubt the suggestion was false, and a 100% likelihood if they had no doubt that the suggestion was true. Intermediate levels of perceived likelihood were given intermediate ratings. The experimenter suggested the correct category for one member of each skewed pair, and suggested the overlap category for the second member. There are, therefore, 12 items (three one-away items X four categories) cued to the correct category and 12 items cued to the overlap category. These suggestions were intended to induce the participants to consider either the correct or overlap category before making their classification.

Following the suggestion, the experimenter displayed the item, and participants rated the probability that the item was a member of the suggested category. Participants assigned a probability that corresponded to their confidence that the suggestion was true. If participants were certain the suggestion was false they assigned it a probability of 0%,

and a probability of 100% if they were certain it was true. Participants then identified the item by writing the initial of the category they selected on their response sheet. They could put down more than one answer with the constraint that they rank-order their answers, putting the most likely answer down first, second most likely down second, and so on. Unbiased participants simply identified the items, and then rated their confidence in their answer on the same scale as biased participants to equate their decisions for complexity and cognitive load. Participants were given a maximum of 30 seconds to respond to each item, or until everyone was finished. All participants finished before the 30-second deadline for the overwhelming majority of trials.

Analysis

We scored responses as correct, overlap or other according to the first category listed. Throughout all the studies, other response rates were constant and low for both biased and unbiased participants. Responses, therefore, were essentially binomially distributed (overlap and correct). Except for Experiment 5, there are more than ten responses per person in all the studies, and we can treat subject means as being normally distributed, permitting the use of parametric tests. Overlap responses showed less change across cueing conditions than correct responses for the biased group, and thus we chose overlap responses to be the DV to be as conservative as possible. Analyses of correct responses produced convergent, but more liberal, results.

The main analysis used a 2 X 2 mixed-design ANOVA with bias condition (biased, unbiased) as a between-subjects factor and cueing (cued to correct, cued to overlap) as a within-subject factor. For the unbiased condition, of course, cueing was a dummy factor. The presence of a categorical biasing effect is indicated by a significant

effect of cueing characterized by an increase in overlap responses when cued to the overlap category. The magnitude of the categorical biasing effect is given by the difference between cueing conditions, or cueing effect.

Results

After removing an inherently ambiguous test stimulus from the cued-overlap condition, and a stimulus from the cued-correct condition to balance observations[2], we found a significant effect of cueing, $F(1,38) = 5.16$, $MSE = 0.97$, $p < .05$. Participants made more overlap responses in the cued-overlap condition (1.38 [12.5% of cued-overlap items]) than in the cued-correct condition (0.88, [8.0% of cued-correct items]). The Cueing X Bias interaction was marginally significant, $F(1,38) = 3.30$, $MSE = 0.97$, $p < .08$. Although it appears from Table 1 that there is no biasing effect for unbiased participants and a moderate biasing effect for the biased participants, the marginal nature of the interaction makes interpretation unclear. Simple effects analyses (paired t-tests on cueing differences at each level of bias) confirm this interpretation, however. The differences between cueing condition are significant only for the biased group, $t(19) = -2.23$, $p < .05$. The cueing effect (cued to overlap – cued to correct) is graphed in Figure 4.

Table 1 also shows that participants in both groups were more than 96% correct when identifying training items at the end of training (assessment round 1), and when identifying training items after test (assessment round 2). A 2 X 2 mixed-design ANOVA, with assessment round as a within-participants factor and bias condition as a between-participants factor reveals only a main effect of assessment round, $F(1,38) = 6.05$, $MSE = 0.35$, $p < .025$. Not surprisingly, people did slightly less well on identifying

the training items after an intervening test task than immediately at the end of training. However, this is a very slight drop for both groups, again pointing to the members of both groups having learned the categories at a high level and equally well.

Discussion

Given earlier evidence for flexibility of feature representations, classification strategies and the use of heuristic decision processes, we expected to be able to bias people's categorization decisions even with simple, unambiguous materials. We fulfilled this expectation, demonstrating that categorization decisions display flexibility not unlike that found for similarity judgments (e.g., Medin, Goldstone & Gentner, 1993; Tversky, 1977; Tversky & Gati, 1978).

We elicited a categorical biasing effect that was not due to degraded or noticeably ambiguous stimuli, vague classification criteria, the criteria for the correct category being unavailable, or rushed judgment. This categorical biasing effect occurred despite explicit instruction regarding relevant features, which implies that it cannot be due to participants abstracting feature descriptions that do not transfer well to novel items. The high level of accuracy on test items shown by unbiased participants rules out the possibility that ambiguity of the test items[3] is critical to the effect. The high level of accuracy on training items by the biased participants rules out the possibility that confusion stemming from poor knowledge of the categories is critical. This effect emerged with only four categories, after only 40 minutes of instruction and using stimuli with features that are not noticeably ambiguous. Categorical biasing, therefore, requires neither the level of training required for successful medical diagnoses nor the complexity of materials found

in such diagnostic tasks, nor does it require the vague criteria often encountered in social judgments.

We next need to demonstrate that the categorical biasing effect is linked to the type of feature representation and with the type of decision-making process employed. Hannah and Brooks found that people reliant on instantiated features also seemed to weight features by their recognizability (a feature-goodness heuristic), and produced only a list of features when asked for their decision strategy at the end of test. People reliant on informational features, however, seemed to sum equally weighted features, and produced an explicit counting rule when asked for their strategy at the end of test.

We propose that the categorical biasing effect shown in Experiment 1 partially depends on a suggestion recruiting prior feature instantiations, and a person using a feature-goodness heuristic. Such recruitment of prior instantiations would help with processing the suggestion-consistent features of the stimulus, making them more readily recognizable. For anyone relying on a feature-goodness heuristic, this would enhance their perceived goodness at the expense of their rivals. However, if the features are poorly recognizable to begin with, the additional help may not result in them outweighing their rivals. People using a feature-goodness heuristic, as indicated by a feature-listing strategy (or, more simply, a listing strategy) report, may show a smaller categorical biasing effect when the lure feature is less recognizable than in the materials used in Experiment 1. People reliant on informational features, as indicated by a counting strategy report, should show either no biasing effect or one that is constant regardless of the quality of the lure feature. It is possible that our categorical biasing effect has more than one cause, and thus people reliant on informational features may be biased for very

different reasons than for those that explain biasing among people reliant on instantiated features.

By demonstrating an interaction between biasing suggestions and the familiarity of lure features, we would show that stimulus-specific feature representations are critical to the effect. By demonstrating that this interaction holds only for those participants giving listing strategies, we would provide additional evidence for the existence of a feature-goodness heuristic tied to the reliance on instantiated features.

Experiment 2: Control by perceptual similarity to previously learned features

In Experiment 2, one set of test materials possessed PO lure features, as were used in Experiment 1. For a second set of items, the lure features are taken from the transfer set of features. Although the lure features still come from the overlap category, they too are skewed, or modified (Modified Overlap, MO), like the rule-consistent features, and thus are no more familiar than the rule-consistent features. By changing the perceptual familiarity of lure features across participants, we should be able to vary the size of the categorical biasing effect across participants among those reliant on instantiated features. If people reporting a counting strategy in the biasing task are largely insensitive to the quality of features, then there will be no difference in biasing across such participants, regardless of the familiarity of the lure features. Only people reporting a listing strategy, therefore, will show a modulation of biasing by feature familiarity (a Cueing X Lure Group interaction). To assess the role of strategy in producing a categorical biasing effect, Experiment 2 will use a post-hoc segregation of participants by their strategy statements.

Methods

Participants

A total of 138 participants contributed data in this study. All participants were McMaster Undergraduates participating in exchange for credit in an introductory psychology course, and all spoke English as their first language.

The number of participants was influenced by two factors: the need to get a sufficient number of participants in each listing strategy condition (PO and MO) and the use of cohorts in experimental sessions. Thus we ran until we had at least 20 participants in each listing strategy condition, potentially truncating the condition with the larger number of participants to equalize group size. No truncation ended up being necessary for this condition, although the counting group was truncated in analysis to equalize group size. We ended up with 21 McMaster University undergraduates in each Listing X Lure condition.

In addition to these listing participants, another 96 participants supplied data. In the MO condition a total of 46 people using other strategies supplied usable data; 42 people gave a counting strategy, and four people gave some other strategy. In the PO group a total of 50 people using a strategy other than a listing strategy supplied viable data; 44 people gave a counting strategy, and six people gave some other strategy.

In the MO condition six people were replaced for: (a) failing to achieve learning criterion (3 persons), (b) being extreme outliers (biasing effects > 3 standard deviations above mean, 2 persons), and (c) for not following directions (1 person). In the PO condition, five people were replaced for: (a) not having English a first language (1 person), (b) failing to reach learning criterion (1 person), (c) being extreme outlier (1

person), (d) not following directions (1 person), and (e) for using a strategy for identifying training items based on counting the number of items in each category (1 person). A total of 149 people participated in the experiment.

Stimuli

The same stimuli used in Experiment 1 were used here, with two modifications. Item 1 had proved to be ambiguous because of its torso, and had to be dropped from analysis. This torso was modified to eliminate the ambiguity. As Item 15 shared this torso, it was given also the new torso, even though responses to this item in Experiment 1 were close to the averages for items. Second, we created a new set of stimuli by replacing the PO lure with the modified versions of these features (these features were the rule-consistent features of the test items for the overlap category). See Figure 5 for examples of training and test.

Procedure

Training and test procedures were largely identical to those in Experiment 1. No unbiased control was used because the only changes from Experiment 1 were to introduce a second test condition in which a lure feature was less familiar than in the standard case, and to replace an ambiguous stimulus with an unambiguous equivalent. Neither change should increase the likelihood of a stimulus bias. The skewed lure features used in the MO group were already used in the previous experiment, but not as lure features.

Analysis

After segregating participants into counting, listing and other strategy groups, we analyzed differences between lure groups (PO, MO) in overlap response rates separately

for listing and counting groups (counting-PO group truncated from $n = 44$ to $n = 42$ for equal group sizes). For each strategy group, we analyzed overlap responses using a 2 X 2 mixed-design ANOVA, with lure type (PO, MO) as a between-subject factor, and cueing (cued to correct, cued to overlap) as a within-subject factor.

Results

Mean overlap response rates for each lure group within listing and counting strategies are summarized in Table 2. Cueing effects across lure groups for both listing and counting strategy groups are depicted in Figure 6. Accuracy on training items on both assessment rounds (end of training and after test) was examined within each strategy type using the same 2 X 2 ANOVA design used to analyze test overlap responses. Analysis within each strategy type revealed no reliable effects. All groups achieved greater than 90% on both assessment rounds, and usually above 95% accuracy.

Listing participants

There were reliable main effects of both lure—$F(1,40) = 5.79$, $MSE = 4.35$, $p < .025$—and of cueing, $F(1,40) = 28.27$, $MSE = 0.97$, $p < .00025$. Both groups showed an increase in overlap responses when cued to overlap (3.31 [27.6% of cued-overlap responses]) compared to when cued to correct (2.17, [18.1%]). Overall, the PO group made more overlap responses (3.29 [27.4%]) than the MO group (2.19 [18.3%]). This replicates the perceptual overlap effect. Most importantly, we found a reliable Cueing X Lure interaction, $F(1,40) = 4.91$, $MSE = 0.97$, $p < .05$. For PO participants, suggesting the overlap category increased overlap responses by an average of 13 percentage points, but there was little more than a five-percentage-point difference for MO participants. Simple effects analyses within each Lure group showed that there was a reliable main

effect of cueing for the PO participants $F(1,20) = 33.41$, $MSE = 0.824$, $p < .00025$. The main effect of cueing was marginally significant for the listing MO participants, $F(1,20) = 4.18$, $MSE = 1.12$, $p < .06$.

Counting participants

There was a main effect of cueing, $F(1,82) = 13.79$, $MSE = 0.50$, $p < .0005$, as suggesting the overlap category increased mean overlap responses (1.02 [8.5%]) compared to suggesting the correct category (0.62 [5.2%]). Importantly, there was no Cueing X Lure interaction, $F(1,82) = 0.43$, $MSE = 0.50$, $p > .5$ Although suggesting the overlap category tends to slightly increase overlap responses for counting participants, this slight cueing effect is constant regardless of the familiarity of the lure feature.

Discussion

In Experiment 2 we found that reducing the perceptual familiarity reduced the categorical biasing effect for those following a listing strategy, but made no difference among those who used a counting strategy. Although counting participants still showed a categorical biasing effect, it was both very small and constant across levels of lure familiarity. Most of the effect of a biasing suggestion, therefore, relies on people using instantiated features when making categorization decisions.

We suggested in the introduction that biasing effects might reflect in part the operation of a feature-goodness heuristic, that is, the evaluation of the reliability of features based on their ease of recognition. We expected that any interaction between biasing suggestions and lure familiarity, therefore, would hold only for those giving a listing strategy the end of test, indicating to us a reliance on instantiated features. This is exactly the pattern we found.

Biasing and attention

Although people reliant on informational features showed a very small biasing effect, this effect was still real for these people and seemed independent of feature familiarity. A biasing effect could arise from factors having nothing to do with the level of feature representation, such as context affecting the retrieval of alternatives to the suggestion. Features consistent with a particular semantic context may be more likely to come to mind than inconsistent features, biasing the allocation of processing resources within some shared area of attention. Thus, if bleeb is suggested, this may call bleeb features to mind, and bias the allocation of resources to the bleeb-consistent features. This biased resource allocation may increase the likelihood that the bleeb features will be noticed before the non-bleeb features, increasing the likelihood of the non-bleeb features being neglected. In Hannah and Brooks (2004), people reporting counting strategies tended to make errors involving the neglect of rule-consistent information.

Alternatively, the biasing could result from concept-specific feature search patterns, or, attentional routines. Prior to developing their counting rule during training, counting participants may place more emphasis on one or two of the three features, and tend to neglect others. Thus, they may find the pentagonal torso of the ramus and the conical torso of the prin especially salient, and tend to ignore the tail and feet in both. On many ramus and prin training trials, they will develop a pattern of examining only the torsos, unless the torsos are overlap features. This could lead to suboptimal search patterns that are preserved as part of the ramus and prin concepts (attentional routines). When the one-away test prin with the pentagonal torso is cued to ramus, this would tend to re-instantiate these suboptimal attentional routines, leading to the neglect of the rule-

consistent features when the expected torso is discovered. When the item's skewed partner is cued to prin, the discovery of the suggestion-inconsistent torso leads to the corrective search that occurred in training when encountering the overlap feature, leading to the discovery of the rule-consistent features.

Attention and superordinate organization. A prominent feature of real-world categories is that they are clustered into superordinate groups. In medicine, for example, there are superordinates based on causal mechanisms—such as genetic disorders, infectious diseases, cancers. Other superordinates are organized around physical structures and systems—such as, cardiac diseases, respiratory diseases, and kidney diseases. Hierarchical organizations may establish separate contexts that influence how we attend to information.

The causal mechanisms organizing some medical superordinates could gate the accessibility of alternatives by establishing different semantic contexts, shaping what features are come to mind and receive priming. The research into encoding specificity and the role of semantic context in memory (e.g., Light & Carter-Sobell, 1970; Tulving & Thomson, 1973) has established that information inconsistent with a semantic context is less likely to be retrieved than that which is consistent with a current context. Similarly, Biederman Mezzanotte and Rabinowitz. (1982) showed that processing of items in a complex scene is impaired if items are inconsistent with the general theme of scene (e.g., a fire hydrant on the counter of a diner). A doctor may miss signs of a lung infection if he or she is thinking about the problem in terms of trauma, generating feature representations consistent with trauma and therefore biasing the processing of perceptual information towards trauma-consistent features.

Diseases of different systems are also distinct in terms of the pattern of attention deployed across the patient. Part of the concept of lung disease may be knowledge about where to look for signs. If this knowledge includes preserved records of prior searches and attention shifts—i.e., attentional routines—a suggestion would not only activate propositional information, but also prime attentional routines. This would favor the execution of category-specific searches at the expense of a broader search, leading to the neglect of information that fell outside of such search. A doctor may miss signs of a lung infection if he or she is thinking about the problem in terms of kidney disease, and searches only those areas of either patient or the medical records that are consistent with features of kidney disease.

Experiment 3: Superordinate structure and biasing

In Experiment 3, we explore the role of superordinate structure in mediating categorical biasing effects by putting our four categories into two superordinates: zoots (bleebs and crooms) and soots (ramus and prins). These classes were distinguished structurally and semantically. For zoots, the relevant features were in the upper half of the body; for soots, they were in the lower half of the body. In addition to this structural distinction, different types of evolutionary stories were given for the different classes. The evolutionary stories for zoots were based on social structures, while the stories for soots involved adaptations to terrain and climate.

If either semantic or physical organization influences categorical biasing, then the size of the categorical biasing effect will vary depending on whether competition occurs among members of the same superordinate or among members of different superordinates. That is, there will be a Cueing X Superordinate interaction.

Method

Participants

Forty participants supplied the data reported here, with 20 participants in each of the biased and unbiased groups. In the biased group, three participants failed learning criterion and were replaced; in the unbiased control, one participant was replaced for failing to meet learning criterion. Thus, a total of 44 McMaster undergraduates participated in this experiment, receiving course credit in either a first- or second-year psychology course. Participants were run in cohorts ranging in size from one to ten participants, although most ranged in size from six to ten participants.

Stimuli

Training stimuli. We modified the stimuli used in Experiment 1 to create two imaginary genuses, each consisting of two species. The zoot genus consisted of bleebs and crooms, and the soot genus consisted of prins and ramuses. For zoots, the diagnostic features for classification were switched from tail, torso shape and number of legs to horns (crooms = forward curving, bleebs = backward curving), head shape (croom = triangular, bleeb = oval) and neck length (croom = long, bleeb = short). The soots only had their nondiagnostic features (horns, head shape and neck length) modified from Experiment 1. Overlap occurred on both diagnostic and nondiagnostic features. For prototypes, all three nondiagnostic features took on novel values. For each one-away exemplar, two nondiagnostic features informationally matched features characteristic of separate species from the rival genus, with the third being novel. For example, a one-away bleeb's tail may match that of the prin, its feet take the number of ramus feet, while

its torso was a novel value. Examples of the prototypes and one-away training exemplars for all four species are shown in Figure 7.

In addition to these structural changes, the evolutionary stories involving the relevant features were amended. For the soots, the features were explained as adaptations to different social structures (solitary animals, highly territorial for crooms, versus gregarious, social animals for bleebs). The cover stories created in Experiment 1, that emphasized terrain and climate, were preserved for the zoots. Examples of the stories, taken verbatim from the experimental protocol are given in Appendix A.

Test stimuli. Test items were created the same way as before, with modifications made to allow for perceptual overlap and cueing either within the same genus or across genuses. For each one-away training item, we created four versions by skewing the rule-consistent features 20° clockwise and counterclockwise. For two of these items (one clockwise-skewed feature set, one counterclockwise-skewed set), we replaced the diagnostic overlap (from the same genus) feature by a training feature from the overlap category (same-superordinate items). For the remaining two items, one of the two nondiagnostic overlap features (from the rival genus) was replaced with the corresponding feature that occurred in training in the overlap category (different-superordinate items). Examples of the test items are shown in Figure 8. This resulted in 48 stimuli being created, with 24 same-superordinate items and 24 different-superordinate items. For biased participants, half of the items within each superordinate condition were cued to the correct category and half were cued to the overlap category.

Procedure

Procedures were identical to experiments 1 and 2. These stimuli are more complex than those used in Experiment 1 and 2, and thus it is possible that new stimulus

biases may emerge. We therefore used an unbiased control group again.

Analysis

We tested for the biasing of overlap responses with a 2 X 2 X 2 mixed-design ANOVA. Bias (biased, unbiased) was a between-subjects factor. Superordinate (same-superordinate items, different-superordinate items) and cueing (cued correct, cued overlap) were within-subject factors. As this experiment was actually run before Experiment 2, we did not do systematic probes into decision rules, and consequently cannot break participants into groups based on strategy statements, as we did in Experiment 2.

Results

We find again a main effect of cueing, $F(1,38) = 8.09$, $MSE = 3.78$, $p < .01$. Overall, suggesting the overlap category increased overlap responses to 2.86 (11.9%) from 1.99 (8.3%) when the correct category was suggested. However, this cueing effect varied by bias group, yielding a significant Cueing X Bias interaction, $F(1,38) = 15.22$, $MSE = 3.778$, $p < .0005$. For biased participants, considering the overlap category first increased overlap responses to 3.78 (15.8%) from 1.7 (7.1%) when considering the correct category first. For unbiased participants, however, the items dummy coded as cued-overlap actually elicited slightly fewer overlap responses (1.95 [8.1%]) than did items dummy coded as cued-correct (2.28 [9.5%]). Most importantly, our analysis revealed a three-way interaction of Cueing X Bias X Superordinate, $F(1,38) = 13.50$, $MSE = 1.45$, $p < .001$. Looking at the data in Table 3 and the cueing effect depicted in Figure 9, it appears that for biased participants suggesting the overlap category seems to have an even bigger effect for different-superordinate items than for same-superordinate

items. For unbiased participants, however, the different-superordinate items seem to produce a negative cueing effect. It looks as if there is a stimulus bias, but one that runs counter to the intended experimental effect, suggesting that the effect of the cue may be even larger than indicated by the data.

To clarify the interaction, we performed simple effects analyses consisting of two 2 X 2 repeated measures ANOVAs conducted within each bias group. Among biased participants, the main effect of cueing again was significant, $F(1, 19) = 13.68$, $MSE = 6.30$, $p < .0025$. Participants made more overlap responses when cued to the overlap category (3.78 [31.4%] as compared to when the correct category is suggested (1.70 [14.2%]). The Cueing X Superordinate interaction was marginally significant, $F(1, 19) = 4.13$, $MSE = 1.89$, $p < .06$, This would seem to confirm that there is a real cueing effect for the biased participants, and this effect is larger for different-superordinate items than for same-superordinate.

For unbiased participants, the only main effect is that of superordinate, $F(1, 19) = 6.45$, $MSE = 0.56$, $p < .05$. Participants made more overlap classifications for same-superordinate items (2.33 [9.7%] as compared to different-superordinate items (1.90 [7.9%]). The data in Table 3 indicates the presence of a stimulus bias operating in the opposite effect of the experimental bias, and the Cueing X Superordinate interaction this implies is significant, $F(1,19) = 11.86$, $MSE = 1.01$, $p < .005$. For the same-superordinate items, there is little difference in overlap responses across the items dummy coded as correct-cued and those items dummy coded as overlap-cued. However, for the different-superordinate items, those items dummy coded as cued to the correct category produced a higher rate of overlap items than those dummy coded as cued to the

overlap.

For performance on training items, only the main effect of assessment round was significant, $F(1, 38) = 4.59$, MSE = 0.61, $p < .05$. Accuracy again dropped slightly (after training = 94% correct, after test = 91% correct). Despite the greater complexity of the materials, therefore, both groups performed at a high level on the training items, and equally well.

Discussion

The categorical biasing effect is much larger when the correct and alternative categories reside in different superordinates than when rivals reside in the same superordinate. For different-superordinate items, the categorical biasing effect was equivalent in size to some of the effects seen in the medical literature (Leblanc et al, 2001). This increase in the biasing effect due to overlap between members of different superordinate classes, or superordinate cueing effect, could be due to the different semantic contexts being established by the different kinds of evolutionary stories accompanying feature instruction (adaptations as a result of social pressures, and adaptations as a result of demands in the physical environment). Alternatively, it could happen because features of both rival categories were usually attended to in the same-superordinate condition, being adjacent to one another, but the spatial segregation of the lure and correct features in the different-superordinate condition encouraged the neglect of one set of features.

The categorization literature has steadily increased its focus on the role of background knowledge, theories and other narrative organizations in concept formation since Murphy and Medin (1985) published their seminal paper on theory theory. Ahn

and colleagues (Ahn, 1998; Sloman, Love & Ahn, 1998) have argued that causal beliefs shape what features are seen as important. Rehder and Hastie (2001) argued that causal beliefs shape inferences. Given that people seem to attach great importance to causal information, it is possible that the different causal stories surrounding the classes produced distinctly different contexts surrounding their respective categories.

However, it could be that the superordinate cueing effect also reflects the differences in the spatial separation of rival information, and resulting differences in the learned patterns of attention associated with each superordinate. The mere physical separation of rival information alone cannot account for the effect because participants seem to inspect items much more thoroughly when no suggestion is made, even though the same physical separation exists. If concepts are formed by encoding a set of instances, this could include the distribution of attention across features. Suggesting a category, and thus activating that concept, would reinstate this instance-based procedural knowledge triggering a concept-specific distribution of attention. However, if an item is encountered without any prior expectation regarding its categorical identity, then attention should be deployed in a more diffuse fashion until an expectation is formed from processing the features.

Untangling these two possibilities is the focus of Experiment 4. Testing the hypothesis that our superordinate cueing effect is due to the different types of evolutionary stories is fortunately quite simple: we simply train one group of participants in the same manner as before but omit the evolutionary stories.

Experiment 4: Semantic or physical contexts

In Experiment 4 we essentially repeat the biased test condition of Experiment 3,

but change feature instruction to a simple verbal listing of the diagnostic features. If the evolutionary stories are providing a substantial component to the superordinate cueing effect, then the different-superordinate condition should show a marked reduction in the biasing effect as compared to that found Experiment 3. This should yield a three-way interaction of Instruction X Cueing X Superordinate.

Methods

Participants

Data were collected from 20 McMaster undergraduates enrolled in an introductory psychology course. This excludes the data from a participant who was replaced for failing to reach learning criterion. A total of 21 people participated for course credit in this experiment.

Stimuli

The same training stimuli used in Experiment 3 were used in Experiment 4.

Procedure

Training differed from Experiment 3 only in that no evolutionary story was given to explain the appearance of the features in the first round of the first training block. Instead, participants were simply told what features were relevant, and had this feature list reinforced with each display throughout the first round of the first training block. An excerpt from the protocol outlining feature instruction is given in Appendix A. Test procedures were identical to those used in Experiment 3, except no unbiased control was necessary.

Results

Potential differences in the frequency of overlap responses due to instructional changes were analyzed by a 2 X 2 X 2 mixed-design ANOVA. Instruction (causal story, feature list) was a between-subjects factor, and superordinate (same-superordinate, different-superordinate) and cueing (cued to correct, cued to overlap) were within-subject factors. Only the cueing and Cueing X Superordinate factors proved significant. For the main effect of cueing, $F(1,38) = 30.59$, $MSE = 4.48$, $p < .001$; cueing the overlap category increased overlap responses (3.38 [14.1%), compared to cueing the correct category (1.53 [6.4%]). For the Superordinate X Cueing interaction, $F(1,38) = 7.54$, $MSE = 1.46$, $p < .01$. As can be seen in Table 4, for both groups there is a larger cueing effect in the different-superordinate condition than in the same-superordinate condition. There were no differences between test groups. Importantly, that means that the slight reduction in the cueing effect for different-superordinate items for participants receiving the feature-list instruction is more apparent than real, as the three-way interaction of Instruction X Cueing X Superordinate is nonsignificant, $F(1,38) = 0.27$, $MSE = 1.46$, $p > .60$.

Discussion

Not only did we replicate the superordinate cueing effect, we found no reliable reduction of it after removing the causal stories, although we did find a slight trend in this direction. That suggests that the bulk, or entirety, of the superordinate cueing effect is due to differences in the physical structure of the classes. It must be acknowledged that the names zoot and soot were retained in Experiment 4, and this labeling distinction could have been sufficient to create distinct semantic contexts. It strikes us as unlikely that a mere naming effect could generate such a substantial effect when elaborate causal

stories have no effect. Nonetheless, further research is necessary to completely rule this out.

The results of the last two studies support the idea that concepts include knowledge of mental operations performed on class members, such as the deployment of attention (see Kolers & Roedigger,1984, and Whittlesea,1997, for similar descriptions of memory as the preservation of specific mental operations). The preservation of specific knowledge has been discussed previously in terms of instantiated features. Here, we are merely extending this from perceptual responses to attention-search responses.

If attention or search information existed only in propositional form, then it would easy to understand how suggesting a category could bias the starting point of a search, but it is not clear why it would tend to confine attention to suggestion-consistent areas. This problem is resolved, however, if we think in terms of a suggestion priming a search pattern stored from prior searches. Attentional routines have ends built into them as part of the pattern or routine.

There is little evidence that the semantic framework established by causal beliefs exerts substantial control over the accessibility of alternatives to a classification in our study. Of course, a straightforward classification task involving unambiguous material may not be a good place to look for such influences. Nonetheless, while we have been successful in finding evidence linking stimulus-specific factors to the categorical biasing effect, we have found nothing that links the accessibility of alternatives to it. Before we surrender the idea that the accessibility of alternatives is an important factor in mediating categorical biasing, however, we can readily test one more accessibility-related factor.

Experiment 5: The Number of Alternative Categories

While the semantic contexts created by establishing different evolutionary stories failed to turned these into isolated categories (Goldstone, 1996), some degree of isolation may result merely from the use of multiple alternative categories. Robinson and Hastie (1985) have shown that when the number of mutually exclusive alternatives to consider exceeds three, complementarity breaks down so that evidence for one alternative is no longer taken as evidence against the others, even with contrastive training such as we have used. With four categories to track, a suggestion pointing to one category may fail to cause participants to consider all of the alternatives, leading people to fail to consider the significance of conflicting features.

Experiment 5 was designed to test the possibility that our biasing effect is dependent in part on multiple categories reducing the accessibility of alternatives. With only two categories, it is unlikely that our participants could think of one category without the other category coming to mind. If the accessibility of alternative categories is critical to the production of a biasing effect, then this reduction to two categories eliminate the effect, or at least decrease it relative to Experiment 1.

Method

Participants

Twenty McMaster University undergraduates enrolled in a first- or second-year psychology course contributed, but two were replaced for failing to follow instructions during test. A total of 22 participants took part in this study. All participants received course credit for participation. Participants were run in cohorts ranging in size from one to eight participants, and all spoke English as their first language.

Stimuli

Stimuli consisted of modified versions of the bleeb and ramus items used in Experiment 1. These were modified such that feature overlap involved only features from these two species. The training and test features composing items, however, came from the training and test feature sets used in Experiment 1. See Figure 10 for examples of the training and test stimuli.

Procedure

Training and test procedures were identical to those used in Experiment 1. Learning criterion was adjusted to 80% accuracy on both assessment rounds to reflect the difference in chance performance. Although we have again changed the stimuli slightly, we have done so by making the set simpler than in Experiment 1, but with a subset of the same features. We can see no plausible argument as to how this could increase the probability of a stimulus bias, and therefore we have dispensed with an unbiased control group.

Results and Discussion

We can no longer assume that participant responses are normally distributed in each condition because there are fewer than 10 items in each cueing condition, precluding the use of parametric tests. A sign test performed on the number of overlap responses in each cueing condition found a significant difference between cueing conditions (paired sign test, $p < .01$). As can be seen in Table 5, participants made more overlap responses when cued to the overlap category (1.5 [25.0%]) than when cued to the correct category (1.0 [16.7%]). Table 5 also shows that the cueing effect is almost identical in size to that found in Experiment 1 (with items 1 and 8 removed), and a 2 (Experiment) X 2 (cueing

condition) χ^2 analysis found no differences between experiments in the pattern of overlap responses across cueing conditions, χ^2 (1) = 2.11, $p > .10$. No differences were found for performance on proportion correct for training items[4] using the usual 2 (group) X 2 (assessment round) mixed-design ANOVA.

The categorical biasing effect persisted without any meaningful reduction from that found in Experiment 1. Given that there are only two possibilities on any trial, and their exemplars are presented together in the first 16 training displays, it seems improbable that our participants failed to think of ramus as a possibility after bleeb had been suggested, and vice versa. The categorical biasing effect, therefore, cannot be due to a loss of accessibility of alternatives because of there being four categories, a number Robinson and Hastie (1985) found sufficient to observe an apparent reduction in the accessibility of alternatives when assigning probabilities of guilt to characters in mystery stories.

General Discussion

In each of our five studies, we have demonstrated a categorical biasing effect using materials consisting of a small number of categories with well-known and unambiguous features, under unhurried decision conditions. While the magnitude of the effect produced here is smaller than the range found in medical diagnosis by LeBlanc, Norman and Brooks (2001), this is likely due to the elimination of factors such as ambiguous features that may well contribute to actual medical classification situations. Even so, in some of our conditions we produced a biasing effect that approached the bottom of the range found by LeBlanc et al (different-superordinate condition, Experiments 3 and 4).

We suggest that three themes are necessary to account for these biasing effects.

1. Instantiated features: Perceptually familiar features are critical for producing this categorical biasing. As the perceptual familiarity of the lure features diminished, so did the size of the biasing effect for those dependent on instantiated features.

2. Concept-specific attentional routines: Procedural knowledge in the form of attention patterns preserved from past interactions with members of a category can explain nonrepresentational components of the categorical biasing effect and the superordinate cueing effect of Experiments 3 and 4. In Experiment 3, and 4 the biasing effect was larger when the rival information came from different categories distinguished by different configurations of relevant features (upper half of torso versus lower half of torso). It was only when a category was suggested that the physical segregation of features interacted with cueing to increase the bias effect.

3. Feature-goodness heuristic: Not only is there a biasing effect in all our experiments, but this effect interacts with feature familiarity (Experiment 2) implying that familiar features and suggestions each influence a common mechanism. This is readily understandable if those people relying on instantiated features are using the ease of feature recognition to judge how significant each feature is. The priming arising from a suggestion may aid feature recognition, enhancing the perceived significance or goodness of a suggestion-consistent feature, resulting in people discounting the suggestion-inconsistent features.

The importance of instantiated features.

Our work shows the importance of the specific manifestations of features, an issue that has been raised recently as an important topic. Barsalou has challenged

traditional treatment of feature representations as language-like descriptions (Barsalou, 1999; Goldstone & Barsalou, 1998; Solomon & Barsalou, 2001). Markman and colleagues' recent works also confirms the importance of specific feature manifestations (Markman & Maddox, 2003; Yamauchi & Markman, 2000). They show that categories characterized by highly variant perceptual manifestations of features are more difficult to learn than those characterized by relatively homogeneous feature manifestations. However, interference from perceptual variability implies sensitivity to perceptual variability within the context of category learning, implying that perceptual form is taken as relevant by the participants. Furthermore, Markman and Maddox found that perceptual variability interfered only when the specific perceptual values associated with one category occurred in exemplars of another, essentially replicating Brooks and Hannah's (2000; 2004) perceptual overlap effect.

We agree with Markman and colleagues that the stability of conceptual use across a variety of conditions strongly implies that features are represented in forms that are insensitive to contingent perceptual variability, that there are informational description of features as well as instantiated descriptions. We also agree that different tasks have different relations with perceptual representations and with perceptual similarity. However, we believe that their treatment of feature variance as disruptive in categorization may produce a misleading view of the utility and importance of perceptual representations, and obscure the need for both perceptually rich and perceptually sparse feature representations (see Pothos, in press, and Gentner, 2003, for two intriguing discussions of the importance of representational sparseness and richness). By creating categories with high within-category variability, Markman and colleagues create

situations that violate the strong association between feature appearance and categorization normally found in real-world categories. We suspect that they are underestimating not only the stability of surface manifestations, but also their systematic association with deep structure.

The stronger association between instantiated features and categorical identity than between informational features and categorical identity is important. Generic feature descriptions, such as "[has (paw, furry)]", do not distinguish between the furry paws of cats, dogs or monkeys, and are thus not sufficient to allow identification based on a single feature. Their concrete counterparts often do allow such identification based on a single feature. All a person needs to see is the paw of a cat feeling around a corner to know that there is cat around that corner. In that sense, arguments regarding the lack of sufficient features in common categories apply only to informational features, not to instantiated features. Because a single feature considered in its particular instantiation often permits ready identification, we are able to rapidly categorize hundreds of familiar objects. Furthermore, if all we need is a glimpse of one particular feature to successfully categorize an item, then all that is needed is for one feature to survive perceptual degradation.

However, we do successfully apply what is learned in a particular situation to very different situations. We do successfully communicate important information to people who do not have the experiences in which the information was originally embedded. To account for both this ability to generalize and communicate we need representations that have wide scope: we do need informational representations. However, for all the reasons discussed above we must also have highly specific

representations: concepts include instantiated features. We need both forms to account for the flexibility and stability humans display in applying knowledge. In a sense, cognition is governed by Marxist principles: From each (representation) according to its abilities, to each (task) according to its needs.

Perceptual versus specific as descriptors of knowledge

Our core distinction is between specific and generic representations. Throughout this paper we have used the term perceptual in reference to feature specificity because it is the perceptual form that we manipulated in this research. However, we can show similar effects using entirely verbal material (Dore, Weaver, Norman, Brooks & Hannah, 2004), where specific wordings influence categorization. The particular vocabulary and forms of verbal communication used by a patient in the course of an interview can be the source of specific prediction by physicians, even though for some purposes they are translated into the "medicalese" of textbooks and professional discourse (Eva, et al). Perceptual representations are an important form of specificity, but they are not the only ones.

We are, however, sympathetic to the view of Barsalou and colleagues that much of thought is perceptually grounded (Barsalou, 1999; Goldstone & Barsalou, 1998; Solomon & Barsalou, 2001). As Kolers and colleagues pointed out (Kolers & Roediger, 1984; Kolers & Smythe, 1984), percepts are dense symbols; that is, they convey a detailed amount of information. This makes them very suited to conveying specific information about items in some context; given the sensitivity to specific experience that people display, it seems plausible that much of thought involves perceptual symbols.

Restricted access: number of alternatives, semantic contexts and causal beliefs

Neither the number of categories to be learned (Experiment 5) nor having the categories embedded in a set of causal beliefs (Experiment 4) had any influence on the categorical biasing effect. Both may play roles in categorization phenomena, but neither is necessary to produce a categorical biasing effect. Causal stories, for example, may have a larger role in some real-world situations than are found here because part of their influence may be as an inferential guide regarding what to attend to (Heit & Bott, 2000). That role may have been supplanted in our studies by the direct instruction regarding relevant features and the unambiguous nature of the stimuli.

Obviously there is an important role for background knowledge in categorization, but we question how much of that necessarily takes on propositional form, let alone something that could reasonably be called a theory (Murphy & Medin, 1985; see Margolis, 1999, for an interesting critique of theory theory). The role of superordinate information in Experiment 3 and 4 regarding where to find diagnostic features, for example, could be accounted for propositionally, assuming that attached to the label zoot is a mental statement about where to find the relevant features. However, it strikes us as simpler to assume that the experiences of attending to ramuses are preserved and simply reinstantiated upon the generation of an expectation that a ramus is about to appear. Much conceptual knowledge may reflect a form of upward inheritance, where the responses made to individuals are compounded at the time of retrieval of multiple processing instances. Upward inheritance extends Hintzman's (1986) notion of structural abstraction via the pooling of multiple instances at retrieval to the generation of class-level procedural knowledge.

Categorization and attention

A second theme in our work is the preservation of specific mental operations, such as the specific patterns of attention deployed when inspecting a training item. We believe interesting parallels exist between the evidence supporting concept-specific attentional routines and item-specific attentional control. Using a Stroop paradigm, Jacoby, Lindsay & Hessels (2003) found that congruency manipulations usually performed in block-wise designs could be produced on an item-wise level. This implies that in learning about an item we learn item-specific response strategies that are automatically triggered by the item in its context ("automatic control of automatic processes"). If the properties of concepts are related to the properties of items via the upward inheritance alluded to earlier, then it would be expected that item-specific attentional control should give rise to concept-specific attentional control

The superordinate cueing effect is also a demonstration, to some extent, of inattentional blindness in a very different context than that in which it is normally discussed (Mack, Tang, Tuma, Kahn & Rock, 1992; Rock, Linnet, Grant & Mack, 1992; Mack and Rock, 1999; Simons & Chabris, 1999; Simons & Levin, 1998). We can interpret the superordinate cueing effect as implying that our participants were inattentionally blind to the features opposite to the region picked out by the biasing suggestion.

This opens up some interesting lines of inquiry. For example, if we combined the approaches of Experiments 2 and 3, would we find that diminishing the familiarity of the lure feature reduced the superordinate cueing effect at least for listing participants? Would we see that attentional processing is sensitive to familiar cues? Some of Mack

and Rock's work suggests this, especially the finding of reduced IB for one's own name over nouns or nonsense words (see Mack & Rock, 1999, for a review). It is possible that such a familiarity-driven reduction would arise for both listing and counting participants. We have been describing the latter as "reliant on informational features," but this reliance may extend only to evaluative or decision-making procedures. Processes of attention and feature search may be orthogonal to decision-making processes, and so decision-making rules may not reflect the influences guiding attention. Alternatively, counting rules may not render a person immune to featural familiarity even at an evaluative level, so much as allow them to override the influence of familiar features. With Judy Shedden, we have started reaction-time and ERP studies exploring whether people possessing counting rules are sensitive to similarity in ways that accuracy judgments do not reveal.

The superordinate cueing effect is also reminiscent of Goldstone's (1993) "regional saliency" bias. Goldstone showed that people's estimate of the percentage of black or white dots in a display can be biased by the clustering of items such that heavily clustered displays produce overestimation of the target items. He attributed this regional saliency bias to the graded nature of visual attention, which favors the processing of items that fall within a spotlight. However, Goldstone's findings raise many questions. Items that fall within the focus of attention should be better processed than those that fall outside, but it is not clear why being better processed should distort frequency judgments. Regional saliency may reflect another source of fluency (via attentional focus), and frequency estimation may share the same inferential basis we are arguing influences judgments of feature goodness.

Embodied content: Concepts and specific procedural knowledge

Both the idea that perceptually specific feature representations are used in categorization decisions and the idea that concept-specific attentional routines influence feature processing are consistent with recent moves to describe cognition in terms of embodied experience and action (Barsalou, 1999; Glenberg & Kaschack, 2002; Glenberg & Robertson, 2000; Lakoff, 1987; Wilson, 2002). The argument that concepts included specific procedural knowledge that reflects the preservation of operations performed on category members is also consistent a number of empirical findings.

The work by Schyns and colleagues (Schyns & Murphy, 1994; Schyns & Rodet, 1997; Schyns, Goldstone & Thibaut, 1998) on feature parsing provide another example of concept-specific procedural routines. Using stimuli that could be parsed in a variety of ways, they found at test that people would parse novel items using the parsing scheme that they had learned earlier. How they parsed features in training, therefore, was preserved with the other records related to a given category, and then retrieved later when encountering similar novel items.

Bub, Masson and Bukach (2003) showed Stroop-like effects after teaching participants gesture-color associations, and then showing common objects displayed in colors linked to object-congruent or incongruent gestures. Imaging research has shown that recognizing manipulable objects activates motor areas in the prefrontal cortex (Martin, Wiggs, Ungerleider & Haxby, 1996). Both sets of findings point to the ready association of motor routines with object identity. If we take the claims of premotor theory seriously (e.g., Bennett & Pratt, 2001), then the properties and patterns of visual

attentional patterns are based on those of motor (eye-movement) routines, and should be just as readily associated with specific items, and form part of object-category concepts.

Feature goodness as a judgment heuristic

We suspect that biasing in medical material may arise partly because feature manifestations sometimes overlap across disease categories, an occurrence uncommon in everyday categories. However, this alone might produce merely a high rate of errors of the type found in artificial materials by Brooks and Hannah (2000, 2004). For a biasing suggestion to systematically change conclusions, decision-making processes must access information in a highly flexible way. Such flexibility in decision-making is well accounted for by heuristic decision processes.

Hannah and Brooks (2004) found that participants reliant on instantiated features, as indicated by their use of a feature-list strategy, were more likely to discount rival, corrective information than participants reliant on informational features, and who gave a counting rule when asked for their strategy. Yoked control participants given the feature lists of a subset of participants in the first experiment, but no perceptual training, were much less likely to make these persistence errors. Instead, their error pattern resembled that of the counting-rule participants, although their overall accuracy rate was the lowest of all the groups. We argue that this means that people using specific feature instantiations use them in part to evaluate features for their reliability as evidence based on easy they are to recognize, discounting less recognizable features in favor of the more recognizable features present. That is, people weight features by their perceived goodness, or recognizability. Whittlesea and Leboe (2000) also found evidence that people used heuristics when making classification decisions, and related these to the

coherency of processing of features and items. We suspect that Whittlesea's discrepancy-attribution hypothesis may illuminate some of our findings (Whittlesea, 1997; Whittlesea & Leboe, 2000; Whittlesea & Williams, 2001a, 2001b)

Whittlesea and colleagues (Whittlesea, 1997; Whittlesea & Leboe, 2000; Whittlesea & Williams, 2001a, 2001b) argue that many decisions reflect inferences based on the fluency of processing (Jacoby & Dallas, 1981), relative to situation-specific expectations of such processing. In terms of our results, highly familiar features should recruit more prior instances of feature processing than unfamiliar features, increasing the fluency of processing these features relative to less familiar surrounding features and leading to an attribution of feature goodness or reliability. A suggestion regarding category identity may also recruit past instances of processing features consistent with the suggestion. This recruitment of previous similar instantiations would enhance the processing of features consistent with the suggestion relative to surrounding features inconsistent with the suggestion. Such greater fluency for the suggested features may be enhance the subjective goodness of the cue-consistent feature(s).

Regardless of how a feature-goodness heuristic may be implemented, its use makes sense when considering the conditions and constraints that people ordinarily deal with. Consider a strange tool: it has the head of a claw-hammer, but the head is attached to a bizarre shaft is vaguely similar to the grips of a pair of pliers. Is it reasonable to go with the familiar feature, and conclude that it is a bizarre type of claw hammer, or to rely on the less familiar feature, and conclude it is strange pair of pliers? Given that the perceptual appearance of features is normally not accidental, we argue it is more reasonable place one's trust on the readily recognizable feature over the strange manifestations.

Alternative explanations

Before we close this discussion, let us briefly consider two features of our design that limit alternative approaches to explaining our results. Nosofsky's Generalized Context Model (1986) could treat our biasing effect as resulting from the suggestion directing attention to suggestion-consistent dimensions or features, and away from those relevant to the actual stimulus. This can only work if categories are nonaligned (Gentner & Markman, 1997; Markman, 1996; Markman & Gentner, 1993; Markman & Gentner, 1997). At a minimum, categories must have different diagnostic features, such as the members the different superordinates in Experiments 3 and 4. However, in three of the five experiments in which a categorical biasing effect was obtained, the categories were fully aligned (Experiments 1, 2 & 5).

Tversky's (Tversky, 1977; Tversky & Gati, 1978) contrast model also has problems accounting for the categorical biasing effect, even though it is based on evidence demonstrating intrinsic flexibility in similarity judgments. Tversky and Gati (1978), for example, reported that similarity ratings and classifications varied depending on the surrounding context. This, however, is a variant of the accessibility-of-alternatives account that was tested and failed repeatedly throughout these experiments. Also, just as with Nosofsky's approach, the contrast model would also only work if there were different features diagnostic for different categories. If, however, we take the central concept of "saliency" in Tversky's account to mean something synonymous with "fluency of processing" (both are influenced by things such as familiarity, frequency, good form, etc.), then Tversky's account of similarity hinges on the fluency of processing of common features relative to the fluency of processing of distinctive features. This

brings it close to accounts such as Whittlesea's (e.g., Whittlesea, 1997), and becomes quite compatible with our understanding of a feature-goodness heuristic. In the terms used by Tversky and colleagues, the feature-goodness heuristic would reflect different weightings of features due to different saliencies, manifest in ease of recognition.

Conclusion

We have shown that a categorical biasing effect similar to that found in medical diagnosis can be produced using simple, well-defined and well-learned artificial materials. To account for this effect, we need to assume that much conceptual content is instantiated content. Concept representations include representations of the specific manifestations of encountered features, or instantiated features. Concept-specific attentional routines also contribute to concept representations. Such instantiated feature knowledge is often applied via a decision process based on the ease of feature recognition, or a feature-goodness heuristic.

References

Ahn, W-k. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. Cognition, 69, 135-178.

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577-660.

Bennett, P. J., & Pratt, J. (2001). The spatial distribution of inhibition of return. Psychological Science, 12, 76-80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. Cognitive Psychology, 14, 143-177.

Brooks, L. R., & Hannah, S. D. (2000, Nov). Relation between perceptual and informational learning of family resemblance structures. Paper presented at the 41st Annual Meeting of the Psychonomics Society, New Orleans. LA.

Brooks, L. R., & Hannah, S. D. (2004). Feature lists and rules: The case for two levels of feature representation. Manuscript submitted for publication.

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. Science, 144, 424-425.LeBlanc, V. R., Norman, G. R., & Brooks, L. R. (2001). Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. Academic Medicine, 76, S18-S20.

Bub, D. N., Masson, M. E. J, Bukach, C. M. (2003). The use of functional knowledge in object identification. Psychological Science, 14, 467-472.

Dore, K., Weaver, B., Norman, G. R., Brooks, L.R., & Hannah, S.D. (2004). [The effect of instantiated wording on diagnosis of pseudo-psychiatric cases]. Unpublished raw data.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), Advances in the study of language and thought (pp. 195-235). Cambridge, MA: MIT Press.

Gentner, D., & Markman, A.B. (1997). Structure mapping in analogy and similarity. American Psychologist, 52, 45-56.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. Psychonomic Bulletin and Review, 9, 558-565.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. Journal of Memory and language, 43, 379-401.

Godden, D.R., & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: on land and underwater. British Journal of Psychology, 66, 325-331.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. Psychological Review, 109, 75-90.

Goldstone, R. L. (1993). Feature distribution and biased estimation of visual displays. Journal of Experimental Psychology: Human Perception and Performance, 19, 564-579.

Goldstone, R. L. (1996). Isolated and interrelated concepts. Memory and Cognition, 24, 608-628.

Goldstone, R.L., & Barsalou, L.W. (1998). Reuniting perception and conception. Cognition, 65, 231-262.

Hannah, S. D., & Brooks, Lee R. (2004). Feature appearance as a determiner of feature importance in classification. Manuscript submitted for publication.

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D.L. Medin (Ed.), The psychology of learning and motivation, vol. 39 (163-198). San Diego, CA: Academic Press.

Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411-428.

Jacoby, L. L., Baker, J. G., & Brooks L. R. (1989). Episodic effects on picture identification: Implications for theories of concept learning and theories of memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 275-281.

Jacoby, L.L, & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 110, 306-340.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. D. (2003). Item-specific control of automatic processes: Stroop process dissociations. Psychonomic Bulletin and Review, 10, 638-644.

Kolers, P.A., & Roediger, H.L. III (1984). Procedures of mind. Journal of Verbal Learning and Verbal Behavior, 23, 425-449.

Kolers, P.A., & Smythe, W.E. (1984). Symbol manipulation: Alternatives to the computational views of mind. Journal of Verbal Learning and Verbal Behavior, 23, 289-314.

Lakoff, G. (1987). Fire, women, and dangerous things: What categories reveal about the mind. Chicago: University of Chicago Press.

LeBlanc, V.R., Norman, G.R., & Brooks, L.R. (2001). Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. Academic Medicine, 76, S18-S20.

Light, L. L., & Carter-Sobell, L.. (1970). Effects of changed semantic context on recognition memory. Journal of Verbal Learning and Verbal Behavior, 9, 1-11.

Mack, A., & Rock, I. (1999). Inattentional blindness: An overview by Arien Mack and Irvin Rock. Psyche [On-line], 5. Available: http://psyche.cs.monash.edu.au/v5/psyche-5-03-mack.html.

Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual organization and attention. Cognitive Psychology, 24, 457-501.

Margolis, E. (1999). What is conceptual glue? Minds and Machines, 9, 241-255.

Markman, A. B., (1996). Structural alignment in similarity and difference judgments. Psychonomic Bulletin and Review, 3, 227-230

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. Cognitive Psychology, 25, 431-467.

Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. Psychological Science, 8, 363-367.

Markman, A. B., & Maddox, W. T. (2003). Classification of exemplars with single- and multiple-feature manifestations: The effects of relevant dimension variation and category structure. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 107-117.

Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J.V. (1996). Neural correlates of category-specific knowledge Nature, 379, 649-652.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. Psychological Review, 100, 254-278.

Murphy, G. L., & Medin, D, L (1985). The role of theories in conceptual coherence. Psychological Review, 92, 289-316.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115, 39-57.

Pothos, E. M. (in press). The rules versus similarity distinction. Behavioral and Brain Sciences.

Rehder, B., & Hastie, R.. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. Journal of Experimental Psychology: General, 130, 323-360.

Robinson, L. B., & Hastie, R.. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. Journal of Experimental Psychology: Human Perception and Performance, 11, 443-456.

Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: Results of a new method. Cognitive Psychology, 24, 502-534.

Schyns, P. G., Goldstone, R. L., & Thibaut, J-P. (1998). The development of features in object concepts. Behavioral and Brain Sciences, 21, 1-54.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D.L. Medin (Ed.), The psychology of learning and motivation: Advances in research and theory, vol. 31, (305-349). San Diego, CA: Academic Press.

Schyns, P. G., and Rodet, L. (1997). Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory and Cognition, 23, 681-696.

Simons, D.J., & Chabris, C.F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception, 28, 1059-1074.

Simons, D.J., and Levin, D.T. 91998). Failure to detect changes to people during a real-world interaction. Psychonomic Bulletin and Review, 5, 644-649.

Sloman, S. A., Love B. C., & Ahn, W-k. (1998). Feature centrality and conceptual coherence. Cognitive Science, 22, 189-228.

Solomon, K.O. & Barsalou, L.W. (2001). Representing properties locally. Cognitive Psychology, 43,129-169.

Snodgrass, J. G., & Hirshman, E.. (1991). Theoretical implications of the Bruner-Potter (1964) effect. Journal of Memory and Language, 30, 273-293.

Tulving, E., & Thomsom, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch and B.B. Lloyd [Eds.], Cognition and categorization, 79-98. Hillsdale, NJ: Lawrence Erlbaum.

Whittlesea, B. W.A. (1997). Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In D. L. Medin (Ed.), The psychology of learning and motivation: Advances in research and theory, Vol. 37 (211-264). San Diego, CA: Academic Press

Whittlesea, B. W.A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. Journal of Experimental Psychology: General, 129, 84-106.

Whittlesea, B.W.A., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. Journal of Experiment Psychology: Learning, Memory and Cognition, 27, 3-13.

Whittlesea, B.W.A., & Williams, L. D. (2001b). The discrepancy-attribution hypothesis: II. II. Expectation, uncertainty, surprise, and feelings of familiarity. Journal of Experiment Psychology: Learning, Memory and Cognition, 27, 14-33.

Wilson, M. (2002). Six views of embodied cognition. Psychonomic Bulletin and Review, 6, 625-636.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. Cognitive Science,18, 221-281.

Yamauchi, T. and Markman, A. B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. Memory and Cognition, 28, 64-78.

Appendix A

Excerpts From Protocol for Experiment 3.

Material that in italics must be read verbatim, material in quotes is intended to be read near verbatim, and bullet points may use any phrasing.

Excerpt 1 (bullet points are condensed from original for space):

> "Zoots, that is bleebs and crooms, are known by their heads, horns and necks. These have evolved in different directions as a result of different social structures."
>
> - Bleebs highly social, many conflicts over social status,
> - Bleebs have evolved ways to settle these conflicts while causing minimal harm to the animals involved in such conflicts.
>
> "Therefore, males and females settle their disputes by head butting, and have evolved backward curving horns to this effect. Their heads have evolved into an oval shape which helps to dissipate the force of the blows well, and the short, muscular neck further acts as a shock absorber."

Excerpt 2:

> "Soots, that is prins and ramuses, are known by their tails, torsos and feet, and these have evolved in different ways as a result of adaptations to different terrains and climate.
>
> - Prins live on flat grasslands
> - Mainly at the edges of lakes, large ponds and large rivers. They are semi-aquatic.

" Their cone-shaped torso gives them buoyancy in water, while the curly tail can be contracted and expanded, to help propel them through the water. Their two-legged posture gives them great running speed on the flat prairie."

Excerpts from protocol for Experiment 4.

Excerpt 1

"Zoots, that is bleebs and crooms, are known by their heads, horns and necks.

- Bleebs

"Have backward curving horns, their heads have an oval shape, with a short, muscular neck."

Excerpt 2

"Soots, that is prins and ramuses, are known by their tails, torsos and feet.

- Prins

" Have a cone-shaped torso, with a curly tail, and have adopted a two-legged posture."

Author note

Samuel Hannah, Dept. of Psychology, McMaster University, and Lee R. Brooks, Dept. of Psychology, McMaster University.

Funding for the first author has been provided by the Ontario Graduate Scholarship Fund, and for the second author by the National Science and Engineering Research Council. We would like to thank Seth Chin-Parker, Karin Humphrys, Brian Ross, Allison Sekuler, John Vokey, Scott Watter and Bruce Whittlesea for very useful criticisms and suggestions.

This work was done as part of the first author's Ph.D. thesis. Some portions of this work were summarized in a brief review article prepared by the first author for a special edition of the Canadian Journal of Experimental Psychology in honor of Lee Brooks.

Send correspondence concerning this work to: Sam Hannah, c/o Dept. of Psychology, McMaster University, Hamilton, Ont. L8S 4K1. Send e-mail to: hannahsd@mcmaster.ca

Footnotes

[1-] The designation of the correct category as correct is somewhat arbitrary, as it assumes the use of our counting rule to classify items. Such rules may be so foreign and unusable in ordinary categorization that it may be argued that it is unrealistic to expect people to apply such a rule. Nonetheless, the term is useful for communication within this article.

[2] Examination of the control groups responding showed a majority (>60%) of unbiased participants classifying the first item as a member of the overlapping category. This item for the experimental participants was cued to the overlapping category, producing a stimulus bias in our experiment. Therefore, we analyzed the data with this item removed. To keep the numbers of observations equal across conditions, we removed an item from the cued-correct condition. This was the first item in the cued-correct condition, which produced perfect responses from control participants; this was as much of a stimulus bias in the opposite direction as could be achieved. The descriptive results and analyses are therefore conducted using eleven items per condition, rather than all twelve items.

[3] While one test item with a clearly ambiguous feature was found (see note above), the effect persisted even when this item was taken out of the analysis.

[4] Proportions were used instead of raw number correct because there are different numbers of training items between the two groups. Because of our learning criterion, all the proportion of correctly classified training items are above .7. The data were also analyzed, therefore, after applying an arcsine transformation to the subject proportions, but this is convergent on the analysis reported here, and therefore the more familiar analysis is reported.

Table 1

Mean Training Accuracy And Test Classification Responses (Standard Deviations in Parentheses), Experiment 1 (Initial Demonstration), for Biased and Unbiased Participants.

	Assessment Round			Classification Responses		
Bias Condition	After Training	After Test	Cued to	Correct	Overlap	Other
Biased	99.1% (3.1)	97.5% (6.2)	Overlap	85.5% (13.3)	**12.3%** **(12.9)**	2.3% (5.8)
			Correct	95.0% (13.0)	**4.1%** **(12.3)**	0.9% (2.8)
N = 20			Cueing Effect	-9.5% (14.9)	**8.2%** **(16.4)**	1.4% (5.3)
Unbiased	99.4% (1.9)	96.9% (6.9)	Overlap	86.4% (17.1)	**12.7%** **(17.0)**	0.9% (2.8)
			Correct	86.8% (19.7)	**11.8%** **(172)**	1.4% (3.3)
N = 20			Cueing Effect	-0.4% (7.3)	**0.8%** **(5.3)**	-0.4% (4.0)

Note. Classification responses represent the percentage of test items in each cueing condition classified as a member of the correct, overlap or other category A total of 16 training items are identified in each assessment round, and 11 test items in each cueing condition at test. Analysis was performed on raw overlap responses.

Table 2

Mean Classification Responses (Standard Deviations in Parentheses) within Cueing Condition by Lure Group and Strategy, Experiment 2 (Perceptual Familiarity).

Strategy	Lure	Cued to	Correct	Overlap	Other
Feature list	PO n = 21	Overlap	60.7% (14.5)	**34.1% (13.2)**	5.2% (6.7)
		Correct	76.2% (16.1)	**20.6% (13.6)**	3.2% (6.7)
		Cueing Effect	-15.5% (15.4)	**13.5% (10.7)**	2.0% (7.9)
	MO n = 21	Overlap	77.4% (15.2)	**21.0% (15.5)**	1.6% (3.4)
		Correct	81.7% (12.8)	**15.5% (11.9)**	2.8% (4.0)
		Cueing Effect	-4.4% (11.7)	**5.6% (12.5)**	-1.2% (4.0)
Counting	PO n = 42	Overlap	89.3% (11.1)	**9.5% (10.0)**	1.2% (3.5)
		Correct	94.2% (8.9)	**5.6% (8.8)**	0.2% (1.3)
		Cueing Effect	-5.0% (9.7)	**4.0% (9.1)**	1.0% (3.3)
	MO n = 42	Overlap	91.7% (12.3)	**7.5% (10.2)**	0.8% (3.1)
		Correct	93.5% (10.3)	**4.8% (7.2)**	1.8% (4.7)
		Cueing Effect	-1.8% (8.1)	**2.8% (7.5)**	-1.0% (3.8)

Note. Classification responses represent the percentage of test items in each cueing condition classified as a member of the correct, overlap or other category. There are 12 items per cueing condition. Analysis was performed on raw overlap responses.

Table 3

Mean Classification Responses (Standard Deviations in Parentheses) within Cueing and Superordinate Conditions, by Bias, Experiment 3 (Superordinate Structure).

		Same-superordinate			Different-superordinate		
Bias	Cued to	Correct	Overlap	Other	Correct	Overlap	Other
Biased N = 20	Overlap	56.3% (23.4)	**30.0% (22.0)**	13.8% (11.9)	49.2% (23.7)	**32.9% (17.4)**	17.9% (12.8)
	Correct	70.8% (17.0)	**17.9% (11.6)**	11.3% (12.8)	76.7% (19.0)	**10.4% (12.9)**	12.9% (9.5)
	Cueing Effect	-14.6% (25.1)	**12.1% (21.7)**	2.5% (10.5)	-27.5% (32.7)	**22.5% (25.8)**	5.0% (13.6)
Unbiased N = 20	Overlap	65.0% (23.8)	**21.3% (16.3)**	13.8% (14.1)	71.7% (20.5)	**11.3% (13.0)**	17.1% (11.9)
	Correct	69.6% (24.5)	**17.5% (138)**	12.9% (142)	65.4% (24.1)	**20.4% (136)**	14.2% (13.5)
	Cueing Effect	-4.6% (10.6)	**3.8% (13.4)**	-0.8% (12.7)	6.3% (14.5)	**-9.2% (11.8)**	2.9% (9.5)

Note. Classification responses represent the percentage of test items in each cueing condition classified as a member of the correct, overlap or other category. There are 12 items per cueing condition. Analysis was performed on raw overlap frequencies.

Table 4

Mean Classification Responses (Standard Deviations in Parentheses) within Cueing and Superordinate Conditions, by Instruction Group, Experiment 4 (Causal Story).

		Same-superordinate			Different-superordinate		
Instruction	Cued to	Correct	Overlap	Other	Correct	Overlap	Other
No story N = 20	Overlap	63.8% (22.3)	**24.6% (19.2)**	11.7% (9.1)	61.3% (213)	**25.0% (17.3)**	13.8% (10.2)
	Correct	77.9% (18.8)	**14.6% (9.7)**	7.5% (11.9)	81.3% (15.0)	**10.8% (9.2)**	10.85 (10.8)
	Cueing Effect	-14.2% (18.0)	**10.0% (14.7)**	4.2% (12.5)	-20.0% (21.5)	**17.1% (17.2)**	2.9% (10.2)
Causal story N = 20 (Ex. 3)	Overlap	56.3% (23.4)	**30.0% (22.0)**	13.8% (11.9)	49.2% (23.7)	**32.9% (.17.4)**	17.9% (12.8)
	Correct	70.8% (17.0)	**17.9% (11.6)**	11.3% (12.8)	76.7% (.19.0)	**10.4% (12.9)**	12.9% (9.5)
	Cueing Effect	-14.6% (25.1)	**12.1% (21.7)**	2.5% (10.5)	-27.5% (32.7)	**22.5% (25.8)**	5.0% (13.6)

Note. Classification responses represent the percentage of test items in each cueing condition classified as a member of the correct, overlap or other category. There are 12 items per cueing condition. Analysis was performed on raw overlap frequencies.

Table 5

Mean Training Accuracy and Test Classification Responses for Experiment 5 (Number Of Categories), by Domain Set Size (Standard Deviations In Parentheses).

	Assessment Round			Classification Responses	
Domain Size	After Training	After Test	Cued to	Correct	Overlap
Two Species N = 20	98.1% (2.3)	97.5% (3.3)	Overlap	75.0% (3.3)	**25.0% (3.3)**
			Correct	83.3% (2.8)	**16.7% (2.8)**
			Cueing Effect	-8.3% (13.1)	**8.3% (13.1)**
Four Species (Ex.1) N = 20	99.1% (3.1)	97.5% (6.2)	Overlap	85.5% (13.3)	**12.3% (12.9)**
			Correct	95.0% (13.0)	**4.1% (12.3)**
			Cueing Effect	-9.5% (14.9)	**8.2% (16.4)**

Note. Classification responses represent the percentage of test items in each cueing condition classified as a member of the correct, overlap or other category. For the two-category group, there were eight training items in each assessment round and six test items in each cueing condition. For the four-category group there were 16 training items in each assessment round, and 11 test items in each cueing condition. Analysis was performed on raw overlap frequencies.

Figure Captions

Figure 1: Items from one experiment in Brooks & Hannah (2000; 2004).

Figure 2: Examples of the training stimuli in Experiment 1. Prototypes for each of the four species are shown in the top two rows, with a description of the characteristic values for the definitional features for each species. The bottom two rows depict one-away exemplars for each species; in this case, the tail of each item overlaps with the informational value of a tail from another species.

Figure 3: Examples of the test stimuli used in Experiment 1. The test items are skewed versions (skewed 20 ° clockwise or counterclockwise) of the training one-away items, shown inset, with the informational overlap replaced with perceptual overlap. The torsos are shown in bold for illustrative purposes only.

Figure 4: Cueing effects on overlap responses for biased and unbiased participants, Experiment 1. The cueing effect is the difference in classification responses, as a percentage of the number of items in each cueing condtion, between cueing conditions (cued to overlap – cued to correct). Error bars = 1 SE.

Figure 5: Examples of the training and test stimuli used in Experiment 2.

Figure 6: Cueing effects for PO and MO participants, within each strategy type. The cueing effect for feature-list strategy users is on the left, and for feature count strategy

users on the right, Experiment 2. The cueing effect is the percent difference in classification responses between cueing conditions (cued to overlap – cued to correct). Error bars = 1 SE.

Figure 7: Examples of the training items used in Experiment 3, organized within superordinate (zoot and soot). Shown are prototypes (left column) and one-away exemplars (right column). Informational overlap happens with members of the same superordinate, and with both members of the rival superordinate, illustrated with the prin one-away.

Figure 8: Samples of test items used in Experiment 3 (bottom, left and right), with equivalent training one-away (top right), prototypes of two of the three species it shares features with (top left).

Figure 9: Cueing effects across same-superordinate and different-superordinate overlap items, biased and unbiased participants, Experiment 3. The cueing effect is the percent difference in classification responses between cueing conditions (cued to overlap – cued to correct). Error bars = 1 SE.

Figure 10: Examples of the training and test stimuli in Experiment 5.

Figure 1

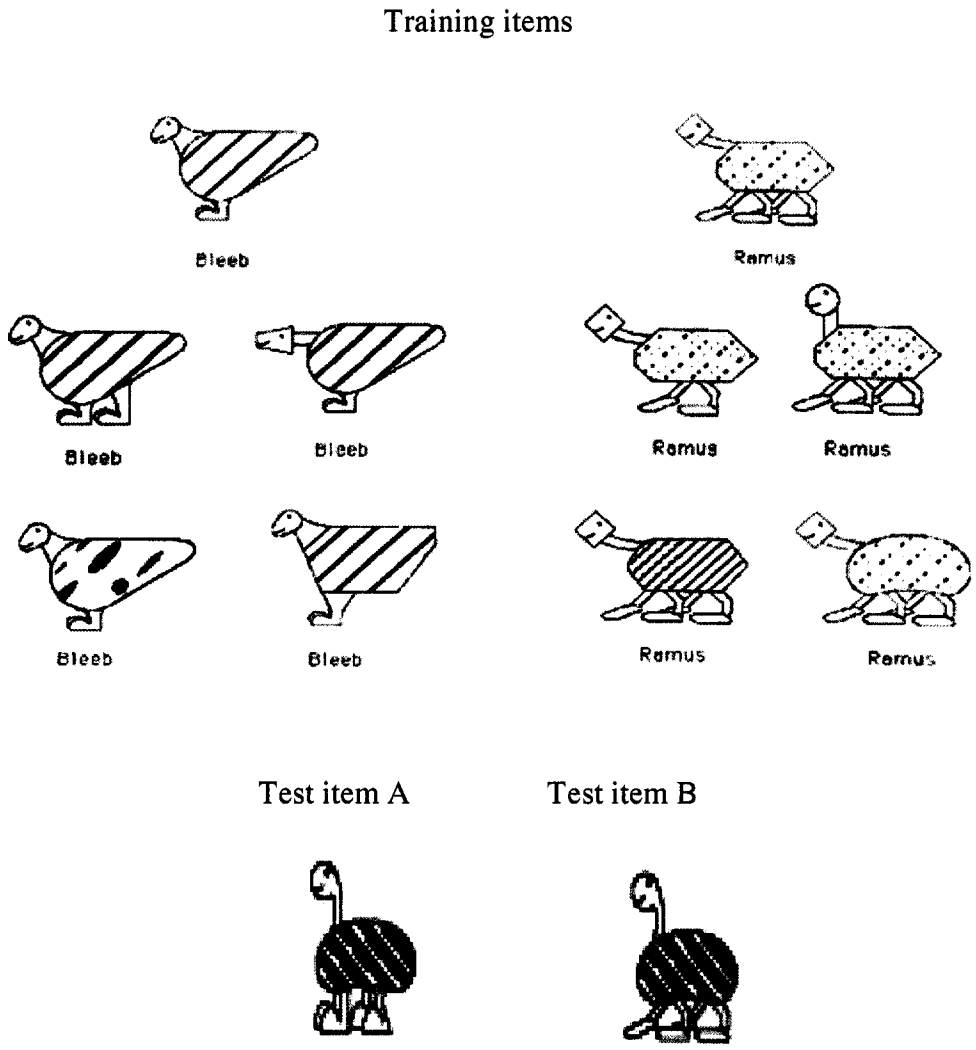

Training items
Bleeb
Ramus
Bleeb
Bleeb
Ramus
Ramus
Bleeb
Bleeb
Ramus
Ramus
Test item A
Test item B

Figure 2

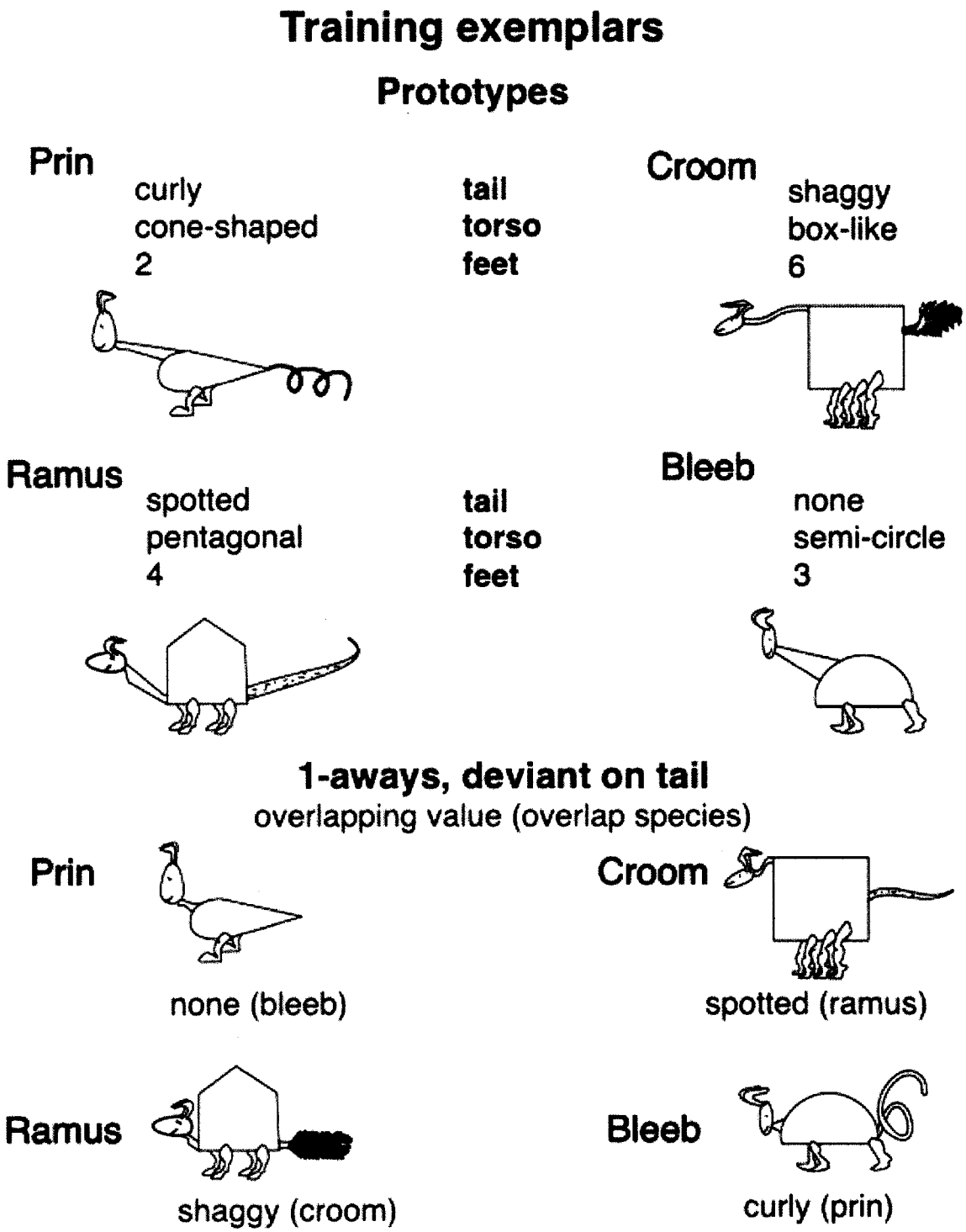

Training exemplars
Prototypes
Prin
curly
cone-shaped
2
tail
torso
feet
Croom
shaggy
box-like
6
Ramus
spotted
pentagonal
4
tail
torso
feet
Bleeb
none
semi-circle
3
1-aways, deviant on tail
overlapping value (overlap species)
Prin
none (bleeb)
Croom
spotted (ramus)
Ramus
shaggy (croom)
Bleeb
curly (prin)

Figure 3

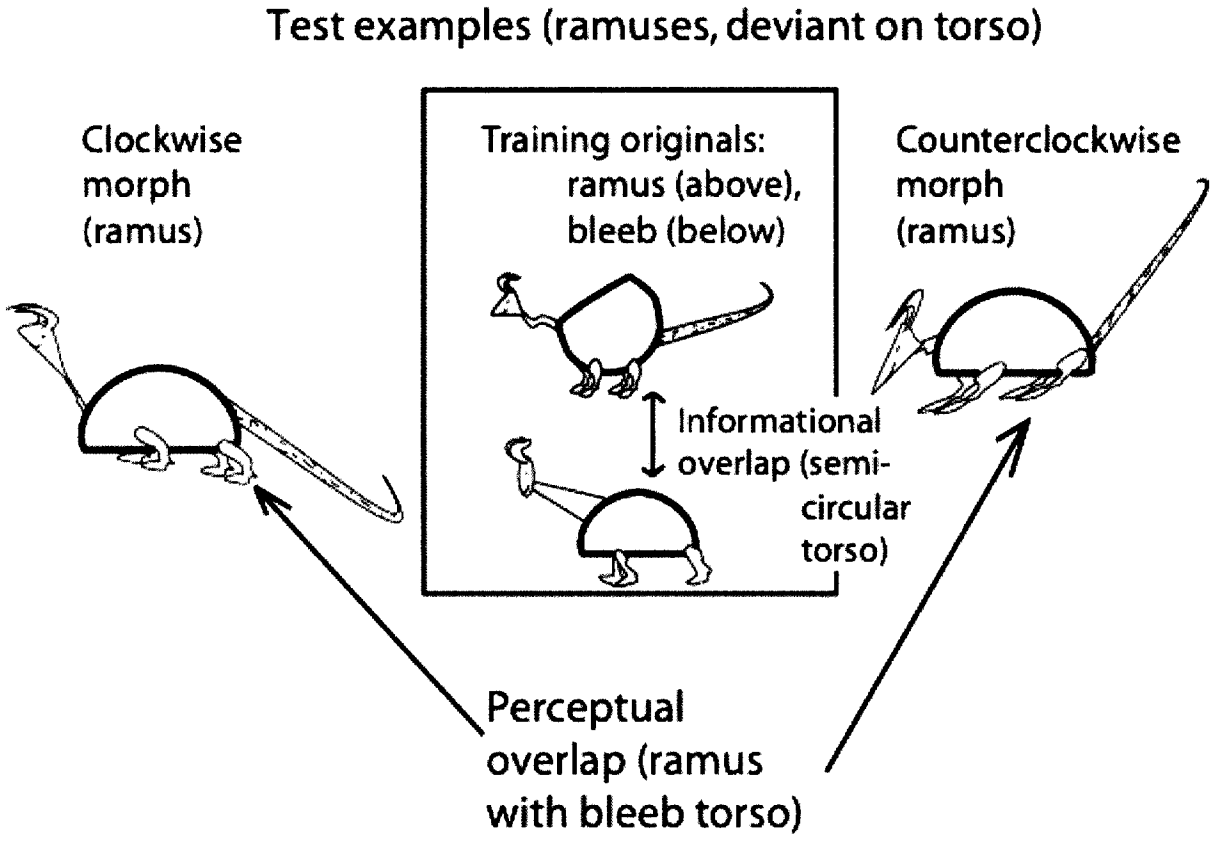

Test examples (ramuses, deviant on torso)
Clockwise morph (ramus)
Training originals: ramus (above), bleeb (below)
Counterclockwise morph (ramus)
Informational overlap (semi-circular torso)
Perceptual overlap (ramus with bleeb torso)

Figure 4

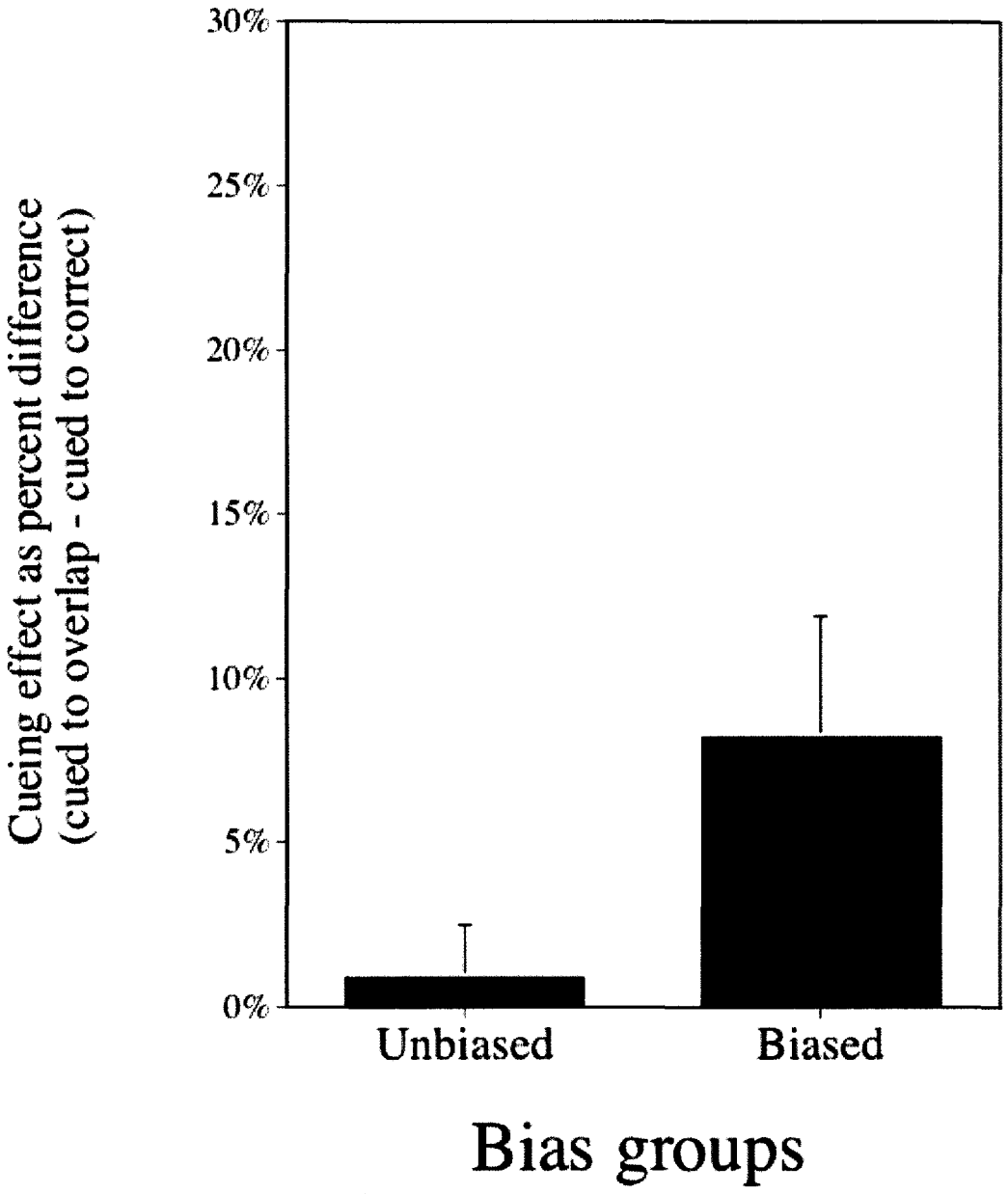

30%
25%
20%
15%
10%
5%
0%
Cueing effect as percent difference
(cued to overlap - cued to correct)
Unbiased
Biased
Bias groups

Figure 5

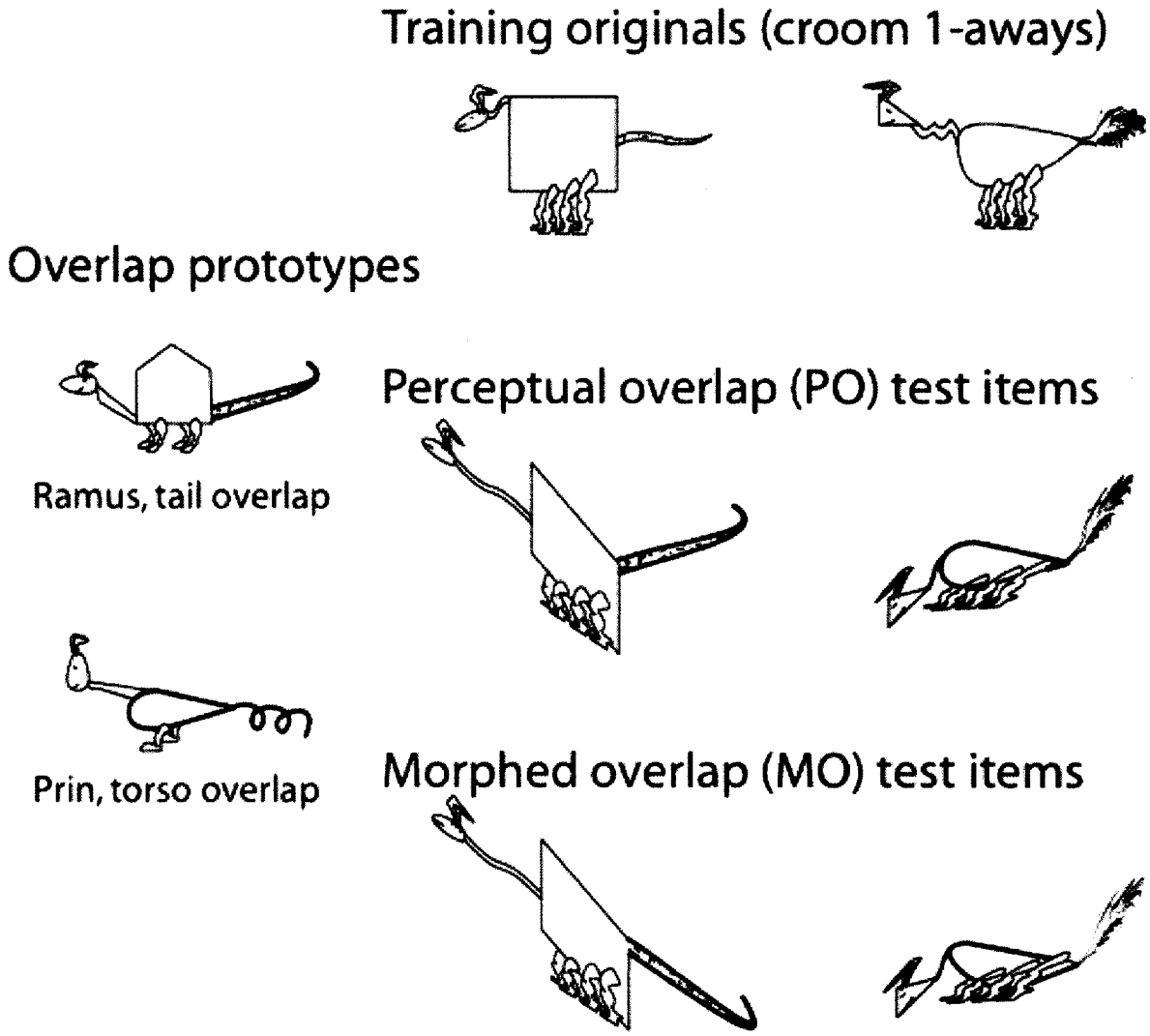

Training originals (croom 1-aways)
Overlap prototypes
Ramus, tail overlap
Prin, torso overlap
Perceptual overlap (PO) test items
Morphed overlap (MO) test items

Figure 6

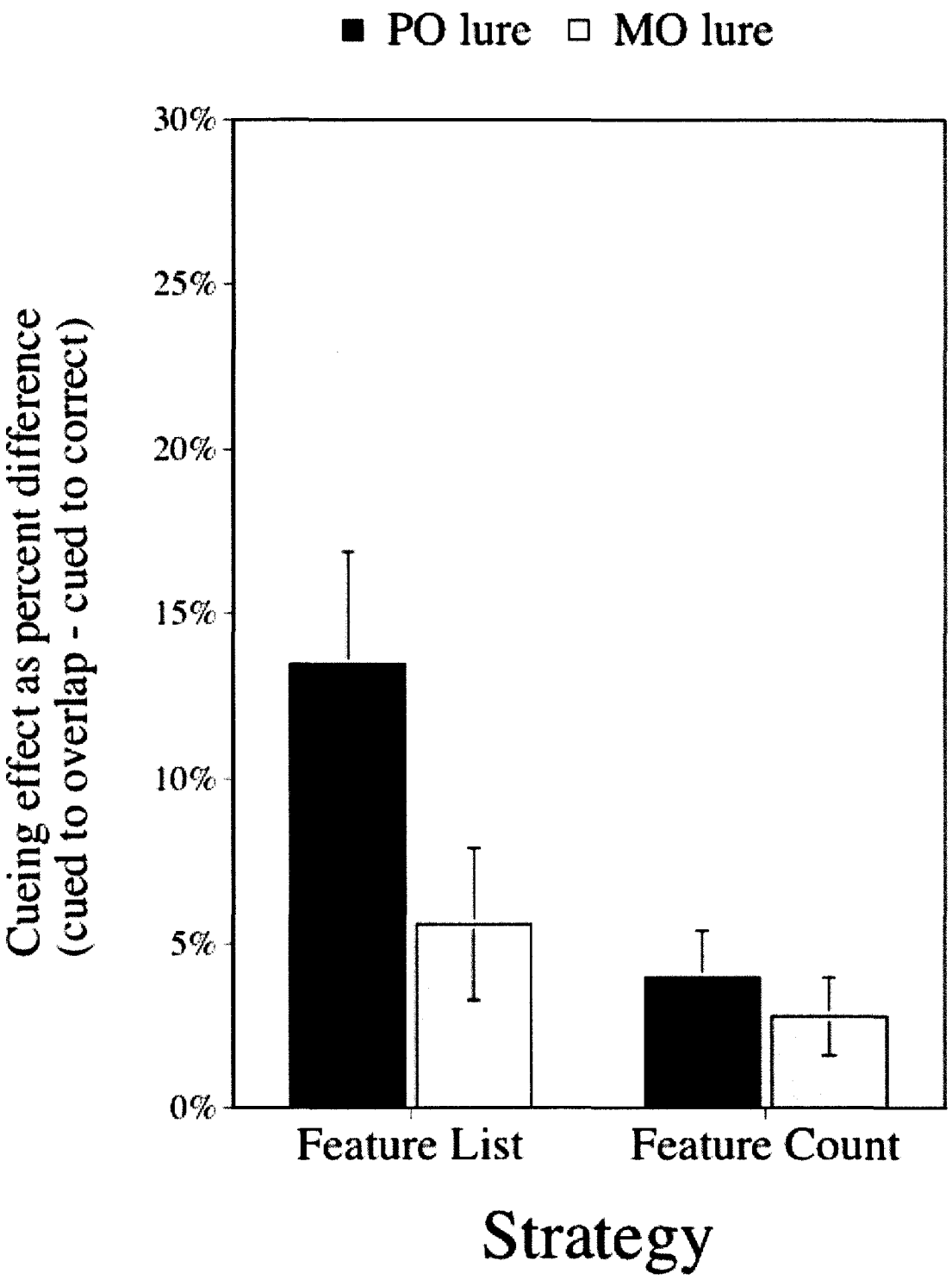

■ PO lure □ MO lure
30%
25%
20%
15%
10%
5%
0%
Cueing effect as percent difference
(cued to overlap - cued to correct)
Feature List
Feature Count
Strategy

Figure 7

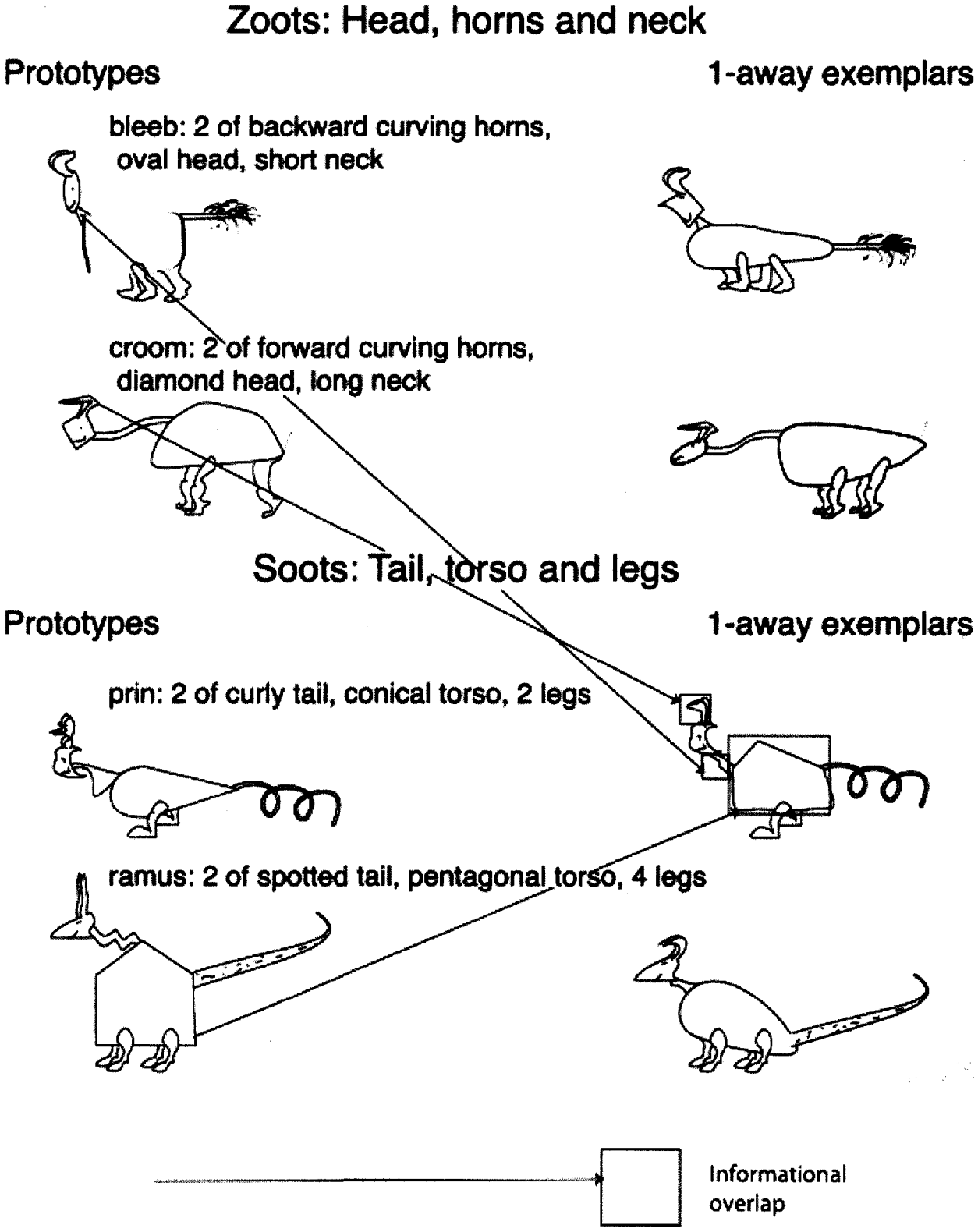

Zoots: Head, horns and neck
Prototypes
1-away exemplars
bleeb: 2 of backward curving horns, oval head, short neck
croom: 2 of forward curving horns, diamond head, long neck
Soots: Tail, torso and legs
Prototypes
1-away exemplars
prin: 2 of curly tail, conical torso, 2 legs
ramus: 2 of spotted tail, pentagonal torso, 4 legs
Informational overlap

Figure 8

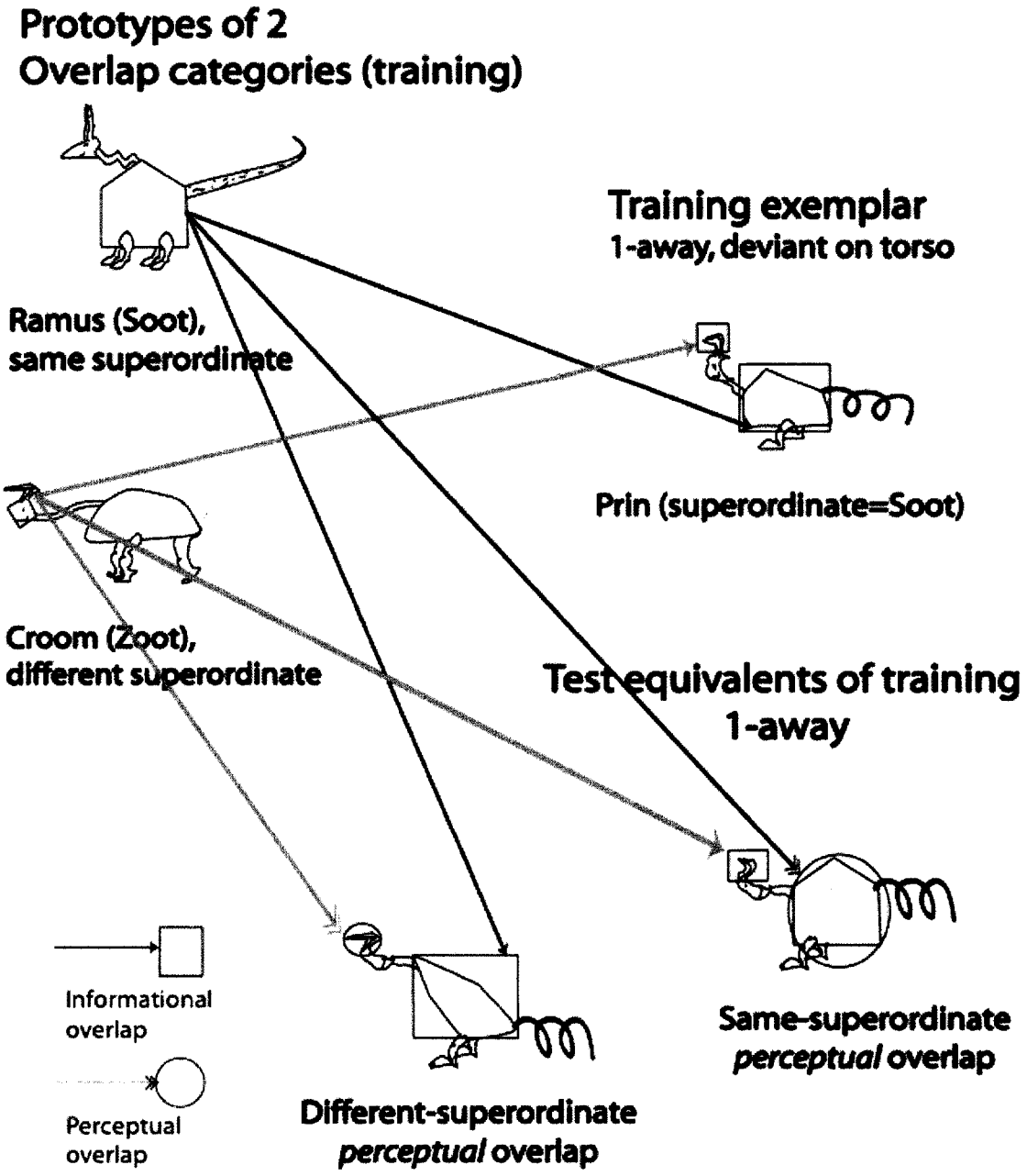

Prototypes of 2
Overlap categories (training)
Training exemplar
1-away, deviant on torso
Ramus (Soot),
same superordinate
Prin (superordinate=Soot)
Croom (Zoot),
different superordinate
Test equivalents of training
1-away
Informational
overlap
Perceptual
overlap
Same-superordinate
perceptual overlap
Different-superordinate
perceptual overlap

Figure 9

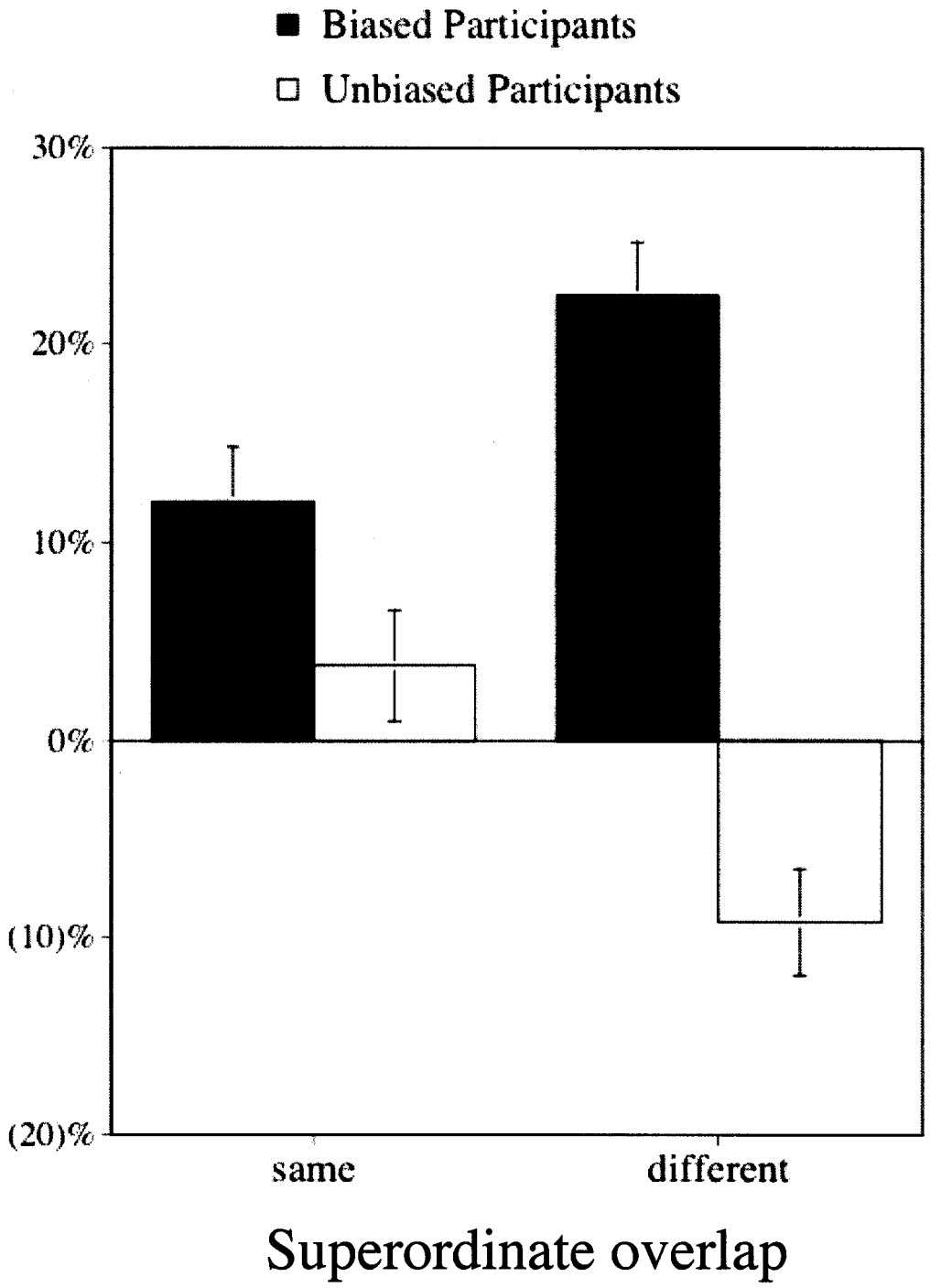
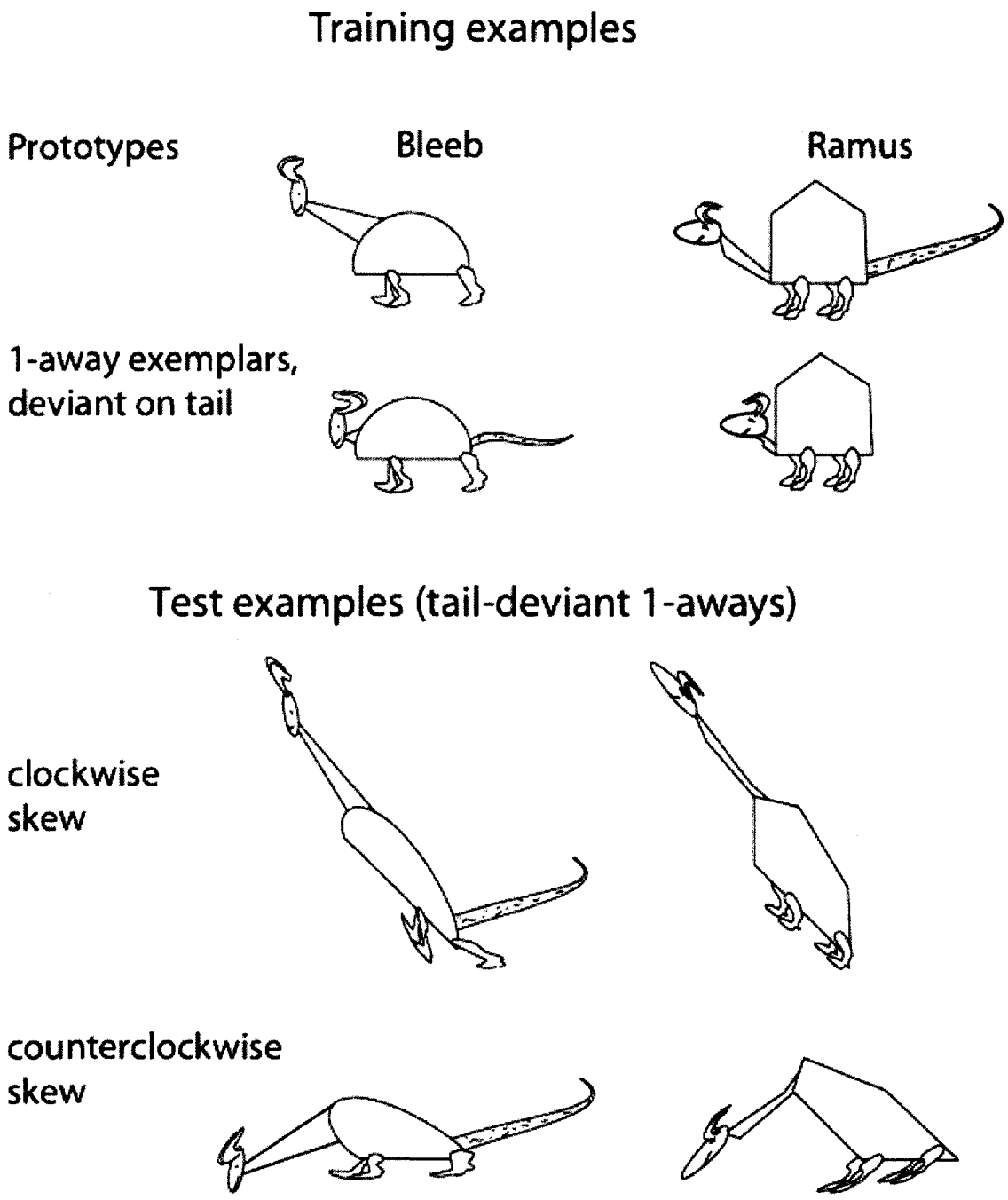

Biased Participants
Unbiased Participants
30%
20%
10%
0%
(10)%
(20)%
same
different
Superordinate overlap

Figure 10


Training examples
Prototypes
Bleeb
Ramus
1-away exemplars, deviant on tail
Test examples (tail-deviant 1-aways)
clockwise skew
counterclockwise skew

Chapter 5

Reducing the Susceptibility to Biasing

This research began with findings from an applied domain, and it seems reasonable that we should attempt to find something about biasing that might be useful for practitioners in areas such as medicine. In Experiments 1A and 1B I attempted to find ways to reduce the susceptibility to biasing. Not only does this provide potential benefits to those working in applied domains, but it also provides some theoretical insights into the robustness of the categorical biasing effect. In 1A, I attempted to reduce susceptibility by a manipulation at test; in 1B, I attempted to reduce susceptibility to biasing by a training manipulation.

The test manipulation consisted of inverting the display of test items. I hoped this would disrupt overall similarity between the test items and the training items. Many contemporary theories of categorization assume that an item currently being processed is compared to some global or holistic representation, usually a prototype or an exemplar representation. Such global or configural information could work in conjunction with information about specific features. A single feature, for example, may drive categorizationand categorical biasing effectswhen things look overall familiar; however, when an item appears strange globally, people may grow skeptical and broaden their search. That is, the weight of individual features may rely in part on similarity to some global representation.

In Experiment 1B, I attempted to "immunize" people from placing a heavy reliance on individual features by introducing perceptual overlap in training, and by modifying the training instructions slightly to point out the overlap. I expected that having been exposed to the ambiguity inherent in any single feature, participants would be disinclined at transfer to make decisions based on a single feature, resulting in a reduced biasing effect.

If the inversion manipulation works to reduce susceptibility to biasing, this would be a technique that could be used in clinical settings, at least with visual materials such as radiograms. If the

training manipulation worked, it would suggest a means of reducing susceptibility to biasing via medical training. The susceptibility to biasing might be reduced by routinely pointing out perceptual "lure" features in training cases.

5.1 Experiment 1A. Inverted Test Display: Overall Familiarity

5.1.1 Methods

Participants

Results are based on the data of 20 participants. All participants were recruited from a first-year psychology course at McMaster University, and received course credit for their participation. All participants spoke English as their first language. Data from four participants were dropped as some items were inadvertently presented in upright positions during the course of the transfer section, and these participants were replaced. Twenty-four participants took part in the experiment. Performance in the inverted display condition was tested against the first 20 participants recruited in Experiment 3, reported in the previous chapter.

Materials and Procedure

All materials were identical to those used in Experiment 3. The only change in procedures from Experiment 3 was that transfer items were displayed upside down.

Design and Analysis

Differences in the mean number of items assigned to the overlap category across Cueing and Display (upright vs. inverted) were analysed by a 2 X 2 mixed design ANOVA with Display as a between-subjects factor and Cueing as a within-subject factor.

5.2 Experiment 1B. Immunized Training: Alerting to Featural Overlap

5.2.1 Methods

Participants

Twenty participants recruited from a first-year undergraduate psychology course at McMaster University contributed data to this experiment. All spoke English as their first language, and all received course credit as compensation for their participants. Performance in the immunized training condition was tested against the first 20 participants recruited in Experiment 3, reported in chapter 4.

Materials and Procedure

All materials were largely identical to those used in Experiment 3, except that the informational overlap present among the training items was changed to a perceptual overlap. Training procedures were nearly identical to Experiment 3 except that the overlap nature of deviant features was pointed out to participants during the first round. Transfer procedures were identical to those used in Experiment 3.

Design and Analysis

The same tests of the differences in the mean percentage of items assigned to the overlap category across Cueing and Training (standard and immunized) were conducted here as in 8A. Training replaced Display as the between-subjects factor.

5.2.2 Results

The ANOVA on the number of items assigned to the Overlap category found only a main effect for Display and Cueing, $F(1,38) = 12.00$, $MSE = 4.82$, $p < .0025$, for Display, and $F(1,38) = 39.20$, $MSE = 1.18$, $p < .0001$, for Cueing. As can be seen in Table 5.1, although the effect of the suggestion was constant across both groups, inverting the display items seems to enhance the saliency of the PO features, as inverted participants were more likely to follow the PO feature than were standard participants regardless of cueing conditions. A reduction in the cueing effect should lead to an interaction between Display and Cueing factors, but this interaction was not significant, $F(1,38) = 1.52$, $MSE = 1.18$, $p > .2$. If there were any true interaction that the test failed to find, it would seem to be in the opposite direction expected, as the effect of the bias was nominally larger for the inverted display condition ($+14.2\%$ points) than for the standard display ($+9.2\%$ points).

The ANOVA on the mean number of items assigned to the Overlap category by participants in the standard training and immunized training conditions found only a main effect for Cueing, $F(1,38) = 31.93$, $MSE = 0.76$, $p < .0001$. The expected interaction between Cueing and Training was unreliable, $F(1,38) < 1$, $MSE = 0.76$, $p = 1.0$. As the data in Table 5.1 show, direct experience with perceptual overlap in training produced no resistance to the biasing suggestions, and no effect on overall performance difference either.

5.2.3 Discussion

Neither the inversion of transfer materials nor exposure to perceptual overlap during training mitigated the effects of a biasing suggestion. The reliance on specific features when categorizing physical objects is therefore highly robust, suggesting that the processes involved are operating at a very basic level.

The inversion manipulation was chosen because I assumed that the processing holistic or global or configural information would contribute to estimates of fluency. Many contemporary accounts

Table 5.1: Percentage of items assigned to the correct, overlap and other response categories across cueing conditions for participants in the standard training and transfer condition, the inverted display condition (Experiment 1A) and the immunized training condition (Experiment 1B). N = 20 for all groups. (Standard deviations in parentheses.)

		Classification response		
Group	Cueing	Correct	Overlap	Other
Standard (Ex.3, chapter 4)	overlap	80.0% (21.1)	16.7% (17.1)	3.3% (6.8)
	correct	91.3% (12.8)	7.5% (9.7)	1.3% (4.1)
	Cueing effect	− 11.3% (15.8)	9.2% (11.8)	2.1% (6.6)
Inverted display (Ex.1A)	overlap	61.7% (16.3)	33.3% (16.2)	5.0% (4.2)
	correct	78.8% (15.2)	19.2% (13.5)	2.1% (3.7)
	Cueing effect	− 17.1% (13.9)	14.2% (13.8)	2.9% (4.9)
Immunized training (Ex. 1B)	overlap	81.3% (17.9)	14.2% (14.3)	4.6% (7.4)
	correct	92.1% (13.4)	5.0% (8.3)	2.9% (5.6)
	Cueing effect	− 10.8% (9.0)	9.2% (8.5)	1.7% (6.4)

treat concept and exemplar representations as being global or holistic representations (e.g., Medin & Schaffer, 1978; Regehr & Brooks, 1993). In the animal learning literature, arguments over featural (or elemental) versus configural accounts of associative memory are still debated (e.g., Rescorla, 2003). The idea of configural processing is critical to most account of face processing (e.g., Diamond & Carey, 1986; Gauthier & Tarr, 1997; Gauthier et al., 2000; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Not only did inverting the display of test materials not reduce the reliance on individual features, it seemed to have enhanced their salience as indicated by the increased levels of Overlap responses across both cueing conditions. All of this suggests that in the paradigm used here there is no or little reliance on holistic or global information. Kruschke and Johansen (1999) also found little effect of global information in a multiple-cue probability-learning paradigm using artificial materials, and Brooks and O'Brien (2003) in a categorization experiment using artificial materials found only a very small effect of configural information.

This raises the possibility that the role of global or configural information has been overstated in the categorization literature, at least in early learning situations. We note that the face inversion effect (e.g. Collishaw & Hole, 2002; Tanaka & Sengco, 1997) hinges on the greater impact of inversion on face recognition than on recognition of ordinary objects, implying that there is a much less of a role of configural information for processing ordinary objects—even highly familiar items such as cars—than for face processing. This also implies that for a wide range of items, configural or holistic encoding may be quite limited. One reason may be because, as I have pointed out in the introduction, exact or near-exact feature overlap is quite rare in most physical objects, and so first-order knowledge alone (knowledge of individual features) is sufficient for identification for many objects. People may

only begin to attend to, and therefore learn about, configural or holistic information when perceptual overlap causes response conflict. Among common physical object categories, such perceptual overlap may be common only among faces.

All of this leads to a very simple principle of learning within concept tasks, that people only learn about that information that seems necessary to them to learn. If there is no reason, within the task, to learn about the relations or combinations of features, then people will not learn such information, and there will be little or no emergence of configural information. That is, people will expend as little effort in learning as they believe is necessary. This *principle of minimal effort* is nothing more than an idea that has been discussed among attention researchers for decades, namely that processing resources are limited, and people will allocate resources based on what they believe will be optimal in a task (e.g. Norman & Bobrow, 1975; Pashler, 1994).

The immunization condition (experience with perceptual overlap in training) produced almost identical results as the standard training condition. The explicit warnings and demonstrations of perceptual overlap were clearly not enough to make our participants wary about the role of individual features. In hindsight, this should have been expected. My reasoning was that having experienced ambiguity because of the overlap, participants would become more computational in their processing of features, seeking out information about feature combinations. This computational processing would be preserved, and then reinstated by the task and stimuli encountered in the transfer stage.

This prediction depended, however, on the participants experiencing response conflict when the perceptual overlap is pointed out. The overlap instruction, however, was done during the first pass of the first training block. During this block, participants did not have to make a response or generate a hypothesis about the identity of the items. There was, therefore, no response conflict, and no motivation to engage the materials in a more analytic fashion. In later stages of responding, the single PO feature is overwhelmed by many correct features that are equally familiar, and so response conflict is again minimal, and there is no need to attend to the PO feature in any detail. The failure to elicit an effect can again be put down to a principle of minimal effort. The PO feature caused no response conflict and received no special emphasis in task instructions, and so it was not relevant to responding in the training tasks; this resulted in little or no resources being allocated to learning about the distribution of features across categories, and therefore little learning about the existence of perceptual overlap.

It might be argued that such a principle of minimal effort would only apply only under conditions of heavy load on the grounds that it is unlikely that we actually know little about most of the things we encounter. However, the principle of minimal effort does not imply that we necessarily know little, only that we will only allocate as much resources as need and interest dictate. We could redescribe the principle in lay language in two ways: "People are lazy," or "People are smart." Under the first, cynical interpretation, the minimization of effort is stressed. We may know quite a bit of detail about the things we work with because our ordinary experience of them requires few resources to learn about them. Under the second, more optimistic interpretation, the strategic aspect of resource allocation is

stressed. We may know a lot about the features of an item of interest because we believe it necessary to learn about them to achieve some goal.

However, all of the above is predicated upon the notion that we do know a fair bit about the things we work with. Such intuitions are often wrong. Intuitively, we may think that if a person in a gorilla suit walked through the middle of a ball-toss game we were watching, we could not fail to notice such a large-scale and strange event. Simons and Chabris (1999) found, however, that over a third of participants viewing such a situation failed to notice the gorilla. Brooks, Squire-Graydon and Wood (2004) found that most participants were unaware of the lack of necessary and sufficient features for both common semantic categories (e.g., "table", "dog") and for artificial categories learned incidentally. Thus, the ability to use and recognize items does not entail detailed knowledge of the items.

Chapter 6

General Discussion

This thesis has been concerned with flexibility in conceptual use and variety in conceptual content. The categorical biasing effect is a demonstration that categorization is a more variable process than has been previously shown, and that such plasticity cannot easily be accounted for under established approaches. The studies in this thesis have attempted to demonstrate that concepts include a mixture of feature representation types (informational and instantiated), as well as specific procedural knowledge (e.g., attentional routines). Not only are categorization decisions susceptible to biasing, there are also a variety of decision processes available as a result of the variety of content, even without considering issues of deliberate strategy. In that sense, the extreme variability demonstrated here might be said to be inherent in concepts, that concepts are fundamentally flexible structures because of the richness of their contents.

The categorical biasing effect arose not only with the complex and ill-defined categories characteristic of medical diseases (LeBlanc et al., 2001), but even with a small set of well-defined, simple categories characterized by distinctive, nonambiguous features (Experiment 1, Chapter 4). Experiment 2, Chapter 4 demonstrated that the categorical biasing effect was influenced by the perceptual familiarity of features, and Experiment 3 of the same chapter demonstrated the effect was influenced by concept-specific procedural knowledge, in the form of attentional patterns.

Experiment 1, Chapter 3, provided evidence for a feature-goodness heuristic being used by those people giving a feature list when asked for their strategy. Experiment 2 of Chapter 4 showed that perceptual familiarity modulated the biasing effect only for these listing participants. Experiment 2 of Chapter 3 provided confirmatory evidence that the listing participants left perceptual experience out of their rule statements. This finding is consistent with the argument that these people used terms that were conceptually grounded in a different way from that of counting participants. This demonstrates that decision-making procedures are bound up with feature representations, and thus the content of a concept is inextricably connected to the later application of the content. Decision procedures are not neutral with regard to representation form, and later application of conceptual knowledge is tied to

the details by which it was acquired and encoded.

Experiment 1, Chapter 3, also demonstrated that counting participants made systematic errors involving neglect of critical features. This is likely the source of the small, but nonzero, biasing effects found in similar participants in Experiment 2, Chapter 4, and partly explains the superordinate biasing effect (Experiment 3, Chapter 4). When rival information is physically adjacent (same-superordinate condition) then the biasing effect observed is likely coming mainly from people reliant on a feature-goodness heuristic. When the rival information is physically separate (different-superordinate condition), then the suggestion can manipulate both attention as well as feature evaluation, eliciting substantive biasing from people of all representational strategies, producing a larger biasing effect. This is consistent with other research suggesting that concept also comprise other forms of instantiated, concept-specific procedural knowledge, such as motor routines (Bub & Masson, 2003; Martin, Wiggs, Ungerleider, & Haxby, 1996) and feature-parsing responses (e.g., Schyns et al., 1998).

However, at this point I wish to go further, and argue that not only do concepts support a variety of contents and processes, and thus categorization is highly variable, but that such variability is *productive*. That is, this variability is linked to pragmatic goals and tasks, and because of this conceptual richness, we can behave flexibly in applying concepts. This is not to say such flexibility arises directly from explicit, conscious strategies, but that expectations and understandings of tasks shape what features forms are relied on and what decision processes become salient. This view of categorization as a involving highly flexible conceptual structures, operating under the indirect control of pragmatic goals makes this a substantially different vision from many common accounts of categorization and concept formation, and situates this closer to the embodied cognition perspective.

6.1 Relations Between this Work and the Field: Biasing and Embodied Cognition

The data and conclusions regarding the nature of concepts and categorization presented here deviate in important ways the abstract, structuralist description of categorization and concept formation given by Shepard et al. (1961), and from standard, contemporary descriptions of categorization that have emerged under Shepard et al.'s influence. The standard approach has been succinctly described by Shanks (1991, p. 433):

> Traditionally, such theories [of categorization]... have assumed a clear division of labor between two stages in the categorization process. In the first stage, a featural description of the object is generated. The object is analysed by the sensory system, whose output is a list of the features the object possesses ... In the second stage, the input is the list of features, and the output is the category

Although Shanks' thumbnail description may leave out the kind of relational knowledge Shepard et al. (1961) argued was necessary to explain categorization, it still shares some important aspects with this and other more sophisticated descriptions. First, most descriptions of concepts and categories focus on concepts as consisting of linguistic or propositional entities, whether that takes the form of a feature list, as in Shanks' caricature, or more detailed representations binding predicates with entities, or indicating relations among entities (e.g. Yamauchi & Markman, 2000; Anderson, 1992). Even Ashby et al.'s (1998) account, which includes a perceptual representation system independent of a verbal system, treats linguistic representations as having priority over the perceptual. Second, a majority of accounts in categorization treat concepts as consisting of one type of representation (Ashby et al.'s being a notable exception).

What emerges from my research is a picture in which concepts include a variety of representation types (instantiated and informational representations). Abstract, proposition-friendly information contribute to most concepts, but concepts are also composed of highly specific perceptual representations that are difficult to treat as propositional units. Not only is a there a variety of representation types contributing to conceptual representations, but nonrepresentational knowledge in the form of specific procedural knowledge also contributes. Even the segregation of information collection and decision-making presented in Shanks' sketch seems questionable given the results reported in this thesis. Both decision procedures and domain structure seem dependent on feature description, and therefore intertwined with the information collection process.

This interdependence of processing and representation, however, is very consistent with certain theories of memory, minority views of conceptual structure, and with a newly emerging perspective often called embodied cognition (Glenberg & Kaschak, 2002; Glenberg & Robertson, 2000; Wilson, 2002). Kolers and Roediger (1984) pointed out that if information is not abstracted automatically, then the mental operations producing that information are preserved, and the independence of content and processing collapses. Part of their argument is that what mental operations do is to generate content, and because details of the generating processes are preserved, mental contents generated by different processes will always be different.

Koler and Roediger's (1984) argument is influenced by a memory theory called transfer-appropriate processing (TAP) (Morris et al., 1977). In this account, memory preserves everything that is encoded, including records of the processing conducted on items. Although all processing is preserved, the manifestation of such preservation may be more or less apparent, according to the TAP argument, depending on how a transfer task taps that processing. Morris et al., for example, showed that although semantic encoding of a word list provided superior performance on a recognition task than phonetic analysis did , making it appear as if semantic processing provides a stronger "trace", this advantage was reversed when the task was changed to a phonetically cued recognition task.

Applying this type of approach to categorization leads to a view of concepts as involving all representations and processes engaged in the course of manipulating category members. A concept as a

category representation, then, is not a compact or well-defined entity, but a highly distributed collection of both representations and procedures related to category members. This is not only consistent with my thesis, but with the descriptions of conceptual structure produced by both Brooks (1987, 1990) and Barsalou (1989, 1999).

Barsalou (1989), for example, showed that typicality judgments, taken as reflective of very stable aspects of conceptual structure are highly variable, even within individuals. The flexibility Barsalou found for typicality judgments is highly reminiscent of the flexibility I have found for categorization decisions. Barsalou's finding is important because typicality judgments are taken as subjective measures of highly central, and, therefore, supposedly highly stable, aspects of conceptual structure. The presumption that typicality judgments should be measuring stable, central characteristics presupposes that concepts are, however, highly coherent and well-structured entities. If, however, concepts are distributed collections of specific experiences related to interacting with category members, then different subsets of records could be active when estimating typicality in different situations. As a result, what items are highly central, and therefore typical, should also vary substantially across testing situations.

My work differs from the accounts presented in Brooks (1987, 1990) and Barsalou (1989, 1999) in that these latter works focus largely on issues of mental representation, whereas my work has emphasized the variety of conceptual content, including procedural knowledge. My thesis further adds to this decentralized view of category structure the conclusion that decision processes and conceptual structures are both dependent on the nature of feature descriptions.

6.1.1 Embodied Cognition

Part of the argument presented in this work is that flexibility of feature encoding is a productive component of cognition, allowing people to respond in a sensitive yet systematic manner to the constantly changing demands of the world. As tasks, goals and priorities change, we need to be able to change the resources we bring to bear in our action within the world. This emphasis on the importance of task and context—or, more broadly, situations—situates this work closely to a relatively recent movement called *embodied cognition* (see Wilson, 2002, for a recent review). Most generally, embodied approaches to cognition argue that even the most abstract thoughts are grounded in our perceptual and motor experiences, and thus are "situated" (Barsalou, 2000) in the specifics of our experience and our plans for action. Embodied cognition accounts see mental contents as grounded in specific experience of the world, and see mental processes as organized around the tasks and goals we engage in.

Lakoff's (1987) account of how linguistic concepts such as *container* may originally be grounded in nonlinguistic perceptual experience of our body was an early example of this approach. Other examples are Barsalou's perceptual symbols approach (Barsalou, 1999; Goldstone & Barsalou, 1998) to symbolic reasoning, and Glenberg's account of the grounding of symbols and semantics in knowledge of the perceptual and action affordances of objects (Glenberg & Kaschak, 2002; Glenberg & Robertson, 2000). Because of the importance of action and acting on the world to these accounts, task demands

and action affordances are often central concepts. Thus, Barsalou's (1983) account of ad hoc categories as representing ordinary categories in their early stage can be thought to be a very early treatment of categories from an embodied perspective. Whittlesea's (1997) SCAPE model also captures some of the flavour of the embodied cognition approach, with its emphasis on the task and context as shaping the encoding of a stimulus, and its depiction of memory as construction rather than retrieval.

As living organisms, we have a variety of goals to fulfill and a constantly shifting landscape to cope with. To act successfully on the world requires a great deal of flexibility. Flexible behaviour implies flexible cognition, as well, and demonstrating that concepts are highly flexible structures and categorization a surprisingly flexible behaviour has lain at the heart of this thesis. In taking an embodied approach to the learning and application of concepts, we might ask what type of conceptual knowledge would best help fulfill goals, which should immediately elicit the response, "Which goals?" Goals are context-specific (situated), and vary substantially across situations, as do optimal strategies for achieving the same goal.

In contrast, the view of cognition promoted by Shepard et al. (1961) implies a kind of mechanistic rigidity, even though it reflects a reaction against the even greater mechanistic views of Behaviorist theorists that dominated North American psychology's treatment of concepts in previous decades. For many theorists since the time of Shepard et al., concept learning is the acquisition of abstract knowledge, and only abstract knowledge, by selective attention to dimension under the direction of verbally mediated hypothesis testing. People are treated like computers, sifting through heaps of data to pull out a complete pattern. They have but one goal: to learn as much as possible about the underlying structure of the domain, and this leads largely to one type of knowledge: sparse propositional descriptions manifested as verbal rules or verbal descriptions.

6.2 Three Paths to Embodied Concepts

The most general theme of this thesis—that categorization possesses the same kind of variability as does similarity judgments—is thus compatible with embodied cognition's perspective. So too, however, are three more specific themes that have emerged in the course of this thesis. Therefore, to further illustrate the linkages between this work and embodied cognition, and to emphasize ideas that this thesis can contribute to that perspective, I want to expand on these three themes: minimal effort, encoding-dependent domain structure and concept-specific procedural knowledge.

6.2.1 Minimal Effort

The experiments of Chapter 5 failed to reduce susceptibility to biasing. This was attributed to a principle of minimal effort, by which I meant that people would allocate limited resources in learning

and applying concepts only to the extent such learning seemed necessary to perform the task.[1]

Inverting items produced no diminishment of biasing, presumably because holistic representations were not used because the demands of training and transfer could be performed with simpler feature representations. Similarly, when perceptual overlap was introduced in training, this did not seem to change encoding. I argued that this was because the overlap caused no difficulties in identifying the items: during the initial discovery of the overlap, the items were labeled and identified by the experimenter. In later blocks, the redundancy among the rule-consistent features overwhelmed any conflict produced by lure feature. Because identification could still proceed easily while ignoring the overlap and its significance, it did not change how or what people learned about the concepts.

What is learned about a concept and its members is limited to what a person interacting with members of a category believes they need to know about it, and no more. This is a reasonable strategy to take, given that cognitive resources are limited, and must be allocated optimally to ensure the best results from our actions. This idea that cognitive resources are limited, and that their allocation is a critical part of mental processing is common in attention research (e.g. Norman & Bobrow, 1975). In studies of concepts and categories, however, there seems to be an assumption that learning concepts means learning the highest-level of organization, the most abstract description possible (Shepard et al., 1961; Alfonso-Reese et al., 2002). That is, that people learn in such a way as to maximize their knowledge, rather than minimize their effort.

Applying this principle of minimal effort to concept learning, this leads us to the conclusion that people will learn the simplest representation that they can, given what their goals demand of them. The goal of concept learning is to acquire optimal representations, not maximal representations. We direct our learning to getting the biggest bang for our cognitive buck, not getting the biggest bang. Because what is optimal depends on what we believe our goals require, once again the task and the goals it embodies is exerting control over cognitive processing.

In some sense, the notion that concept learning varies in how elaborate it is according to how we believe we can best allocate our resources has some connections with recent work done by Goldsmith, Koriat, and Weinberg-Eliezer (2002) on retrieval grain. They have argued that the "grain", or degree of detail, a person reports when recalling reflects a strategic compromise between accuracy and the desire to maximize the informativeness of a verbal report. The principle of minimal effort points to an equivalent compromise in the degree of encoding scope, although this time the compromise is between informativeness and resource minimization. Such trade-offs in learning of features and properties have, to the best of my knowledge, not been discussed in works arising from the embodied cognition perspective, nor have they examined issues of resource limits as shaping strategies in learning or responding.

[1] An exception to this is information picked up in the course of learning about some property, attribute, etc. that is thought necessary.

6.2.2 Domain Structure and Feature Encoding

Shepard et al. (1961) argued that concept formation in some domain of knowledge was primarily a matter of the abstraction of the relational knowledge regarding the distribution of features or the variability along common dimensions defining that domain. The picture that emerges from Shepard et al. is one in which people try to learn as much as possible about the high-level relations among members of a domain. However, Shepard et al. did not permit their participants to learn about features because there was no systematic feature variance about which to learn. Features either overlapped exactly across categories, or not at all.

My work follows the direction of Brooks and Hannah (2000, 2004), and involves stimuli in which feature variance is systematically related. The results suggest that finding optimal feature encodings is *more* important than is abstracting structural relations. This is because domain structure is dependent on feature description, and this in turn depends on how we expect to use our knowledge.

By definition, general feature representations, or informational features, overlap more often and across more categories in a domain than do specific feature representations. This means that the covariance structure (pattern of feature overlap) describing a domain is more complex when described in terms of informational features. However, we can find instantiated features that are uniquely associated with categorical identity, and are thus sufficient for identification. If we can find a small set of such sufficient features, then we can change our description of the domain, describing each category in terms of a set of independent features or dimensions. To use the terminology of Shepard et al., by changing from informational features to instantiated features we change the domain structure from a Type II or higher structure to a Type I structure.

Because of the flexibility in feature description, there is not a single covariance structure defining a domain, but a family of such structures that vary in complexity depending on how the features composing them are defined. The nature of a particular structural description of a domain is a reflection of feature overlap, but feature overlap depends on how features are described. Categorical structure is dependent on feature encoding.

Thus much of learning a concept, at least for physical object categories, entails learning how best to encode features. However, what counts as "best" depends on what we expect to do with the knowledge. A small set of sufficient (strongly associated) instantiated features is very useful for identification, and identification is a necessary prerequisite for acting on some object. However, if we expect that we are going to have to engage in some kind of transfer task—to recognize many disparate forms, to do some inferential reasoning, or extract some deeper set of relations or principles from the material—then we need the very pattern of overlap we circumvent in learning instantiated forms. This pattern of overlap is the higher-order relations that not only support but define abstract knowledge. To get this pattern of overlap, features have to be described at an informational level. Structure is controlled by feature description, and feature description is driven by how we plan to use our knowledge.

6.2.3 Concept-specific Knowledge and Upward Inheritance

This research has demonstrated that categorical biasing effects can operate through at least two distinct routes: via the manipulation of feature evaluation, and via the manipulation of attention. This fits well with the evidence already mentioned regarding strategies and types of errors from Experiment 1. There we saw different error patterns for people with heuristic and algorithmic decision procedures, with the former exhibiting a pattern of dominated by evaluative errors and the latter a pattern of dominated by attentional errors.

This evidence for an influence of lure features and biasing suggestions upon attention, I argue, points to a reliance on instantiated knowledge in the form of concept-specific procedural knowledge. In Experiment 3 and Experiment 4, Chapter 4, participants' abilities to notice information was reduced if they had first considered the possibility it was a member of the other superordinate class, and therefore had its diagnostic features located in that half of the stimulus opposite to their actual location. When expecting a zoot (ramus or prin), for example, participants were more likely to confine their inspection to the lower half of the torso, missing rival information in the upper half. They were less likely, however, to use a genus-specific distribution of attention when not expecting any particular category. That implies that the expectation of a ramus activated the pattern of attention used when learning about ramuses in training (attentional routine), biasing attention when for the transfer item.

Concepts consist not only of instantiated features and specific whole items, but also of the specific mental operations engaged in when interacting with exemplars (Jacoby et al., 1989; Kolers & Roediger, 1984; Whittlesea, 1997). Concept-specific attentional routines would be one example of such concept-specific procedural knowledge. Another example comes from the work of Schyns and colleagues (Schyns & Murphy, 1994; Schyns & Rodet, 1997; Schyns et al., 1998). They have shown that the parsing of items into features is a highly flexible process that is based on the parsing of similar previously encountered items. That is, the parsing routines engaged in on one item are preserved and available to guide the parsing of items encountered later. Some evidence exists that object-specific motor routines are also part of category representations. Martin et al. (1996) found, for example, that viewing tools, but not animals, activated areas involved in motor programming that were also activated when imagining actions (Decety et al., 1994). Bub and Masson (2003) found that a motoric Stroop effect could be elicited by pairing colours with gestures, and then presenting people with coloured photographs of several common objects. When asked to perform the gesture associated with the colour of the objects, people were quicker and more accurate when the gesture associated with the colour and the functional gesture associated with the object were congruent than if they were incongruent. Bub et al. also found some evidence was found for the activation of motor information by naming, but this seemed to arise only with regard to gestures related to handling of items, and eliciting shape and identity information.

Upward Inheritance

One question that springs to mind is how such instantiated procedural knowledge may emerge at the class level. Highly relevant in this regard are the findings of Jacoby, Lindsay, and Hessels (2003). They found that the degree of cognitive control could be set on an item-wise basis. Items frequently congruent (colour words displayed in the same colour they name) produced large Stroop effects, while items that were frequently incongruent (colour words displayed in a different colour than the one they name) produced much smaller Stroop effects, even though over a block as a whole the items were just as often congruent as incongruent. Such results are easily understood if we assume that knowledge of proportion congruent, and the optimal degree of cognitive control necessary to maintain good performance, exists at an item-specific level. Encountering the same item within the same task and setting elicits the responses made to it in previous encounters. If little attentional control has been exerted over the item in early prior encounters, and no error has arisen, then minimal attention will be deployed in later encounters, leading to large Stroop effects when the item is re-encountered. If successive encounters has taught the person that the item is a difficult one because the colour word and its display colour are usually incongruent, then the display of the word will elicit a high degree of attentional control, reducing the size of the Stroop effect. In some situations, therefore, strategies for attentional control are under automatic control by the item interacting with episodic memory for past cognitive responses.

Transforming such item-specific procedural knowledge into concept-level knowledge is relatively straightforward. If a number of similar items share a cognitive or motor response, then an encounter with a new item that is interpreted as similar is likely to recruiting these instances, resulting in the shared responding being reconstructed . That is, a concept would inherit the instantiated procedural knowledge embedded within the processing of individual instances. More formally, a model like that Hintzman (1986) proposed for the generation of abstract representations could well capture such upward inheritance. Hintzman proposed that transient schemas or prototypes could be generated on-line by the retrieval of multiple specific exemplars—or, a "chorus of instances"—which would be pooled at retrieval, forming an abstract representation that would persist only as long as the component instances are active. Upward inheritance is, therefore, much like representational abstraction via the pooling of multiple instances at retrieval, but extended to procedural knowledge.

6.3 Current Events and Next Steps

6.3.1 Modeling Work

Currently, several studies are under way exploring the balance of competition between informational and instantiated features. This is partly motivated by recent findings from a simple two-layer heteroassociative neural network model developed in collaboration with Damian Jankowicz and Lee Brooks.

The model is largely the same as that used to model categorization by Gluck and Bower (1988) and Shanks (1991), and essentially implements associative learning as described by the Rescorla and Wagner (1972) model (Abdi, Valentin, & Edelman, 1999).

Critically, there is an informational and instantiated feature in the model for any single "objective" feature. In training, the model's informational representations are activated by some members of both categories because of the informational overlap characterizing the training stimuli. Its instantiated representations, however, are activated only by the members of one category. There is thus disparity in the degree of association between category label and feature representation types. The model suggests that this competition between instantiated and informational descriptors at learning results in a cue-interaction-driven diminishment of learning about the informational features. This produces a strong association between the instantiated features and label responses at the expense of the associative connection between informational features and label responses. It is because of this cue interaction that performance on all-novel test items is around 80%, even though the three correct informational features should jointly outweigh the single lure feature on almost every item, leading to near perfect performance.

The model correctly predicted that adding additional features would produce no increase in improvement on all novel test items, despite increasing the cue validity of informational features, but would weaken the effect of perceptual overlap at test. The first result demonstrates that it is not merely the cue validity of feature types that affects responding. Instead, responding is driven by the relative disparity in feature-category associations among representational forms. No matter how many features are added, instantiated features are always more strongly associated with category identity than are informational features. The pool of instantiated features, therefore, always has an advantage over the pool of informational features, resulting in a constant blocking of learning of informational features, and a stable level of performance with additional features. However, an increase in the number of features should produce an increase in competition among features, leading to weakened associations between any category label and any feature. Increasing the number of features, therefore, weakens the association between any instantiated feature and any category label, weakening the interfering effect of an instantiated lure feature. Recent tests of these predictions have born them out. Use a six-featured, two-category training set, we found that overall accuracy on all-novel test items held at about 74%, while performance on perceptual overlap test items yielded an accuracy level of about 66%.

Interestingly, the model fails when there is a single training feature that is perfectly correlated with category identity. If, for example, all training bleebs have two legs and all training ramuses have four legs, the model predicts that the perfect correlation between both instantiated and information representations simply cancel each other out. When confronted with test items containing a perceptually old lure feature with legs still correlated with identity, although now perceptually novel, a normal sized perceptual overlap effect is predicted. In humans, however, there seems to be little difficulty in spotting the two-legs-versus-four-legs distinction, and when confronted with items with rule-consistent

novel legs and an old lure feature, the familiar feature is largely ignored, and they achieve an accuracy rate of about 90%. The majority of such participants produce a two-legs-versus-four-legs rule when asked at the end of training. Thus humans, unlike our model, behave as if they were actively looking for features that were simple and maximized both cue validity and scope (range of items applied to). In some sense, as Shepard et al. (1961) talked about, people are actively testing hypotheses, but about optimal features, not structural relations.

6.3.2 Feature Goodness, Discrepancy Attribution and Strong Rules

In developing the idea of a feature-goodness heuristic, I have suggested it is based in a mechanism like that described by Whittlesea's (1997, 2001a, 2001b) discrepancy-attribution hypothesis. More direct support that the feature-goodness heuristic is linked to discrepancy attribution would be desirable. If it is discrepant processing—i.e., unexpectedly fluent processing—that underlies the feature-goodness heuristic, then a simple experiment can provide such direct support using the materials used here. After induction-based training (to maximize the percentage of participants using a feature-goodness heuristic) participants can be asked to classify new items, giving their classification in the form of a confidence rating (e.g., 5 = definitely a bleeb, 1= definitely a ramus). If the items are entirely perceptually novel or have one rule-consistent feature that is taken from training (facilitative familiarity, rather than interfering familiarity), then rule-consistent classifications and confidence ratings should be higher for items with the perceptually old rule-consistent feature than for the all-novel items. Importantly, these differences should be even greater for items are shown with one rule-consistent feature initially missing, with the missing feature added a short time (about 500 milliseconds) later. Temporally separating the feature from the rest should increase the salience of any expectation; when the delayed feature is perceptually older than surrounding features, its fluency should be more surprising because of the stronger expectation, and so confidence that the feature is pointing to the correct category should be even greater than when all features are presented simultaneously [2]. Similarly, the biasing effect, perseverance responses and perceptual overlap effect should be heightened by presenting lure features with a delay. Additionally, breaking up all the familiar spatial relations by presenting items as scrambled features should lower the coherency of processing for the display and should reduce the relative rate of persistence errors and biasing effects among Listing participants by reducing the coherency of fluency associated with the item and its features.

Along with Lee Brooks and Judy Shedden, I am exploring whether the feature-goodness heuristic is replaced entirely by the generation of an algorithmic decision approach, such as a counting rule, or whether it is simply overridden by such a rule. If we accept that there is a profound distinction between perceptual and verbal processing, such as assumed by Ashby et al.'s (1998), then we might expect to see the strong verbal rules displace reliance on a perceptually mediated heuristic. However, if there are not such qualitative differences, then we might expect that strong verbal rules such as the counting rules

[2]Thanks to Bruce Whittlesea for this suggestion

used by many of our participants would only result in them overriding a feature-goodness heuristic. We have nearly completed an initial behavioural study in which induction training was provided to teach participants two categories of four-featured animals. The members of one group were given a two-out-of-four counting rule prior to test, and the members of the other group were left to their own devices (such rules are rarely discovered with induction training). Despite the use of the rule, a slight effect of feature familiarity seems to emerge for the rule group on accuracy, and an even larger effect emerges in measures of reaction time. Next we plan to examine ERP measures to see whether the effect on the counting rules users points to a requirement to translate perceptual experiences into verbal ones, which is quicker for familiar as compared to unfamiliar features, or whether the application of the rule itself is influenced when applied to perceptually familiar elements as compared to perceptually novel elements.

6.3.3 Attentional Routines and Conceptual Structure

The argument for the existence of attentional routines also could use support that is more direct. Rehder and Hoffman (submitted) have shown that learners of Shepard et al. (1961) Type I categories (single relevant dimension) undergo systematic shifts of eye-movements during training, distributing gaze relatively equally across all dimensions early in learning, and then suddenly changing to fixate only on the relevant dimension. This implies that there are distinct patterns of eye movements that change as conceptual knowledge changes, supporting the contention that patterns of eye movements, and presumably attentional patterns, are a procedural part of conceptual content. Eye movements could be studied using more complexly structured categories, to see whether exemplars in training were inspected using a common inspection pattern (attentional routine), and whether this transferred to perceptually novel test exemplars. Using something like the two-superordinate materials of Experiments 5 and 6, which had relevant features in different regions, would allow us to determine whether such routines could be elicited by biasing suggestions, suggesting that the production of an attentional pattern was directly tied to the category label, in a similar fashion to the linking of feature representations and category label.

In section 6.2 above, I described briefly the distributed view of conceptual structure laid out by Brooks (1987, 1990) and Barsalou (1989, 1999). As noted in that section, this thesis shares many of the assumptions about conceptual structure found in those accounts, and exploring this view of conceptual structure as a set of associations among highly distributed experiences and responses is of interest. Judy Shedden and I have taken some preliminary steps in this direction.

The research follows the logic that if concepts are distributed records of experiences and responding, then the distinction between conceptual and perceptual processing is largely arbitrary. If concepts are distributed collections of experiences, and include all forms of information linked to category members, then concepts occupying different levels of a taxonomic hierarchy should embody different experiences. For physical object categories, these differences would include the different perceptual experiences in

processing items at different levels of taxonomic level. For example, a basic-level concept such as *dog* may involve all the perceptual information involved in distinguishing dogs from nondogs, such as outline, or large-scale features such the head, coat texture and so on. The subordinate-level concept *collie*, however, would include the percepts generated in distinguishing collies from other noncollie dogs. This latter information would presumably contain much more fine-grained perceptual information. Thus, activating the concept *dog* may involve recruiting prior perceptual experiences that are more dominated by global information than is the case for activating the concept *collie*, which would result in the recruiting of more instances of processing local features. If this is true, we might find that the efficiency of global or local processing as measured by reaction time and accuracy, and by the latency and amplitude of ERP components would vary depending on the level of abstraction on an interleaved concept verification task.

6.4 Conclusion

This thesis has shown that categorization decisions display wide variability just as do judgments of similarity and typicality, and that this variability is systematically related to the variability in the contents of concepts, and to the instantiated nature of some contents in particular. Concepts must include as contents both propositional and procedural knowledge, and representations of features must include both instantiated and informational representation forms. Such variety in conceptual content not only accounts for the variability in error patterns or biasing patterns, but also means that much of conceptual learning involves learning to form and deploy the optimal representations, rather than simply the extraction of high-level statistical relations mapping features to categories in some domain.

References

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural networks.* Thousand Oaks, Cal.: Sage.

Ahn, W.-K. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition, 69*, 135-178.

Alfonso-Reese, L. A., Ashby, F., & Brainard, D. H. (2002). What makes a categorization task difficult? *Perception and Psychophysics, 64*, 570-583.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409-429.

Anderson, J. R. (1992). Automaticity and the act* theory. *American Journal of Psychology, 105*, 165-180.

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 259-277.

Anderson, J. R., & Ross, B. H. (1980). Evidence against a semantic-episodic distinction. *Categorization and sensitivity to correlation, 6*(441-466).

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsycholgical theory of multiple systems in category learning. *Psychological Review, 105*, 442-481.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*(3), 211-227.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (p. 76-120). Cambridge, Mass.: Cambridge University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Barsalou, L. W. (2000). Being there conceptually: Simulating categories in preparation for situated action. In N. Stein, P. Bauer, & M. Rabinowitz (Eds.), *Representation, memory, and development: Essays in honor of jean mandler.* Mahwah, NJ: Lawrence Erlbaum.

Brooks, L. (1990). Information, language, and cognition. In P. P. Hanson (Ed.), (Vol. 1, p. 141-160). Vancouver: UBC Press.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd, Barbara (Eds.), *Cognition and categorization* (p. 169-211). Hillsdale, NJ: Lawrence Erlbaum.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorizations* (Vol. 1, p. 141-174). New York: Cambridge University Press.

Brooks, L. R., & Hannah, S. D. (2000). Relation between perceptual and informational learning of family resemblance structures. In . Paper delivered at 41st Annual Meeting of the Psychonomics Society, New Orleans.

Brooks, L. R., & Hannah, S. D. (2004). Feature lists reflect the learning of instantiated features: The case for two levels of feature representation. *Manuscript submitted for publication.*

Brooks, L. R., LeBlanc, V. R., & Norman, R., Geoffrey. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychological Science, 11*, 112-117.

Brooks, L. R., & O'Brien, V. (2003). Unpublished raw data.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* New York: John Wiley and Sons.

Bub, D. N., & Masson, C. M., Michael E. J.and Bukach. (2003). The use of functional knowledge in object identification. *Psychological Science, 14*, 467-472.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-247.

Collishaw, S. M., & Hole, J., Graham. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception, 31*, 287-296.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., R., W., Mazziotta, J. C., & Fazio, F. (1994). Mapping motor representations with positron emission tomography. *Nature, 371*, 600-602.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117.

Edelman, S. (1998). Representation is representation of similiarities. *Behavioral and Brain Sciences, 21*, 449-498.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107-140.

Fodor, J. A. (1985). Précis of 'the modularity of mind'. *Behavioral and Brain Sciences, 8*, 1-42.

Gauthier, I., & Tarr, M. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*, 1673-1682.

Gauthier, I., Tarr, M., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12*, 495-504.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558-565.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*, 379-401.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227-247.

Goldsmith, M., Koriat, A., & Weinberg-Eliezer. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General, 131*, 73-95.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65*, 231-262.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34*, 686-708.

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 39, p. 163-198). San Diego, CA: Academic Press.

Hintzman, D. L. (1986). "§chema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

Jacoby, L. L., Baker, J. G., & Brooks, L. R. (1989). Episodic effects on picture identification: Implications for theories of concept learning and theories of memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 275-281.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In *The psychology of learning and motivation: Advances in research and theory* (Vol. 18). New York: Academic Press Inc.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. D. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin and Review, 10*, 638-644.

Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M., & DeWitt, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General, 119*, 397-411.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136-153.

Kolers, P. A., & Roediger, H. L., III. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23*, 425-449.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45*, 812-863.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review, 7*, 636-645.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learming, Memory, and Cognition, 25*, 1083-1119.

Lakoff, G. (1987). *Fire, women, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

LeBlanc, V. R., Norman, R., Geoffrey, & Brooks, L. R. (2001). Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Academic Medicine, 76*, S18-S20.

Love, B. C., & Markman, A. B. (2003). The nonindependence of stimulus properties in human category learning. *Memory and Cognition, 31*, 790-799.

Mackintosh, N. (1997). Has the wheel turned full circle? fifty years of learning theory, 1946-1996. *The Quarterly Journal of Experimental Psychology, 50A*, 879-898.

Maddox, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization. *Perception and Psychophysics, 63*, 1183-1200.

Maddox, W. T., & Bogdanov, S. (2000). On the relation between decision rules and perceptual representation in multidimensional perceptual processes. *Perception and Psychophysics, 62*, 984-997.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649-652.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 700-708.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*, 266-300.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53-79.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220-244.

Posner, M. I., & Keele, S. W. (1968). Retention of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic streucture: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122*(1), 92-114.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*, 323-360.

Rehder, B., & Hoffman, A. (submitted). Eyetracking and selective attention in category learning. *Manuscript submitted for publication.*

Rescorla, R. A. (2003). Elemental and configural encoding of the conditioned stimulus. *The Quarterly Journal of Experimental Pscyhology, 36B*, 161-176.

Rescorla, R. A., & Wagner, A. R. (1972). In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: current research and theory* (p. 64-99). New York: Appleton-Century-Crofts.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of endel tulving* (p. 3-41). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johson, D. M., & Boyes-Bream, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*, 1-54.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 31, p. 305-349). San Diego, CA: Academic Press.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 681-696.

Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experiment Psychology: Learning, Memory and Cognition, 17*, 433443.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science, 237*, 1317-1323.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classification. *Psychological Monographs, 75*(13), Whole No. 517.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28*, 1059-1074.

Sloman, S. A., Love, B. C., & Ahn, W.-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189-228.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 1411-1430.

Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology, 43*(2), 129-169.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Society, 46A*, 225-245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition, 25*, 583-592.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381-391.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Whittlesea, B. W. (1997). Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 37, p. 211-264). San Diego, CA: Academic Press.

Whittlesea, B. W., Brooks, L. R., & Westcott, C. (1994). After the learning is over: Factors controlling the selective application of general and particular knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 259-274.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feeling of familiarity and perceptual quality. *Journal of Memory and Language, 29*, 716-732.

Whittlesea, B. W., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129*, 84-106.

Whittlesea, B. W., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. the heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 3-13.

Whittlesea, B. W., & Williams, L. D. (2001b). The discrepancy-attribution hypothesis: Ii. expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 14-33.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 6*, 625-636.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18*, 221-281.

Wittgenstein, L. (1953/1994). Philosophical investigations. In A. Kenny (Ed.), *The wittgenstein reader* (p. 35-49). Oxford: Blackwell.

Yamauchi, T., & Markman, A. B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory and Cognition, 28*, 64-78.