

**CATEGORIZATION OF PARTIALLY OCCLUDED VISUAL
STIMULI: BRIDGING THE GAP BETWEEN
COMPLETION AND CLASSIFICATION**

by

Michael J. Wood
B.A., University of Victoria, 2006

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

In the
Department
of
Psychology

© Michael J. Wood, 2009
SIMON FRASER UNIVERSITY
Summer 2009

All rights reserved. However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for *Fair Dealing*. Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

APPROVAL

Name: Michael J. Wood

Degree: Master of Arts (Department of Psychology)

Title of Thesis: Categorization of Partially Occluded Visual Stimuli:
Bridging the Gap Between Completion and
Classification

Examining Committee:

Chair: Dr. Kevin Douglas
Associate Professor

Dr. Mark Blair
Senior Supervisor
Assistant Professor

Dr. Richard Wright
Supervisor
Associate Professor

External Examiner: Dr. John Alderete
Associate Professor
Department of Linguistics
Simon Fraser University

Date Defended: August 26, 2009



SIMON FRASER UNIVERSITY
LIBRARY

Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website <www.lib.sfu.ca> at: <<http://ir.lib.sfu.ca/handle/1892/112>>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Simon Fraser University
Burnaby, BC, Canada

ABSTRACT

Many current models of categorization assume full knowledge of the properties of the stimulus to be categorized. To remedy this situation, it is first necessary to understand how humans categorize stimuli with missing information. To that end, two visual category learning experiments were conducted using an inverse base-rate effect paradigm. In the second experiment, transfer trials included stimuli in which a category-diagnostic present/absent feature was occluded. Response proportions showed that people tend to treat occluded features as being absent from the stimulus, suggesting a more general tendency to assign default values to features of unknown status at the time of categorization. This pattern of results could not be replicated by several computational models – EXIT, SUSTAIN, or EXALT, a modification of EXIT implementing additive similarity.

Keywords: categorization; visual cognition; visual completion; object recognition; occlusion; missing information

ACKNOWLEDGEMENTS

Many thanks are due to John Kruschke, Todd Gureckis, Brad Love, Koen Lamberts, and Tom Verguts for their expert consultation and provision of modelling code without which this project would not have been possible.

Of vital importance were the contributions of the many pilot subjects, lab members and others who provided much-needed grounding and advice on methodology, procedures, and stimulus creation: Aaron Ancell, Jordan Barnes, Adam Burnett, Bill Chen, Angelina Katz, Hugh MacIntosh, Kim Meier, Gordon Pang, Jennifer Schneider, Calen Walshe, Marcus Watson, Denise Weisgerber, and any others I may have regretfully neglected to mention.

I owe a particular debt of gratitude to my supervisors, Drs. Mark Blair and Richard Wright, external examiner Dr. John Alderete, and defence chair Dr. Kevin Douglas for their penetrating questions and insights throughout the process of this project.

Finally, I must thank my parents, for their continuing support, inspiration, and, when necessary, prodding to get the job done.

TABLE OF CONTENTS

Approval	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Figures	vii
List of Tables	viii
1: Introduction	1
2: Experiment 1	14
2.1 Method	14
2.1.1 Participants	14
2.1.2 Materials	15
2.1.3 Procedure	17
2.2 Results	18
2.3 Discussion	19
3: Experiment 2	22
3.1 Method	24
3.1.1 Participants	24
3.1.2 Materials	24
3.2 Results	24
3.2.1 IBRE replication	24
3.2.2 Occluded stimulus	25
3.2.3 Recognition test	26
3.3 Discussion	26
4: Computational Modeling	29
4.1 Method	33
4.2 Results	36
4.2.1 Fit of EXIT	36
4.2.2 Fit of EXALT	37
4.2.3 Fit of SUSTAIN	39
4.3 Discussion	40
5: General Discussion	43
Appendices	50
Appendix A – EXALT Model Specification	50
Appendix B: Supplemental Modelling Information	52
EXIT/EXALT	52

SUSTAIN.....	54
Reference List.....	56

LIST OF FIGURES

Figure 1: The Kanizsa illusory square, an example of modal completion (left); a square partly occluded by a circle, exemplifying amodal completion (right).....	3
Figure 2: Composite training stimuli showing all six features for Experiments 1 (left) and 2 (right). Category-diagnostic features include a feather on the head, neck ridge, claws, shark-like fin, body spikes, and tail. Wing, beak, feet, and other features are category-invariant.....	16
Figure 3: A sample transfer stimulus from Experiment 2 with the head feature occluded.....	23
Figure 4: Response proportions (consistent categories only) to the three critical transfer stimuli in Experiment 2. Error bars represent standard error.....	26
Figure 5: Architecture of the EXALT connectionist model.....	34
Figure 6: Each model's best-fitting proportion of consistent common category responses to the three critical transfer stimuli in Experiment 2, with human data for comparison.....	39

LIST OF TABLES

Table 1: Sample category structure for Experiment 1 training phase.....	17
Table 2: Categorization response proportions to transfer stimuli.	19
Table 3: Human categorization response proportions to transfer stimuli (Experiment 2).....	25
Table 4: Best fit of EXIT to human transfer data from Experiment 2	37
Table 5: Best fit of EXALT to human transfer data from Experiment 2.....	38
Table 6: Best fit of SUSTAIN to human transfer data from Experiment 2.....	40

1: INTRODUCTION

Imagine that you are driving down the street toward a busy intersection when you see a red, rounded metal object with windows, doors, two wheels, and a bumper with a winch on it poking out from behind the corner of a nearby building. You would probably recognize this as the front end of a pickup truck, waiting for a break in traffic to turn right at the intersection. The building that blocks your view of the rest of the pickup probably hides two more wheels, a truck bed, and perhaps a canopy – but how do you know this? Did you infer what the rest of the vehicle might look like from your classification of it as a pickup truck, or did you decide that a truck bed and wheels are most likely present due to the jacked-up suspension and winch on the front, and only then categorize it as a pickup based on that inference? This question concerns the nature of the interaction of two distinct psychological processes: categorization and completion. Categorization is the process by which objects are classified as members of one category or another – for instance, the object in the above example is classified as a member of the category “truck.” Completion is the perception of partially visible objects as a unified whole through extension and interpolation of available visual information, such as contours and surfaces. With a few notable exceptions, previous research in visual categorization has been limited to stimuli with all their relevant features visible; similarly, the majority of the completion literature has focused on simple stimuli, such as lines and

geometric shapes, rather than complex real-world objects. While these are both eminently sensible approaches, they appear to have led to some neglect of the connection between the two processes. Exactly how we come to a categorization decision for a partially visible complex object remains largely uninvestigated.

Completion is broadly necessary for adaptive functioning in the world, as visual input is often fragmentary and incomplete. Completion takes this imperfect sensory input and shapes it into unified object representations. Although there is some debate as to whether completion is a unitary process or is actually served by separable neural subsystems (e.g. Kellman, 2003; Albert, 2007), for the purposes of phenomenology it is generally subdivided into the component processes of modal and amodal completion. A familiar example of modal completion is the Kanizsa illusory square (see Figure 1, left) – it seems as though the contours of the “square” outside of the notched circles are actually visible, in spite of the fact that they have no physical counterpart in the image itself. Modal completion, then, is the interpolative process by which separate object contours and surfaces are connected with one another in a perceptually salient manner.

In contrast, amodal completion does not induce illusory contours; rather, it allows partially occluded objects (see Figure 1, right) to be perceived as a whole object, rather than as an object fragment (Kellman, 2003). Thus, a square partially occluded by a circle is not perceived as a square with a rounded chunk missing; rather, it is seen as a whole square with a circle in front of it (Johnson & Olshausen, 2005). The contours of the square behind the circle are not visible, but we infer that they are there regardless, and can easily and reliably determine

where they would lie. Amodal completion effects are observed with a high degree of reliability in the presence of relative depth cues indicating that one object occludes another, and has obvious adaptive value for real-world functioning. It has been observed not only in adult humans, but also in pigeons (Nagasaka, Lazareva, & Wasserman, 2007), and its presence has been demonstrated in young children after the age of five or six months (Otsuka, Kanazawa, & Yamaguchi, 2006).

Figure 1: The Kanizsa illusory square, an example of modal completion (left); a square partly occluded by a circle, exemplifying amodal completion (right).



As alluded to above, much of the previous work on completion processes has focused on low-level perceptual stimuli, such as elementary geometric shapes or simple line drawings (e.g. Albert, 2007; He & Nakayama, 1993; de Wit, Schlooz, & Hulstijn, 2007; Rauschenberger, Lu, Slotnick, & Yantis, 2006; Scheessele & Chaaban, 2008; Kellman, Gutman, & Wickens, 2001). One attempt to extend amodal completion beyond simple contour interpolation involves the

idea of global completion – completion based on broader stimulus properties, such as symmetry, rather than on the extension of local contours. While Kellman (2003) expressed skepticism regarding the amenability of amodal completion to such relatively high-level perceptual processes, Plomp and van Leeuwen (2006) found evidence to the contrary: in a primed same/different task, test pairs consistent with a global amodal-completion interpretation of ambiguous composite primes led to faster responses than did test pairs consistent with local amodal completion interpretations. However, maximizing symmetry is quite clearly not the usual approach to completion in the environment; seeing the head of a tiger poking out from behind an occluding tree does not lead us to believe that there is an identical tiger head lurking behind the tree!

Despite the extensive research on both categorization and completion processes, it remains unclear how partially visible complex objects are categorized. That question is the focus of the present study – specifically, what inferences do people make about hidden object components that are useful for categorization? To return to an earlier example, a truck bed is highly relevant in categorizing a vehicle as a pickup truck versus a minivan, but how do we factor the presence or absence of a truck bed into a categorization decision if it is unseen? Previous work on this issue has suggested that missing information is used in categorization decisions by inferring a kind of *mean value* for unknown features (e.g. Ganzach & Krantz, 1990; White & Koehler, 2004). For instance, if we wish to categorize a particular vehicle as being a truck or a minivan, the size of the engine may be relevant. If we do not know the engine size, we might

somehow accord it the mean value of all the engine sizes we've seen, and then use that inferred value as the basis for a categorization decision.

However, it is not clear how widely this finding can be applied. As a representative example, White and Koehler (2004) found evidence for mean-value inference in a probability-based disease-diagnosis paradigm in which it was impossible to obtain perfect accuracy, and no cue was universally category-diagnostic. This may have encouraged a hedging strategy, while a more deterministic system of feature-category associations could induce a stronger assumption as to the value of a missing feature. In addition, disease diagnosis uses abstract symptom descriptions as features rather than visual stimuli, and recent work by Johnson and Olshausen (2005) indicates that there is a qualitative difference between how missing information and occluded information are processed. Since occlusion can induce completion, it seems plausible that occluding a stimulus feature could lead to a stronger inference of a particular feature value than would be elicited by simply stating that its value is unknown.

At any rate, there should be some involvement of categorization or recognition processes in the completion of complex visual stimuli in a naturalistic setting, otherwise completion would likely be of limited practical use. Kellman (2003) proposed a system of Recognition by Partial Information (RPI), by which a familiar occluded object is recognized or judged to most likely be symmetrical based on whatever information is visually available, and subsequently completed based on this inference. Thus, in Kellman's view, recognition is an intermediate stage between initial presentation and completion of complex visual stimuli.

Although RPI was invoked specifically as an explanation for post-perceptual global completion effects in the absence of symmetry as a contributor to initial amodal completion, it could quite plausibly be extended to more complex visual stimuli. A partially occluded natural object could be categorized on the basis of its visible category-diagnostic features, and then completed by “filling in” the missing information based on previously stored exemplars of the chosen category. This would be a case of categorization informing the completion process, although completion in this case would be quite different from modal completion of a Kanisza square or amodal completion of a simple occluded geometric shape.

RPI, then, has some promise as a model for completion of complex stimuli. The question of how exactly to categorize the occluded stimuli, however, was not specified by Kellman (2003). Although there are many candidates for a plausible model of categorization or recognition, it is an unfortunate assumption of many such current models that all properties of a visually presented stimulus are known and available for making a categorization decision – that is, the models do not account for missing information (e.g. ALCOVE, Kruschke, 1992; EXIT, Kruschke, 2001). One model that is explicitly equipped to account for missing data is ADDCOVE (Verguts, Ameel, & Storms, 2004), an extension of Kruschke’s (1992) ALCOVE model which computes the similarity of a stimulus to stored exemplars additively rather than through a standard geometric (Euclidean or city-block) distance function. In ADDCOVE, the similarity of two exemplars is roughly equal to the weighted sum of all matching components; missing features are simply not included in this summation. In contrast, ALCOVE computes

similarity according to the distance between the two exemplars in a multidimensional space. In spite of the fact that it does not incorporate previous findings regarding the inference of mean values for unknown features, ADDCOVE accounts for human data from tasks with missing information much better than does ALCOVE, and performs at least as well as its predecessor in fitting data from a standard categorization task (Verguts et al., 2004).

A second possible route for the completion of complex occluded stimuli involves using visible features to infer the structure of unseen stimulus features, and categorizing the object on the basis of a combination of visible and inferred components – perhaps the partial stimulus is compared to previously stored exemplars, and is completed using values for the missing features taken from its nearest neighbour in a collapsed multidimensional space. Rauschenberger et al. (2006) found that amodal completion occurs fairly early in the visual stream, which makes it at least somewhat plausible that completion may take place before categorization. Indeed, there are indications that this may be the case for complex objects; for instance, Johnson and Olshausen (2005) found that images of common objects, such as chairs or violins, are more easily recognized when parts of them are occluded than when the same components are simply deleted from the image.

ERP studies have demonstrated the plausibility of completion informing categorization of complex objects. Johnson and Olshausen's Experiment 2 (2003) found differences in occipitoparietal cortical activation patterns between partially occluded and partially deleted images as early as 130ms after initial

stimulus presentation, suggesting that completion effects occur quite early in the visual stream. Animal studies appear to support this assessment; single-cell recordings of macaque V1 cells show that completion effects may occur as early as 100ms after stimulus onset (Lee, 2003), though in more naturalistic settings this would likely be delayed somewhat by the necessity of processing depth cues in the scene in order to determine appropriate targets for completion (Johnson & Olshausen, 2003). In contrast, the neural correlates of categorization or recognition processes seem to occur a good deal later: Piliastides and Sajda (2005) found that the major occipitotemporal EEG component thought to be involved with the consideration of evidence in categorizing a visual stimulus occurred between 300 and 450 milliseconds after stimulus presentation. A related component, the well-known N170 involved with face processing, did appear much earlier in the visual stream (around 170ms); however, this component was highly stimulus-specific and did not appear to be correlated with the latency of psychophysical performance on the categorization task. The apparent temporal separation of completion and categorization processes, even for complex objects, lends plausibility to the hypothesis that the former may inform the latter.

Why would it be at all advantageous to infer the presence of category-diagnostic features rather than simply applying a category label to an object on the basis of what is immediately visible? Making inferences about the unseen features of an object before categorizing it does not extract any new information from the stimulus itself, and would at first seem to be an unnecessary

intermediate step that could be easily dispensed with. Although this is undoubtedly true in many cases, there are situations in which inferences about occluded category-diagnostic features could change the categorization decision made about the object – perhaps for the better. For instance, in the inverse base-rate effect (IBRE; Medin & Edelson, 1988), the presence or absence of an imperfectly diagnostic but previously omnipresent feature in a transfer task has a considerable effect on categorization decisions. If the existence of such a feature were made ambiguous through occlusion, a completion effect that infers the presence of the feature would change the way in which the stimulus is categorized.

The IBRE is a rather perplexing effect in which people categorize an ambiguous stimulus counter to the principles of normative Bayesian reasoning. A particular combination of features that provides equal evidence for membership in a common category and a rare one leads, paradoxically, to a preponderance of rare-category classifications. A simplified example (Kruschke, 1996) involves a simulated medical diagnosis task with two diseases: one common and one rare (the usual ratio of cases of the common disease to cases of the rare disease is 3:1, though this varies – see Shanks, 1992). Subjects are asked to diagnose a patient with one of these diseases, based on the presence or absence of three symptoms. Symptom “I” (Imperfectly diagnostic) is always present in both diseases, symptom “PC” (Perfectly diagnostic of the Common disease) is always present in the common disease, and symptom “PR” (Perfectly diagnostic of the Rare disease) is always present in the rare one. Thus, the common disease is

always characterized by symptoms I+PC (say, headaches and dizziness), and the rare one is characterized by symptoms I+PR (headaches and nausea). I is thus an imperfectly diagnostic cue, as it occurs in both diseases, and would presumably be irrelevant to a categorization decision. Subjects are trained in this category structure until a learning criterion is reached, at which point they are tested on a number of novel symptom combinations, including three ambiguous combinations: I, I+PC+PR, and PC+PR. Patients with symptom I only or with all three symptoms are most often diagnosed as having the common disease, in accordance with base rates. However, when participants are asked to diagnose a patient with symptoms PC+PR only, they choose the rare disease more often than the common disease, contrary to the principles of normative Bayesian reasoning. This is the inverse base-rate effect.

It should be noted that Medin and Edelson (1988) actually used three such pairs of rare and common diseases, for a total of six. In contrast, Johansen, Fouquet, and Shanks (2007; Experiment 1) had only one such pair, while Kruschke (1996) had two. The issue of exactly how many categories to use is a complex one; in a medical diagnosis task with verbally or textually presented stimuli, one pair of diseases seems to suffice. However, having more than two categories serves to increase task difficulty, and is necessary in versions of the IBRE paradigm that employ visual stimuli. For instance, Lamberts and Kent (2007) used images of microorganisms rather than descriptions of diseases. Microorganism species C_1 and R_1 were characterized by PC_1+I_1 and PR_1+I_1 respectively, whereas species C_2 and R_2 were characterized by PC_2+I_2 and

PR_2+I_2 . Having multiple category pairs is necessary when the stimuli are presented visually, so that the imperfectly diagnostic features are distinguishable from the invariant structure of the stimulus during training, and are thus seen as potentially useful features for categorization. Including multiple category pairs does influence the diagnosticity of the imperfectly diagnostic feature to some degree; the presence of a certain I-feature will effectively narrow the field of possible categories from some arbitrary number to only two, and could potentially be considered useful for a kind of superordinate categorization. Fortunately, this does not seem to impact the reliability of the IBRE; in accordance with previous work, Lamberts and Kent found a robust difference between PC+PR and I+PC+PR.

Why does the IBRE occur? Kruschke (1996, 2001) explained the effect as the result of cue competition between symptoms and rapidly shifting attention, as exemplified by models such as RASHNL (Kruschke & Johansen, 1999). This explanation is supported by the fact that the IBRE only arises when the imperfectly diagnostic cue (I) is present. The difference in base rates causes the common disease to be learned more quickly, and to be associated with symptoms I and PC. When the rare disease is learned, attention rapidly shifts away from symptom I as it is now recognized as not being category-diagnostic; however, due to cue competition (Kruschke, 1996), the association with the common disease is still split between I and PC. Thus, the association between PR and the rare disease will be stronger than the association between PC and the common disease, and the stimulus PC+PR will be more likely to evoke a

“rare” response. In this account of the IBRE, then, base rates in themselves are not an essential component; rather, the effect arises as a result of divided attention and learning one category before another (however, base rates are taken into account when presented with the I+PC+PR stimulus, resulting in a greater proportion of common responses). Indeed, in a paradigm with equal base rates, Bohil, Maddox, and Markman (2005) observed an effect analogous to the IBRE – a preference for one category over another when presented with both perfectly-diagnostic features – when the PR-equivalent cue was made much more salient than the others. In the same study, the equivalent of I+PC+PR also yielded a greater proportion of common responses. This result, among others, appears to provide evidence against the strictly rule-based eliminative-inference explanation for the IBRE put forward by Winman and colleagues (e.g. Winman, Wennerholm, & Juslin, 2003; Winman, Wennerholm, Juslin, & Shanks, 2005).

Theoretical explanations for the IBRE aside, it is abundantly clear that an ambiguous stimulus will elicit very different responses when symptom I is present compared to when it is not. The stimulus I+PC+PR is most often categorized as the common disease, while PC+PR is most often classified as the rare disease. This abstract category structure, translated into a visual medium along the lines of Lamberts and Kent (2007), affords a unique opportunity to investigate the nature of the completion-categorization relationship. If participants are trained in the IBRE category structure and presented with a transfer stimulus in which the PR and PC cues are present but the area where the I cue usually appears is occluded, their categorization decisions will reveal the presence or absence of

completion before categorization. If complex stimuli are completed based on previously seen stored exemplars before a categorization decision is made, participants should complete the stimulus as possessing feature I, since I would have been present in all training exemplars containing PC or PR. Categorization of the completed stimulus I+PC+PR would then yield a preference for the more common category. In contrast, if the stimulus is not completed prior to the categorization decision, participants would categorize based on only the visible features PC+PR, leading to an observable inverse base-rate effect – a preference for the rare category. If participants infer a mean value for the I- feature, in accordance with the findings of White and Koehler (2004), categorization responses to an occluded stimulus should fall somewhere between the two. There are reasons to expect that this latter possibility is not the case, however – a mean value for a present/absent feature is unlikely to make much sense, and in general people make predictions based on what they feel is most likely to be the case, while completely discounting plausible alternatives (Ross & Murphy, 1996).

2: EXPERIMENT 1

Although several studies have been conducted on the IBRE, very few (e.g. Lamberts & Kent, 2007) have used visual stimuli rather than abstract disease descriptions. It was thus unclear whether the stimulus set developed for the present study would be appropriate for the task. As such, it was deemed prudent to demonstrate that this particular stimulus set could indeed produce an IBRE. If no IBRE was observed in a simple task involving the new stimuli, using the IBRE to determine the nature of the relationship between categorization and completion would be quite fruitless. Experiment 1, then, was an attempt at a simple replication of the effect with the current stimulus set. A positive result would ensure that the stimuli were appropriate for the larger task at hand.

2.1 Method

2.1.1 Participants

Participants were 40 undergraduate students from Simon Fraser University who participated in exchange for course credit in introductory psychology courses. No demographic variables were thought likely to have any impact upon the validity of the proposed research, and as such no restrictions on participation were imposed except for a requirement of normal or corrected-to-normal vision, as Experiment 1 was centered on making judgments regarding the category membership of visually presented images.

2.1.2 Materials

2.1.2.1 Equipment

All testing was conducted using iMac computers running Windows XP, located in individual, dimly lit, sound-dampened rooms. The computer program used to carry out the experiment was written using E-Prime 2.0 by the author. Participants seated themselves at a comfortable distance from the monitor, and responses were given using the computer mouse.

2.1.2.2 Stimuli

The visual stimuli used were created using the computer game Spore (Wright, 2008), and comprised semi-realistic computer-generated images of walking, ostrich-like birds approximately 300 pixels in width by 400 pixels in height (see Figure 2, left). The birds varied on six unique binary-valued dimensions, determined by the presence or absence of certain physical features: a feathered plume on the bird's head, a ridged protrusion on the neck, a shark-like fin on the back, a pair of claws on the lower legs, a cluster of spikes on the body, and a large, rounded tail.

Figure 2: Composite training stimuli showing all six features for Experiments 1 (left) and 2 (right). Category-diagnostic features include a feather on the head, neck ridge, claws, shark-like fin, body spikes, and tail. Wing, beak, feet, and other features are category-invariant.



The category structure followed Kruschke (1996). There were four separate bird species, each of which had its own perfectly diagnostic cue and shared an imperfectly diagnostic cue with one other bird (see Table 1). The base rates with which each species was presented during training followed previous work on the inverse base-rate effect: two species were designated as common and two as rare, with the common species appearing three times as often as the rare species. There was no variation within species in the training task, and feature assignment was counterbalanced such that each feature represented each abstract cue an equal number of times across subjects (for instance, the claws represented I to some subjects, PC, to others, and PR to the rest). Three different images of each exemplar were used in training, showing the same stimulus against a plain black background with subtle differences in posture; for the transfer phase, two different images of each bird were used, each of which appeared against a background of grass and trees.

Table 1: Sample category structure for Experiment 1 training phase.

Species	Frequency	Nondiagnostic feature (I)	Diagnostic feature (PC/PR)
Peruvian	Common	Feather on head	Spikes on body
Bolivian	Rare	Feather on head	Rounded tail
Chilean	Common	Ridged neck	Fin on back
Mexican	Rare	Ridged neck	Curved claws

2.1.3 Procedure

2.1.3.1 Training

Participants were instructed that four new, related species of flightless bird have been discovered in Latin America, and that they must learn which birds belong to which species: Bolivian, Peruvian, Mexican, and Chilean. Training took the form of a self-paced supervised category learning task, in which the participant would see a bird, select an answer after making a decision about the bird's category membership, and be given corrective feedback before moving on to the next trial. The training phase was divided into twelve randomized blocks of eight trials each for a total of 96 trials; each block consisted of three exemplars of each common species and one of each rare species. An early learning criterion was implemented, whereby 24 consecutive correct answers would result in an early termination of the learning phase.

2.1.3.2 Transfer

Once the training phase ended, the transfer phase would begin. Participants were informed that they would see a series of photographs of birds taken by an amateur photographer somewhere in South America, and must decide on the basis of these photos which species the birds belong to. The transfer stimuli were all novel, and had the abstract structures I, PC, PR, PC+PR,

and I+PC+PR. This mirrors the transfer phase of Kruschke's (1996) Experiment 1, excluding the stimuli that combine cues from multiple category pairs (I_1+PC_2 , PC_1+PR_2 , etc.). For instance, consider a case where feature I is a feathered plume on the bird's head, feature PC is a shark-like fin on the back, and feature PR is a rounded tail. Thus, the stimulus PC+PR would be a bird with a fin, a tail, and no feather on its head. Each transfer stimulus was presented twice for each category pair, for a total of 20 transfer trials. Response options were the same as in the learning phase – Bolivian, Chilean, Mexican, and Peruvian – but no feedback was given after the answer was provided for each trial..

2.2 Results

All statistical analyses were conducted using SPSS 15.0 statistics software. The mean response proportions for each abstract transfer stimulus were obtained. Category pair membership was taken into account such that the response to each stimulus was classified as common versus rare, and consistent versus inconsistent with the category pair implied by the visible features. For example, categorizing a bird with features $I_1+PC_1+PR_1$ as species C_1 or R_1 would be consistent with implied category pair membership, whereas categorizing the same bird as a member of species C_2 or R_2 would be inconsistent. This allowed response proportions to be collapsed across category pairs, greatly simplifying analysis. Following Medin and Edelson (1988), an IBRE was considered to exist if the stimulus characterized by the ambiguous I+PC+PR feature combination produced a higher proportion of common category responses than the equally

ambiguous PC+PR stimulus; this was tested using a paired-samples one-tailed t -test.

Consistent with the presence of an IBRE, the I+PC+PR stimulus was categorized as a member of the consistent common category 40.6% of the time, while the PC+PR stimulus was categorized as such only 23.8% of the time (see Table 2). The difference was statistically significant, $t(39) = 4.281$, $p < .001$. It could potentially be argued that this difference was an artifact of a discrepancy in the proportion of consistent-category responses, either rare or common, between the two stimuli; while PC+PR was categorized as a member of either consistent category 86.9% of the time, I+PC+PR was categorized as such 92.5% of the time. However, this difference is quite small in magnitude, and in any event did not reach significance, $t(39) = 1.548$, $p > .15$. The proportion of consistent common categorizations for the I+PC+PR stimulus was not significantly different from 50%, $t(39) = -1.891$, $p = .066$; at best, it is a marginal difference, and not in the normative direction.

Table 2: Categorization response proportions to transfer stimuli in Experiment 1.

Stimulus	Consistent common	Consistent rare	Inconsistent
I	63.1	14.3	22.6
I+PC+PR	40.6	51.9	7.5
PC+PR	23.7	63.1	13.2

2.3 Discussion

The robust difference between category responses to I+PC+PR and PC+PR indicates a strong IBRE, an encouraging sign for the use of this stimulus

set in Experiment 2. However, there appears to be a systematic bias in the responses - in previous research, I+PC+PR has generally been categorized as a member of the common category more often than as a member of the rare category (e.g. Lamberts & Kent, 2007; Kruschke, 2001). This discrepancy was likely due to a systematic salience advantage of the PR feature. While this would not do any great injury to the validity of a study focusing on the IBRE, equating salience across features would likely provide a significant advantage in reducing error variance across category structures and in more closely fitting the assumptions of computational models such as EXIT (Kruschke, 2001).

Further examination of the data revealed consistent differences in response proportions across the three different feature assignments. When the head and tail were the two PC features, the proportion of common category responses was quite high – 64.3% for I+PC+PR and 51.8% for PC+PR. When the tail or head was used as an I feature, the difference between PC+PR and I+PC+PR became quite large – a mean of 19.3%, compared to 12.5% when the I features were the claws and body-spikes. This seems to suggest that the head and tail possessed a much higher initial salience than the other stimulus features.

In an attempt to more closely equate salience across stimulus features, two modifications were made to the stimulus set – the size of the head-feather feature was decreased, as was the size of the tail (see Figure 2, right). It was suspected that these changes would give the head and tail features less initial salience in Experiment 2, though it seems likely that a significant amount of their salience originated from their position on the bird. Being at the extreme ends of

the bird, they were inherently quite noticeable, perhaps much more so than medial features such as claws.

3: EXPERIMENT 2

Experiment 2 had the same general structure as Experiment 1, though its purpose was to directly test the relationship between categorization and completion rather than to validate the stimulus set. This was achieved by adding an additional transfer stimulus, ?+PC+PR, which included the PC and PR features, but with an opaque occluder, such as a tree or a rock, blocking the participant's view of the I feature (see Figure 3). With the imperfectly diagnostic feature thus occluded, the nature of the responses to ?+PC+PR would provide important insight into how we categorize partially visible objects. If ?+PC+PR were categorized in a similar manner to I+PC+PR, it would suggest that people attempt to make informed inferences about the likely value of hidden features from previously stored exemplars, and take the inferred feature values into account when making a categorization decision. Conversely, a close resemblance in response proportions between PC+PR and ?+PC+PR would suggest that people simply categorize based on immediately available information, and would treat missing features as though they were absent. A result somewhere between the two would indicate that people remain agnostic regarding the value of hidden features – or infer a mean value (White & Koehler, 2004) – and adjust their categorization decisions accordingly. The second option would be the outcome predicted by RPI (Kellman, 2003), though the first seems plausible as well.

Figure 3: A sample transfer stimulus from Experiment 2 with the head feature occluded.



To further elucidate the processes underlying high-level completion of complex objects, a recognition test was also administered as part of Experiment 2. Although the imperfectly diagnostic feature would not have been seen during the learning phase, there is a qualitative difference between an absent feature and a feature whose presence or absence is unknown (Verguts et al., 2004). Thus, chance responding on the recognition test would suggest either a lack of completion effects, as both outcomes would seem equally familiar, or mean-value completion as suggested by White and Koehler (2004).

3.1 Method

3.1.1 Participants

42 undergraduate students from Simon Fraser University agreed to participate in Experiment 2 in exchange for either course credit or a cup of coffee and a doughnut. Three of the participants did not yield useable data due to computer-related difficulties; as such, the sample size for statistical purposes was 39.

3.1.2 Materials

The materials for Experiment 2 were essentially identical to those used in Experiment 1, with the exception of the novel transfer stimulus ?+PC+PR and the changes in the head and tail features. The learning phase was substantially the same as in Experiment 1, but the transfer task was split into three distinct phases: a categorization task involving the occluded stimuli alone, a recognition test, and finally, a series of transfer trials identical to those in Experiment 1.

3.2 Results

3.2.1 IBRE replication

To detect the presence or absence of an IBRE, it suffices to test whether the proportion of common-species responses was greater for the I+PC+PR stimulus or for the PC+PR stimulus. Consistent with the hypothesis, PC+PR was categorized as a member of the common species less often than was I+PC+PR (28.2% vs. 39.7%), $t(38) = 2.262$, $p < .05$ (see Table 3).

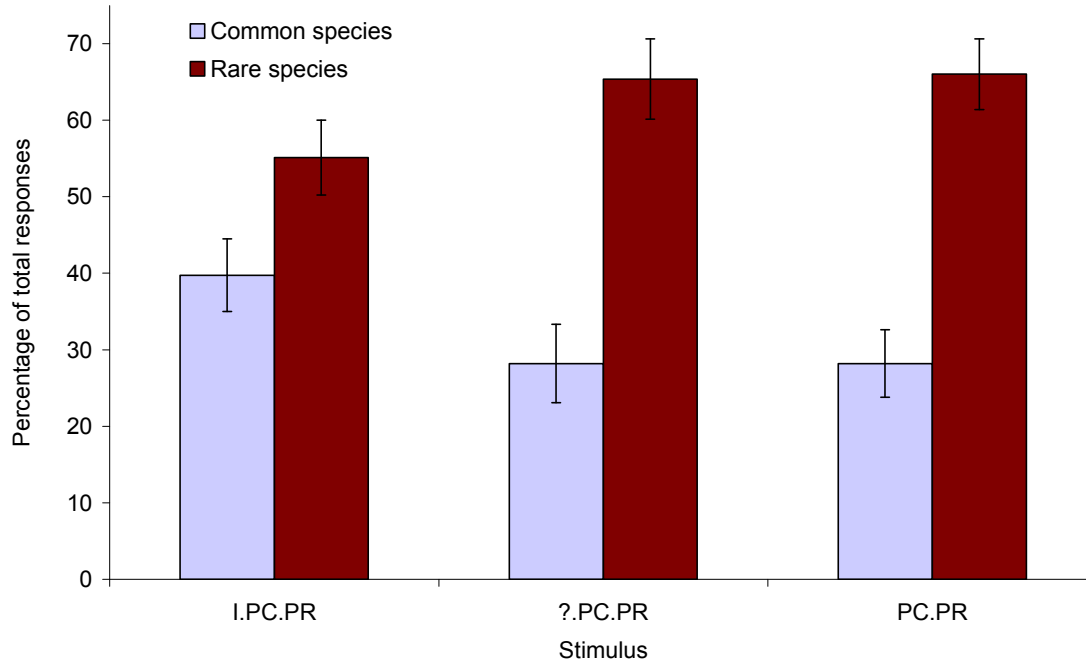
Table 3: Human categorization response proportions to transfer stimuli in Experiment 2.

Stimulus	Consistent common	Consistent rare	Inconsistent common	Inconsistent rare
I	60.3	21.2	14.1	4.5
I+PC+PR	39.7	55.1	3.8	1.3
PC+PR	28.2	66.0	1.9	3.8
?+PC+PR	28.2	65.4	1.3	5.1
PC	73.1	4.5	16.7	5.8
PR	7.7	82.7	5.1	4.5

3.2.2 Occluded stimulus

The primary concern of the present study was whether the partially occluded stimulus ?+PC+PR was categorized in a more similar fashion to PC+PR or to I+PC+PR (see Figure 4). In fact, the proportions of consistent common responses to PC+PR and ?+PC+PR were identical (28.2% and 28.2%), $t(38) = 0.00, p = 1.00$, while ?+PC+PR was categorized as common significantly less often than was I+PC+PR (28.2% vs. 39.7%), $t(38) = 1.780, p < .05$ (one-tailed).

Figure 4: Response proportions (consistent categories only) to the three critical transfer stimuli in Experiment 2. Error bars represent standard error.



3.2.3 Recognition test

It was originally predicted that the recognition test, administered just after the occluded-stimulus transfer trials, would reveal a general preference for the I+PC+PR stimulus. In fact, this was not the case; participants responded that the PC+PR bird looked more familiar than the I+PC+PR bird 57.7% of the time. However, this was not a statistically reliable departure from chance, $\chi^2(1) = 1.846, p > .15$.

3.3 Discussion

In one sense, the occluded-stimulus transfer trial results are in accordance with the prediction made by the RPI framework. When a stimulus feature was

occluded, the stimulus was categorized almost exactly as though it were a fully visible, otherwise identical stimulus from which the same feature was absent. People do not appear to make informed inferences about the value of hidden features, and in this sense, the stimulus is categorized strictly on the basis of partial information. In another sense, though, RPI is not a good description of the completion of complex occluded objects. When a category-diagnostic feature is occluded, people tend to assume the absence of unseen features – to extrapolate the general structure of the stimulus, but to infer the presence of nothing beyond that – and to use that inference in categorization.

The pattern of responses to the recognition trials is somewhat unexpected, and may be due to demand characteristics. In order to avoid showing participants possible completions of the occluded stimulus before the recognition test, the recognition trials came immediately after the occluded stimuli. This provision may have rendered the purpose of the recognition test rather obvious; participants, seeing a bird hidden behind a tree and then being asked to choose two possible completions of that bird, could very easily guess the purpose of the test. At any rate, the results are somewhat equivocal; there was no significant preference for either bird.

A perplexing discrepancy between this result and those of earlier IBRE studies is the fact that both I+PC+PR and PC+PR elicited comparatively few common-species responses; indeed, they were both categorized as rare a majority of the time (55.1% and 66.0%, respectively). Again it seems likely that this is an effect of salience; such an effect could result from a much stronger

association between PR and the rare category than between PC and the common category. As was suspected to be the case in Experiment 1, such a difference in associative strength could well be the result of systematic salience differences between features. Although Experiment 1 was devised with the specific aim of remedying this issue, it would appear that the stimulus modifications did not do much to mitigate the salience problem. At any rate, a global deviation in response proportions is of little importance to the central aim of Experiment 2; as long as there is a reliable difference in response proportions between I+PC+PR and PC+PR, the two stimuli can be used as a fruitful basis for comparison with the partially occluded stimulus ?+PC+PR.

4: COMPUTATIONAL MODELING

The question of modeling human performance in Experiment 2 is a complicated one. In recent years there has been a great expansion in the number and scope of available models of human category learning, allowing models to accurately predict increasingly diverse human datasets. For instance, ALCOVE (Kruschke, 1992), though one of the most robust exemplar-based connectionist models available for many years (Kurtz, 2007), cannot account for base-rate effects (Kruschke, 1996) and has no obvious way to handle incomplete input – stimuli with one or more features of unknown valuation. The base-rate issue was addressed by Kruschke with the ADIT model (1996), whose extension, EXIT, includes a selective attention mechanism and has been shown to provide a good fit for data from several category learning experiments in animals and humans (Kruschke, 2001). Incomplete information, however, remained a problem for the ALCOVE lineage of models until the development of ADDCOVE several years later (Verguts et al., 2004).

The problem of handling unknown feature values is arguably inherent to the distance metric used in models such as ALCOVE and EXIT (Verguts et al., 2004). Stimuli in these two exemplar-based models are encoded as discrete points in an n -dimensional psychological space (where n is the number of object features), such that an exemplar's position along a certain dimension is equivalent to its value on the corresponding dimension. The perceptual similarity

between two points in the space is said to be equivalent to the distance between them, computed either using a Euclidean (as in ALCOVE) or city-block (as in EXIT) metric. Tversky (1977) presented several arguments against the utility of the geometric distance assumption and proposed two alternatives: the *contrast model*, which instead computes the similarity between two stimuli based on the number of shared and distinctive features they possess, and the *ratio model*, which expresses similarity as the ratio of matching to nonmatching features. These two approaches fall under the umbrella of *additive* methods of computing similarity – each additional common feature adds to the total similarity between the two stimuli being compared. This is in contrast to geometric distance-based methods of measuring similarity, in which adding a common feature to two stimuli produces no change in the perceived similarity between them.

Among the problems remedied by an additive approach to similarity is the issue of comparing two stimuli with different numbers of known feature values. For instance, assume that stimulus A has feature values [1 0], while stimulus B has feature values [1 ?] (“?” representing an unknown feature valuation – perhaps the second feature in stimulus B is occluded). A model using geometric distance to compute similarity would have no way of directly comparing these two stimuli, as the position of stimulus B along the second dimension is indeterminate. The only solution would seem to be to collapse the space such that the second feature is not taken into account, and compare the stimuli on the basis of the first feature alone. Since the distance between 1 and 1 is zero, the stimuli should be judged to be the same. This leads to the curious prediction that

there should be no difference in perceived similarity between [1 ?] and [1 0] on the one hand and [1 0] and [1 0] on the other, as the distance is zero in both cases. Verguts et al. (2004) tested exactly this prediction in their Experiment 1, and unsurprisingly found it to be false.

To address this shortcoming, Verguts et al. (2004) devised the ADDCOVE model, a modification of ALCOVE. Whereas ALCOVE uses a Euclidean distance computation to determine the similarity of two exemplars, ADDCOVE employs an additive feature-matching similarity metric reminiscent of Tversky's ratio model (Tversky, 1977). By design, ADDCOVE is much better equipped than its predecessor to handle stimuli with missing data. In the situation described above, [1 0] and [1 ?] have only one matching feature out of a possible two, while [1 0] and [1 0] have two matching features out of a possible two, making the latter pair the more similar of the two. This is made possible by, among other things, doubling the number of nodes at the input layer. In ALCOVE and EXIT, the presence or absence of each feature is indicated by the activation or nonactivation of a single node. In contrast, each feature in ADDCOVE is represented by two nodes: one is active when the feature is present, the other when it is absent. If it is not known whether the feature is present or absent, neither node is activated, rendering the feature unable to contribute to additive similarity.

Unfortunately, no current model of categorization is known to be able to both handle missing features and reproduce base-rate effects observed in human data. While EXIT's rapid attentional shifting and attention to base rates

produce a close fit to human data in the inverse base-rate effect and apparent base-rate neglect paradigms (Kruschke, 1996), its geometric distance-based method of computing similarity will most likely render it unable to fit human performance on trials with incomplete stimulus information, as in the transfer phase of Experiment 2. Conversely, ADDCOVE solves the problem of missing information, but is quite similar to ALCOVE in its general structure (Verguts et al., 2004). Like its predecessor, then, it is unlikely to produce an acceptable fit to human data when base rates are uneven, as is the case in the present study.

One possibility for a model that can solve both of these problems is SUSTAIN (Love et al., 2004), which employs an additive method of comparing probes to stored stimulus clusters. Although SUSTAIN is a connectionist model, it employs a clustering-based architecture quite distinct from that of the ALCOVE lineage of exemplar models, and may conceivably be able to accurately model human performance in the inverse base-rate effect and apparent base-rate neglect paradigms. To the authors' knowledge, however, no data yet exists regarding SUSTAIN's effectiveness in modeling either of these effects.

Another solution is to develop a new model that is to EXIT as ADDCOVE is to ALCOVE – that is, a version of EXIT with a feature-matching rather than a geometric-distance similarity measure. Such a model would provide a solution to both of the above issues, and if found to be a good model of performance, may aid in generating useful hypotheses for future work in this area. Following the tradition of rather strained acronyms for models of category learning (for instance, ALCOVE stands for Attention Learning COVERing map and EXIT

stands for EXtended adIT), the new model is entitled EXALT, for EXit with Additive simiLariTy.

Thus, the goodness of fit of all three models – SUSTAIN, EXIT, and EXALT – to human performance data in Experiment 2 was examined. It was predicted that SUSTAIN and EXALT would provide better fits than EXIT due to their reliance on additive similarity.

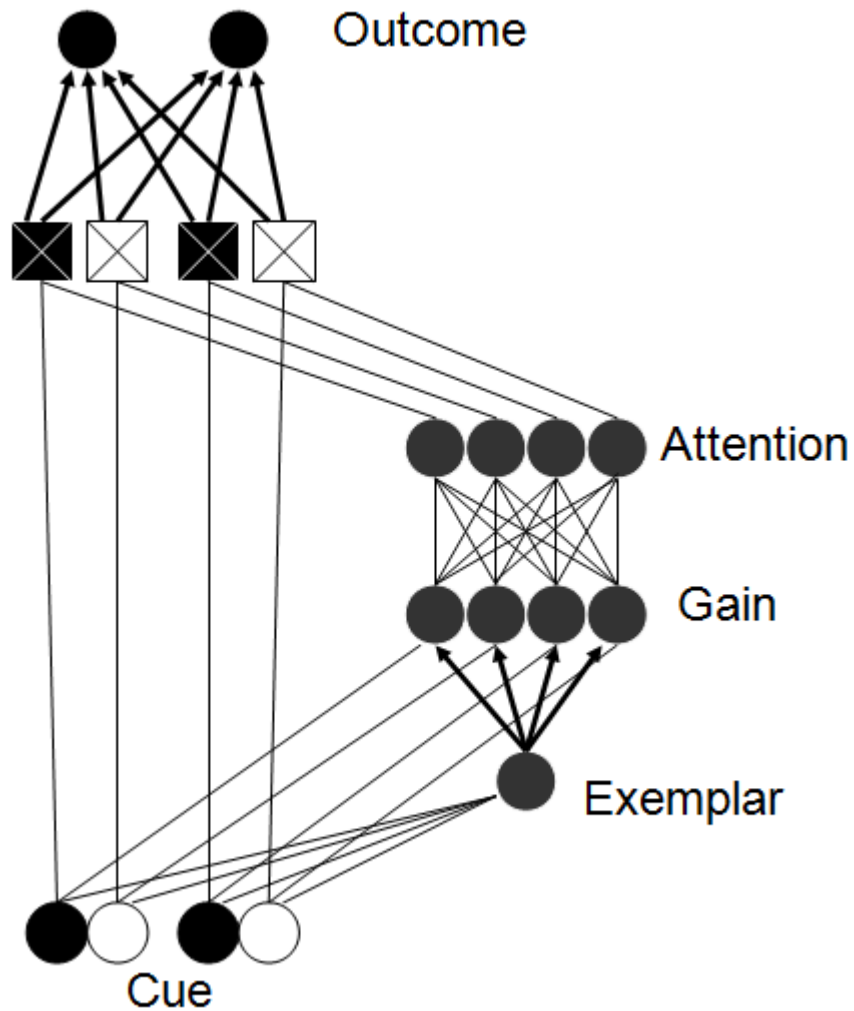
4.1 Method

The present study's implementation of the EXIT model, done in MATLAB, was essentially identical to the version used by Kruschke (2001) as a comparison against the eliminative-inference model of Juslin et al. (ELMO; 1999). Aside from the data used as input, the only substantive difference between the two versions lay in the exemplar-input similarity computation; since the original EXIT code did not have a specific provision for dealing with missing data, a conditional statement was added such that any missing input feature would not contribute to the city-block distance between exemplars and inputs, effectively treating missing features as matching features (following Verguts et al., 2004).

EXALT was developed through minor modifications to the original EXIT code: first, an additive similarity metric replaced EXIT's city-block distance computation; second, the format of the input was changed such that each feature was represented by two nodes – one of which was active when the feature was present, the other active when it was absent, and neither active when the feature value was unknown (see Figure 5 for an illustration of the model's architecture;

see also Appendix A for a full specification of the differences between EXIT and EXALT).

Figure 5: Architecture of the EXALT connectionist model.



Both EXIT and EXALT were fit to human data using constrained function-minimization with a Nelder-Mead simplex algorithm. The models both had seven free parameters, which were constrained such that none was allowed to approach zero or to be too large (see Appendix B for details of the constraints).

This decision was motivated by two factors: first, by considerations of model plausibility (for example, an unconstrained fit of EXIT to the data produced a good qualitative fit, but only when the attentional capacity parameter was allowed to take an extraordinarily high value); second, by the fact that a zero value for many of the parameters would obviate large chunks of model architecture (for instance, an exemplar-specificity of zero in EXIT or EXALT would cause all exemplars to be equally activated by any stimulus, while a zero value for bias salience would render both models unable to account for base-rate effects). Model fit for each iteration of the function-minimization algorithm was determined by computing the root mean-squared deviation (RMSD) between the model's response proportions on the transfer task and the corresponding human data.

The code for the SUSTAIN model, written in Python 2.4, was originally developed to model learning curves from the seminal category-learning study of Shepard, Hovland, and Jenkins (1966; Love et al., 2004). Substantial modifications to the provided program were required in order to bring the model in line with the requirements of the present study; although the basic architecture of SUSTAIN remained unchanged, it was necessary to rewrite segments of the code in order to allow for the existence of a transfer phase, and to implement a comparison between computer- and human-generated response proportions rather than an examination of learning rates and block-by-block accuracy. As with EXIT and EXALT, a simplex function-minimization algorithm was used to find local minima, with appropriate constraints applied to SUSTAIN's four free parameters.

For all three models, best-fitting parameters were obtained through minimization of RMSD. Since there is no general solution whereby a global function minimum can be found, simplex minimization tends to get caught in local minima; as such, multiple initial parameter values were used in order to attempt a best-fit convergence from several independent directions. A full listing of parameter constraints, function minimization starting points, and resulting RMSD values is provided in Appendix B.

4.2 Results

4.2.1 Fit of EXIT

An unconstrained fit of EXIT was obtained using the initial parameters in Experiment 2 of Kruschke (2001); in addition, two constrained simulations were run. One started from the eventual best-fitting parameters in the same experiment, the other from the minimum values for each parameter. Local minima were found in each case, though a different result was obtained each time.

The best fitting parameters ($c = 0.152$, $P = 3.932$, $\psi = 3.192$, $\lambda_g = 4.787$, $\lambda_w = 0.357$, $\lambda_x = 0.247$, $\sigma_1 = 1.4862$), derived from a constrained minimization starting at the final parameters used in Experiment 2 of Kruschke (2001), resulted in a total RMSD of 4.01 (see Table 4). In spite of the relatively close fit, the model was not able to capture some of the qualitative patterns of the data. Of greatest interest is the preponderance of common categorizations of the occluded stimulus ?+PC+PR; it was categorized as common more often (31.1%)

than either PC+PR (27.7%) or I+PC+PR (28.8%), which were quite close to one another. While most of the parameter values in this fit are eminently reasonable, one that stands out is the bias-salience free parameter (σ_1) which represents the salience of the response prompt relative to the features of the stimulus. Here, the best-fitting value for bias salience was 1.49, which is considerably greater than the salience of each cue (1.0 by default) – a seemingly unlikely scenario, and a value over a hundred times greater than the best-fitting value of 0.0143 obtained by Kruschke (2001).

Table 4: Best fit of EXIT to human transfer data from Experiment 2

Stimulus	Consistent common	Consistent rare	Inconsistent common	Inconsistent rare
I	55.8	23.4	11.3	9.5
I+PC+PR	28.8	62.9	4.2	4.2
PC+PR	27.7	62.9	4.7	4.6
?+PC+PR	31.1	62.5	3.1	3.2
PC	76.9	6.6	8.8	7.8
PR	4.7	83.0	6.3	5.9

4.2.2 Fit of EXALT

As with EXIT, the EXALT model was run using several different initial parameter values. One unconstrained run was performed starting at the initial parameter values used by Kruschke (2001), as well as several constrained runs with different starting points.

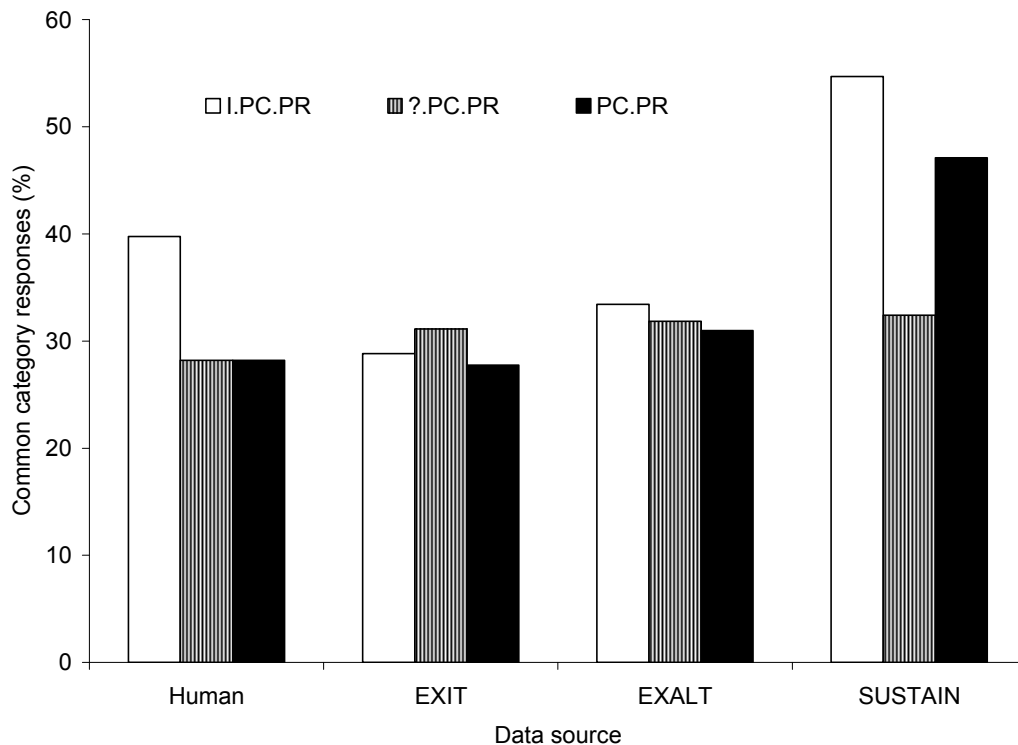
The best fitting parameters in EXALT ($c = 0.406$, $P = 2.822$, $\psi = 6.553$, $\lambda_g = 0.587$, $\lambda_w = 0.123$, $\lambda_x = 0.102$, $\sigma_1 = 0.015$) yielded an RMSD of 3.34, a slightly better fit than was obtained using EXIT (see Table 5). However, the relative response proportions for PC+PR, I+PC+PR, and ?+PC+PR are not quite in line

with the human data (see Figure 6); all three are quite similar to one another, and ?+PC+PR actually lies in between the other two. The parameters required to obtain this fit are in line with what one would expect from the model; while the choice-decisiveness parameter is somewhat high, it is not overly so, and the best fitting bias salience is nearly identical to the result obtained by Kruschke (2001).

Table 5: Best fit of EXALT to human transfer data from Experiment 2

Stimulus	Consistent common	Consistent rare	Inconsistent common	Inconsistent rare
I	64.6	14.5	15.1	5.7
I+PC+PR	33.4	60.6	2.9	3.1
PC+PR	31.0	60.4	4.7	3.9
?+PC+PR	31.8	55.3	6.2	6.8
PC	61.7	2.6	20.3	4.2
PR	8.5	78.4	8.1	5.0

Figure 6: Each model's best-fitting proportion of consistent common category responses to the three critical transfer stimuli in Experiment 2, with human data for comparison.



4.2.3 Fit of SUSTAIN

Although the SUSTAIN model was successfully implemented, function minimization proved quite difficult. In spite of the fact that SUSTAIN has only four free parameters, the shape of the function was such that the minimization program was unable to find a consistent local minimum. Attempts to solve this issue were in vain; increasing the number of simulated subjects did not improve the situation, and other minimization algorithms built into the same module yielded the same result. An alternative minimization module was not forthcoming for a compatible version of the Python language. Due to this unforeseen difficulty, SUSTAIN failed to converge to a solution; however, the simulation was run from

several different starting points in an attempt to find a local minimum.

The resulting best fitting parameters ($r = 1.388$, $\beta = 2.635$, $d = 3.480$, $\eta = 0.556$) did not produce response proportions approaching the accuracy of EXIT or EXALT, with an RMSD of 54.023 (see Table 6). The proportion of consistent common responses was considerably lower for ?+PC+PR than for either of the other ambiguous transfer stimuli, indicating that SUSTAIN provides neither a good qualitative nor quantitative fit to the data.

Table 6: Best fit of SUSTAIN to human transfer data from Experiment 2

Stimulus	Consistent common	Consistent rare	Inconsistent common	Inconsistent rare
I	47.3	26.0	14.3	12.6
I+PC+PR	54.7	26.6	9.5	9.3
PC+PR	47.1	27.4	12.8	12.8
?+PC+PR	32.4	45.3	10.8	11.6
PC	78.9	6.5	7.5	7.1
PR	8.5	74.1	8.5	8.9

4.3 Discussion

The relative inability of SUSTAIN, EXIT, and EXALT to provide a good qualitative fit for the critical data indicates that there exists a discrepancy between the architecture of each model and the way in which humans respond to missing data. Indeed, this could be anticipated by looking at the provisions that each model makes for unknown feature values. Due to its reliance on city-block distance as a way of calculating exemplar activation, EXIT counts a missing feature as a matching feature across all exemplars – that is, for the occluded transfer task in the present study, the missing feature did not serve to differentiate the input from any of the four previously stored exemplars. As such,

EXIT's response proportions varied rather unpredictably; in one simulation, the proportion of common responses for ?+PC+PR was higher than that of either I+PC+PR or PC+PR; in another, it was the lowest of the three; in yet another, all three of those transfer stimuli elicited an approximately equal number of consistent common categorizations.

In contrast to its predecessor, EXALT produced a fairly consistent pattern of responses to the occluded transfer stimulus across several model fitting attempts. Unlike EXIT, however, EXALT produces no additional exemplar activation when a feature value is unknown – that feature simply does not factor into exemplar similarity computations at all. As such, EXALT effectively remains agnostic as to the presence or absence of the unknown feature, which resulted in a proportion of consistent common responses to ?+PC+PR that fell in between the values for PC+PR and I+PC+PR.

Finally, the technical difficulties with SUSTAIN was make it difficult to draw strong conclusions about the model's suitability for tasks involving missing data. What results were obtained were not encouraging, however, and much like EXALT and ADDCOVE before it, it essentially implements a form of additive similarity. Each cluster has a receptive field for each cue, and the closer a cue's value is to the center of that receptive field, the more activation it produces for the cluster. A missing feature would produce no activation at all; indeed, a provision for this exact situation is built directly into the model. Like EXALT, then, SUSTAIN essentially remains agnostic about the identity of missing features. The results of the present study indicate that this is not a good approximation of how

people categorize objects with unknown features; rather, they seem to make the positive assumption that the unseen features are absent.

5: GENERAL DISCUSSION

The results of the present study, particularly of Experiment 2, point to a previously unexpressed tendency in human visual categorization. Although Kellman's (2004) formulation of complex-object completion as a kind of recognition by partial information could be argued to be correct, the way in which that recognition happens is something that existing models simply cannot approximate: partially visible stimuli are generally categorized as though unseen category-diagnostic features are absent. This only appears to be the case at a particular value of feature diagnosticity, however. To illustrate this, it would perhaps be instructive to return to an example from the beginning of the present study, of a pickup truck whose back end is occluded by a building.

The object can immediately be categorized as some kind of motor vehicle, based on the visible features – headlights, wheels, and a driver sitting behind a steering wheel. From the results of Experiment 2, it seems likely that the gross stimulus structure – the presence of wheels, windows, and so on – is inferred based on what is visible, with help from low-level completion processes such as contour and surface extension. If it is not immediately obvious to which subordinate category the vehicle belongs (pickup truck, minivan, sports car, etc.), the precise identity of those features, as well as other identifying markings that might lead to subordinate categorization, may be set to particular *default values*. This can be seen in the results for Experiment 2; there is no indication that

people perceived the partially occluded stimuli as belonging to a novel stimulus class, such as headless birds or wildlife bisected by trees. Recognizing that they had the general structure of other birds seen in the experiment, participants categorized the birds on a superordinate level, completed the overall structure of the birds accordingly (assuming the presence of a head, of feet, etc.) and subsequently set the appropriate unknown features to their default values before making a subordinate categorization decision (were the birds Chilean, Bolivian, Peruvian, or Mexican?). Completion of a stimulus region containing category-diagnostic features could thus be termed default-feature completion (DFC) – the general structure is inferred based on superordinate categorization, and important features are set to some default value.

If DFC is a generally accurate description of the categorization of partially occluded objects, further elucidation of the process by which it takes place could be fruitfully dealt with in future research. The precise identity of the default values is one question – if the feature values of the bird stimuli used here were “inverted” such that features were present by default, and their *absence* was significant for category membership, the default value for completed features would not be immediately clear *a priori*. If such an experiment was conducted and features were perceived as present when occluded, it would suggest that the least informative feature value is used as the default. On the other hand, if the stimuli were categorized as though the occluded features were absent, it would suggest that the default assumption is absence, regardless of how informative or uninformative such information would be. The latter seems the more probable

option, as the imperfectly diagnostic feature in the training stage of Experiment 2 was equally informative regardless of its value – exactly one of the two imperfectly diagnostic features was always present, such that one feature’s absence was perfectly correlated with the other’s presence. This contrasts sharply with previous work regarding categorization based on missing information – White and Koehler (2004) suggested that the “mean” value is generally inferred for missing features, but the mean value for the imperfectly diagnostic features in Experiment 2 would be half-present, half-absent. Perhaps the tendency to use the mean as a default value is restricted to continuously valued features, or to probability-based learning paradigms in which no single feature is perfectly category-diagnostic.

Thus, there are some unanswered questions regarding the generality of the present study’s findings. It is not clear whether the tendency to perform DFC is particular to visual categorization or whether it might apply to other modalities as well. For instance, much of the literature on the IBRE to date has focused on using abstractly described symptoms to settle upon a medical diagnosis; indeed, the only previous work looking at missing data in the IBRE (Verguts et al., 2004) used just such a task. White and Koehler (2004) also used abstract disease descriptions in their investigation of the contribution of missing information to categorization; it may be that there is something particular to visual stimuli that makes the absence of a feature more likely to be the default value. An important difference here is that visual stimuli are subject to low-level completion process, while abstract descriptions are not. For instance, in the present study, the

category-diagnostic features of the birds were mostly discontinuous surfaces or appendages that protruded from the bird's body. If the section generally containing one of those features were occluded, the bird would be amodally completed across the occluder – and the completed bird would not include the feature, as its presence would require a deviation from the visible contours or surfaces on either side of the occluder, counter to what one might expect from amodal completion. However, the head and tail features may constitute a counterexample to this account of the apparent DFC effect; when Experiment 2's head feature was occluded, the entire head and a good portion of the neck were blocked out by the tree, rather than just the feather – this results in a significant discontinuity between the last visible portion of the neck and the tip of the beak visible on the other side of the tree, rendering it seemingly unlikely that the general contours of the head could be meaningfully interpolated at all (see Figure 3). The tail feature in Experiment 2 was a rough continuation of the contours of the bird's body, so it seems as likely as not that a tail would be included in any amodal completion of the bird. The same is true for the claws; the contours of the claws are essentially a continuation of the contours of the bird's upper legs, so one would expect amodal continuation to capture the claw-contour in the occluded condition, as the knee joint, lower legs, and feet were all blocked by the occluder.

Of course, contour relatability, as explained by Kellman et al. (2001), is only a restriction on local completion effects; a large contour discontinuity does not preclude global completion from affecting the participants' perception of the

occluded stimulus. While the status of global completion is a matter of some debate (e.g. Kellman, 2003; Plomp & van Leeuwen, 2006), the findings of the present study may suggest that DFC is a specific kind of global completion – one based not specifically on symmetry, but on assumptions regarding gross stimulus structure as a result of superordinate categorization.

The possibility that low-level amodal completion may account for the apparent DFC effect may be fruitfully examined in future research, perhaps by restricting category-diagnostic visual cues such that amodal completion of an occluded section would not produce a firm prediction as to a feature's presence or absence. It may also be instructive to perform an IBRE task similar to that used in Experiment 2, but, following Johnson and Olshausen (2005), with “deleted” stimuli used in addition to occluded ones. As deletion of sections of a visual stimulus does not produce amodal completion effects in the same way that occlusion does, the results of such a study would provide further insight into the processes underlying DFC – is it dependent upon depth and occlusion cues, or is it simply a reaction to missing information?

Along with the behavioural results from the transfer task of Experiment 2, the failure of the tested models to provide a good fit for the data indicates a systematic problem in the way that missing information is conceptualized in the current categorization literature. By and large, existing connectionist models do not specify how missing data is dealt with (e.g. ALCOVE, EXIT), while those few models that include provisions for missing data make assumptions about how unknown feature values play into the categorization of complex objects that, on

the basis of the present study, are now known to be faulty (SUSTAIN, ADDCOVE).

ADDCOVE and EXALT, along with SUSTAIN, deal with missing data in essentially the same way – they take missing features out of exemplar similarity computations, so that anything that is unknown does not contribute to the activation of previously stored items. This is equivalent to assuming that missing features are universally nonmatching and do not increase additive similarity, no matter the fully-informed exemplar with which they are compared. Paradoxically, then, though the new models sought to improve upon the way in which ALCOVE and EXIT dealt with missing data, they fell into the same trap of assuming that missing features would affect similarity to all stored exemplars in the same way. While the newer models assume that missing information is perceived as universally discrepant from items in memory, the older models treat missing information as a universal match. The present experiment, along with previous work (e.g. Ganzach & Krantz, 1990; White & Koehler, 2004), has demonstrated that both of these approaches are fundamentally flawed. People do not to treat missing information about a particular category-diagnostic feature in a qualitatively different way from other features; rather, they simply accord some sort of default value, and use that default value in making subordinate categorization decisions.

It is premature to attempt to describe a model that incorporates DFC as a method of dealing with unknown data. The factors contributing to the assignment of default values are currently unknown, as is the generality of the effect – is DFC

common to all categorization problems, only those involving visual stimuli, only occluded visual stimuli, or only occluded visual stimuli with discretely valued features? Regardless, the present study has made it abundantly clear that current models of categorization are simply inadequate, and that further investigation into DFC is necessary if future models are to deal with unknown feature values in a coherent, empirically supported way.

APPENDICES

Appendix A – EXALT Model Specification

EXALT, or EXit with Additive simiLariTy, is, as its name implies, an extension of the connectionist model EXIT (Kruschke, 2001). While it mimics its predecessor in most respects, EXALT uses an additive rather than a distance-based similarity metric and possesses twice as many input nodes for each stimulus feature, allowing it to deal with missing data in a different way from the original model. A broad overview of the architecture of EXALT for a structure with two input features, one exemplar, and two possible category outputs is shown in Figure 4; bold arrows represent learned weights.

Each input cue is represented by two nodes; one is active when the feature is present in the stimulus, the other is active if the feature is absent, and when the status of the feature is unknown neither one is activated. When a stimulus is presented, activation propagates to each output node according to a weighted sum of the input cues, and the model's output is chosen from the available category outputs via the Luce choice rule.

The weights accorded to each input node are determined by activation in the attentional system, which includes exemplar-specific attentional gatings that weight cues differently according to the similarity of the input to previously stored

exemplars. In EXIT, the activation of a given exemplar follows an inverse exponential curve based on the city-block distance between exemplar and input,

$$a_x^{ex} = \exp(-c \sum_i \sigma_i |\psi_{xi} - a_i^{in}|), \quad (\text{Eqn. 3; Kruschke, 2001})$$

where a_x^{ex} is the activation of exemplar x , c is a free parameter that determines the specificity of exemplars, σ_i is the perceptual salience of cue i , ψ_{xi} is the activation of cue i in the exemplar, and a_i^{in} is the activation of cue i in the input.

EXALT computes activation in a somewhat different way,

$$a_x^{ex} = \exp(c \sum_i \sigma_i \psi_{xi} a_i^{in}),$$

such that exemplar activation follows a positive exponential curve depending on the number of matching features between exemplar and input (a match is defined as a cue being present or absent in both input and exemplar).

Attention then propagates to the gain nodes, where exemplar weights and input data are combined and then normalized to form attentional weights that can be applied to the original input to determine which categorization decision is made. When feedback is received, weights are adjusted via gradient descent on error.

The exemplar-specificity equation and the number of input nodes are the only differences between EXIT and EXALT; further information on any of the components of EXALT is available in the original specification of the EXIT model (Kruschke, 2001).

Appendix B: Supplemental Modelling Information

Response proportions generated by EXIT, EXALT, and SUSTAIN were all fit to human data from multiple starting points, both with and without constraints. Constraints were chosen somewhat arbitrarily, based on the plausibility of particular parameter values within each model. For the majority of these analyses, model fit was calculated for all six transfer stimuli; however, for each model a three-stimulus case was also attempted, in which only response proportions for the three critical transfer stimuli (I+PC+PR, PC+PR, and ?+PC+PR) were recorded and compared. This restriction did not produce substantially different response proportions in any of the models, and as such the results of the three-stimulus case are not reported in detail here.

EXIT/EXALT

The exemplar specificity parameter, c , determines the tolerance of the exemplar similarity computation to discrepant values. It was constrained to the range (0.1, 5).

Attention capacity, symbolized by P , affects the model's ability to distribute attention. A low value of P indicates that an increase in attention to one cue must be accompanied by a large decrease in attention to other cues. P was constrained to the range (0.1, 10).

ϕ , the choice decisiveness parameter, determines how likely the model is to pick the winning output. The lower the decisiveness, the larger the stochastic component in determining output. Decisiveness was constrained to fall within

(0.1, 10).

The attention shifting rate, λ_g , determines how quickly the model can shift its attention from one cue to another. It was constrained to the range (0.1, 5).

λ_w , the learning rate for output weights, was kept within the range: (0.1, 5).

The learning rate for associative weights from exemplars to gain nodes, λ_x , was constrained to the range (0.1, 5) as well.

Finally, bias salience, σ , represents the salience of the response prompt relative to the features of the stimuli to be categorized. σ is the parameter that allows EXIT to model base rate effects; as such, it was constrained to the range (0.01, 1.5).

For EXIT, the starting points used were ($c=0.01$, $P=2.3881$, $\varphi=3.9175$, $\lambda_g=0.3633$, $\lambda_w=0.0502$, $\lambda_x=0.0167$, $\sigma=0.0143$), from the best-fitting values of Experiment 2 in Kruschke (2001); ($c=0.5$, $P=2.4$, $\varphi=5.0$, $\lambda_g=0.36$, $\lambda_w=0.05$, $\lambda_x=.018$, $\sigma=0.01$); ($c=0.1$, $P=0.1$, $\varphi=0.1$, $\lambda_g=0.1$, $\lambda_w=0.1$, $\lambda_x=0.1$, $\sigma=0.1$); and ($c=0.5$, $P=0.5$, $\varphi=0.5$, $\lambda_g=0.5$, $\lambda_w=0.5$, $\lambda_x=0.5$, $\sigma=0.5$). The best-fitting values ultimately resulted from the simulation with the Kruschke values as a starting point (see Table 4 for response proportions). In general, RMSD values ranged from 4-10, and the response proportions for the critical transfer stimuli I+PC+PR, PC+PR, and ?+PC+PR were quite close to one another, with ?+PC+PR typically eliciting fewer common-category responses than the other two.

EXALT was run with the same starting points; however, the best fit ultimately resulted from the second, ($c=0.5$, $P=2.4$, $\varphi=5.0$, $\lambda_g=0.36$, $\lambda_w=0.05$, $\lambda_x=.018$, $\sigma=0.01$; see Table 5 for response proportions). Model fit values were in

the range 3-10, and response proportions were consistent across starting points, with I+PC+PR eliciting the most common-category responses, PC+PR having the fewest, and ?+PC+PR falling somewhere in between.

SUSTAIN

The attentional focus parameter r , which determined the speed of attentional shifting, was constrained to fall within the range (0.1, 20).

β , the cluster competition parameter, governed the degree to which different clusters inhibited one another. It was restricted to the range (0.1, 10).

Decision consistency, symbolized by d , affected the probability of making an odd or counterintuitive categorization decision. It was set to fall within (0.1,20).

Finally, η was the learning rate, which determined the speed with which cluster-output connections changed in response to corrective feedback. It was constrained to be within the range (0.1, 5).

SUSTAIN was run from the starting points ($r=9.012$, $\beta=1.252$, $d=16.924$, $\eta=0.092$), ($r=5.0$, $\beta=2.0$, $d=8.0$, $\eta=0.01$), ($r=1.0$, $\beta=1.0$, $d=1.0$, $\eta=0.1$), ($r=4.0$, $\beta=4.0$, $d=4.0$, $\eta=0.5$), and ($r=0.1$, $\beta=0.5$, $d=18.0$, $\eta=0.5$). The best fitting value resulted from the third simulation, ($r=1.0$, $\beta=1.0$, $d=1.0$, $\eta=0.1$), with an RMSD of 54.023 (see Table 6 for response proportions). Due to the model-fitting algorithm's tendency to crash, this value is derived from a rather small number of iterations; as such, the exact response proportions cannot be expected to serve as a good approximation to human behavioural data. Model fit values ranged widely, from 54 to 90, and the pattern of responses tended to be less consistent with SUSTAIN than with EXALT and EXIT. While I+PC+PR elicited the greatest

number of common-category responses across most parameter values, there was a higher degree of variance in the responses to PC+PR and ?+PC+PR.

REFERENCE LIST

- Albert, M.K. (2007). Mechanisms of modal and amodal interpolation. *Psychological Review*, 114, 455-469.
- Bohil, C.J., Markman, A.B., & Maddox, W.T. (2005). A feature-saliency analogue of the inverse base-rate effect. *Korean Journal of Thinking & Problem Solving*, 15, 17-28.
- D'Errico, J. (2006). fminsearchbnd (Version 2.1) [Computer software]. Rochester, NY.
- Ganzach, Y. & Krantz, D.H. (1990). The psychology of moderate prediction I: Experience with multiple determination. *Organizational Behavior & Human Decision Processes*, 47, 177-204.
- Johansen, M.K., Fouquet, N., & Shanks, D.R. (2007). Paradoxical effects of base rates and representation in category learning. *Memory & Cognition*, 35, 1365-1379.
- Johnson, J.S. & Olshausen, B.A. (2005). The recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Research*, 45, 3262-3276.
- Kellman, P.J. (2003). Interpolation processes in the visual perception of objects. *Neural Networks*, 16, 915-923.
- Kellman, P.J., Guttman, S.E., & Wickens, T.D. (2001). Geometric and neural models of object perception. In T.F. Shipley & P.J. Kellman (Ed.), *From Fragments to Objects – Segmentation and Grouping in Vision* (pp. 183-244). Philadelphia: North Holland.
- Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Kruschke, J.K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 22, 3-26.
- Kruschke, J.K. (2001). The inverse base-rate effect is not explained by eliminative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1385-1400.

- Kruschke, J.K. & Johansen, M.K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1083-1119.
- Kurtz, K.J. (2007). The divergent autoencoder (DIVA) model of category learning. *Psychonomic Bulletin and Review*, *14*, 560-576.
- Lamberts, K. & Kent, C. (2007). No evidence for rule-based processing in the inverse base-rate effect. *Memory & Cognition*, *35*, 2097-2105.
- Lee, T.S. (2003). Computations in the early visual cortex. *Journal of Physiology*, *97*, 121-139.
- Love, B.C., Medin, D.L., & Gureckis, T.M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309-322.
- Medin, D.L. & Edelson, S.M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, *117*, 68-85.
- Nagasaka, Y., Lazareva, O.F., & Wasserman, E.A. (2007). Prior experience affects amodal completion in pigeons. *Perception & Psychophysics*, *69*, 596-605.
- Otsuka, Y., Kanazawa, S., & Yamaguchi, M.K. (2006) Development of modal and amodal completion in infants. *Perception*, *35*, 1251-1264.
- Philiastides, M.G. & Sajda, P. (2005). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, *16*, 509-518.
- Plomp, G. & van Leeuwen, C. (2006). Asymmetric priming effects in visual processing of occlusion patterns. *Perception & Psychophysics*, *68*, 946-958.
- Rauschenberger, R., Liu, T., & Slotnick, S.D. (2006). Temporally unfolding neural representation of pictorial occlusion. *Psychological Science*, *17*, 358-364.
- Ross, B.H. & Murphy, G.L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 736-753.
- Scheessele, M.R. & Chaaban, I. (2008). Spatiotemporal context influences perception of an ambiguous target in visual search. *Perception & Psychophysics*, *70*, 190-198.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.
- Verguts, T., Ameel, E., & Storms, G. (2004). Measures of similarity in models of categorization. *Memory & Cognition*, *32*, 379-389.

- Winman, A., Wennerholm, P., & Juslin, P. (2003). Can attentional theory explain the inverse base-rate effect? Comment on Kruschke (2001). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1390-1395.
- Winman, A., Wennerholm, P., Juslin, P., & Shanks, D.R. (2005). Evidence for rule-based processes in the inverse base-rate effect. *The Quarterly Journal of Experimental Psychology*, *58A*, 789-815.
- White, C.M. & Koehler, D.J. (2004). Missing information in multiple-cue probability learning. *Memory & Cognition*, *32*, 1007-1018.
- Wright, W. (2008). Spore (Version 1.0) [Computer software]. Emeryville, CA: Maxis Software.
- de Wit, T.C.J., Schlooz, W.A.J.M, & Hulstijn, W. (2007). Visual completion and complexity of visual shape in children with pervasive developmental disorder. *European Child & Adolescent Psychiatry*, *16*(3), 168-177.