City Research Online

City, University of London Institutional Repository

Citation: Close, J. & Pothos, E. M. (2012). "Object Categorization: Reversals and Explanations of the Basic-Level Advantage" (Rogers & Patterson, 2007): A simplicity account. Quarterly Journal of Experimental Psychology, 65(8), pp. 1615-1632. doi: 10.1080/17470218.2012.660963

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/4700/

Link to published version: https://doi.org/10.1080/17470218.2012.660963

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

RUNNING HEAD: Reversals of the Basic-Level Advantage

"Object Categorization: Reversals and Explanations of the Basic-Level Advantage" (Rogers & Patterson, 2007): A Simplicity Account

James Close

Max Planck Institute for Evolutionary Anthropology

&

Emmanuel M. Pothos

Swansea University

Number of text pages: 40 **Number of Figures:** 3

Please address correspondence regarding this article to James Close, Max Planck Institute for Evolutionary Anthropology, Department of Developmental and Comparative Psychology, Deutscher Platz 6, D-04103 Leipzig, Germany, or to Emmanuel Pothos, Department of Psychology, Swansea University, Swansea, SA2 8PP, UK. Electronic mail may be sent at james_close@eva.mpg.de (tel.: +49 (0)341 3550 421) or e.m.pothos@swansea.ac.uk (tel.: +44 (0)1792 295606).

Abstract

T. T. Rogers and K. Patterson [Rogers, T. T., & Patterson, K. (2007). Object Categorization: Reversals and Explanations of the Basic-Level Advantage. *Journal of Experimental Psychology: General, 136,* 451-469] reported an impressive set of results demonstrating a reversal of the highly robust basic level advantage in both patients with semantic dementia and in healthy individuals engaged in a speeded categorisation task. To explain their results, as well as the usual basic level advantage seen in healthy individuals, the authors employed a parallel distributed processing theory of conceptual knowledge. In this paper, we introduce an alternative way of explaining the results of Rogers and Patterson, which is premised on a more restricted set of assumptions born from standard categorisation theory. Specifically, we provide evidence that their results can be accounted for based on the predictions of the simplicity model of unsupervised categorisation.

Keywords: Basic level; categorisation; object recognition; PDP; semantic dementia.

Rogers and Patterson (2007): a Parallel Distributed Processing account

Human categorisation can take place at a number of different levels of abstraction: people may classify a set of objects at the *superordinate level* (e.g., animal, furniture), at the *basic level* (e.g., dog, chair), and/ or at the *subordinate level* (e.g., Labrador, armchair). In a seminal paper by Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976), these authors reported a robust preference in humans for classification at the basic level, establishing that this level of abstraction has a ‘special status’ in human categorisation.

For patients with the neurological disorder *semantic dementia* (SD), however, basic level and subordinate level conceptual knowledge is found to degrade, leaving only superordinate level knowledge relatively intact (see Warrington, 1975). In an intriguing paper by Rogers and Patterson (2007), these authors first replicated the robust basic level superiority effect in a healthy population (e.g., Hoffmann & Ziessler, 1983; Murphy & Brownell, 1985; Rosch et al., 1976; Tanaka & Taylor, 1991; see also Malt, 1995, for a cross-cultural perspective) and then demonstrated a reversal of the basic level advantage in four patients with severe SD: that is, superordinate level categorisation was found to be superior to that of basic level categorisation (henceforth, we refer to this as a superordinate level > basic level advantage; see also Hodges, Graham, & Patterson, 1995; Patterson & Hodges, 2000). Rogers and Patterson (2007) explained their results based on a parallel distributed processing (PDP) theory of conceptual knowledge (see also Rogers et al., 2004; Rogers & McClelland, 2004). One novel prediction born from this explanation was that the usual basic level advantage found in healthy individuals should actually reverse under conditions of speeded categorisation. Testing this prediction using a tempo-matching experimental procedure adapted from Kello (2004), Rogers and

Patterson (2007) observed the anticipated superordinate level > basic level advantage in healthy individuals.

According to PDP theory, “knowledge about the meanings of words and objects emerges from the interactive activation of perceptual, motor, and linguistic representations across different modalities of reception and expression” (Rogers & Patterson, 2007, p. 456). It has been argued that these different kinds of sensory-motor information are coded in neuroanatomically distinct cortical regions, which converge in the anterior temporal cortex (the focus of the neuropathology in SD; e.g., Nestor, Fryer, & Hodges, 2006). That is, the anterior temporal lobes are seen to function as a kind of cross-modal “hub” for the interaction between these different types of representations. Semantic representations, then, are considered not to encode any explicit or directly interpretable content *per se*. Rather, the combination of our perceptual, motor, and linguistic representations give rise to the content of our semantic memory (Barsalou, Simmons, Barbey, & Wilson, 2003; Rogers & Patterson, 2007).

A PDP account of the basic level advantage

The similarity structure of the patterns of activation generated by any single modality of perception or expression may be considerably different from those arising after cross-modal matching (Rogers & Patterson, 2007). Akin to *differentiation theory* (see, e.g., Murphy & Brownell, 1985), Rogers and Patterson (2007) argue that exemplars that are members of the same basic level category will be represented by similar patterns of activation within the “hub” – as they will share many attribute matches across modalities – while exemplars from different basic level categories will be represented by rather different patterns of activation within the “hub” – as they will share fewer attribute matches across modalities. With respect to the different levels of categorisation, therefore, while basic level categories will correspond to relatively tight and widely separated clusters in representational space, subordinate categories will correspond to smaller and less well-separated clusters, and superordinate categories will correspond to more inclusive but sparsely populated clusters. As a direct consequence of the similarity-based generalisation that is assumed to occur during name acquisition and retrieval, a basic level categorisation advantage is predicted (Rogers & Patterson, 2007).

More specifically, members of one subordinate level cluster within a basic level category will share a very similar pattern of cross-modal activation to members of a second subordinate level cluster within the same basic level category. For example, while humans may treat a Robin as distinct from a Blue Tit, these two birds share many similarities with each other. This high degree of cross-category similarity (and associated cross-category similarity-based generalisation) will create interference, which will slow subordinate level name acquisition and retrieval. In contrast, superordinate clusters (e.g., animal) will be formed from the representations of many, relatively diverse category members (e.g., birds, cats, cows, dogs, horses, etc.), and therefore, from many relatively different patterns of cross-modal activation. Consequently, similarity-based generalisation of the associated superordinate level name will proceed poorly between category members at this level of abstraction. Again, this will create interference, which will slow superordinate level name acquisition and retrieval. At the basic level of abstraction, however, the within category similarity of cross-modal patterns of activation will be high, allowing for effective name generalisation within categories, while the between category similarity of cross-modal patterns of activation will be low, limiting cross-category generalisation and interference. As such, while superordinate level structures (e.g., animal) are assumed to be activated before basic level ones, due to the influence of similarity-based generalisation, the name of basic level structures (e.g., bird) will become fully activated more quickly. Accordingly, this situation will allow for fastest name acquisition and retrieval at the basic level, resulting in a basic level > superordinate level advantage.

A PDP account of the superordinate level advantage in patients with SD

To explain the superordinate > basic level advantage found in patients with SD, Rogers and Patterson (2007) argue that, when considering the similarity structure of semantic representations, superordinate information will simply be more robust against the brain damage associated with SD. This brain damage involves a deterioration of the anterior temporal cortex, which is the region of the brain that is assumed to represent the “hub” for cross-modal interactive activation. The reason for this, they argue, is that damage to the anterior temporal cortex will result in distortion of the patterns of activation that arise in response to the presentation of different stimuli: this distortion in a given stimulus’ pattern of activity results from the disease having “destroyed some of the neurons that coded the healthy pattern” (Rogers & Patterson, 2007, p. 460). As a consequence of these small distortions in stimulus representation, the spread of activation (i.e., generalisation) through the semantic network will become less specific, meaning that very specific properties of a stimulus will be less likely to be activated.

Taking the example used by Rogers and Patterson (2007, p. 460), what happens when a healthy network tries to name a picture of a *canary*, for example? Critically, such subordinate level categorisation requires that the hub produce an almost veridical pattern of activation of a *canary*. The reason for this is that the patterns of activation for the category *birds* will all be quite similar, but of course, the name *canary* does not apply to all birds. Given such similar representations, small distortions of the *canary* pattern will disturb the network's ability to generate the name *canary* as output. Subordinate level naming is, therefore, highly specific. Naming at the basic level, however, can tolerate the instantiation of less exact patterns of activation in the hub, as the name *bird* applies to all birds, meaning any bird-like pattern of activity will allow for the name *bird* to be generated as output. By extension, if the pattern of activation that is instantiated in the hub for a *canary* is so heavily distorted that its representation falls closer to a different species of animal, then where basic level naming will now fail, naming at the superordinate level (i.e., *animal*) will be successful. In summary, superordinate level naming in the network can tolerate the highest levels of distortion to stimulus representations. Such distortions can be brought about by SD, which causes neuronal loss in the anterior temporal cortex. When SD is severe, therefore, more specific (i.e., subordinate and basic) level naming will be heavily disrupted, leading to a superordinate level > basic level advantage.

A PDP account of the superordinate level advantage in healthy individuals engaged in speeded categorisation

As stated above, one novel prediction of Rogers and Patterson's (2007) PDP theory is that if a healthy individual were to engage in a speeded version of the standard basic level categorisation task, then a reversal of the usual basic level > superordinate level advantage should be found. This prediction sits in contrast to other theories of basic level categorisation. For example, based on the entry-level account of the basic level advantage (Jolicoeur, Gluck, & Kosslyn, 1984), the speed of categorisation should have no impact on the basic level advantage, as the first category representations to become activated by a stimulus are always those at the basic level. Equally, differentiation theory (Murphy & Brownell, 1985) offers no obvious insight into why speeded categorisation should impact on the usual basic level > superordinate level advantage; its aim is simply to explain what has been regarded as the ubiquitous basic level advantage under standard task (time) constraints (Rogers & Patterson, 2007). As noted earlier, this novel prediction of PDP theory was empirically confirmed by Rogers and Patterson (2007).

According to Rogers and Patterson, a reversal of the standard basic level > superordinate level advantage during speeded categorisation in healthy individuals occurs because the network has to produce "a "best guess" proportional to the activation of the name unit" (2007, p. 461). As noted earlier, a critical assumption of Rogers and Patterson's (2007) model is that because subordinate level structures are encompassed fully within superordinate level structures in representational space, the network will necessarily start to activate the superordinate (e.g., *animal*) region of space – and consequently the name *animal* in the output – before it starts activating the basic (e.g., *bird*) region of space – and consequently the name *bird* in the output. While subsequent activation of the name *bird* will proceed more rapidly than the name *animal* – due to similarity-based generalisation – at very short response latencies, threshold levels for naming will not be exceeded, resulting in the model having to make a "best guess". At very short response latencies, therefore, the *animal* (superordinate level) name unit will have a greater level of activation than any basic level name unit encompassed within its structure (e.g., *bird*), because the *animal* region of space was activated before the *bird* region of space. As a consequence of the greater level of activation of superordinate level names at this early stage in stimulus representation, rapid responding will lead to the network favouring a superordinate level category response, resulting in a superordinate level > basic level advantage (Rogers & Patterson, 2007).

Comments on Rogers and Patterson (2007)

Through employing a PDP theory of conceptual knowledge, Rogers and Patterson (2007) provide compelling accounts of the standard basic level advantage, the superordinate level advantage found in patients with SD and the superordinate level advantage found in healthy individuals engaged in a speeded classification task. Moreover, with regard to the PDP model presented, a number of contentious assumptions made in other spreading-activation approaches are not present. For example, while some spreading-activation networks assume that category structures are hierarchical (i.e., taxonomically organised), with activation spreading in a top-down fashion from superordinate, through basic, to subordinate levels (Collins & Loftus, 1975; Collins & Quillian, 1969), the PDP approach is not constrained in this manner (Rogers & McClelland, 2004). This is a clear strength of the PDP approach. Furthermore, the PDP model contains elements of biological plausibility, mapping its assumptions to specific areas of the brain: the "hub" (where representational information from different modalities is integrated), for example, is mapped directly to the anterior temporal lobes.

In the present paper, our goal is not to question the general validity of Rogers and Patterson's (2007) assumptions; they are all well motivated and psychologically plausible. Rather, we wish to explore the extent to which standard categorisation theory can offer an alternative way of explaining their results. Why is this important?

First, the PDP model is primarily a process model, emphasising the particular algorithms which are involved in basic level categorisation. By contrast, most categorisation theories are specified at the computational level. In general, a computational level theory may be compatible with an algorithmic level one, even if superficially the two theories are very different (indeed, eventually we would want to converge our best algorithmic theory with the best computational one; Marr, 1982). With regard to the basic level categorisation results considered here, if particular standard categorisation models can accommodate them, then it becomes meaningful to consider further their relation to the PDP account. Second, and relatedly, even though most standard categorisation models offer some account of basic level categorisation, they have not been explored in terms of the particular results of Rogers and Patterson (2007). Thus, these results present a potential challenge to standard categorisation theory. Conversely, the PDP model makes a range of assumptions, which are, as far as we can tell, all neurologically plausible. However, from a modelling point of view, it is important to ask whether these assumptions are essential in accounting for the SD and speeded categorisation results. Considering this issue will partly determine the extent to which the assumptions made by PDP theory can be justified on the basis of just the SD/ speeded categorisation results, or whether further justification needs to be sought.

For our demonstration, we selected the simplicity model of unsupervised categorisation (Pothos & Chater, 2002). To foreshadow our results, we demonstrate that the simplicity model readily predicts both the usual basic level > superordinate level advantage in healthy individuals engaged in a standard classification task, and the superordinate level > basic level advantage in patients with SD and in healthy individuals engaged in a speeded classification task.

Rogers and Patterson (2007): a simplicity approach

In explaining Rogers and Patterson's (2007) results, a promising candidate model would initially be required to offer an account of the basic level advantage under standard task constraints. While a number of models provide an explanation for this basic level advantage (e.g., Corter & Gluck, 1992; Gosselin & Schyns, 2001; Jones, 1983), the model we have chosen to employ stems from research into unsupervised categorisation, namely, the simplicity model of unsupervised categorisation (Pothos & Chater, 2002). Our reasoning for this choice was two-fold: First, the simplicity model was directly motivated from the proposal for basic level categorisation put forward by Rosch and Mervis (1975) – this is not the case with other influential models of unsupervised categorisation (e.g., SUSTAIN, Love, Medin, & Gureckis, 2004; or the Rational Model, Anderson, 1991). That is, the simplicity model is premised on the suggestion that categories should maximise within-category similarity and minimise between-category similarity (this suggestion generalises the definition for basic level categories Rosch and colleagues adopted; e.g., Rosch & Mervis, 1975). Second, the simplicity model provides a straightforward quantification of the 'intuitiveness' of different classifications. As will shortly be shown, this property of the simplicity model is particularly useful in the examination of Rogers and Patterson's (2007) findings. While other models of unsupervised categorisation can often be modified to produce estimates of category intuitiveness, this computation is particularly straightforward with the simplicity model (models of unsupervised categorisation can typically predict the classification that should be most intuitive for a set of stimuli, though predictions of category intuitiveness are more challenging; Pothos et al., 2011).

The simplicity model of unsupervised categorisation

For a given stimulus set, the simplicity model computes the *codelength* required to describe all the information in the similarity relations between the items (call this information *S*). That is, codelength(*S*) effectively corresponds to the 'raw' information in the data, before any attempt to look for category structures has been made. The second step in the simplicity model is to examine whether codelength(*S*) can be reduced by imposing categories, *C*. Categories are defined as postulating that all within category similarities are greater than all between category similarities. Therefore, if a classification for the stimuli is identified that specifies numerous and accurate *constraints* in the aforementioned manner, codelength (*S*|*C*) is going to be small. When there are incorrect constraints, the codelength required to correct these needs to be taken into account. If there are *u* constraints in total and *e* errors, then since there are ${}_uC_e = \frac{u!}{e!(u-e)!}$ ways to select *e* items out of *u*, the codelength required to correct the erroneous constraints is given by $\log_2(u+1) + \log_2\left({}_uC_e\right)$ (this function is slightly modified so that it is monotonically increasing). Finally, the chosen classification needs to be specified as well. This is done by considering all possible classifications of *r* items into *n* categories, which is given by Stirling's number, $\sum_{v=0}^{n} (-1)^v \frac{(n-v)^r}{(n-v)!\,v!}$. The contribution from this term is typically very small. To sum up, the simplicity model predicts that a classification will be psychologically obvious to the extent that codelength(*S*) – codelength (*S*|*C*) is large. In general, simplicity model predictions are typically specified as the ratio of codelength (with categories) / codelength (without categories), expressed as a percentage; therefore, the lower this percentage, the greater the 'simplification' of the code achieved by imposing a classification, and the more psychologically intuitive (obvious) the classification is predicted to be. For brevity, this percentage is referred to as 'codelength'.

Classification codelengths typically vary between 50% and 100% (as noted, lower values indicate a more psychologically intuitive classification). The computation of the different codelength terms specified above is effectively an application of the formal simplicity framework of Minimum Description Length (Rissanen, 1989). The simplicity model is run in a straightforward way: its input is the coordinates of a set of stimuli when represented in an assumed psychological space, out of which the model generates information about pairs of similarities (typically using the Euclidean metric). The model employs a search algorithm to identify the best possible classification for the set of items. The algorithm is akin to agglomerative clustering ones, which initially assume that all items belong to separate categories, and then gradually combine items to try to improve this classification. Unlike many prominent models of categorisation (whether they model supervised or unsupervised categorisation), the simplicity model is *parameter free* (for a more detailed description of the simplicity model, see Pothos & Chater, 2002).

What is the psychological intuition behind the model? If a lot of simplification can be achieved by a specific categorisation, then this means that there is a lot of classification 'structure' in the similarity relations of the stimuli (that is, there are well-separated categories). Such structure would be in the form of certain similarity relations consistently being less than other similarity relations. With real life categories, for example, a child may note that a Labrador and a Rottweiler are consistently more similar to each other than a Labrador and all kinds of cats. For the simplicity model, this is a clue that certain items should be in the same category (i.e., the Labrador and the Rottweiler, as opposed to the Labrador and a cat), with the end result being that categories are organised on the basis of exemplar similarity. This is exactly Rosch and Mervis's (1975) proposal for basic level categorisation; the simplicity model is just a formal way to computationally implement their idea.

Evidence showing that simplicity can provide a suitable framework for cognitive modelling has come from a number of sources (e.g., Chater, 1999; Feldman, 2000; Pothos & Wolff, 2006). Moreover, simplicity inference is analogous to Bayesian inference under particular choices of priors (Chater, 1996; Tenenbaum, Griffiths, & Kemp, 2006). It is important to note here, though, that category intuitiveness is potentially influenced by other considerations as well, such as consistency with general knowledge (Heit, 1997; Murphy & Allopenna, 1994; Murphy & Medin, 1985; Wisniewski, 1995). In principle, the simplicity model could be modified to take general knowledge factors into account, but, as many authors have discovered, computationally incorporating effects of general knowledge is extremely hard, if not logically impossible (see, e.g., Fodor, 1983; Heit, 1997; Lewandowsky, Roberts, & Yang, 2006; Murphy, 2002; Pickering & Chater, 1995). Therefore, in all modelling work employing the simplicity model to-date, we have examined the scope of the model in its form based solely on similarity, and this approach is adopted in the present paper.

A simplicity account of the basic level advantage

With respect to explaining the basic level advantage using the simplicity model, the situation is rather straightforward: we suggest that, given a hierarchy of classifications, the basic level will correspond to the classification that is 'most intuitive' (according to the model). Such a statement readily follows both by analogy from Rosch and Mervis's (1975) early formulation of basic level categorisation, and from the several subsequent proposals for understanding basic level categorisation (e.g., Corter & Gluck, 1992; Gosselin & Schyns, 2001; Jones, 1983).

As an example, consider the stimulus structure presented in Figure 1 (in these and all subsequent examples, we assume that both dimensions are equally salient; see Lamberts, 2002; Milton, Longmore, & Wills, 2008; Pothos & Close, 2008). The codelength for the classification predicted by the simplicity model to be most intuitive (optimal) is 68.92%; we assume this to be the basic level of categorisation. The codelength associated with a more general (superordinate) level of classification is 93.05% and the codelength associated with a more specific (subordinate) level of classification is 88.05%. Therefore, the basic level of classification is readily predicted to be most intuitive by the simplicity model.

Why is this the case? First, classification at the superordinate level is prejudiced: although the larger categories generate more constraints (that is, there are a greater number of comparisons required between both within-category and between-category similarities), the categories are not very coherent (cf. Murphy & Medin, 1985). Therefore, there are many erroneous constraints that need to be corrected. As a consequence of this, the overall codelength for classification at this level is high, reflecting the fact that using superordinate categories provides only a minor simplification in the description of the 'raw' similarity information. Classification at the subordinate level is also prejudiced: while in this case most constraints are correct, there are (relatively) few of them, as the clusters are generally smaller. Basic level categorisation, on the other hand, achieves the right balance between numerous and correct constraints; hence, classification at this level is favoured (i.e., predicted as most intuitive by the simplicity model). At the conceptual level, the simplicity formulation of the basic level advantage closely mirrors previous discussions in terms of category specificity and informativeness (e.g., Komatsu, 1992; Medin, 1986; Murphy & Brownell, 1985; Murphy & Smith, 1982).

How do the predictions of the simplicity model map onto the standard empirical finding that people are generally faster and more accurate to categorise a presented stimulus at the basic level (e.g., bird) than at either the superordinate (e.g., animal) or subordinate (e.g., Robin) levels (Rogers & Patterson, 2007; Rosch et al., 1976)? The assumption is that, given their high level of intuitiveness, the category structures that represent classification at the basic level will simply be more readily accessible than the category structures that represent classification at the superordinate and subordinate levels (which are not very intuitive): that is, superordinate and subordinate level categorisations will be cognitively *more effortful* and *time-consuming* to perceive. Consequently, under standard task constraints, people will be best able to assess a presented stimulus in terms of basic level category structures. This will mean that, in general, basic level classification will be fastest and most accurate, and this will be reflected in people's response behaviour. In terms of the predictions of the simplicity model, therefore, the smaller the codelength associated with a specific category structure, the faster and more accurate stimulus classification should proceed with respect to that structure. Note, however, an important qualification: while we can assume a general association between intuitiveness and speed of categorisation, the simplicity model is not a process model and therefore cannot make as detailed predictions concerning categorisation speed as the PDP model.

----------Figure 1 about here----------

A simplicity account of the superordinate level advantage in patients with SD

A common assumption when modelling a range of neurological disorders is that there exists an increased amount of noise within patients' encoded memory signal, and their assessment of this memory signal. Indeed, the introduction of noise in this manner has proven to be an effective tool for modelling amnesia (e.g., Berry, Shanks & Henson, 2008; Malmberg, Zeelenberg, & Shiffrin, 2004). In the present paper, we also assume that a larger degree of noise will be associated with the storage and processing of exemplars in patients with SD, relative to healthy individuals. As we will show, this increased level of *representational noise* in patients with SD leads naturally to the prediction of a superordinate level classification advantage by the simplicity model.

To recapitulate, the simplicity model predicts that the most intuitive level of categorisation for a healthy individual is the basic one. Let us assume that the brain damage associated with the anterior temporal cortex of patients with SD (Garrard & Hodges, 2000; Mummery et al., 2000; Nestor et al., 2006) results in noise being introduced into these patients' underlying stored exemplar representations. Noisy exemplar representations result in random distortion of the category structures that were in place before the onset of dementia. Therefore, if one starts with well-defined, well-separated categories and a noise signal is introduced into the corresponding exemplar representations, the resulting categories start encompassing less similar items (that is, in psychological space, categories could be seen as covering a greater surface area or volume). In other words, introducing noise into pre-existing category structures makes categories less specific, and as we will show, this can lead to a shift from favouring basic level categories to superordinate level ones. As the pathology worsens and becomes more acute, so too does the amount of noise present in SD patients' stored category representations, resulting in category structures losing specificity. Again, the loss of representational specificity would be associated with a preference for categorisation at a more general level, culminating in a superordinate level > basic level advantage in patients with severe SD (e.g., Hodges et al., 1995; Rogers, Lambon Ralph, Garrard, et al., 2004; Rogers, Lambon Ralph, Hodges, & Patterson, 2003, 2004; Warrington, 1975).

To assess the above predictions, we generated 50 random stimulus structures using MATLAB (Mathworks, 2007) that all resulted in an initial optimal classification of six clusters being identified by the simplicity model.[1] Each randomly generated stimulus structure was composed of 12 stimulus points. Computationally, six cluster stimulus structures were chosen so as to create a situation where, in principle at least, distortion of these structures could equally lead to an increase or decrease in the number of clusters that are subsequently considered optimal by the simplicity model. In contrast, if, for example, the initial optimal classifications were composed of only two clusters, then intuitively any distorted classification would be more likely to have more than two clusters – in the simple sense that there are more possible classifications with more than two clusters than ones with fewer. Of course, real life category structures would be consistent with a range of classifications at the basic level. Nonetheless, we felt that the particular choice of stimulus sets with 12 items, consistent of six cluster classifications, was a reasonable choice that balanced the realism of the simulation with the constraints of the computational procedure. Finally, the coordinates of each stimulus could vary along two arbitrary dimensions,

[1] Initially, the only substantive requirement for the stimulus structures was that their optimal classification consisted of at least five clusters prior to the application of noise. In practice, the vast majority of stimulus structures identified in this way were associated with optimal classifications of six clusters. To standardise our analyses, therefore, we chose to further analyse only those stimulus structures that were associated with a six cluster optimal classification.

from 0 – 10. The randomisation of the coordinates for each stimulus point was based on MATLAB's 'rand' function, and a simple linear transformation was employed (separately for each stimulus set) to stretch the dimensional coordinates to a range of 0 to 10.

For each of our 50 stimulus structures, we separately applied a random amount of noise to each coordinate. This random noise could vary between 0% of the coordinate range and a specified maximum noise level of 5%, 15% or 30% of the coordinate range. Noise level was manipulated over three separate simulations to explore the effect of increasing noise on stimulus classification, which under our analysis, is analogous to worsening SD. For example, for a given stimulus structure subjected to the highest level of noise, each coordinate of each stimulus point would be distorted by a number between -30% and 30% of the range of the coordinate. Following the introduction of a specified level of noise into our 50, six cluster base stimulus structures, we reassessed the optimal classification of these 'noisy' stimulus structures using the simplicity model.

The question of primary interest is whether increased noise implies that the original optimal classification becomes less specific (so that it is composed of fewer clusters), thereby leading to a shift from a basic level advantage to a superordinate level advantage. Accordingly, the predictions of the simplicity model are considered only in terms of the number of clusters that the optimal classification consists of. Our results clearly show that the application of noise leads to a shift in category intuitiveness away from the original six cluster classifications to classifications consisting of fewer clusters (see Figure 2). More specifically, following the introduction of a 5% level of noise into our base stimulus structures, the optimal classification identified by the simplicity model consisted of an average of 4.86

clusters. With increasing levels of noise (i.e., 15% and 30%), the classifications identified as optimal by the simplicity model consisted of even fewer numbers of clusters: the introduction of a 15% level of noise resulted in an average of 4.34 clusters in the optimal classification, while the introduction of a 30% level of noise resulted in an average of 3.82 clusters in the optimal classification. Relative to the original 6 cluster classifications, these reductions in the number of clusters in the classifications identified as optimal were highly significant (5% noise level: $t(49) = -7.82$, $p < .001$; 15% noise level $t(49) = -13.85$, $p < .001$; 30% noise level: $t(49) = -16.78$, $p < .001$; all these *t*-tests were single sample tests against 6).

In summary, following the introduction of noise, the simplicity model identified as optimal classifications with fewer clusters, a finding which we interpret as a bias for favouring category structures that are superordinate to the original 6 cluster classifications. Moreover, this bias to identify fewer clusters as optimal in the 'noisy' stimulus structures became more pronounced with increasing levels of noise (at least up to a maximum noise level of 30%; though note that the psychological meaning of higher degrees of noise is unclear). Overall, we interpret the above results as demonstrating a clear shift from a basic level advantage to a superordinate level advantage, as has been extensively documented in patients with SD (e.g., Hodges et al., 1995; Rogers, Lambon Ralph, Garrard, et al., 2004; Rogers, Lambon Ralph, Hodges, & Patterson, 2003, 2004; Rogers & Patterson, 2007; Warrington, 1975). It is worth noting that in a number of cases following the introduction of a 5% or 15% level of noise into a base stimulus structure, the original 6 cluster classification was still identified as optimal by the simplicity model. This finding fits well with the data provided by Rogers and Patterson (2007; Figure 2, 'More severe'), since it is not the case that accuracy for basic level categorisation was completely suppressed in their severe SD patients.

One potential criticism of the above results, however, regards whether the simplicity model is simply identifying large but strange categories as optimal following the introduction of noise into our base stimulus structures. This would contrast sharply with what superordinate categories actually represent, which are agglomerates of basic level categories. That is, if one has the basic level categories Dogs, Cats, Cars and Motorbikes, then having a superordinate category in which dogs and cars are grouped together and another superordinate category in which cats and motorbikes are grouped together is, of course, nonsensical. [2] To assess whether the simplicity model is identifying large, nonsensical categories as optimal in our 'noisy' stimulus structures, we adopted the following approach: First, we computed the similarity between the original optimal classification of a given stimulus set and its subsequent optimal classification following the introduction of noise. Second, we computed the similarity between the original optimal classification of a given stimulus set and a random classification that was composed of the same number of clusters as the optimal classification for the associated noise-distorted stimulus set. In this way, we were able to infer whether or not the optimal classifications for the 'noisy' stimulus sets were sensible and not just random, nonsensical structures. We measured the similarity between two stimulus classifications using the Rand Index (Rand, 1971). The Rand Index can be used to compare two classifications by determining the ratio of pairs of stimulus items that are both in different clusters, or both in the same cluster, in the two classifications, divided by all pairs of stimulus

[2] We thank Tim Rogers for noting this possibility.

items. A Rand Index of 0 denotes totally different classifications, whereas a Rand Index of 1 denotes identical classifications.

Overall, a higher degree of Rand similarity was found between the original optimal classifications of our base stimulus sets and the optimal classifications of our 'noisy' stimulus sets, than between the original optimal classifications and the random (as described above) classifications (see Table 1). Specifically, following the introduction of a 5% level of noise, the average Rand similarity between the optimal classifications for the original stimulus sets and the optimal classifications for the 'noisy' stimulus sets was 0.88. In contrast, the average Rand similarity between the optimal classifications for the original stimulus sets and the random classifications was 0.66 (this difference was significant, $t(49) = 17.75$, $p < .001$). A similar pattern of results was obtained with increasing levels of noise: following the introduction of the 15% level of noise, the corresponding Rand similarities were 0.81 and 0.64, respectively, ($t(49) = 12.81$, $p < .001$), and following the 30% level of noise, they were 0.72 and 0.62, respectively, ($t(49) = 8.43$, $p < .001$).

----------Table 1 about here----------

Therefore, the superordinate classifications identified by the simplicity model in our 'noisy' stimulus sets appear to be sensible generalisations, sharing a high degree of classification similarity with the classifications identified as optimal in the original stimulus sets. Given the similarity-based nature of the simplicity model, this makes sense: categories that are similar in kind (e.g., are composed of, say, mammals) will be situated closer together in representational space than categories that are not similar in kind (e.g., mammals versus automobiles). The probability that it will be basic level categories of a similar kind that will (partly) combine together following the introduction of noise is, therefore, much greater.

----------Figure 2 about here----------

A simplicity account of the superordinate level advantage in healthy individuals engaged in speeded categorisation

In the previous section, we showed how the introduction of noise into stored stimulus representations can lead to a superordinate level categorisation advantage. If we accept this explanation for superordinate level classification superiority in patients with SD, one can reasonably ask what the role of noise would be in a speeded categorisation task. A critic might claim that, for healthy individuals, internal representations will always be wholly intact (indeed, we would agree with this position) and that a speeded categorisation task could not undermine this intactness. However, speeded processing can influence exemplar representation in a different way. Specifically, when engaged in a speeded categorisation task, we propose that exemplar representations will be distorted at the time of initial processing, and that this distortion will *produce the same effect* on classification as the noise present in SD patients' stored stimulus representations.

In what way will a speeded categorisation task lead to distortion in healthy individuals' internal representations? A number of authors have proposed feature-sampling to be a key aspect of the time-course of categorisation (e.g., Lamberts, 1995, 1998, 2000, 2002; Lamberts & Freeman, 1999; see also Nosofsky & Palmeri, 1997). Intuitively, if the sampling of the features of a presented stimulus is interrupted, then only a reduced representation of that stimulus will be available for any further processing. Theorists examining the time course of categorisation have proposed three components to this process: First, features of a stimulus must be processed and then integrated into a coherent representation; second, this stimulus representation must be assessed in relation to retrieved information about categories and category members; and finally, category membership must be decided (Lamberts, 2002). There are slight variations on this theme; for example, some authors have argued for sequential ordering effects when assessing the dimensions of a stimulus during a categorisation task (i.e., processing dimension 1, followed by dimension 2, followed by dimension 3, etc.; see Schyns, Petro, & Smith, 2007). The key insight, though, is that when there are constraints which limit an observer's ability to fully assess a stimulus, sampling of the stimulus' dimensions or features will be interrupted, and a reduced stimulus representation will be formed. This may simply mean a reduction in the number of dimensions sampled, or equally a sampling-related bias that reflects the weighting of the stimulus' dimensions (e.g., Lamberts, 1995). Such possibilities have been observed for categorisation judgments at RT latencies which are consistent with those considered in Rogers and Patterson's (2007) Experiment 3 'Fast' condition (i.e., around 450 ms; e.g., Lamberts, 1995).

In summary, there is extensive theoretical and empirical work suggesting that under speeded categorisation conditions the representations formed of stimuli are reduced. Accordingly, when a stored stimulus is invoked for comparison with a presented one, this comparison will presumably have to take place in terms of the reduced dimensionality of the presented one. For example, for items normally represented in N dimensions in psychological space, a speeded categorisation task might lead to a new stimulus being represented in $M < N$ dimensions. As a consequence of this dimensional impoverishment, the comparison of the new stimulus with stored stimuli would have to take place with respect to the reduced *M-dimensional* subspace of the full psychological space, resulting in distortion of stimulus representations. We propose that, as a result of this process, corresponding categorisations would become less specific, with an associated shift from a basic level advantage to a superordinate level one. While the cognitive processes outlined here may be viewed as qualitatively different from those outlined in the previous section, at the root of each account is a distortion of stimulus representations (of course, where the distortion process comes from differs between the two approaches).

The simplicity model can be utilised in conjunction with the assumption of reduced dimensionality to explore the plausibility of our proposal: We randomly constructed a new set of initial, base stimulus structures in nearly the same way as that described above. Specifically, we once again requested that a 6 cluster classification should be identified as optimal by the simplicity model in our initial, base stimulus structures (each composed of 12 stimuli). However, while in our previous demonstration the stimuli were composed of two dimensions of variation, in the present case, the stimuli were composed of 10 dimensions of variation, each ranging from 0 – 10. The reason for this is that, as we are now interested in examining classification predictions within dimensionally reduced representations, the initial stimulus structures needed to provide the scope for dimensional loss.

In the present demonstration, we assessed the impact of eliminating 1, 3, 5, and 7 of the 10 initial dimensions of variation on the optimal categorisation. For each set of analyses, we randomly generated 50, 10 dimensional stimulus structures and specified a ‘knock-out’ level. This ‘knock-out’ level specified the number of dimensions of variation that were to be eliminated. For example, a ‘knock-out’ level of 3 would mean that only 7 of the original 10 dimensions of variation would be available for categorising the stimulus set of reduced dimensionality. The manner in which the dimensions of variation were selected for elimination was determined randomly.

Overall, a reduction in the number of dimensions from which the stimuli were represented resulted in the simplicity model identifying optimal classifications with fewer clusters than previously identified when all 10 dimensions were available. Moreover, this pattern became more pronounced as the stimulus structures became more dimensionally impoverished. At a 'knock-out' level of 1, the optimal classification consisted of an average number of clusters of 5.16; at a 'knock-out' level of 3, 4.64 clusters; at a 'knock-out' level of 5, 4.52 clusters; and at a 'knock-out' level of 7, 4.10 clusters (see Figure 3). The number of clusters identified as making up the optimal classifications of the dimensionally reduced representations was consistently significantly lower than the number of clusters identified as making up the optimal classifications of the initial stimulus structures (smallest $t(49) = -6.86$, $p < .001$; all single sample *t*-tests against 6). One interesting point to note here is that out of the 200 simulations that made up our modelling work at all levels of dimensional 'knock-out', we found only two instances of an increase in the number of clusters in the classification identified as optimal by the simplicity model (in these two cases, a classification of 7 clusters was identified as optimal). It is clear, therefore, that when a stimulus representation is impoverished, the subsequent optimal classification identified by the simplicity model will typically consist of fewer clusters, relative to the optimal classification identified by the simplicity model when the same stimuli are fully represented.

----------Figure 3 about here----------

As for the ‘noisy’ stimulus structures in the previous section, we further assessed whether the optimal classifications identified for the impoverished stimulus structures ‘made sense’. To do so, we again computed the Rand similarity between the optimal classifications for the original stimulus structures and the corresponding reduced dimensionality ones. Furthermore, we also computed the Rand similarity between the optimal classifications for the original stimulus structures and random classifications, which contained the same number of clusters as the optimal classifications for the corresponding reduced dimensionality stimulus sets. At all levels of ‘knock-out’, the Rand similarity between the optimal classifications for the original stimulus structures and the optimal classifications for the corresponding reduced dimensionality stimulus structures was significantly higher than the Rand similarity between the original classifications and the random (as described above) ones (smallest $t(49) = 4.81$, $p < .001$; all *t*-tests were paired samples *t*-tests; see Table 1). As one would expect, as more dimensions are eliminated, there is a monotonic decrease in the similarity between the optimal classifications for the original stimulus structures and the optimal classifications for the dimensionally impoverished stimulus structures. Critically though, Rand similarity was high at all levels of ‘knock out’, suggesting that these more general (superordinate) classifications were consistent with the original, more specific (subordinate) ones.

In summary, the present results show that when the representation of a given stimulus structure is impoverished due to a reduction in the number of dimensions being sampled, this leads to the identification of optimal classifications that are superordinate to the classifications identified with complete representations. Our demonstration is based on well-established accounts of the time course of object categorisation and the influence of interruption in feature-sampling during speeded perception (e.g., Lamberts, 1995, 1998, 2000). In conclusion, the somewhat counterintuitive finding that healthy individuals show a superordinate level > basic level advantage during speeded categorisation (Rogers & Patterson, 2007) can be predicted by the simplicity model with straightforward assumptions concerning the influence of time constraints in feature-sampling during speeded categorisation (see Lamberts, 2002).

----------Table 2 about here----------

General Discussion

The purpose of the present paper was to assess whether the findings of Rogers and Patterson (2007) could be explained using a standard, computational-level model of categorisation. Specifically, we utilised the simplicity model of Pothos and Chater (2002; Pothos & Close, 2008) to show that there is a shift to more general (superordinate) categories under conditions that we propose simulate SD and speeded categorisation. In both cases, this superordinate level shift was assumed to result from the distortion of an individual's original stored stimulus representations. However, the manner in which this distortion came about was different: in SD patients, distortion was assumed to result from noisy stimulus representations, in terms of random perturbations in the position of stimulus items in psychological space. In healthy individuals observing stimuli under speeded conditions, distortion was manifested in terms of reducing the dimensionality of the stimulus representations, relative to those representations under conditions of unlimited observation time.

In explaining Rogers and Patterson's (2007) data, a number of similarities can be identified between the PDP account and the simplicity account. For example, both accounts assume that categories should maximise within-category similarity and minimise between-category similarity, and that basic level categories provide the optimal balance between these two factors. Moreover, in the case of speeded categorisation, both the PDP account and the simplicity model similarly assume that speeded categorisation results in classifications based on an impoverished/ incompletely-specified stimulus representation. While noting these similarities, the account we offer for Rogers and Patterson's (2007) findings – in terms of the simplicity framework – allows a consideration of the assumptions made in their modelling approach. First, the simplicity model requires no assumption to be made about how different modal elements of a stimulus are combined within the human brain, but this is a major component of Rogers and Patterson's (2007) model. While the assumptions Rogers and Patterson (2007) make appear highly plausible, the question we raised here is whether they are essential to explain the superordinate level > basic level advantage in SD patients, and in healthy individuals engaged in speeded categorisation. Second, the PDP model of Rogers and Patterson (2007) involves various free parameters or architectural choices (e.g., activation levels, or the way such activation generalises to other, proximal concepts). The simplicity model account we provide is parameter free and is based on only two assumptions: 1) that stimulus exemplars are stored in representational space; and 2) that representations can be distorted, either by introducing noise or by reducing their dimensionality.

The fact that the simplicity model can reproduce the two key empirical findings of Rogers and Patterson (2007) suggests there is room to examine the possible convergence between standard, computational-level categorisation theory and the PDP process model of Rogers and Patterson (2007). Note that the simplicity model is not a process model and so its coverage of the data is not as detailed as that of the PDP model. As presently implemented, when given a specific classification, the simplicity model provides no ready means for assessing the probability of generating a yes/no response in a categorisation experiment. This contrasts with the PDP model which has been designed to readily provide a measure of participants' likelihood of producing yes/no responses, given a specific classification scheme. In Rogers and Patterson's (2007) paper, for example, *d*' scores are reported as a measure of the likelihood of discriminating between category members and distractors. One further important weakness of the simplicity model (and of standard categorisation theory in general) is that it links poorly with the known pathology of SD; in other words, there are no assumptions in the simplicity model that can associate increased distortion due to noise with damage of the anterior temporal lobe. Of course, a key feature of Rogers and Patterson's (2007) model is its specification in terms of a proposal for the neuroscience of categorisation.

Within the categorisation literature, only a handful of computational models have risen to the challenge of including assumptions about the processes that underlie categorisation at the neural level (e.g., Ashby & Maddox, 2005; Ashby, Alfonso-Reese, Turken, & Waldron, 1998). However, such models are typically models of supervised categorisation and so can offer relatively little insight regarding basic level categorisation. Very provisionally, the process of spontaneous categorisation possibly reflects the spontaneous reorganisation of perceptual information in later visual areas (such as the lateral occipital cortex; Op de Beek, Torfs, & Wagemans, 2008). If so, this raises the question of how the particular pathology of SD (i.e., damage to the anterior temporal cortex) can affect the categorisation process. Exploring this possibility will clearly require extensive additional research. Our general point is that by appreciating that the superordinate level > basic level advantage can be understood with respect to a standard model of categorisation, new possibilities can emerge for understanding the corresponding neuroscience of SD deficits as well.

Experimentally, the present application of the simplicity model has a number of intriguing implications. For example, according to the simplicity model, not all superordinate level category structures should be resistant to the pathology of SD. Specifically, those superordinate level category structures that contain a large amount of variability, and are therefore not especially intuitive from the outset, should be less resistant to SD than those superordinate level category structures that are associated with little variability, and are therefore rather intuitive from the outset. With future research we hope to address this issue, noting, however, the difficulty of quantifying the intuitiveness of real-life categories.

In conclusion, we have shown that a standard model of categorisation, the simplicity model, which is theoretically more restricted than PDP theory, predicts the main findings of Rogers and Patterson (2007). It is important to note, however, that the empirical coverage that the simplicity model provides is not as detailed as that of the PDP model. The account we offer is premised on the notion of category intuitiveness, which is embodied in the simplicity model (and most models of unsupervised categorisation), and on the way distorting a representation (either by noise or by reducing dimensionality) can disrupt this intuitiveness. Our results may motivate a possible re-examination of which assumptions in the PDP model are strictly necessary for explaining the superordinate level > basic level advantage when it occurs. More generally, our findings provide some novel perspectives on the neuroscience of SD.

Acknowledgements

This research was partly supported by a School of Psychology Studentship, Cardiff University. We would like to thank Tim Rogers, Ulrike Hahn and two anonymous reviewers for their thoughtful and constructive criticism of earlier versions of this paper.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409-426. doi: 10.1037/0033-295X.98.3.409

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.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149-178. doi: 10.1146/annurev.psych.56.091103.070217

Barsalou, L. W., Simmons, W. K., Barbey, A., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences, 7,* 84-91. doi: 10.1016/S1364-6613(02)00029-3

Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008). A single-system account of the relationship between priming, recognition, and fluency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 97-111. doi: 10.1037/0178-7393.34.1.97

Chater, N. (1996). Reconciling Simplicity and Likelihood Principles in Perceptual Organization. *Psychological Review, 103,* 566-581.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology, 52A,* 273-302. doi: 10.1080/027249899391070

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Corter, J. E., & Gluck, M. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin, 111,* 291-303. doi: 10.1037/0033-2909.111.2.291

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407,* 630-633. doi: 10.1038/35036586

Fodor, J.A. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Garrard, P., & Hodges, J. R. (2000). Semantic dementia: Clinical, radiological, and pathological perspectives. *Journal of Neurology, 247,* 409-422. doi: 10.1007/s004150070169

Gosselin, F., & Schyns, P. G. (2001). Why Do We SLIP to the Basic Level? Computational Constraints and Their Implementation. *Psychological Review, 108,* 735-758. doi: 10.1037/0033-295X.108.4.735

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts, & D. Shanks (Eds.), Knowledge, concepts, and categories (pp. 7-41). London: Psychology Press.

Hines, P., Pothos, E. M., & Chater, N. (2007). A non-parametric approach to simplicity clustering. *Applied Artificial Intelligence, 21,* 729-752. doi: 10.1080/08839510701526541

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organization of semantic memory. *Memory, 3,* 463-495.

Hoffmann, J., & Ziessler, C. (1983). Objectidentifikation in kunstlichen begriffshierarchien [Object identification in artificial concept hierarchies]. *Zeitschrift fur Psychologie, 194,* 135-167.

Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16,* 243-275. doi: 10.1016/0010-0285(84)90009-4

Jones, G. V. (1983). Identifying basic categories. *Psychological Bulletin, 94,* 423-428. doi: 10.1037/0033-2909.94.3.423

Kello, C. (2004). Control over the time-course of cognition in the tempo-naming task. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 942-955. doi: 10.1037/0096-1523.30.5.942

Komatsu, L.K. (1992). Recent Views of Conceptual Structure. *Psychological Bulletin, 112,* 500-526. doi: 10.1037/0033-2909.112.3.500

Lamberts, K. (1995). Categorization Under Time Pressure. *Journal of Experimental Psychology: General, 124,* 161-180. doi: 10.1037/0096-3445.124.2.161

Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 695-711. doi: 10.1037/0278-7393.24.3.695

Lamberts, K. (2000). Information accumulation theory of categorization response times. *Psychological Review, 107,* 227-260. doi: 10.1037/TO33-295X.107.2.227

Lamberts, K. (2002). Feature sampling in categorization and recognition of objects. *The Quarterly Journal of Experimental Psychology, 55A,* 141-154. doi: 10.1080/02724980143000208

Lamberts, K., & Freeman, R. P. J. (1999). Building object representations from parts: Tests of a stochastic sampling model. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 904-926. doi: 10.1037/0096-1523.25.4.904

Lewandowsky, S., Roberts, L., & Yang, L. (2006). Knowledge partitioning in categorisation: boundary conditions. *Memory & Cognition, 34,* 1676-1688.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review, 111,* 309-332. doi: 10.1037/0033-295X.111.2.309

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the Noise or Turning Down the Volume? On the Nature of the Impairment of Episodic Recognition Memory by Midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 540-549. doi: 10.1037/0278-7393.30.2.540

Malt, B. C. (1995). Category Coherence in Cross-Cultural Perspective. *Cognitive Psychology, 29,* 85-148. doi: 10.1006/cogp.1995.1013

Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information. San Francisco: W. H. Freeman.

Medin, D.L. (1986). Commentary on “Memory storage and retrieval processes in category learning”. *Journal of Experimental Psychology: General, 115,* 373-381. doi: 10.1037/0096-3445.115.4.373

Milton, F., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 676-692. doi: 10.1037/0096-1523.34.3.676

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic dementia. *Annals of Neurology, 47,* 36-45. doi: 10.1002/1531-8249(200001)47:1<36::AID-ANA8>3.0.CO;2-L

Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Murphy, G.L., & Allopenna, P.D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 904-919. doi: 10.1037/0278-7393.20.4.904

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 70-84. doi: 10.1037/0278-7393.11.1.70

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316. doi: 10.1037/0033-295X.92.3.289

Murphy, G.L., & Smith, E.E. (1982). Basic-level superiority in picture categorisation. *Journal of Verbal Learning and Verbal Behavior, 21,* 1-20. doi: 10.1016/S0022-5371(82)90412-1

Nestor, P. J., Fryer, T. D., & Hodges, J. R. (2006). Declarative memory impairments in Alzheimer's disease and semantic dementia. *NeuroImage, 30,* 1010-1020. doi: 10.1016/j.neuroimage.2005.10.008

Nosofsky, R. M., & Palmeri, T. J. (1997). An Exemplar-Based Random Walk Model of Speeded Classification. *Psychological Review, 104,* 266-300. doi: 10.1037/0033-295X.104.2.266

Op de Beeck, H., Torfs, K., & Wagemans, J. (2008). Perceived Shape Similarity among Unfamiliar Objects and the Organization of the Human Object Vision Pathway. *The Journal of Neuroscience, 28,* 10111-10123. doi: 10.1523/JNEUROSCI.2511-08.2008

Patterson, K., & Hodges, J. (2000). Semantic dementia: One window on the structure and organization of semantic memory. In J. Cermak (Ed.), *Handbook of neuropsychology: Vol. 2. Memory and its disorders* (pp. 313-333). Amsterdam: Elsevier Science.

Pickering, M., & Chater, N. (1995). Why cognitive science is not formalized folk Psychology. *Minds and Machines, 5,* 309-337. doi: 10.1007/BF00974748

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26,* 303-343. doi: 10.1016/S0364-0213(02)00064-2

Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *Quarterly Journal of Experimental Psychology, 58A,* 733-752. doi: 10.1080/02724980443000322

Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition, 107,* 581-602. doi: 10.1016/j.cognition.2007.11.007

Pothos, E. M., & Wolff, J. G. (2006). The Simplicity and Power model for inductive inference. *Artificial Intelligence Review, 26,* 211-225. doi: 10.1007/s10462-007-9058-x

Pothos, E. M., Perlman, A., Bailey, T. M., Kurtz, K., Edwards, D. J., Hines, P., & McDonnell, J. V. (2011). Measuring category intuitiveness in unconstrained categorization tasks. Cognition, 121, 83-100.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association, 66,* 846-850.

Rissanen, J. (1989). Stochastic complexity and statistical inquiry. Singapore: World Scientific.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). The structure and deterioration of semantic memory: A computational and neuropsychological investigation. *Psychological Review, 111,* 205-235. doi: 10.1037/0033-295X.111.1.205

Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2003). Object recognition under semantic impairment: The effects of conceptual regularities on perceptual decisions. *Language and Cognitive Processes, 18,* 625-662. doi: 10.1080/01690960344000053

Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2004). Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology, 21,* 331-352. doi: 10.1080/02643290342000366

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

Rogers, T. T., & Patterson, K. (2007). Object Categorization: Reversals and Explanations of the Basic-Level Advantage. *Journal of Experimental Psychology: General, 136,* 451-469. doi: 10.1037/0096-3445.136.3.451

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605. doi: 10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439. doi: 10.1016/0010-0285(76)90013-X

Schyns, P. G. (1991). A Modular Neural Network Model of Concept Acquisition. *Cognitive Science, 15,* 461-508. doi: 10.1207/s15516709cog1504_1

Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of Visual Information Integration in the Brain for Categorizing Facial Expressions. *Current Biology, 17,* 1580-1585. doi: 10.1016/j.cub.2007.08.048

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder. *Cognitive Psychology, 23,* 457-482. doi: 10.1016/0010-0285(91)90016-H

Tenenbaum, J.B., Griffiths, T.L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10,* 309-318. doi: 10.1016/j.tics.2006.05.009

Warrington, E. K. (1975). Selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology, 27,* 635-657. doi: 10.1080/14640747508400525

Wichmann, B. A., & Hill, I. D. (1982). An Efficient and Portable Pseudo-random Number Generator. *Applied Statistics, 31*, 188-190.

Wisniewski, E.J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 449-468. doi: 10.1037/0278-7393.21.2.449

Figure Captions

Figure 1. A stimulus structure for which the simplicity model predicts a 5 cluster classification as most intuitive when classification is undertaken assuming the stimuli are represented in terms of both dimensions, and without noise. This classification represents the basic level, with a codelength of 68.92%.

Figure 2. Overall clustering results following the introduction of different levels of noise into the original, base stimulus structures.

Figure 3. Overall clustering results following different levels of dimensional ‘knock-out’ to the original, base stimulus structures.

Figures

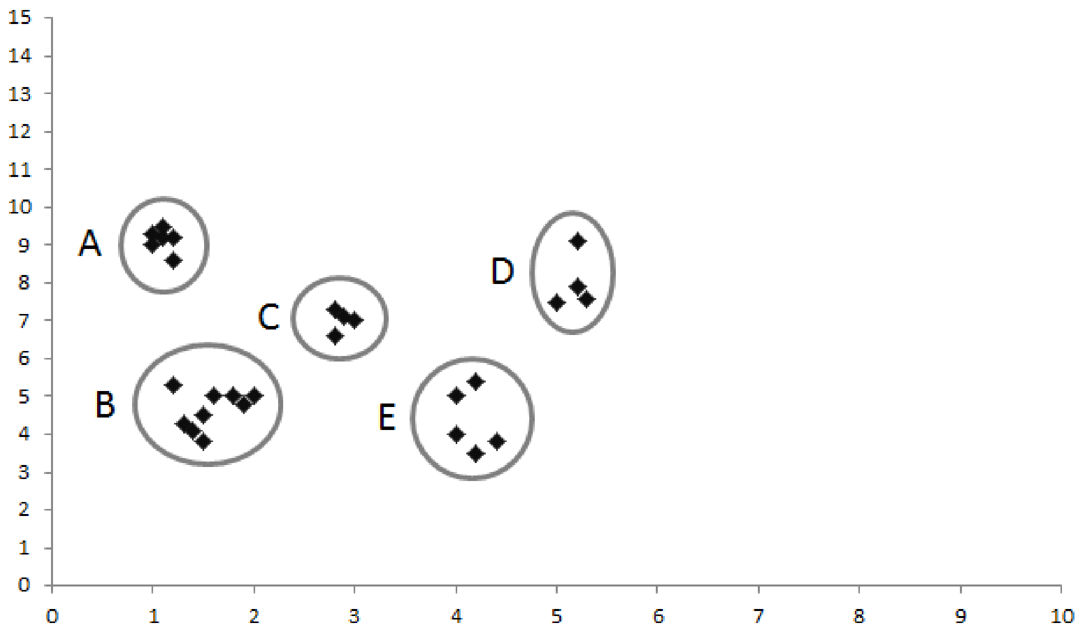

Figure 1.

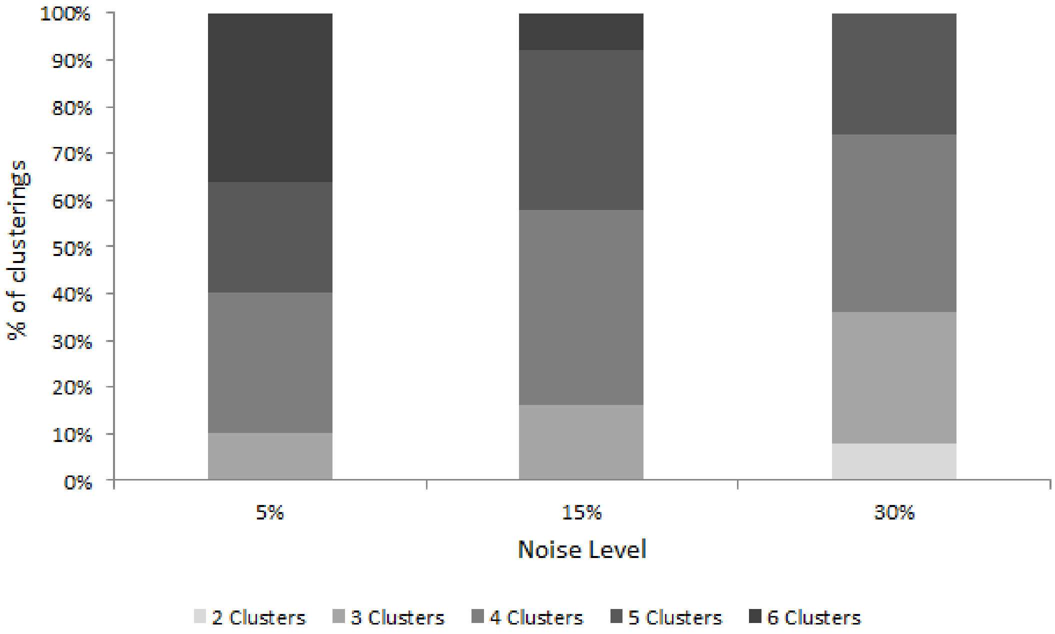
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
% of clusterings
5%
15%
30%
Noise Level
2 Clusters
3 Clusters
4 Clusters
5 Clusters
6 Clusters

Figure 2.

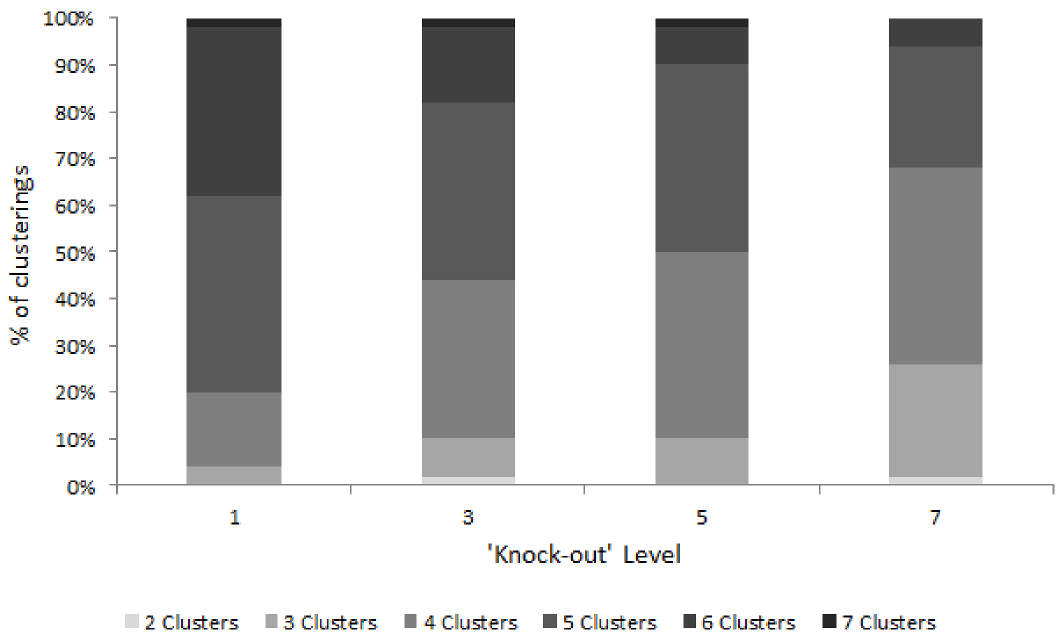
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
% of clusterings
1
3
5
7
'Knock-out' Level
2 Clusters
3 Clusters
4 Clusters
5 Clusters
6 Clusters
7 Clusters

Figure 3.

Table 1
Rand similarities between the optimal classifications of the original stimulus structures and the corresponding noisy ones and random ones.

	Noise Level		
	5%	15%	30%
Noisy Structure	0.88	0.81	0.72
Random Structure	0.66	0.63	0.61

Table 2

Rand similarities between the optimal classifications of the original stimulus structures and the corresponding dimensionally impoverished ones and random ones.

	'Knock-out' Level			
	1	3	5	7
Impoverished Structure	0.88	0.81	0.78	0.73
Random Structure	0.69	0.67	0.66	0.67