Knowledge Selection in Category-Based Inductive Reasoning

CRISP-BRIGHT, AIMEE,KAY

Knowledge Selection in Category-Based Inductive Reasoning

Aimée Kay Crisp-Bright

A thesis submitted to Durham University for the degree of

Doctor of Philosophy

2010

Durham University

Department of Psychology

Science Site

Durham, UK

Aimée Kay Crisp-Bright

Thesis Title: Knowledge Selection in Category-Based Induction

Abstract

Current theories of category-based inductive reasoning can be distinguished by the emphasis they place on structured and unstructured knowledge. Theories which draw on unstructured knowledge focus on associative strength, or temporal and spatial contiguity between categories. In contrast, accounts which draw on structured knowledge make reference to the underlying theoretical frameworks which relate categories to one another, such as causal or taxonomic relationships. In this thesis, it is argued that this apparent dichotomy can be resolved if one ascribes different processing characteristics to these two types of knowledge. That is, unstructured knowledge influences inductive reasoning effortlessly and relatively automatically, whereas the use of structured knowledge requires effort and the availability of cognitive resources. Understanding these diverging processes illuminates how background knowledge is selected during the inference process.

The thesis demonstrates that structured and unstructured knowledge are dissociable and influence reasoning in line with their unique processing characteristics. Using secondary task and speeded response paradigms, it shows that unstructured knowledge is most influential when people are cognitively burdened or forced to respond fast, whereas they can draw on more elaborate structured knowledge if they are not cognitively compromised. This is especially evident for the causal asymmetry effect, in which people make stronger inferences from cause to effect categories, than vice versa. This Bayesian normative effect disappears when people have to contend with a secondary task or respond under time pressure.

The next experiments demonstrate that this dissociation between structured and unstructured knowledge is also evident for a more naturalistic inductive reasoning paradigm in which people generate their own inferences.

In the final experiments, it is shown how the selection of appropriate knowledge ties in with more domain-general processes, and especially inhibitory control. When responses based on structured and unstructured knowledge conflict, people's ability to reason based on appropriate structured knowledge depends upon having relevant background knowledge and on their ability to inhibit the lure from inappropriate unstructured knowledge.

The thesis concludes with a discussion of how the concepts of structured and unstructured knowledge illuminate the processes underlying knowledge selection for category-based inductive reasoning. It also looks at the implications the findings have for different theories of category-based induction, and for our understanding of human reasoning processes more generally.

List of Figures

List of Tables

Declaration

None of the data or material contained in this thesis has been submitted for previous or simultaneous consideration for a degree in this or any other university.

Work from this thesis has resulted in the following publications:

Crisp-Bright, A.K. and Feeney, A. (2010). The Effects of Domain and Type of Knowledge on Category-Based Inductive Reasoning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 33nd Annual Conference of the Cognitive Science Society* (pp. 67-73). Portland, Oregon: Cognitive Science Society.

Statement of Copyright

Acknowledgements

I would like to thank my supervisors Aidan Feeney, who continued to supervise me from a distance in Belfast, and David Over, who provided support and advice here at Durham.

Love and special thanks to my wonderful husband Daniel, for his untiring support and for putting up with the inevitable chaos that characterizes a postgraduate's household!

Love and thanks also to my lovely mum, Sally, for looking after my four-legged "children" and my Dad, Stephen, for the stimulating discussions and suggestions that enhanced the quality of my thesis. Without these special people this thesis would have never happened.

The research was funded by the Economic and Social Research Council.

Chapter I

Introduction

The human ability to generalize from past experience is extraordinary. Imagine a friend tries fresh prawns for the first time, and becomes ill after eating them. He knows that they were fresh, so his conclusion is most likely to be that prawns contain a substance which does not agree with him. Consequently, in future he might avoid meals containing shrimps, lobster or crab, reasoning that if prawns made him ill, it is quite likely that other shellfish would have the same effect. In contrast, imagine that the prawns which made him ill had passed their expiration date. In this case, he might avoid eating chicken, beef or vegetables that are past their expiration date, but he might be quite willing to eat fresh prawns in future. Although in both cases he ate prawns which made him ill, based on his knowledge about contextual factors, the inferences he draws will result in quite different behaviour.

Such everyday examples illustrate the crucial role knowledge plays in inductive reasoning. Yet what is the nature of this background knowledge and which aspects are important in influencing inductive reasoning? What cognitive processes are involved in the use and selection of background knowledge? In this thesis, we endeavour to provide a comprehensive insight into how background knowledge shapes our category-based inferences.

The thesis proceeds as follows: it begins with a review of early attempts to explain inductive inferences in knowledge-rich environments. It then focuses more specifically on one class of inductive inferences, category-based induction. It outlines several theories and frameworks of category-based inductive reasoning and evaluates how they conceptualize knowledge. In doing so, it motivates one of the central proposals, arguing that there is an implicit theoretical controversy regarding the emphasis placed on two different types of background knowledge. This is followed by an attempt to clearly delineate these two contrasting types of knowledge. The thesis then reviews a more general framework of reasoning, dual process theories, and explores whether ascribing contrasting processing characteristics to different types of knowledge in category-based induction can offer a solution to the aforementioned theoretical dichotomy. The final section outlines the precise questions explored throughout the experimental chapters.

1.1 Early Approaches to the Role of Knowledge in Inductive Reasoning

Inductive inferences come in many forms. Inferences can be specific, that is, we use one experience to infer something about another particular instance, or general, that is, we abstract a general rule about the state of the world following a particular observation. For example, having cut yourself with a sharp knife, you might infer that the other knife in your drawer could have similarly hazardous properties and that this is probably true of all sharp knives. Furthermore, inferences can operate bi-directionally. That is, inferences can be tools for making predictions about future events or diagnoses about past incidences. What all inductive inferences have in common though is that they go beyond our immediate experience, elaborating our semantic knowledge beyond the directly observable.

The simplicity with which we base our actions on inductive inferences conceals the importance of our ability to learn from past experience and apply this knowledge to novel situations. In contrast to the ease with which we make inductive inferences, scientific enquiry is still struggling to come to a consensus on what inductive reasoning is, how it can be logically justified and what mental processes underlie induction.

This struggle can be traced back to some of the great philosophical thinkers such as Hume, who was one of the pioneers in making the "problem" of induction explicit (Howson, 2000). In his view, induction is one of the great riddles of thinking because it cannot be logically justified (e.g. Hume, 1748 in *Enquiry*, section 4, part 2). However, this is also the major strength of our ability to reason inductively: it is hugely flexible, allowing us to abstract any number of general rules from specific observations. Whilst we cannot prove a conclusion derived inductively to be true, the conclusions we draw greatly reduce the uncertainty in our environment, making the world a more predictable and thus controllable place. In fact, Bayesian theorists (Chater & Oaksford, 2007; Oaksford & Hahn, 2007) argue that most reasoning is inductive in nature, providing a pertinent tool for facilitating adaptive behaviour and acquiring new knowledge in an inherently probabilistic world.

Early philosophical approaches raised many questions about the function and nature of knowledge in inductive reasoning. However, one of the most critical questions is what characteristics we are willing to generalize. Many have noted that some features seem to have more inductive potency than others. For example, when making inferences about the characteristics of blackbirds from observing a blackbird in our garden, the property '*has an orange beak'* is intuitively more projectable than *'has dirt stuck to its beak'*. According to Goodman (1955), projectable properties tend to be those that support law-like hypotheses. We are likely to observe many instances of blackbirds which have orange beaks, but many instances of blackbirds that do not have dirt stuck to their beaks. This makes the former a coherent hypothesis, whilst stripping the latter of its law-like essence. We are far more likely to make projections of the first kind, than of the second kind. In Goodman's own words, based on such successful past projections, having an orange beak is a far more *entrenched* predicate than having a dirty beak.

Quine (1977) offered a slightly different answer to the question of which predicates are projectable. In his view, predicates of entities that can be grouped into similar kinds tend to be projectable. This leads to two questions, firstly what aspects of similarity lead to grouping into kinds and how well do these subjective kinds match up to the objective structure of the world? Quine (1977) appeals to a Darwinian justification: strategies for grouping cases into kinds would be reinforced if the alignment of entities along certain similarity relations offered an evolutionary advantage. Note how this foreshadows the later emergence of two apparently dichotomous traditions of explaining induction. One approach places emphasis on the actual similarity and association between instances, whereas the other highlights the importance of peoples' knowledge about the structure that results from grouping similar kinds, such as category membership. As we will see later, this divergent emphasis on different types of knowledge has been especially apparent in early models of one specific type of induction, namely category-based reasoning (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sloman, 1993b). This split is also implicit in the majority of later frameworks and models dealing with the effects of knowledge on category-based inductions (Kemp & Tenenbaum, 2009; Rogers & McClelland, 2004; Shafto, Kemp, Bonawitz, Coley, & Tenenbaum, 2008; Sloutsky & Fisher, 2004a; Tenenbaum, Griffiths, & Kemp, 2006).

1.2 Category-Based Induction

Category-based inductions cover a class of inferences in which category membership of an object supports people's inferences about properties shared by other category members. For example, classifying an animal as a rabbit allows us to infer that it probably lives in a burrow. Furthermore, if we observe that the animal we have classified as a rabbit eats carrots, we are likely to infer that other rabbits also eat carrots.

Rips (1975) was the first to systematically investigate how category membership and similarity between entities shaped peoples' patterns of inductive reasoning. After establishing the dimensions of similarity between various animal categories by using multidimensional scaling, he asked participants to estimate the proportion of different kinds of animals (the *'target'* species) on an isolated island that would share a novel contagious disease with a specific category of animal (the *'premise'* species). Peoples' category-based inductive estimations were predicted along two gradients, the similarity between the premise and target species and the typicality of the premise species carrying the disease. The higher the similarity between the premise and target species, the more members of the target species were thought to have the disease. For example, when told that *eagles* had the contagious disease, people thought that more *hawks* would also have the disease than would *geese* or *ducks.* Typicality can be quantified by using multidimensional scaling techniques and refers to the geometric distance between the category instance (e.g. *horse*) and its superordinate category (e.g. *mammal*). For example, people thought that relatively more of the other animal species would have the disease if it was present in a typical mammal category such as *horses*, than when it was present in a more atypical species such as *mice*.

Following on from this early work, similarity and category membership became the hallmark dimensions for explaining patterns of category-based induction. Using the now-standard argument evaluation paradigm, people are presented with a premise statement informing them that one or more base categories have a certain property. They are then given a conclusion which includes a target category. It is the participants' task to judge how likely it is that this conclusion category shares a property with the category mentioned in the premise. To minimize judgements that reflect mere knowledge retrieval about the presence of specific properties rather than inductively derived inferences, researchers use so-called 'blank' properties (e.g. *has 'talio-cells'*). Furthermore, if the conclusion category encompasses all the categories presented in the premises, as in examples (1a) and (1b) below, the argument is general. In contrast, if the conclusion is an instance from the same category level as the premise categories, the argument is said to be specific, as in examples (2a) and (2b) below. Consider examples of each type of argument:

Moths have talio-cells

Butterflies have talio-cells

Therefore, all Insects have talio-cells *(1a)*

Moths have talio-cells

Cockroaches have talio-cells

Therefore, all Insects have talio-cells *(1b)*

Tigers have talio-cells

Lions have talio-cells

Therefore, Rabbits have talio-cells *(2a)*

Tigers have talio-cells

Horses have talio-cells

Therefore, Rabbits have talio-cells *(2b)*

There are two kinds of tasks participants might be asked to carry out. They may either be asked to rate their belief that the target conclusion category does indeed share the property with the base premise(s) on a Likert-scale. Alternatively, they may be asked to choose the stronger of two arguments which have identical conclusions but include different premise categories. In the above examples (1a) and (1b), this would give an indication as to whether a property found in "*moths* and *butterflies*" or in "*moths* and *cockroaches*" offers more support that *all insects* share this property.

Such tasks give rise to a host of inductive reasoning phenomena (Osherson et al., 1990; Sloman, 1993b). Apart from typicality and similarity effects already discovered by Rips (1975), arguably the most important one is the diversity effect (Lopez, 1995; Heit, Hahn & Feeney, 2005). Diversity captures the finding that a more varied set of premise categories lead to stronger inferences than more similar premise categories. To illustrate, arguments (1b) and (2b) should be stronger than (1a) and (2a), as the premise categories in the former arguments are more dissimilar to one another and therefore "represent" the conclusion categories more exhaustively.

Whilst these phenomena imply that similarity is fundamental to the process of category-based induction, evidence suggests that category membership is just as vital. Gelman and Markman (1986) gave adults and children pictures of three animals, illustrated in Figure 1.1 below. The first animal (a *flamingo*) was from the same superordinate category as the third animal (a *blackbird*), but was perceptually dissimilar. In contrast, the second animal (a *bat*) was perceptually similar to the third animal (the *blackbird*) but came from a different superordinate category. Children were then told about the properties of two of the animals. For example, they may learn that the *flamingo's* heart has a right aortic arch only, whereas the *bat's* heart has a left aortic arch only. They then had to decide whether the *blackbird*'s heart had a right aortic arch like the *flamingo*, or a left aortic arch like the *bat*. Gelman and Markman's (1986) results indicated that children tended to believe that the *blackbird* would share a property with the *flamingo* rather than with the *bat*, ignoring perceptual similarity in favour of category membership.

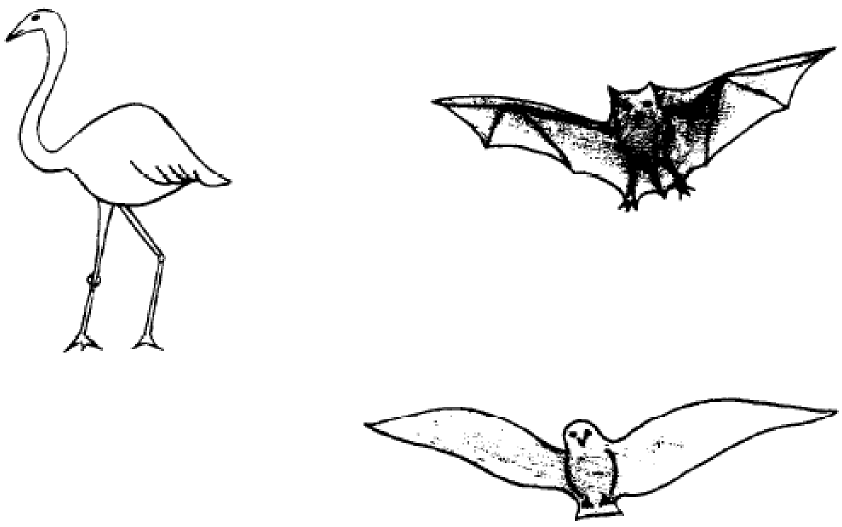

Figure 1.1: Example of stimuli used in Gelman and Markman's (1986) triad task

As the above examples illustrate, similarity and category membership are both crucial for explaining patterns of category-based induction. However, one of the major theoretical controversies to emerge alongside the next generation of models of category-based induction was the diverging emphasis placed on each of these components. For example, Osherson and colleagues' (1990) model emphasises both similarity and knowledge about stable category hierarchies, whereas the other early model by Sloman (1993b) explains these phenomena entirely in terms of similarity.

One way of capturing this apparent dichotomy is to divide theoretical approaches with reference to the contrasting types of knowledge they emphasize. On the one hand, there are approaches which emphasize the role of different types of similarity, such as featural overlap (Sloman, 1993b), perceptual similarity (Sloutsky & Fisher, 2004a) or semantic association (Rogers & McClelland, 2004). We describe such knowledge as being *unstructured*, in that it does not capture higher-order or abstract relations between categories. On the other hand, there are apparently contradictory approaches that place theory-based knowledge at the centre of the inductive process, such as knowledge about stable category-hierarchies (Osherson et al., 1990) and causal relations between categories (Tenenbaum & Griffiths, 2001). We call this *structured* knowledge, as it describes abstract and underlying relationships between categories.

The following sections provide a precise definition of what we mean by unstructured and structured knowledge. It first reviews approaches which emphasize unstructured knowledge, before outlining frameworks which highlight the role of structured background knowledge, evaluating the extent to which models within each approach can explain aspects of category-based induction. The thesis argues that the reason this dichotomy has not been addressed is because most researchers have focused on distinctions between domains of knowledge rather than types of knowledge (e.g. Carey, 1985, Rehder, 2006, Shafto, Coley & Baldwin, 2007). With reference to widely applied dual process theories, we suggest that this apparent conflict can be resolved by examining the processing characteristics of these two contrasting types of knowledge, exploring the roles of cognitive resources and inhibitory control.

1.3 The Role of Unstructured Knowledge in Category-Based Induction

Before proceeding with a review of approaches that put emphasis on unstructured knowledge, it is crucial to clarify exactly what we mean by this concept. We treat frameworks and models of category-based induction in which background knowledge is not described by higher order structure, abstract theories and/or interrelationships as frameworks which draw on unstructured knowledge. Hence, this includes approaches which explain inductive inferences as a result of relations between entities which are based on contiguity, co-occurrence, similarity or associations. Whilst the former two are inherent in the environment, the latter two relations describe unstructured knowledge from a psychological aspect. However, a precise or single psychological definition of the terms *similarity* and *association* are quite hard to come by. With regards to similarity, the simplest form is presumably perceptual similarity (Estes, 2003; Sloutsky & Fisher, 2004). Similarity between two entities (e.g. between *horses* and *cows*) can be determined by weighing up the number of shared features compared to the number of unique features (Tversky, 1977). However, as pointed out by Hahn, Chater & Richardson (2003), most entities and objects cannot be fully represented by a list of discreet features. According to Estes (2003) and Markman and Gentner (1996), the comparison of similarity between two entities occurs by structural alignment, whereby conceptually equivalent features of each entity are juxtaposed. For example, the shape of the *horse* is aligned with the shape of the *cow*, highlighting a commonality, whereas the whinny of the *horse* is aligned with the moo of the *cow*, exposing a difference. By tallying up differences and commonalities, one can calculate the similarity between two categories or concepts.

A related theory of similarity is Representational Distortion (Hahn, Chater & Richardson, 2003; Hodgetts, Hahn & Chater, 2009). This framework holds that similarity is inversely related to the computational complexity of transformations needed to convert one mental representation into another. When two objects can be structurally aligned, it reduces the number of steps needed to carry out this transformation process. Note that whilst this account emphasizes structured representations of the objects or entities to be compared for similarity, no structured knowledge about the relationship *between* the objects is required. For example, one can compare the structured representation of a horse to that of a cow without having to know about structured relations between them, such as shared habitat or taxonomic relatedness.

Wisniewski and Bassok (1999) also identify another kind of similarity based on an integration rather than comparison process and which may explain similarity on a more semantic level. For example, they found that people rated non-alignable categories, but which could be integrated into a thematic scenario, such as "*male* and *tie*", as being more similar than categories which did not share such a thematic relation, such as "*woman* and *tie*". Whilst the thematic relation between *men* and *ties* does not increase the actual perceptual similarity between the two concepts, people may be swayed by the strong association between the two. This might provide a bridge between the notions of similarity and association, with similarity being a specific kind of unstructured knowledge which relies more on a comparison process, whereas unstructured knowledge in the form of associations may be a more overarching concept that captures many kinds of relations, ranging from thematic, to causal and functional, etcetera. The reason we call this knowledge unstructured is because people need not be aware of the nature of the relation, but that the presence (or absence) of such relations increases (or decreases) people's perception of psychological "closeness" between entities.

At a phenomenological level, associations are best described as psychological connections, which relate ideas, events, objects or experiences to one another. Thus, if two concepts or mental images become associated, for example through temporo-spatial contiguity, then activation of one leads to activation of the other (Postman &Keppel, 1969; Tulving 1972; Shanks 2007). In this sense, the way we approach the construct of unstructured knowledge corresponds largely to the way in which knowledge is represented in connectionist network models of semantic cognition (Rumelhart & McClelland, 1986)

The rationale for assuming that associations and similarity must play a crucial role in inductive reasoning comes from a vast developmental literature (e.g. Colunga & Smith, 2005; French, Mareschal, Mermillod, & Quinn, 2004; Jones & Smith, 2002; Sloutsky & Fisher, 2004b; Sloutsky, Kloos, & Fisher, 2007). These studies suggest that early category formation and induction is driven by the statistical properties inherent in the environment, such as co-occurrence and statistical distribution of perceptual features. For example, acquisition of knowledge about which properties are crucial for generalizing specific category names can be explained by associations amongst perceptual properties (Jones & Smith, 2002), forming the foundation on which older children can build up richer structural representations of categories (French, et al., 2004). Similarly, Sloutsky and Fisher's (2004a) model of Similarity, Induction and Categorization (SINC) assumes that children perform categorization and inductive reasoning on the basis of perceptual similarity, in which the category label is simply treated as another feature contributing to increased similarity between different instances. These researchers also claim that there is only a gradual and developmentally late transition from exclusive reliance on similarity to the use of category membership as a basis for induction. This transition is largely seen as the product of explicit instruction and learning about general characteristics of categories (Fisher & Sloutsky, 2005). However, despite category representations being supplemented and expanded by top-down knowledge, proponents of associative approaches to category-based induction advocate that adult categorization and induction is still heavily influenced by similarity (Sloman, 1993a) and associations in semantic memory (Rogers & McClelland, 2004).

1.3.1 Sloman's (1993) Feature-Based Induction Model

One of the early models which explains category-based induction as driven primarily by bottom-up associative similarity is Sloman's (1993b) Feature-Based Induction Model (FBM). The FBM explicates all inductive reasoning phenomena purely in associative terms. Argument strength is determined by the degree to which the presentation of premise instances activates overlapping features of the conclusion instance. Arguments in which premise and conclusion categories share more features are stronger than arguments with little featural overlap between premise and conclusion. Consequently, there is no need to assume a stable category hierarchy. Consider for example how this model explains the diversity effect without making any assumptions about category membership. Thus, *blackbirds* and *penguins* cover the featural space of *birds* better than do the more similar categories *blackbirds* and *chaffinches*. Owing to their high similarity, the second category of the latter pair is largely redundant, failing to add additional features over those contributed by the first category, to cover the *bird* category. In contrast, *penguins* have many additional properties not shared by *blackbirds*, increasing the number of unique feature activations.

Smith and DeCoster (2000) link this associative view to underlying memory processes. A slow-learning system mediated by areas of the neocortex gradually forms stable representations of the environment based on the accumulation of analogous experiences. When similar situations or objects are encountered in the future, there is an automatic activation of these knowledge representations, which deliver information and shape behaviour without the need for deliberate or effortful thought. As such, this kind of reasoning predominantly delivers responses which encode statistical properties of the environment.

1.3.2 Connectionist Models

Smith and DeCoster's (2000) suggestions about the nature of associative induction resonate strongly with the characteristics of connectionist networks. Connectionist models are abundant across many domains of cognition, ranging from semantic memory (Collins & Loftus, 1975), word identification (Seidenberg & McClelland, 1989), acquired reading disorders (Hinton & Shallice, 1991; Plaut, McClelland, Seidenberg, & Patterson, 1996) to cognitive development (Elman, 1996; Morton & Munakata, 2005). Owing to the breath of application to different domains, there seems a certain appeal in being able to implement category-based reasoning in a connectionist architecture. Although the precise architecture can differ between connectionist networks[1], most models represent knowledge in a distributed manner. Thus, what defines a phoneme, word or concept is the precise configuration of activation across simple "neuron-like" processing units. The same units can be used in representations of different concepts. As this invariably encodes statistical properties both within and between concepts, it can encode more complex information than would be possible if each concept had its unique representation within the overall system.

The only connectionist model of category-based induction can be found in Rogers and McClelland's (2004) impressively wide-ranging application of their Parallel Distributed Processing (PDP) Model. It assumes that semantic information is processed through the

[1]For a summary of differences between feed-forward, recurrent and interactive network architectures, see Hadley (1999, 2000)

propagation of activity between neuron-like processing nodes in response to a stimulus input. Initially, the connections between nodes are random and weak, but incremental adjustments are made to the strength of the connection between different nodes in response to the learning input from the environment. Thus, semantic information about an instance is stored as an internal representation encoded by the pattern of distributed activation across processing units. Knowledge is refined by gradually adjusting the weight of connections between concepts (e.g. *sparrow*) and attributes (e.g. *wings*, *chirp, small*) for different kinds of relations (e.g. *has*, *can*, *is*). This means that generalizations from one instance to another will be strong to the extent that the activated distributed representations of the two instances overlap via their shared attributes.

Several predictions follow from the way in which the connectionist model acquires semantic knowledge and makes generalizations. As it acquires knowledge gradually based on experiential input, the internal representations should mirror the structure of the learning environment. For example, if one repeatedly encounters two species in the same context, the internal representations ought to reflect this statistical co-occurrence. Inductive inferences between categories should be stronger to the extent that the categories have repeatedly been simultaneously activated in semantic memory, forming strong associations.

1.4 Structured Approaches to Category-Based Induction

As mentioned above, an opposing approach to explaining inductive reasoning focuses on the influence of structured knowledge. The justification for assuming that structured knowledge can play an important role in category-based induction arises from several reasoning phenomena that cannot be explained exclusively by the use of unstructured or associative knowledge, for example the developmental trajectory of the diversity effect (Lopez, Gelman, Gutheil, & Smith, 1992; Heit & Hahn, 2001) and causal asymmetry effects, in which inferences are stronger from prey to predator than vice versa (Shafto, et al., 2008). These can however be well explained by models that put structured knowledge at the centre of the inductive reasoning process. The next section first outlines some of the formal models that incorporate structured knowledge, and how these can explain the aforementioned reasoning effects.

1.4.1 Osherson et al's (1990) Similarity-Coverage Model

Following Rips (1975), Osherson et al.'s (1990) Similarity-Coverage Model (SCM) posited knowledge about stable taxonomic structure as an important source of information that people rely on when evaluating categorical arguments. In contrast to Sloman's (1993) feature-based notion of similarity, Osherson et al.'s (1990) conceptualization of similarity relies far more on explicit knowledge about similarity relations between category instances. Thus, inductive evaluations reflect the weighted sum of two primary parameters, similarity and coverage, thus allowing for individual differences to be ascribed to the relative importance of these two components. Similarity refers to the maximum average similarity between the premise and conclusion categories. Thus, inductions are better supported when there is a higher resemblance between the premise and conclusion categories. Coverage assumes that people have knowledge about stable category hierarchies and refers to the degree to which the premise categories cover the category space of the inclusive superordinate category. For one-premise arguments, coverage is analogous to typicality, whereby more typical category instances make for stronger inductive inferences than atypical category members. For multiple-premise arguments with general conclusions, coverage refers to the extent to which the premise categories are representative of a diverse range of categories from the superordinate conclusion category.

One of the hallmark phenomenon accounted for by the SCM is the diversity effect. According to this model, people have to generate the lowest superordinate category which includes the premise and conclusion categories. Diverse arguments are stronger because two dissimilar category instances cover the superordinate category space better than two similar category instances. This explanation relies on people having a structured representation of taxonomic relations between species. Although Sloman's (1993b) model can account for the basic diversity phenomenon, other evidence suggests that certain characteristics of this effect are not explicable in terms of associative knowledge, instead requiring reference to structured representations. For example, Feeney (2007) showed that only a minority of people manifest the diversity effect when reasoning about specific conclusions and that this was related to cognitive ability. Similarly, Lopez et al. (1992) showed that several developmental changes seem to depend on knowledge about stable category hierarchies. The primary aim of these researchers was to identify the separate components which comprise the coverage parameter in Osherson et al's (1990) SCM model. However, their findings are also consistent with the notion that the emergence of computationally more complex inductive reasoning phenomena lags behind those that are thought to rely on computationally simpler associative processes. For example, in the case of the diversity effect for specific arguments, children first have to generate an inclusive superordinate category before calculating coverage, a process that requires knowledge about structured taxonomic relations between categories. Lopez et al. (1992) showed that 9-year-old children exhibit sensitivity to diversity with general arguments, but performed at chance levels with specific arguments. This suggests that in order to show adult-like inductive reasoning, there must be a qualitative developmental shift in how children use their category knowledge to make inductive inferences.

Together these findings suggest that generating the superordinate category which includes both the premise and conclusion category before being able to assess coverage (e.g.

having to generate the inclusive superordinate category *mammal* when reasoning from the premise categories *dog* and *sheep* to the conclusion category *buffalo*) involves a computationally demanding processing step that relies on knowledge about stable category hierarchies. In contrast, this superordinate category is already available when reasoning about general conclusions (e.g. reasoning from the premise categories *dog* and *sheep* to the conclusion category *mammal*). If the diversity effect was exclusively explicable in terms of associative knowledge, one would not expect to see differences based on cognitive ability and changing developmental trajectories, nor should there be a dissociation between specific and general arguments.

1.4.2 Causal Knowledge and Structured Bayesian Accounts

Accounts which draw on associative knowledge also have no straightforward means of explaining phenomena that arise through considerations of causal structure. Within the inductive reasoning literature, the influence of similarity and taxonomic group membership has been studied most extensively. In comparison, research on how causal knowledge influences category-based inductions is still in its infancy. Nonetheless, causal knowledge has some unique characteristics that are relevant to the question of whether induction can be explained purely with reference to associative or unstructured knowledge or if it is necessary to draw on structured knowledge.

There is evidence that causal knowledge plays a crucial role in children's category-based inductions early on. Hayes and Thompson (2007) taught five and eight-year old children and adults about features of two artificial base creatures. Figure 1.2 below illustrates Hayes and Thompson's (2007) task. The causal base contained two properties which were causally linked, whereas the attribute base contained three causally unrelated properties. This was followed by a target, which was more similar to the attribute base (two shared properties, no causal link), but shared the antecedent causal property with the causal base. Participants had to infer whether the target was more likely to match the attribute or the causal base on a fourth property, thus creating a choice conflict between featural similarity and causal knowledge. Results indicated that when the causal link was explicit, all age groups preferred to make causal rather than similarity-based inductions. When the causal relation was implicit, 5-year-olds did not yet show a preference for choosing the causally related items, unlike the older children and adults who made predominantly causal choices.

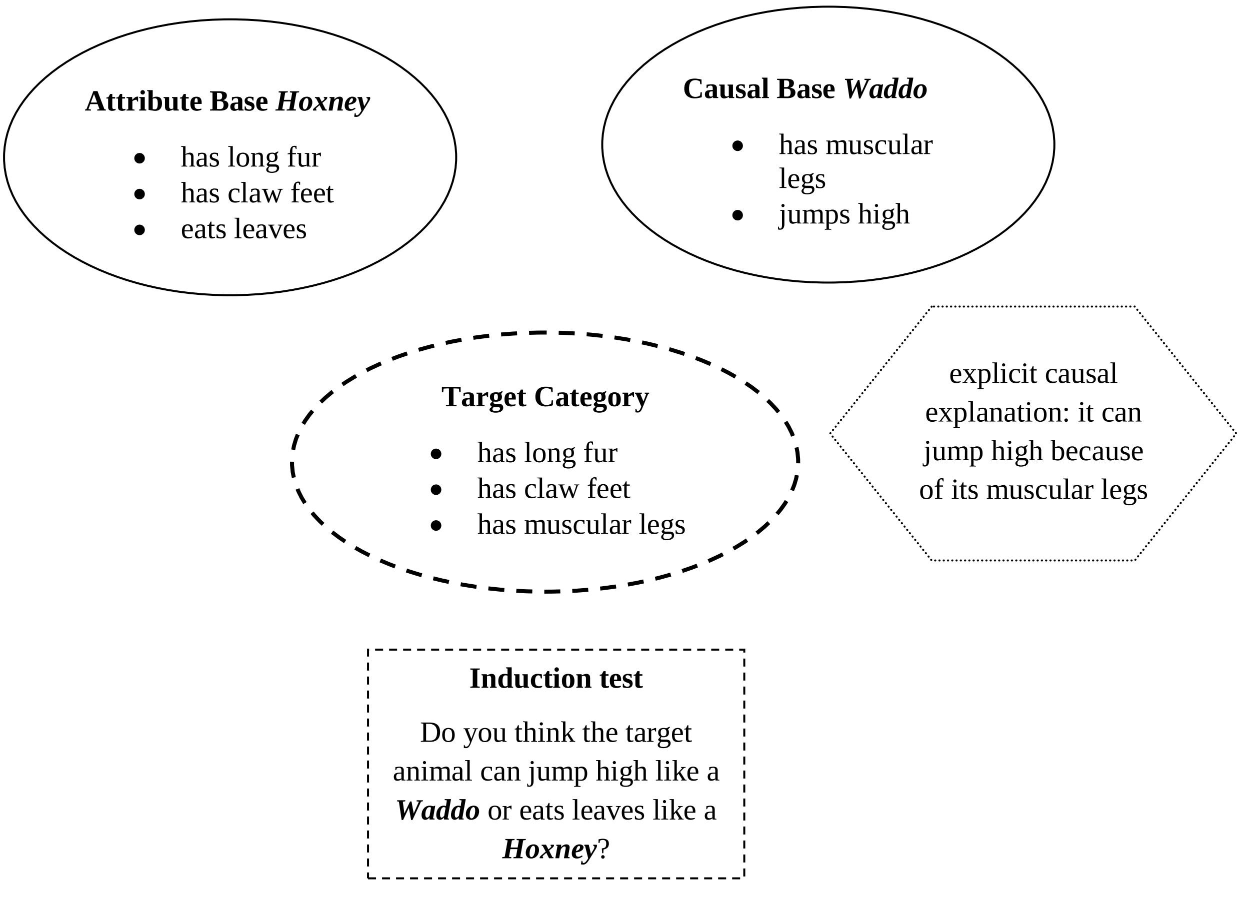

Figure 1.2: Task structure in Hayes and Thompson's (2007) experiment

In a similar vein, a recent study by Opfer and Bulloch (2007) demonstrated that 5-year-olds are capable of ignoring perceptual similarity in favour of relational similarity when the latter had a causal antecedent but not when it was non-causal. As well as demonstrating that children can flexibly choose between different knowledge structures, the above studies challenge the contention that young children's inferences are primarily based on perceptual similarity (Sloutsky & Fisher, 2004b).

The fact that causal knowledge is inductively potent from such a young age explains why it can sometimes supplant some of the most robust similarity effects in adults (Medin, Coley, Storms, & Hayes, 2003; Rehder, 2006; Rehder & Burnett, 2005) However, use of causal knowledge in inductive reasoning is especially interesting as it has several unique characteristics not shared by other kinds of knowledge (Sloman, 2005). Asymmetry is one of its most distinguishing features. Thus, causes always precede or at least coincide with their effects. Causal structures can also vary in their complexity, ranging from simple cause-effect relations, for example unidirectional causal chains, to more elaborate common cause models (Pearl, 2000; Sloman, 2005). People seem to make use of these abstract relations to evaluate inductive arguments as their inferences are highly sensitive to the direction of the causal relation between category features (Rehder, 2009), as well as the direction of the causal link between species when making generalizations about novel properties (Medin, et al., 2003) and disease transmission (Shafto, et al., 2008). For example, people are more likely to endorse the argument that '*lions have retinum*' given that '*gazelles have retinum'* than the diagnostic inference that '*gazelles have retinum*' given that *'lions have retinum'* (Medin, et al., 2003).

Such effects are easily explained by frameworks which draw on domain-specific structured knowledge. Kemp and Tenenbaum (2009) and Shafto et al. (2008) provide

Bayesian accounts of taxonomic and causal induction in which a domain-general Bayesian inference engine operates over a domain-specific theoretical model. The Bayesian inference machine is domain-general, and as discussed in Oaksford and Chater's (2007) work, provides a tool for updating people's subjective degree of beliefs in a hypothesis or similar proposition. These approaches tend to be based on Bayes' theorem, which is expressed mathematically as the conditional probability of A given B, i.e. the *posterior probability*:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P (A) is the *prior of A*, P (B) is the *prior of B* and P (B|A) is the conditional probability of B given A, also known as the *likelihood*.

In the case of category-based inductive reasoning, the Bayesian inference machine tries to compute the probability that a property observed in one or several example categories can be found in one or several other categories. To illustrate the derivation of the posterior probability, take four categories, *giraffes*, *antelopes*, *lions* and *tigers*. The first step is to generate a hypothesis space of all possible extensions of a novel property X, resulting in 16 distinct hypotheses *h*.

	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	h_{10}	h_{11}	h_{12}	h_{13}	h_{14}	h_{15}	h_{16}	Prediction
Giraffes	○	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	●
Antelopes	○	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	●
Tigers	○	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	●
Lions	○	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	●
p(*h*)	0.2	0.05	0.05	0.05	0.05	0.08	0.01	0.01	0.01	0.01	0.08	0.05	0.05	0.05	0.05	0.2	

Figure 1.3: Hypothesis space and prior distributions as displayed by Kemp and Tenenbaum (2009). All possible extensions of novel property X. Black circles signify the presence of the property in a given category.

Each hypothesis provides a unique combination of categories that have the property. For example, hypothesis 2 specifies that only giraffes have the property. The model makes the assumption that the prior probability of each hypothesis is known. In our example, the priors broadly capture the belief that carnivorous African mammals are similar to each other, as are herbivorous African mammals. The greyscale vector on the far right indicates that the probability that one of the categories has the novel property is 0.5, derived by summing the probabilities of the individual hypotheses in which the property is present. Thus, based on these priors alone, the probability that giraffes have the property is 0.05 + 0.08 + 0.01 + 0.01 +0.05 + 0.05 + 0.05 + 0.2 = 0.5.

Assume now that we obtain data indicating that giraffes have the property but lions do not. This can be denoted with X = [giraffe, lions] and the label vector $l_x = [1, 0]$. Based on this data, Bayes' theorem stipulates how the prior distribution p(h) can be updated, resulting in a posterior probability distribution for the hypothesis space h:

$$p(h|\mathrm{l_x}) = \frac{p(l_x|h)p(h)}{\sum_h p(l_x\,|h)\;p(h)}$$

Based on the generic assumptions that X is randomly sampled and the labels l_x are generated without noise, the likelihood term $p(l_x|h)$ is 1 if $h_i = l_x$ and zero otherwise. This simply means that hypotheses which are inconsistent with the observed data (i.e. l_x) are assigned a posterior probability value of zero, as these hypotheses are no longer possible given the data. This Bayesian property induction is exemplified in Figure 1.4 below.

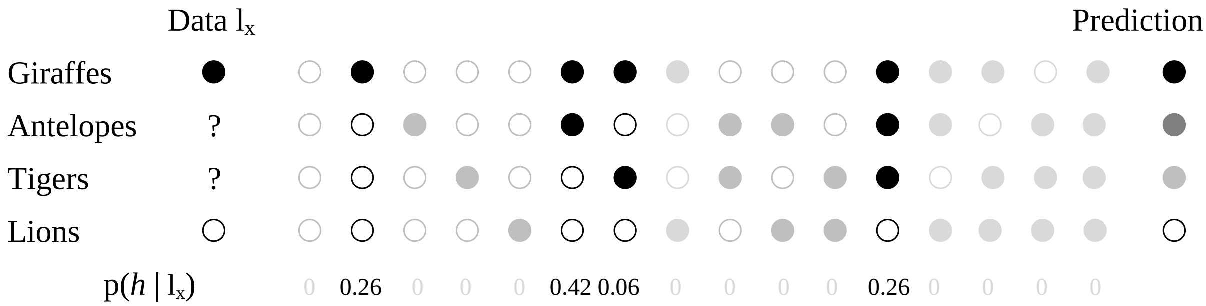

Figure 1.4: Hypothesis space and posterior distributions. Based on the observation that giraffes have the property but lions do not (l_x), 12 hypotheses are inconsistent with the observed data and have been greyed out. The posterior probability distribution $p(h| l_x)$ can be calculated by renormalizing the prior probability distribution $p(h)$ on the 4 remaining hypotheses in the hypothesis space. The prediction vector at the far right shows that Giraffes definitely have the property, Lions definitely do not have the property, and Antelopes (p= .68) are more likely to have the property than Tigers (p=.32).

The posterior probability of a specific hypothesis is hence equal to the proportion of hypotheses that are consistent with the data, weighted by their prior probability. For example, the posterior probability that antelopes have the property given that giraffes have it but that lions do not is: $\frac{0.08 + 0.05}{0.5 + 0.8 + 0.1 + 0.5} = 0.68.$

However, a point not touched on is how the prior probability distribution $p(h)$ was generated. The suggestion of current models is that they reflect people's expectations about a relevant property which can be represented by structured models. Such models might pertain to the hierarchical taxonomic relations or causal links between species. However, what all domain-specific models have in common is that they capture people's beliefs about the prior distribution of properties. Thus, inferences are generated by combining current observations with prior beliefs about which and how properties might be shared amongst categories within a domain. For example, inferences about diseases are often informed by causal knowledge about disease transmission. In this case, the underlying causal model consists of a structured representation describing people's domain-specific knowledge about the causal relations between the categories. Prior probabilities for the categories represented in the structure are generated through a stochastic process. In the case of disease transmission, the stochastic process that generates the priors includes two background knowledge parameters, background rate and transmission probability. This assumes that diseases can potentially be caused by a source external to the food web and that diseases are more likely to be passed from prey to predator. Such Bayesian models hence provide a knowledge-based probabilistic justification as to why predictive inductive inferences from prey to predator ought to be stronger than the reversed diagnostic inference from predator to prey.

Kemp and Tenenbaum (2003), Tenenbaum and Kemp (2009) as well as Shafto et al. (2008) also demonstrated that inductive reasoning patterns about causal transmission (for example when reasoning about a novel disease) can be dissociated from inductive inferences about physiological properties (e.g. when reasoning about novel genes or cells). Whilst the Bayesian inference machine is identical in both cases, knowledge about the distribution of physiological properties is captured by a theory-based model incorporating taxonomic interrelationships between species. Thus, such dissociations suggest that the context or property people are reasoning about prompts them to draw on different and most relevant sources of structured knowledge. It should be noted however that despite their strengths in predicting the output of people’s inferences and their capability to explain *why* such inferences are made, Bayesian approaches are silent with regards to *how* such inferences are achieved.

To summarize thus far, the review suggests that explanations of inductive reasoning can be grouped into two contrasting clusters: on the one hand, there are approaches which try to model inductive inferences with reference to unstructured sources of knowledge based on associations, contiguity, co-occurrence and/or similarity. These approaches assume no higher-order underlying rules or structures that describe the interrelations between different categories. On the other hand there are approaches which emphasize the role of structured knowledge. These approaches assume that people use the most relevant knowledge structure to support their inductive inferences. Examples include taxonomic hierarchies when reasoning about physiological properties and food chain relations when reasoning about disease transmission.

Whilst we have been focusing on how knowledge can be dissociated by type, i.e. associative or unstructured versus structured knowledge, the Bayesian models themselves focus more on the distinction between different domains of knowledge, such as ecological or taxonomic. In fact, the majority of work on knowledge effects in category-based induction has been focused on domains rather than types of knowledge, owing in part to the success of this approach in explaining induction in non-standard populations, patterns of category-based inductions in experts and inductions that are tailored to different reasoning contexts. The next section describes the scope of this work and highlights how a shift in focus towards types rather than domains of knowledge can shed light on the processes underlying induction, thus resolving the apparent theoretical dichotomy between accounts of inductive reasoning which assume unstructured knowledge on the one hand, and accounts which assume highly structured knowledge on the other.

1.5 Domain Distinctions in Category-Based Induction

Knowledge in inductive reasoning comes in many guises. On the one hand, there is domain-general knowledge that relates to normative considerations about making sound inductive inferences, such as sample size, variability of premise and conclusion categories (Nisbett, Krantz, Jepson, & Kunda, 1983). On the other hand, there is more domain-specific knowledge concerning the relations between the premise categories themselves and/or the premise and conclusion categories, as well as knowledge about the mechanisms by which properties come to be shared. A full account of how structured knowledge influences inductive reasoning needs to consider what kinds of knowledge people have about the categories themselves, the relations between categories and the nature of the property, and how this knowledge combines or interacts to influence inductive responses. We turn to the role of such domain-specific knowledge in induction next.

1.5.1 Acquiring Domain-Specific Knowledge

From earliest development, patterns of inductive inferences are strongly shaped by what we know about a domain. Thus, changes in children's underlying domain knowledge leads to qualitative changes in their category-based inductive patterns (Carey, 1985; Hatano & Inagaki, 1997, 2000; Inagaki & Hatano, 1993, 1996; Simons & Keil, 1995). For example, Carey (1985) showed that young children living in urban USA tended to see humans as the prototype on which to base their reasoning. They made stronger inferences from *humans* to *bugs* than from *bees* to *bugs*, violating the principle of similarity. They also tended to make asymmetrical inferences, such that inferences from *humans* to *animals* were stronger than from *animals* to *humans*. Medin and Waxman (2007) provide support for the idea that such non-normative asymmetries in reasoning about natural kinds arise due to lack of domain-specific knowledge.

Atran et al. (2001) further explored the question of how knowledge within a domain influences patterns of inductive reasoning with young 4 and 5-year old Yukatek Mayan children who have a close affiliation with their natural environment. These children did not show the tendency to see humans as the best prototype from which to make their inferences, nor did they show any asymmetry or violation of similarity effects. Coley, Medin and James (1999) and Ross et al. (2003) have obtained similar findings with children growing up in a rural Native American clan (Menominee Indians). Such cultural differences suggest that developing expertise within a domain, in this case familiarity and increased knowledge about the biological world, enables children to make more accurate inductions.

1.5.2 Expertise Effects

Differences in knowledge about how categories are related also leads to distinctive differences in adults' patterns of inductive inferences. Thus, salient relations between categories often support stronger inferences than mere similarity-based measures would predict (Coley, Shafto, Stepanova, & Barraff, 2005). For example, Lopez, Atran, Coley, Medin and Smith (1997) asked Itza Mayan people and US undergraduates to make category-based inferences about local mammals. After establishing the unique taxonomies for their various respective local mammal species, Lopez et al. (1997) compared three hallmark effects predicted by Osherson et al.'s (1990) similarity-coverage model, namely typicality, similarity and diversity effects. To illustrate, American participants might be told that *wolves* and *deer* have a certain disease and that *wolves* and *coyotes* have another disease. Their task was then to evaluate whether it was more likely that all *mammals* had the disease shared by *wolves* and *deer* or whether they were more likely to have the other disease shared by *wolves* and *coyotes*. According to the similarity-coverage model, the former argument should be chosen more frequently, as the premise categories are more diverse, thus providing better coverage of the superordinate *mammal* category. Results indicated that both cultures showed typicality and similarity effects, but the diversity phenomenon was only present in American Undergraduate students. Itza Mayan people's justifications in contrast frequently alluded to ecological or causal justifications for why less diverse species pairs provide more support for inferring that a disease could be present in all mammals. Whilst this suggests that domain-specific knowledge might account for the qualitatively different patterns of reasoning across the two groups, it cannot be ruled out that such differences simply have a cultural origin.

To address this issue, Proffitt, Coley and Medin (2000) compared inductive inferences about novel diseases affecting trees across three American tree expert groups (taxonomists, landscapers and park maintenance workers). When generalizing from a single premise instance, none of the expert groups showed typicality effects and only the taxonomists displayed diversity effects when making generalizations about shared diseases from multiple premises. This suggests that each expert group had specialist knowledge relevant to the goals of their occupations, and that this knowledge had a unique influence on the pattern of inferences they made about disease distribution across tree species. They tended to use a vast array of reasoning strategies ranging from ecological and causal strategies, to local coverage, depending upon the context and features of the trees deemed most relevant to making accurate inductions.

More elaborate knowledge possessed by experts may enable them to draw on different reasoning strategies depending upon context. For example, Shafto and Coley (2003a) compared commercial fishermen's inductive inferences about marine life to those of US undergraduates. When reasoning about completely blank properties, both groups relied on taxonomic similarity between two premise pairs as a basis for making property projections to a third target category. In contrast, when reasoning about novel diseases, fishermen's but not undergraduates' reasoning drew on causal/ecological relations between the premise pairs and target categories. Thus, when reasoning about diseases, experts tended to show asymmetrical patterns of reasoning, making more projections to species higher up the food chain than to species lower in the feeding hierarchy.

1.5.3 Property Effects

As already hinted at by Shafto and Coley's (2003a) study of commercial fishermen, context constrains which domain of knowledge is most relevant. Thus, when reasoning 'in a vacuum' about blank properties, both experienced fishermen and novice undergraduates used default taxonomic relationships between categories to guide their inferences about marine life. However, when reasoning about diseases, commercial fishermen's patterns of inductive inferences changed dramatically. They tended to draw on extensive background causal and ecological knowledge, maximizing the potency of their inferences with reference to a particular contextual domain.

However, even average college students will tailor their inferences to different reasoning contexts if doing so does not require expert knowledge. For example, Ross and Murphy (1999) showed that people generally relied on taxonomic category relations between different foods as a guide to inductive strength when making generalizations about the biochemical composition of food. However, when they made inferences about situational properties (such as *effort to prepare*, *when and where something is eaten* etc.), participants relied on less available thematic category relations as a basis for making their inferences.

Even young children are to some extent capable of tailoring their inferences to different contexts. For example, children were shown triad pictures in which the two animals from different categories looked alike (e.g. *blackbird* and *bat*) whereas the instances from the same category were perceptually dissimilar (e.g. *blackbird* and *flamingo*). Children had to choose whether the target (*blackbird*) was more likely to share a property with the perceptually dissimilar instance from the same category (*flamingo*) or with the perceptually similar instance from a different biological category (*bat*). With these natural kinds, children as young as 4 years of age tended to ignore perceptual similarity in favour of categorical membership as a basis for imputing properties such as feeding or behavioural habits (Carey, 1985; Gelman & Markman, 1986). However, when the nature of the property was varied so that category membership no longer provided good grounds for making accurate generalizations (for example reasoning about weight), reasoning patterns were strikingly altered. Overall, there was less reliance on category membership and some 4-year-olds started to exhibit selective induction, basing their reasoning exclusively on perceptual similarity. Similarly, Springer (1992) found that children were more willing to use kinship (e.g. parent-offspring relations) than perceptual similarity or social ties (e.g. friendship) as a basis for making inductions about biological characteristics. However, this did not extend to more idiosyncratic, non-biological properties, such as transient physical or mental states. However, whilst these studies suggest that some properties are in general more projectable than others, they do not address whether identical properties can be differentially projectable depending upon contextual factors.

The seminal study by Heit and Rubinstein (1994) sought to explore whether dynamic conceptions of similarity based on weighting features according to context (Tversky, 1977) would predict selective category-based induction in adults. That is, they hypothesized that the degree to which people would be willing to make a strong inference about a specific property would be a function of the degree of concordance between the nature of the property (such as a behavioural or anatomical property) and the nature of the similarity relation between the base and target categories (such as being behaviourally or anatomically similar). They found a reliable interaction between type of property and type of relation, with people making stronger inferences for anatomical properties when the premise and conclusion categories were aligned along taxonomic dimensions than for categories matched along the behavioural dimension, whereas this ordering was reversed when reasoning about behavioural properties. When they took people's background beliefs about anatomical and behavioural similarity into account, they found that inferences about behaviour were predicted by both behavioural as well as anatomical similarity ratings, whereas inductive inferences concerning anatomical properties were exclusively predicted by ratings of anatomical similarity between the two categories. This strongly suggests that people recruit structured knowledge that is relevant to how two categories might come to share a specific property. The precise domain of knowledge that is deemed relevant thus seems to differ depending upon the nature of the property itself.

1.5.4 Competing Knowledge

Whilst the above and further studies (Bailenson, Shum, Atran, Medin, & Coley, 2002; Stepanova & Coley, 2003) illustrate that expertise and contextual knowledge can change qualitative patterns of inductive reasoning by making some domain-specific knowledge structures richer or more salient, they do not address potential differences in the underlying processes that mediate use of knowledge in experts and novices in different contexts, nor do they say anything about cases in which knowledge structures from two domains might stand in opposition to one another. Consider a case in which you want to make an inference about the likelihood that *horses* have a disease given that *zebras* have it. On the one hand, you might use causal or ecological knowledge to reason that these two species live in very different geographic regions and have little immediate contact, thus concluding that it is unlikely that the two species share the disease. On the other hand, if you know that *horses* and *zebras* are taxonomically very closely related, you might conclude that they could be genetically susceptible to the same kinds of diseases and therefore believe that it is quite likely that they share a disease. So when knowledge competes, what is the relative status of different knowledge structures? The findings to date are less than clear, with different paradigms often delivering orthogonal results regarding the relative status of different types and domains of knowledge.

From a developmental perspective, causal knowledge seems to supplant similarity-based inductive reasoning from around 5 years of age when the causal relations are explicit (Hayes

& Thompson, 2007). When the causal relations between a category member and its features are implicit, only 8-year-olds and adults reliably use causal relations instead of similarity as a basis for feature induction. Rehder (2006), who takes the position that causal knowledge supplants similarity effects when it is available, drew similar conclusions from his paradigm. He asked participants to reason about shared properties of fictitious creatures. When people based their inferences on the causal explanation for the occurrence of a property, some of the classic similarity effects, such as diversity, typicality and similarity, no longer emerged. However, there were two equally sized subgroups, one of which continued to display similarity effects even when causal knowledge was available. This and the fact that both experiments used artificial stimuli question the generalization that causal knowledge always supplants other knowledge structures.

1.5.5 Processing Differences

Shafto, Coley and Baldwin (2007) addressed the issue of which domain-specific knowledge might be most dominant by looking at processing differences underlying the use of knowledge from different domains. Using real-world categories which were ecologically related, taxonomically related or unrelated, they asked their participants to make generalizations about diseases or genes. Some of their participants were asked to respond as fast as possible, whereas others were forced to delay their responses and were asked to consider their answers carefully before responding. Whereas people's inferences about taxonomically related categories were unaffected by the timing manipulation, they found that people gave lower inductive strength ratings to ecologically related categories when they were under time pressure. This selective restriction of access to different knowledge structures led them to suggest that taxonomic knowledge is the default primary source of knowledge. However, they did not control for level of association between their category pairs. An alternative explanation for their results might be that their taxonomically related categories (e.g. *tiger* and *lion*) were more strongly associated than the ecologically related items (e.g. *tiger* and *parrot*). Rather than arising because of chronic differences in accessibility to knowledge structures per se, apparent differences between ecological and taxonomic inferences might instead be due to the strength of association between categories (Rogers & McClelland, 2004). In this case there would be no need to appeal to differences in the processing effort associated with knowledge from different domains.

It was argued in the first half of the review that an important distinction exists between types of knowledge. Distinguishing between types of knowledge in this way might be more fruitful than the distinction between domains of knowledge for explaining processing differences in category-based induction. Shafto and colleagues' (2007) results would be consistent with this alternative explanation. Hence, people might have been unable to access structured ecological knowledge when under time pressure, whereas inferences based on unstructured associative knowledge would be unaffected by timing manipulations. Similarly, in his recent work, Rehder (2009) taught participants about the causal links between category features of artificial categories. In line with the assumption that people draw on extensive causal knowledge, he demonstrated various phenomena, such as a causal asymmetry effect. However, he also found that there was a substantial minority of people whose patterns of inductions did not adhere to those predicted by his causal-based generalization model. Instead, they seemed to rely more on nondirectional associations between the category features. In Rehder's (2009) experiments, there appears to be one source of structured knowledge about causal relations and another source of unstructured knowledge based on the simple co-occurrence of features.

1.6 A Resource Account of Category-Based Inductive Reasoning

If people can be influenced either by unstructured knowledge such as associative strength, or structured knowledge such as information about causal links or taxonomic relatedness, it is important to demonstrate that these two types of knowledge can be dissociated. Rehder (2009) explicitly suggests that the use of causal knowledge relies on an elaborate, analytical thought processes, whereas associative or unstructured knowledge influences inductive reasoning fairly automatically and without much cognitive effort. Thus, if the use of unstructured knowledge is indeed mediated by effortless processing, whereas the use of structured knowledge requires more elaborate reasoning, then one way of dissociating these two types of knowledge is by exploiting these apparent processing differences. For example, a study looking at reasoning about music in experts (composers and musicians) and novices (Baraff & Coley, 2003; Coley & Barraff, 2003) suggests that whereas both experts and novices have access to associative knowledge, only experts are able to draw on more structured knowledge as a basis for reasoning. These researchers created an index of taxonomic distance by asking participants to sort composers from an array of genres according to similarity of music composition style. They then presented participants with inductive arguments consisting of two premise categories and a conclusion category. They varied the taxonomic strength of the arguments, so that they were either strong (e.g. *Beethoven* and *Bach→ Mozart*) or weak (e.g. *Bob Marley* and *John Lennon→ Mozart*). Participants rated the strength of the argument that the three composers in the premises and conclusion "use technique X in music writing". Whereas novices rated taxonomically strong arguments as more plausible than taxonomically weak arguments, experts showed no difference between the two types of problems, suggesting that they were using more elaborate context-dependent relational knowledge. However, when the induction task was carried out under time pressure, thus decreasing available cognitive processing time, experts' reasoning was indistinguishable from novice reasoning. According to the researchers, their results highlight the primacy of taxonomic similarity, which is only supplanted by relational information when experts have enough time to engage in more complex processing. However, the findings are also consistent with our distinction between structured and unstructured/associative knowledge: The change in expert reasoning suggests that their use of structured relational knowledge, when they have time and resources to do so, is cognitively demanding. Under time pressure, they had to rely more on associative similarity. In contrast, novices simply did not have an alternative source of structured knowledge, forcing them to rely on associative similarity under both speeded and delayed conditions. Thus, such experimental findings would support Rehder's (2009) and our own contention about the role of cognitive resources in category-based inductive reasoning, whereby the use of structured knowledge is cognitively demanding, whereas relying on unstructured knowledge is less effortful.

Medin et al's (2003) relevance theory is the only framework within the category-based induction literature to consider the role of cognitive resources during the reasoning process. Based on work by Sperber and Wilson (1995), this framework assumes that when the cognitive system processes new information, relevance is determined by the cognitive effects this processing has, as well as the degree of cognitive effort it took to achieve these effects. Applied to category-based induction, the relevance framework further posits that people formulate hypotheses about relevant relations between premise categories. Consider the following two arguments:

Magpies have property X

Pandas have property X

Zebras have property X

All animals have property X. *(Consistent argument)*

Brown bears have property X

Pandas have property X

Zebras have property X

All animals have property X. *("Garden path" argument)*

In the consistent argument, all three premises suggest that the property to be projected has to do with black-and-white colouring. In contrast, the first two premises of the "garden-path" argument suggest that being a member of the superordinate category *bear* might be relevant. However, this hypothesis is refuted by the third premise. As this should have large cognitive effects by forcing the reasoner to change his or her beliefs, people should be willing to expend considerable cognitive effort in processing this input. Feeney, Coley and Crisp (2010) showed that when the category in the third premise was inconsistent with a person's likely first hypothesis, reading times for this third premise sentence were significantly longer. In contrast, when the category was consistent with a person's likely hypothesis, reading times for the identical third premise sentence were significantly shorter.

The notion of effort in the relevance framework suggests that the cognitive system cannot process all input homogeneously. As in the above example, formulating certain hypotheses might be more effortful than others. Thus, some input seems to be processed relatively automatically, for example the fact that *carrots* and *rabbits* are causally related, whereas other input may require the deployment of more mental resources in order to impact on the reasoning output, for example noting the fact that *carrots* and *bamboo* are taxonomically related. Note that the reverse is also true. Thus, recognizing that *zebras* and *horses* are taxonomically related does not require much elaborate reasoning. In contrast, identifying a causal relationship between *flies* and *herons* might require the construction of a causal link and the assignment of causal roles in which *frogs* act as the food chain link between the two categories. This assignment of causal roles is likely to be a more effortful and less automatic process. However, whilst this model highlights that differences in how knowledge is processed crucially influence the final reasoning output, it has no means of predicting or explaining which types of knowledge are likely to involve more or less processing resources, nor does it offer an instantiation of the processes that mediate use of different types of knowledge. Auspiciously though, the different processing characteristics we have ascribed to unstructured or associative versus structured knowledge resonate with distinctions made by dual-process models popular in other reasoning domains.

1.6.1 Dual Process Theories

The idea that reasoning is driven by an interplay between effortless and more elaborate reasoning processes is well established across an array of thinking and reasoning disciplines, such as deductive reasoning (Evans, 2006, 2007; Evans & Over, 1996), judgement and decision-making (Kahneman & Frederick, 2005), individual differences (Stanovich & West, 1998), social cognition (Chen, Duckworth, & Chaiken, 1999; Pacini & Epstein, 1999) and cognitive development (Klaczynski, 2004). Such accounts are variants of dual-process theories (for a full review see Evans, 2008). The basic tenet of this approach is that the human mind runs at least two distinctive processes. We will follow Evans' (2006) classification and define one process as being heuristic and the other as being analytical. Each of these is in turn associated with different characteristics. In a recent review, Evans (2008) groups these characteristics into 4 major clusters, covering *Consciousness, Age of Evolution, Functional Attributes* and *Individual Differences*. We focus predominantly on the latter three, as these are most relevant when considering how to apply a dual process account to selective category-based induction and conversely, what phenomena in category-based induction can tell us about the nature of the mental processes purported to underlie reasoning.

Dual process theories all agree that there are two separate and competing processes. Heuristic processing often produces biased output, whereas analytical thinking is more likely to result in normatively sound responses. There are several functional attributes that have been assigned to the two processes, some of which are more robust than others. Table 1.1 below summarizes the most important opposing features ascribed to the two processes.

Table 1.1: Functional Characteristics of Dual Process Theories

Heuristic Process	**Analytical Process**
Fast	Slow
Automatic	Controlled
Effortless	Resource-Demanding
Contextualized	Largely Abstract and Decontextualized

To be evolutionary viable and assist cognitive economy, heuristic reasoning should approximate the response that would follow from more effortful processing. Consequently, the two processes usually deliver congruent output. However, there are instances in which responses from each process stand in opposition to one another, for example in response-incongruent tasks such as Wason’s selection task (Wason, 1966), syllogistic reasoning under belief-bias (Evans & Curtis-Holmes, 2005) and the conjunction fallacy (Crisp & Feeney, 2009; De Neys, 2006a; Stanovich & West, 1998).

The fact that analytical reasoning is cognitively costly, with restrictions imposed by available time, cognitive ability and working memory resources has enabled researchers to dissociate the two processes. For example, Stanovich and West (1998) have demonstrated that people higher in IQ are less likely to succumb to the conjunction fallacy. Similarly, studies have shown that forcing people to carry out a secondary task increases the rate of fallacious responses to the classic Linda problem (De Neys, 2006a), and increases erroneous reasoning on other deductive inference tasks (Gilhooly, Logie, Wetherick, & Wynn, 1993). Finally, forcing people to respond quickly leads to belief-based rather than logic-based responding in syllogistic reasoning (Evans & Curtis-Holmes, 2005). The fact that time pressure and cognitive load decreases people's ability to recruit effortful analytical processes supports the argument that incorrect responses are based on an effortless, heuristic process. Applying this to our hypothesis about the role of knowledge in category-based induction, it might be the case that the use of different types of knowledge is mediated either by heuristic or analytical processing. If the use of structured knowledge is subject to timing and cognitive resource constraints then this would be akin to analytical processing. In contrast, responses based on unstructured knowledge might be mediated by heuristic processing, and would hence not be affected by timing or resource manipulations.

Apart from the nature of the two processes themselves, an important question is how the two processes interact with one another. Although the precise nature of the relationship between the two processes has not been fully resolved (Evans, 2007), one suggestion is that heuristic responses have to be inhibited if analytical reasoning is to dominate the response output (Evans, 2008). For example, theories of deductive reasoning posit that people must inhibit strong background beliefs and knowledge if the answer generated by this conflicts with the logically correct response (De Neys, in press). De Neys and Van Gelder (2009) showed that the ability to resist belief bias on a syllogistic reasoning task exhibits a curvilinear developmental trend, which is in line with the hypothesized developmental pattern of inhibitory control. In line with this, neuro-imaging studies suggest that successful reasoning is related to activation in brain areas known to be involved in inhibitory functions. Thus, areas in the left prefrontal cortex are active when people manage to resist a belief-based incorrect prepotent response in favour of a logical correct answer. In contrast, no such activation is found when the prepotent response does not conflict with the correct answer and thus does not have to be inhibited (De Neys & Van Gelder, 2009; De Neys, Vartanian, & Goel, 2008; Goel & Dolan, 2003). Similarly, Handley, Capon, Beveridge, Dennis and Evans (2004) looked at 10-year-old children's ability to reason logically when the logical answer conflicted with the answer suggested by background knowledge. They found that whereas the ability to reason logically in the absence of conflicting beliefs was predicted by working memory only, logical reasoning skills on conflict problems was additionally related to children's level of inhibitory control as measured by the stop signal task. Thus, those children who were more able to withhold a prepotent response were more likely to resist a belief-based answer in favour of an unbelievable but logically correct response. This suggests that inhibitory control determines whether a response will be based on effortless heuristic processing, or will be based on effortful, analytical processing.

Whilst the distinction between the two processes based on their processing characteristics is very robust, the distinction between the two processes based on their representational characteristics is less convincing (Evans, 2006; Verschueren, Schaeken, & d'Ydewalle, 2005). The suggestion that heuristic processes are concrete, contextualized and domain-specific, whereas analytical processes would have the orthogonal features of being abstract, decontextualized and domain-general (Klaczynski & Lavallee, 2005) is especially questionable when considering how a dual process account might apply to category-based inductive reasoning, which is so heavily based on background knowledge. Recent evidence (Feeney & Crisp, 2010) from a category-based induction paradigm suggests that analytical processes can operate on contextualized knowledge representations. We gave participants the following arguments:

Soil has property X.
How likely is it that Grass has property X? (1)

Soil has property X.
How likely is it that Cows have property X? (2)

Soil has property X.
How likely is it that Grass and Cows have property X? (3)

As originally demonstrated by Medin et al. (2003), people frequently evaluated the conjunctive category conclusion (3) higher than the single causally distant category conclusion (2). This violates a fundamental law of probability, asserting that the conjunctive occurrence of two events can never exceed the lone occurrence of each event. In contrast, only a minority of people tended to rate the conjunctive category conclusion (3) higher than the single causally near category conclusion (1). We showed that a secondary task designed to interfere with analytical processing led to an increase of the conjunction fallacy for causally distant conclusions, whereas it did not increase the fallacy rate for causally near conclusions. This suggests that people were using some effortful process to reconstruct the causal link between the premise and causally distant conclusion category, allowing them to avoid the fallacy when enough mental resources are available. This finding is consistent with the idea that analytical processes can operate on knowledge-rich representations and make structured knowledge available to the reasoning process.

It is conceivable that similar principles might operate in inductive reasoning when there is conflict between different types of knowledge. If this is the case, then in order for one type of knowledge to dominate the response, it might be necessary to inhibit competing knowledge if it is not relevant. Thus, the notion of *effort* in Medin et al's (2003) relevance theory might reflect a) the ease with which a heuristic process makes relevant knowledge available and b) the extent to which mental resources are required to inhibit irrelevant knowledge. The first point a) is likely to be the result of factors such as expertise, frequency of co-occurrence between categories and other over-learnt category relations. To reiterate a point made earlier, knowledge we have described as unstructured in the above review might influence category-based induction in a heuristic manner. In contrast, the latter point b) is more likely linked to domain-general executive processes, such as working memory and inhibitory control.

This analysis maps onto Toates' (2006) distinction between stimulus-bound and higher-order processing. Toates' (2006) framework assumes that stimulus-bound processes are evolutionary older processes characterized by prescriptive, yet rapid and often automatic responses to environmental or internal memory-based stimuli. In contrast, higher-order processing refers to evolutionarily more recent processes, which exert inhibitory control over stimulus-bound processes. In order to be evolutionarily viable, the juxtaposition of different modes of processing and reliance on different knowledge has to offer an adaptive advantage.

The evolutionary rationale for such dual processing is especially salient for selective category-based inductive reasoning. Thus, environmental regularities and co-occurrences create stable mental associations between objects, events and situations that offer an economical and rapid source of information for reducing uncertainty, informing behaviour and making predictions about likely outcomes. However, given the probabilistic nature of our environments, there will be occasions where inductive potency can be maximized by carefully monitoring and if necessary inhibiting inappropriate application of associative knowledge in favour of more effortful reasoning. For example, imagine a fatty chip pan goes up in flames in the kitchen. Our first intuitive reaction is to rush and throw a bucket of water over the burning chip pan. However, in this case, actually acting on this strong association we have between water and putting out fire would worsen the situation. The water sinks beneath the burning fat and turns into high-pressure steam which rapidly expands upwards. This causes the burning oil to explode up out of the pan, engulfing us and the kitchen in a flash fire. Thus, one has to inhibit this inappropriate associative knowledge in favour of more complex causal knowledge about the relationship between fat, fire and water. In this case, we might act by throwing a fire blanket over the fat fire, extinguishing it by starving the fire of oxygen.

To summarize the role of mental processes involved in category-based induction, we suggest that different types of knowledge are subject to divergent processing constraints. Use of unstructured knowledge might be based on a fast and effortless heuristic process. In contrast, if the use of structured knowledge is indeed mediated by analytical processes, then its influence should be constrained by the availability of mental resources and the efficiency of executive functions. The role of executive functions, which include the ability to engage in complex and goal-directed behaviour e.g. (Rabbitt, 1997) has received almost no attention in the category-based induction literature (for exceptions see Feeney, 2007; Feeney, Crisp, &

Wilburn, 2008; Heit & Feeney, 2005). Empirical studies have either been concerned with predicting and explaining output patterns rather than psychological processes that result from the use of structured knowledge (Kemp & Tenenbaum, 2009; Osherson, et al., 1990; Shafto, et al., 2008; Tenenbaum, et al., 2006), or have emphasized the automaticity of inductive reasoning (Rogers & McClelland, 2004) with no need to refer to effortful or executive processing. However, it is unlikely that a process theory of category-based induction would be complete without considering the role of executive functions such as inhibitory control. This mechanism seems especially interesting, as it is likely to be in operation when people have to choose selectively between different domains of background knowledge in inductive reasoning, as well as when there is competition between structured and unstructured or associative knowledge. Thus, one might have to inhibit a response based on unstructured associative knowledge if one wants to consider possible implications of the underlying structure by which categories are related, for example which category is hierarchically higher in a food web.

1.7 Summary of Outstanding Questions and Thesis Overview

There are several issues arising from the above review. One of the fundamental questions concerns the processes that make human thought economical and adaptive yet hugely flexible in the light of an ever-changing environment. This is especially important for category-based inductive reasoning, where maximization of inductive potency depends on interplay between contextual factors and various distinctions to do with background knowledge. The thesis will attempt to resolve several theoretical issues specific to the field of category-based inductive reasoning, whilst also contributing to more fundamental enquiries about the architecture and resulting nature of the processes mediating human thought.

Current theoretical positions about the nature of knowledge that drives category-based inductive reasoning appear to form a dichotomy. At one end of the theoretical scale, one can place models that operate purely on featural similarity (Sloman, 1993b) and associative learning processes (Rogers & McClelland, 2004; Sloutsky & Fisher, 2004a, 2008). We subsume these under models which draw on unstructured knowledge. At the other end of the scale are models which call for structured knowledge representations, such as stable category hierarchies (Osherson, et al., 1990) and domain-specific theories of conceptual structure (Gopnik & Tenenbaum, 2007; Griffiths & Tenenbaum, 2005; Kemp & Tenenbaum, 2003, 2009; Shafto, et al., 2008) and hypotheses about relevant relations between categories (Medin et al., 2003). By distinguishing knowledge by type (i.e. structured or unstructured) rather than by domain (i.e. causal versus taxonomic) we will attempt to demonstrate how it is possible to reconcile the longstanding theoretical conflict between these contrasting frameworks. With reference to dual process theories we argue that these two types of knowledge differ in their processing characteristics, with the use of structured knowledge mediated by an effortful, analytical process, and unstructured knowledge based on effortless heuristic reasoning. This enables us to describe the contextual factors that determine the use of contrasting types of knowledge and explore the nature of the executive processes involved in the use of elaborate structured knowledge.

First Key Objective

The first goal is to demonstrate that structured and unstructured types of knowledge influence category-based induction in line with their unique processing characteristics. In doing so, the thesis will answer the following more specific questions:

- Are unstructured and structured knowledge dissociable, and if so, do they have differential impact on the reasoning output?

- Are apparent differences regarding the importance and availability of knowledge from different domains (Rehder, 2006; Shafto, Coley, et al., 2007) still evident when level of association between the categories has been controlled for?

Second Key Objective

The second goal of the thesis is to show how the characteristics of these two types of knowledge tie in with more domain-general processing capacities and can thus be linked to more general theories of reasoning. With reference to the widely applied dual-process framework we will demonstrate how use of different types of knowledge in category-based induction is tied to the availability of mental resources and to the ability to inhibit prepotent responses. The specific questions we focus on are:

- Do people have to inhibit one type of knowledge in order for the other to dominate the inference process?
- Does the ability to withhold a response based on unstructured knowledge correlate with general measures of inhibitory control? If so, are there differences depending upon the level at which an instrument measures the construct of inhibition?

The thesis is organized into a series of experimental chapters. Chapter 2 begins with an extensive pre-testing of stimulus materials in order to remedy shortcomings of previous research regarding the selection of appropriate stimuli, as well as to operationalize our notion of unstructured knowledge. The chapter describes the rationale and theory behind the way we generated these stimulus materials. By using both a subjective and objective measures of association, we created a database of stimulus materials from which we could pick category pairs from different domains that were equated for level of association.

Chapter 3 reports three experiments which used timing manipulations and secondary task paradigms. These examine in more detail the differential effect unstructured and structured knowledge have on category-based induction with reference to causal asymmetry effects.

Chapter 4 describes two experiments that used a more naturalistic paradigm to explore the effects of different types of knowledge in category-based induction. Using a novel methodology, participants generated their own inferences either under loaded or unloaded conditions rather than being asked to evaluate pre-determined category pairings. It explored whether people's inductive strength ratings were better predicted by strength of association or an index of structural relations between category pairs.

Chapter 5 includes three experiments in which unstructured associative knowledge was put into immediate conflict with structured knowledge. It explores whether and how the ability to inhibit unstructured knowledge in favour of more appropriate structured knowledge is related to measures of inhibitory control.

The final Chapter 6 provides answers to the specific questions posed above. It summarises how this thesis is thought to contribute to our understanding of the mental processes mediating the use of background knowledge in selective category-based induction in particular, and inductive reasoning more generally. The thesis concludes by pointing towards exciting opportunities for future research.

Chapter II

Pretesting for Strength of Association

Addressing the role that knowledge plays in category-based induction necessitates an understanding of what is meant by the construct of knowledge. We conceptualize knowledge as the psychological result of perception, learning and reasoning. Within learning paradigms, it has been demonstrated that acquisition of knowledge is driven by two mechanisms which differ in their temporal characteristics. People’s generalizations are influenced not only by featural or associative similarity, but also by rule-based processing (Shanks & Darby, 1998). Whereas similarity appears to be most influential early on during the learning process, rule-based processing has more impact at a later stage.

Similarly, the notion that retrieval of information from memory can be dissociated into automatic activation of knowledge and more strategic, effortful search processes is popular in the memory literature (Moscovitch, 1995). Thus, some memories are retrieved automatically upon processing a cue, for example retrieving the name of a neighbour’s dog upon hearing its familiar bark. In contrast, some memories might require strategic retrieval, for example trying to remember the names of all James Bond movies in order of their year of appearance. This dissociation is supported by the fact that strategic retrieval is related to individual differences in working memory. Thus, Rosen and Engle (1997) found that people who scored high on the operation span task, an index of working memory capacity, showed superior performance on a test thought to reflect strategic retrieval, the verbal fluency task, in which people have to generate as many instances of a category (e.g. *mammal → antelope*, *lion*,....etc.) as possible in a limited period of time. Similarly, Baddeley, Lewis, Eldridge and Thompson (1984) showed that whereas a concurrent secondary load during the retrieval process decreased people's performance on the verbal fluency task, it appeared to have no impact on tasks tapping automatic retrieval processes, such as paired-associate learning, free recall and sentence verification tasks.

The notion that knowledge might be dissociable with regard to its processing characteristics has thus far not been incorporated into models of category-based induction. However, a recent study by Rehder (2009) suggests that category-based induction can be shaped either by more elaborate causal knowledge or simple associative knowledge. In this study, participants learnt about the features of different of fictitious creatures. He found that when there was a causal explanation for the occurrence of these features, the majority of his participants used this relation as a guide to their inductive reasoning, showing directional asymmetry effects. However, a substantial minority appeared to treat such causal links as nondirectional associations. Whilst he focuses solely on the role of causal structure, the idea that general elaborate structured knowledge, such as causal networks or taxonomic hierarchies, provides an important constraint on inductive reasoning is implicit in several other models. Thus, charting people's structured knowledge can accurately predict people's patterns of inductive inferences (Kemp & Tenenbaum, 2003, 2009; Osherson, et al., 1990;

Shafto, et al., 2008). However, as the review in the previous chapter suggests, another type of knowledge which we refer to as unstructured or associative knowledge (Rogers & McClelland, 2004; Sloman, 1993b; Sloutsky & Fisher, 2004a) appears to have an equally crucial influence on people's category-based inductions.

Using real-world categories in inductive reasoning to examine the roles of different kinds of category relations and different types of knowledge has some potential pitfalls. One of the major shortcomings of previous research is the failure to generate stringently pre-tested materials. It is insufficient to generate categories which share different kinds of relations based exclusively on scientific taxonomies and ecologically supported facts without verifying their psychological reality. To claim that category relations are causal or taxonomic, it is necessary to demonstrate that people do perceive the type of connection we are anticipating. Similarly, if claims are to be upheld about the relative availability of one type of knowledge above another, or to posit the primacy of certain domains of knowledge we have to be certain that this is driven by the domain-specificity of that knowledge rather than by confounding factors such as level of association. So for example, to predict that inferences from *tigers* to *lions* should be stronger than from *tigers* to *parrots* when people are under time pressure because the former pair is related taxonomically and the latter ecologically (Shafto, Coley, et al., 2007) is to ignore the possibility that strength of association rather than domain-specific attributes of knowledge is driving inductive strength ratings. In this case, the associative link between *tigers* and *lions* may be stronger than between *tigers* and *parrots*, explaining the difference without reference to the domain-specific nature of the relation between the pairs. This example illustrates that processing differences attributed to domain-specific relations between categories may be confounded by associative strength between those same category pairs. In order to ascribe processing differences to domain-specific knowledge it is necessary to equate strength of association between category pairs from different domains. Only if processing differences still emerge under such stringent conditions can it be said that knowledge from one domain (e.g. taxonomic) is more available than knowledge from another domain (e.g. ecological). Consider an inference from *mice* to *owls* or from *mice* to *porcupines*. Intuitively, given the strong association between the former ecologically related category pair it seems unlikely that inferences would be unanimously higher for the latter taxonomically related category pair, as the categories are only weakly associated.

To address these concerns, the aim of the current pre-test study is to check that people perceive the purported relation between two categories (e.g. a causal relation between *frogs* and *herons*) and to cross-validate two novel approaches for indexing the strength of associations between categories. As already described, we take the stance that associations are psychological connections between concepts that mirror the statistical co-occurrence of categories within the environment. Thus, if two concepts or mental images become associated, then activation of one leads to activation of the other (Shanks, 2007; Tulving & Donaldson, 1972). As temporo-spatial contiguity is vitally important for the formation, maintenance and adjustment of associations, they are subject to gradual changes over time rather than rapid modifications. Changes in one's personal environment can lead to adjustments in associative strength between two concepts. For example, if one works in a flower shop, *roses* might become strongly associated with other bouquet flowers such as *lilies* or *peonies*. However, for someone who is a garden maintenance worker, *roses* might become more strongly associated with pests such as *greenflies* or *red spider mites*.

However, apart from such specialist associations, most people within a certain culture will be exposed to similar co-occurrences between categories. For example, children tend to have a strong association between *cats* and *mice* owing to television programmes such as *'Tom and Jerry'*. Consequently, words that frequently co-occur in natural language should also have a stronger associative relationship (Spence & Owens, 1990). This has some striking effects on tasks that involve the retrieval or activation of lexical semantic representations or in which contextual knowledge influences responses. For example, people are faster at making a lexical decision about the word *doctor* if it is preceded by the associated word *nurse* than when it is preceded by an unrelated word such as *garden* (Meyer & Schvaneveldt, 1971; for further examples see De Groot, 1989; McNamara, 1992a, 1992b; Neely, 1991; Seidenberg, Waters, Sanders, & Langer, 1984).

Given the potential confounding effects resulting from associative relations between words or concepts across a host of psychological domains, there is a pressing need to have an index of strength of association between different words. We cannot distinguish whether inductions are stronger from *tiger* to *camel* than from *tiger* to *parrot* due to the domain-specific nature of the relation between the two categories or because one pair is simply more strongly associated than the other if we have not controlled for strength of association between them from the outset. The most commonly used method of generating association norms are so-called free associations, whereby people are asked to generate the first word that comes to mind upon encountering a specific category (Brown, 1976; Cramer, 1968; Nelson, McEvoy, & Schreiber, 1998). For example, Palermo & Jenkins (1964) asked for free associations for 200 words from several thousand participants. To illustrate, participants generated over 70 different words upon encountering *doctor*, however it was most commonly associated with *nurse* followed by *sick*, *health*, *medicine*, *hospital*, *man*, *sickness* and *lawyer* etc. The strength of association between two words was then indexed by the proportion of people who spontaneously named one upon encountering the other.

Apart from somewhat clouding our understanding of the psychological meaning of association, this methodology has several other conceptual and practical shortcomings.

Conceptually, Nelson, McEvoy & Dennis (2000) raise concerns about reliability and validity of free association indices. Although the response probabilities form a useful numerical hierarchy from strongly to weakly associated, the measurement does not include error estimations. This questions the reliability and objectivity of these association norms (Landauer & Dumais, 1997). Regarding validity, it is not entirely clear what psychological construct is supposed to be reflected by associative strength in free association paradigms. One interpretation is that association strength reflects the absolute strength of the memory trace between the two categories (Rosen & Russell, 1957). However, as noted by Nelson et al. (2000), the procedure by which the response probabilities are obtained is relative rather than absolute, as the more people produce the primary associate, the fewer people can produce one of the more weakly associated categories. So a weak associative strength rating could either be the result of a truly weak association between the target and a free associate, but could also be the result of one word having many strongly associated words. In the latter case, the richness of the network of associates would necessarily be inversely proportionate to the strength ratings, thus masking absolute associative strength.

Furthermore, the procedure only allows the indexing of a finite number of associations for any one category, based on the answers given by participants. Consequently, we only have association indices for relatively strongly associated word pairs, neglecting the lower end of the association scale. Thus, if the free association procedure did not result in the naming of one upon encountering the other, we have no means of estimating the strength of association between two concepts. For example, the primary free associate of *elephant* in the Birkbeck Word Association Norm Corpus (Moss & Older, 1996) is *trunk*. As none of their participants named *antelope* in response to *elephant*, there is no reference to the strength of association between these two categories, yet most people will agree that there is a definite and probably quite strong association between these two categories.

The above weakness also has some more serious conceptual shortcomings, such as clouding our understanding about whether it is possible to distinguish between semantic relations and associative relations. When free association methods fail to establish an association between two categories, several researchers (Foss & Ross, 1983; Lupker, 1984; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995) would define such word pairs as being purely semantically related but non-associated. They justify this with reference to the finding that non-associated word pairs as defined by free association norms, such as the category pair *pig* and *horse*, do not have priming effects. Thus, the majority of researchers in the priming field would agree that word pairs such as *pigs* and *horses* are exclusively semantically related (category coordinates of *'mammal'*) but would not be associatively related. However, again this seems to be largely a result of how 'association' is traditionally measured. People almost never spontaneously name *horse* in response to *pig*, leading to the conclusion that *pigs* and *horses* are not associated. However, intuitively it seems that *pigs* and *horses* would at least be weakly associated.

Also, more recent research has shown that there do not seem to be any variations in the magnitude of priming effects depending on free associative strength norms. For example, Anaki and Henik (2003) compared priming effects for strong primary associates (word pairs in which the target word was the primary response given with a mean frequency of 42%), weak primary associates (word pairs in which the target word was the primary response but with the lower frequency of 10%) and non-primary associates (word pairs in which the target word was not the most frequent response with a mean frequency of 10%). They found that the priming effect was of equivalent magnitude for the strong and weak primary associate word pairs (31 ms), but there was no priming effect at all for the weakly associated non-primary associates. This casts doubts on the claim that category relations can be strictly separated into ***either*** semantic ***or*** associative relations based on their priming properties.

Rather, it appears that the dominant free association method used to derived association norms has led to an unfortunately narrow definition of what 'association' means.

Summarizing, apart from the aforementioned direct and indirect theoretical issues arising from the use of the free association methodology, for our current purposes the practical shortcomings are more problematic. Thus, it does not enable us to index the extensive network of associations between categories[2], nor does it provide us with an estimate of the absolute strength of these links.

This problem can be addressed by approaches which incorporate measures of objective linguistic criteria, as associations between words tend to be reflected in how we use language. Going back to the definition of association in terms of temporo-spatial contiguity, one strategy of deriving an index of degree of association is to look at the frequency of co-occurrence between two words in a text or sentence. For example, Spence & Owens (1990) looked at the relationship between co-occurrence of 47 word pairs in the Brown corpus, which consists of diverse fiction and non-fiction from different areas and genres, and correlated this with strength of association ratings obtained from free association norms. Whilst they did find a significant correlation between strength of association and frequency of co-occurrence the study suffers from several fundamental shortcomings, ranging from the methodology by which the relationships were established, to the adequacy of the studied word corpuses. Firstly, simply counting the number of times two words co-occur does not address the fact that co-occurrence could be contaminated with the frequency of each individual word in the pair. For example, the word *dog* appears 7846 times in the British National Corpus, whereas *meerkat* only appears four times in the same corpus. A more adequate measure of associative co-occurrence is to calculate the 'conditional probability'

[2] For an example of the extent to which free association methodology can provide a reliable yet limited index of set size, see Nelson & Schreiber (1992)

that the word w_1 is followed by the word w_2. Heylighen (2001) suggests the following formula:

$$Aw_{1\&}w_2 = P(w_1 \mid w_2) = \frac{P(w_1 \& w_2)}{P(w_1)} = \frac{N(w_1 \& w_2)}{N(w_1)}$$

In this equation, $P(w_1 \& w_2)$ represents the probability that a text contains both words w_1 and w_2, $P(w_1)$ represents the probability that it contains w_1 on its own. To calculate the conditional probability, one can simply count the number of times w_1 and w_2 co-occur and divide this by the number of times w_1 occurs by chance in the same text sample. This is similar to Church and Hanks (1990) measure of 'mutual information' given by the following formula:

$$A'w_{1\&}w_2 = \log \frac{P(w_1 \& w_2)}{P(w_1) * P(w_2)}$$

One drawback of this version is that the conditional probability is equal from w_1 to w_2 and vice versa, masking potential asymmetries.

The second problem with studies like those carried out by Spence and Owens (1990) concerns the adequacy of the actual text sample itself, which often suffer from data sparseness. Ide and Veronis (1998) have claimed that corpuses like the Brown corpus (Kucera & Francis, 1967), which contains around 1 million words, are grossly insufficient for reliably measuring co-occurrence. Thus, Rapp and Wettler (1991) suggest that 10 million words is the absolute minimum for gaining a robust index of association from co-occurrence data. With the recent explosion in the availability of online media, prose, and non-fiction amongst many other types of genres, the World Wide Web provides an almost infinite source of written language. Alongside computer algorithms used by search engines such as *Google*, there is an enormous potential for exploiting these two tools for psycholinguistic research. To generate a thoroughly controlled set of stimuli for use in our psychological tasks we

a) checked that people did perceive the purported structural relations between categories

b) indexed the strength of association between category pairs irrespective of the nature of the relation.

For generating association indices, we adopted a dual-approach strategy, using both subjective association ratings from participants and the objective co-occurrence measure outlined above. As the common free association norms only index the association for a handful of the most commonly associated word pairs, we gave people two categories and asked them to provide a rating of associative strength. We then also obtained web-based associative strength indices. The degree to which the two measures correlate provides support for our claim that we are indeed measuring a construct of associative strength, in which the activation of one leads to activation of the other, irrespective of the nature of relation between the two.

2.1 Causal Beliefs Pre-Test

We catalogued people's beliefs about the nature and strength of relations between categories. Firstly, we generated 51 pairs of categories thought to have a causal connection, such as "*salmon* and *grizzly bear*", "*cabbage* and *snail*" or "*steel* and *cook*".

Participants were told that we were interested in their beliefs about how strong the causal link between these two categories was. Instructions stated that this could include any kind of causal connection between the categories. They were given two explicit examples:

> *Properties of "acorns", such as enzymes or any pollutants it may contain, might be transmitted to "squirrels" because the latter category feeds on the former category.*

Similarly, "clay" and "vases" may be related because "vases" can be made of "clay".

10 Durham University students rated the strength of the possible causal link between two categories on a scale from 1 (unrelated) to 9 (very strong causal relation). They were instructed to give a rating closer to 1 if they thought that the causal link was very weak, and a rating closer to 9 if they thought that the causal link was very strong. The order in which the words were presented was counterbalanced across participants.

2.1.1 Results Causal Pre-test

The overall mean rating for the causally related category pairs was 4.9. A mixed-model ANOVA with item as the repeated-measures variable and direction (predictive or diagnostic order of presentation) as the between-subjects variable showed that there were significant differences for the different items, $F_{(50, 400)} = 11.36$, $p < .0005$.

There was no main effect of order $F_{(1, 8)} = 2.86$, $p = .13$. Thus, categories presented in a predictive direction (e.g. *carrot → rabbit*) received a causal strength rating of 4.6 (*SE* = .33) whereas when the order was reversed and diagnostic (e.g. *rabbit → carrot*) people gave a mean causal strength rating of 5.3 (*SE* = .33).

However, there was a significant interaction between direction and item, $F_{(50, 400)} = 1.98$, $p < .0005$, so to be sure that we selected items in which people did perceive a causal connection and to ensure that there were no directional differences, we looked at each individual item by comparing the ratings in both orders with independent-samples t-tests and also carried out one-sample t-tests with a test value of 1 (equivalent to a rating on the scale of no causal link).

The independent-samples t-test showed that direction only had a significant effect for 2 items, *tree* → *book*, and *mercury* → *fisherman*. Similarly, causal strength ratings did not differ significantly from 1 for two pairs, *acorn* → *cat* and *grass* → *sweater*. For the remaining category pairs, people perceived a causal link between all category pairs, albeit of varying strengths. The mean causal strength ratings for all pairs are shown in Appendix 1A.

2.2 Strength of Association

2.2.1 Subjective Association Ratings

Next we assessed people's beliefs about how strongly category pairs were associated. Given that we suggest that associative knowledge exerts its influence relatively automatically without the need for explicit reasoning or activation of a declarative knowledge system (Dienes & Berry, 1997), asking people to provide explicit association ratings has some obvious pitfalls. Thus, people may base their responses on a thorough analysis of potential relations between the categories, thus drawing on structured rather than automatic associative knowledge. However, several precautions were taken to try and counter these weaknesses. Firstly, research has shown that putting people under time pressure increases their reliance on automatic beliefs (Evans & Curtis-Holmes, 2005), as do instructions to respond intuitively (De Neys & Franssens, 2009). Thus, we told participants that the experiment was timed and that they should provide the first intuitive answer that came to mind, responding as fast as possible.

Method

Participants and Design

18 participants from Durham University were given a list with 270 category pairs and asked to rate the strength of association between the pairs. The order in which the two categories were presented was counterbalanced across participants.

Materials and Procedure

Firstly, we generated 49 pairs of categories thought to have a causal connection, such as "*salmon* and *grizzly bear*", "*cabbage* and *snail*" or "*steel* and *cook*" and for which we had verified that people did perceive a causal link (see previous section). For each of these causal pairs, we generated a control pair thought to have no immediate discernable relationship, for example "*carrot* and *fox*" or "*swallow* and *flower*". We also generated 3 or more alternative taxonomically related category pairs for each causal pair, for example "*salmon* and *goldfish*", "*salmon* and *herring*", "*grizzly bear* and *wolf*", "*cabbage* and *cauliflower*", *snail* and *squid*" etc. In total, participants were asked to rate 270 word pairs.

We told participants that we were interested in their beliefs about how strongly associated these two categories were. They were asked to think about all kinds of possible associations and although they were given examples which included structured relations, we emphasized that we wanted them to give their first intuitive response:

> *Please think about all kinds of possible associations, such as causal, functional, categorical, etcetera. Please do* ***not*** *think in detail about the mechanism by which they are related, just give your* ***intuitive*** *response. For example, if you believe that ladybirds and butterflies are strongly associated please give a rating closer to 9. In contrast, if you think cars and ladybirds are unrelated, please give a rating closer to 1. Please give the answer that first comes to mind, as* ***fast*** *as possible.*

Participants were asked to rate how strongly the category pairs were associated on a scale from 1 (unrelated) to 9 (very strongly associated). The order in which the two categories were presented was counterbalanced across participants.

2.2.2 World Wide Web Conditional Co-Occurrence

Given the aforementioned potential pitfalls of asking people for explicit ratings, and the fact that the instructions made some reference to structured relations, we wanted to compare our subjective ratings to a more objective association criterion. Thus, for all the category pairs for which we had obtained participant association ratings, we examined their co-occurrence frequencies in natural language. To this end, we conducted a proximity search on two different search engines, the World's largest and most popular search engine Google, as well as the less well-known Exalead, which explicitly supports co-occurrence search queries[3]. We specified that the categories must co-occur in any order within a window of 6 consecutive words. This fairly stringent window was adopted to maximize the likelihood that the two words would co-occur within the same sentence and also to reflect the fact that we were primarily interested in associations that are activated automatically in psychological tasks. Priming studies (Canas, 1990) have shown that priming between associated words occurs mainly at short stimulus onset asynchronies (SOAs), suggesting that potentiation decays with increasing time or distance between the word pairs. Our narrow window can also be justified by assuming that words which co-occur more frequently within close proximity to one another are more likely to be more strongly associated than words which co-occur over a larger distance. Strong association ratings derived from our subjective measure are hence more likely to reflect associations between words which co-occur in fairly close proximity.

Once we derived the co-occurrence measure, we recorded the Google and Exalead word frequency of each individual category. Using Heylighen's (2001) formula, we calculated the conditional co-occurrence. As we were not interested in potential asymmetries, we calculated the mean conditional co-occurrence for each word pair and correlated this measure with the

[3] For further details about the two search engines, as well as the precise search queries entered, please refer to Appendix 1C

mean associative strength ratings. The standardized co-occurrence Z-scores for the 270 categories can be found in Appendix 1B.

2.2.3 Results: Strength of Association

The mean association ratings for the 270 category pairs can be found in Appendix 1B. Mean ratings ranged from 1.3 (*SD* = .62) for the pair "*carrot* and *fox*" to 8.3 (*SD* = .83) for the category pair "*rabbit* and *hare*".

Co-occurrence ranged from once for the pair "*acorn* and *lychee*" to 12500000 times for the category pair "*paper* and *book*".

Table 2.1: Correlations between Mean Association Ratings and Co-Occurrence Indices

		Z-score Google Co-Occurrence	Z-score Exalead Co-Occurrence
Mean Association Rating	Spearman's rho	.56**	.66**
	N	270	270
Z-score Google Co-Occurrence	Spearman's rho		.74**
	N		270

** Correlation is significant at the 0.0005 level (2-tailed).

As can be seen from Table 2.1, the correlations between the mean conditional co-occurrence measures and peoples' subjective associative strength ratings were highly significant. We attributed our construct of psychological association as arising from the spatial and temporal contiguity of categories within the environment. As the web-based data gives us a more objective indication of the actual co-occurrence, these strong correlations between the objective measures and subjective ratings seem to validate our method for estimating people's psychological perception of associative strength between categories. It suggests that we were successful in tapping one type of unstructured knowledge construct, i.e. that of strength of association between categories.

2.3 Discussion

The goal of the current studies was to create a database of stimulus materials by indexing properties such as causal strength and strength of association. To date, most stimulus materials used in studies investigating the role of knowledge in reasoning have relied on normative structures governing relations between categories from a specific domain, such as scientific taxonomies, or have simply used idiosyncratic criteria for selecting categories, such as living in the same habitat or similar causal relations. However, whilst it is impossible to ensure that all participants will have the same degree of knowledge about the various interrelationships between categories, it is still feasible to create materials that have been equated along certain dimensions, such as degree of association or causal strength. This is especially important when claiming that knowledge from one domain is superior or more easily available than other domain-specific knowledge. As our review makes clear, probably the most important but thus far wholly neglected dimension along which stimulus materials ought to be equated is strength of association between the category pairs from different domains. For example, if people assign a higher inductive strength rating to the inference from *carrots* to *potatoes* (taxonomically related category pair) than from *carrots* to *rabbits* (causally related category pair), one cannot claim that this reflects the superiority of taxonomic knowledge compared to ecological or causal knowledge, as the former pair are more strongly associated (mean association rating of 7.28) than the latter pair (mean association rating of 6.06). Thus, the higher inference rating could simply reflect the stronger association people have between the taxonomically related pair, rather than being due to status or availability differences in knowledge structures between the two category pairs. Had the researcher instead asked participants to make the taxonomic inference from *carrots* to *fennel,* which has a lower mean association rating of 4.64, the results might well have been reversed, with higher inductive strength ratings attached to the causally related category pair.

Obtaining association norms for an array of category instances is not a trivial task, especially from a methodological point of view. The first problem is that current databases only index a finite number of usually fairly strongly associated words for each target. This is a by-product of using the free association method (Brown, 1976; Cramer, 1968; Moss & Older, 1996; Nelson, et al., 1998; Palermo & Jenkins, 1964), in which participants name the first word that comes to mind upon presentation of a base word. Similarly, it provides only a relative rather than an absolute rating of associative strength (Nelson, et al., 2000). Although some advances have been made by using a continuous rather than discrete free association methodology, in which people are asked to generate more than one associated word (De Deyne & Storms, 2008a, 2008b)[4], it cannot index word pairs which lack an obvious association. Thus, there is still a restriction of range when it comes to selecting stimuli that cover the complete associative continuum from unrelated to highly associated.

However, one clear advantage of the discrete free association method is that the first-named category is most likely to reflect the operation of automatic associative retrieval processes from long-term memory rather than a more controlled and effortful strategic retrieval processes that might necessitate recruitment of more structured knowledge (Rosen & Engle, 1997). The risk of asking participants for additional associates, as in the continuous free association method, or asking for explicit association ratings, as in our study, is that responses might shift progressively towards reflecting structured rather than associative knowledge activation. However, we circumvented this by using experimental strategies that

[4] For potential pitfalls of using this methodology, see McEvoy & Nelson (1982)

interfere with or dissuade the use of effortful processing, such as putting people under time pressure (Evans & Curtis-Holmes, 2005; Roberts & Newton, 2001; Shafto, Coley, et al., 2007) and encouraging them to respond intuitively (De Neys & Franssens, 2009). In addition, we tried to verify our subjective association measure against an external and more objective measure, such as co-occurrence of concepts in natural language. The fact that there were highly significant correlations between the co-occurrence measures themselves and between the co-occurrence measures and subjective association ratings suggests that our rating tasks tap a construct that is associative and automatic in nature, rather than reflecting a carefully deliberated assessment of a structured relation between category pairs.

In addition, in all experiments reported in subsequent chapters, we ensured that we could dissociate people's structured knowledge from their unstructured associative knowledge for the category pairs used in the inductive tasks by assessing people's structured knowledge in a post-test. Participants were asked to explicitly state whether they thought that the category pairs were taxonomically and/or causally related. We then calculated taxonomic and causal endorsement proportions across the category pairs. If this measure of structured knowledge correlates with associative strength ratings (i.e. the more people believe that a category pair is taxonomically and/or causally related, the higher the association ratings for this pair) then we cannot claim to be measuring two types of knowledge that each contribute independently to inductive reasoning. Rather, in this case associative knowledge might merely reflect structured knowledge that has become so well learnt that it no longer requires effort to recall. In contrast, if associative strength ratings do not correlate with the measure of structured knowledge, we can be more confident that we are measuring dissociable sources of knowledge which might have differential effects on category-based inductive reasoning.

The notion that there are different types of knowledge underpinned by contrasting processes was suggested by Moscovitch (1995) in the context of information retrieval from long-term memory. On the one hand, associative memory, mediated by hippocampal brain region, allows people to retrieve information automatically. On the other hand, retrieval of information from semantic memory can be strategic, in which a cue serves as input to a thorough, effortful process akin to problem-solving. Such strategic memory retrieval is mediated by prefrontal regions of the brain and semantic memory in areas of the Lateral Anterior Temporal Lobe (Patterson, Nestor, & Rogers, 2007). As Sloman (1998) has already suggested, it is feasible that factors such as available time, mental resources and intentions/motivation will influence the extent to which people will engage in more effortful retrieval of knowledge or rely on simple, cognitively economic associations to guide their category-based inductions. The next chapter will use the pre-tested stimulus materials to explore whether unstructured and structured knowledge sources can have dissociable effects on category-based inductive inferences, in line with their purported unique processing characteristics.

Chapter III

Two Types of Knowledge in Category-Based Induction

Inductive inferences do not occur in isolation, but are made in a complex and uncertain world. Whereas their similarity to a social stereotype might support inferences about a new acquaintance, inferences about possible health hazards rely much more on knowledge about the causal relations between disease agents and organisms. However, as argued in the first two chapters, focusing predominantly on domain-specific category relations has masked the importance of another distinction, namely the distinction between structured and unstructured types of knowledge. For example, people may make the automatic inference that the new Muslim doctor is threatening, owing to the association between acts of terrorism and radical Muslim groups repeatedly highlighted in the media. On the other hand, applying more elaborate structured knowledge, such as the fact that doctors tend to be highly educated and intelligent, might lead to a positive revision of the initially biased inference.

The cognitive processes that underpin our ability to use different kinds of knowledge flexibly depending on context are poorly understood. Furthermore, work to date has predominantly focused on differences between domain-specific category relations. For example, the study by Shafto, Coley and Baldwin (2007) suggests that context selectively primes the use of knowledge from different domains. Thus, reasoning about diseases seems to prime the use of ecological knowledge, whereas reasoning about genes should activate taxonomic knowledge. Shafto et al. (2007) also contend that different domains of knowledge may vary in their relative accessibility and thus the likelihood that this knowledge will come to bear on the final inference. In their study, they showed that taxonomic knowledge was unaffected by manipulating the time people had to make their inferences, whereas access to ecological knowledge seemed to be restricted when participants were put under time pressure. They therefore concluded that people selectively use knowledge from different domains when they have sufficient time to do so, but fall back onto using privileged taxonomic knowledge when under time pressure.

Shafto et al.'s (2007) paradigm can also be used to test our alternative framework which emphasizes the processing differences between structured and unstructured knowledge. As described in the two previous chapters, we define structured knowledge as the theoretical framework that explains how two categories are related, for example taxonomic knowledge (e.g. the relation between *carrots* and *bamboo* → both are species of plants) or causal knowledge (e.g. the relation between *carrots* and *rabbits* → part of the same food chain). Unstructured knowledge on the other hand is defined as the degree to which the two categories are associated, co-occur or are similar to one another regardless of the nature of this relationship.

We hypothesize that when people have plenty of time and cognitive resources, they are able to effortfully process the structural relations between the categories. In contrast, people would be more likely to rely on unstructured knowledge, such as strength of association, as a useful heuristic cue to inductive strength when they are under time pressure or have to contend with a cognitive load. Thus, the major goal of the first experiment was to use resource manipulations to demonstrate that structured and unstructured knowledge can have dissociable effects on people’s category-based inferences.

A promising avenue for generating evidence that the effects of structured and unstructured knowledge can be dissociated is to look at reasoning phenomena that arise from assessing the underlying structural relations between categories. Thus, such phenomena should only arise when people use structured knowledge, but should be absent when people rely on unstructured knowledge. Looking at the role of causal knowledge in inductive reasoning is a prime candidate due to its complex and directional structure (Sloman, 2005), affecting people’s inductive evaluations of category-based arguments. For example, causal transmission of properties is more likely to occur from prey to predator than vice versa (for a Bayesian justification and specification of prior probabilities, see Shafto et al., 2008). People tend to make stronger inferences when the relation is ecologically consistent and predictive than when the causal direction is reversed and diagnostic (Medin, et al., 2003; Tenenbaum, Kemp, & Shafto, 2007).

For causal knowledge to be maximally inductively potent, and manifest itself in phenomena such as the aforementioned causal asymmetry effect, each object or concept has to be assigned an appropriate relational role, suggesting that semantic processing under such conditions may involve more than mere spreading activation within a semantic network (Spellman, Holyoak, & Morrison, 2001) or strength of nondirectional associations (Rehder, 2009). Rather, taking causal structure into account is likely to be based on a more effortful, analytical process which takes time and available mental resources to accomplish.

To test this hypothesis, the following experiment adopts the same time-manipulation paradigm used by Shafto et al. (2007), but some changes were made to both material selection and experimental design. Firstly, we always used the same base categories to form causal and taxonomic pairs. Secondly, based on the pre-test results reported in Chapter 2, we selected category pairs for which we had stringently assessed the strength of association to ensure that causally and taxonomically related category pairs did not differ in their level of association. Finally, we manipulated property within rather between subjects. This should encourage participants to use their knowledge flexibly and remain attentive rather than implementing an automatic overall strategic set (Markman & Gentner, 1993), potentially amplifying selectivity and accessibility effects.

3.1 Experiment 1

The first goal was to test our hypothesis that structured and unstructured knowledge have different effects on category-based induction depending upon available mental resources. Experiment 1 included three types of relations between categories, causal predictive, causal diagnostic and taxonomic, covering the whole range of association levels. People were asked to reason either under time pressure or were forced to delay their responses. The arguments included two different properties, cells and infections. We predicted a causal asymmetry effect, that is, causal inferences should be stronger for categories with a predictive causal link than for those with a diagnostic causal link. However, if the retrieval of a causal knowledge structure and assignment of causal roles is indeed mediated by an effortful process, one might expect this causal asymmetry effect to be weakened when people are under time pressure, in which case they should fall back on unstructured knowledge in the form of non-directional strength of association between the two categories.

Furthermore, we used our measure of unstructured knowledge (strength of association) as well as indices of structured knowledge (beliefs about taxonomic and causal relatedness) to predict people's inferences under contrasting timing conditions. As we believe that access to structured knowledge requires slower and more elaborate processing, it was expected that people would rely more on unstructured knowledge when under time pressure. Thus, unstructured knowledge, i.e. associative strength, ought to predict people's inductive strength ratings for both cells and infections when they were under time pressure. In contrast, under delayed conditions, structured taxonomic knowledge should also be predictive of people's inferences about cells, whereas structured causal beliefs ought to predict people's inferences about infections.

Finally, our causally related categories are analogous to what Shafto et al. (2007) call ecologically related categories. By collapsing across our two types of causal relations (predictive and diagnostic), our design also allowed us to follow-up Shafto et al's (2007) claim that there is a difference in accessibility of knowledge from different domains once we have controlled for level of association between causally and taxonomically related category pairs.

3.1.1 Method

Design

The experiment had a 3 (relation: causal predictive, causal diagnostic, taxonomic) by 2 (timing: speeded or delayed) by 2 (property: cells or infection) by 4 (list) mixed design, with list and timing as between-subjects manipulations.

Participants

Participants were 60 Durham University students (mean age = 20.7 years, *SD* = 4.0 years, 42 females and 18 males) who took part in the experiment in return for course credit.

Materials and Procedure Inference Task

In total there were 16 items. For each item, we created four induction problems: A causal predictive argument and its taxonomic counterpart, as well as a causal diagnostic argument and its taxonomic counterpart.

In predictive causal induction problems people were told that the base category (prey/plant) had either a novel infection or novel cells, such as *infection 9Tye4*, or blank cells, such as *Lo78-cells*. They then evaluated the likelihood that this property was also present in the conclusion category (predator/consumer). The taxonomic counterpart was created by keeping the same prey/plant base category but replacing the conclusion category with a taxonomically related category of equivalent associative strength.

As a variation of causal structure, for each of these items we then reversed the order of the categories for the predictive causal problems (base category: predator/consumer → conclusion category: prey/plant), resulting in a diagnostic causal induction problem. This was matched again with a taxonomic induction problem by keeping the predator/consumer base category but substituting the conclusion category for a taxonomically related alternative. Thus, there were a total of 64 problems. For example, participants might be presented with one of the following induction problems:

Mice have infection 9TT7. (causal predictive/ infection)

How likely is it that Eagles have infection 9TT7?

Mice have infection 7rR4. (taxonomic / infection)

How likely is it that Squirrels have infection 7rR4?

Eagles have 45T-cells. (causal diagnostic/ cells)

How likely is it that Mice have 45T-cells?

Eagles have e2T-cells. (taxonomic / cells)

How likely is it that Parrots have e2T-cells?

Causally related targets were always from different superordinate categories, for example, *plants* and *animals*, or *mammals* and *reptiles*. In contrast, taxonomically related pairs were always from the same superordinate taxonomic category. There were two exceptions in which the causally related categories were both mammals but belonged to different orders. In these cases, the taxonomic alternative was from the same order as the base category.

All our stimulus materials had been extensively pre-tested to ensure that people did perceive a causal link between the causally related category pairs. Similarly, based on the pre-test ratings of strength of association between category pairs (full details and procedure are described in Chapter 2), we ensured that the strength of association was identical for the predictive category pairs and their taxonomic counterparts, as well as for the diagnostic category pairs and their taxonomic alternatives. Across the 64 unique category pairs, the magnitude of the correlation between the subjective measure of associative strength and the

Google web-based measure of conditional co-occurrence was significant, Spearman's rho = .55, $p < .0005$.

Participants rated the likelihood that the target category shared the infection or cells on a 9-point scale. A different infection/cell was used for each problem. For each problem, participants reasoned either about the predictive or diagnostic causal category pairs and the taxonomically matched category pairs. This meant that each participant rated 32 inductive arguments. The items for which participants reasoned about cells or infections in a predictive or diagnostic direction was counterbalanced in an incomplete Latin square design, resulting in four lists. We tried to ensure that there were an equal number of strongly and weakly associated items in both the cell and infection condition. This design is shown in Table 3.1 below.

Table 3.1: Design Stimulus Materials for Experiments 1 and 2

	List 1	List 2	List 3	List 4
Items 1-4	Causal Predictive and Taxonomic Arguments *Cells*	Causal Diagnostic and Taxonomic Arguments *Cells*	Causal Predictive and Taxonomic Arguments *Infections*	Causal Diagnostic and Taxonomic Arguments *Infections*
Items 5-8	Causal Diagnostic and Taxonomic Arguments *Cells*	Causal Predictive and Taxonomic Arguments *Cells*	Causal Diagnostic and Taxonomic Arguments *Infections*	Causal Predictive and Taxonomic Arguments *Infections*
Items 9-12	Causal Predictive and Taxonomic Arguments *Infections*	Causal Diagnostic and Taxonomic Arguments *Infections*	Causal Predictive and Taxonomic Arguments *Cells*	Causal Diagnostic and Taxonomic Arguments *Cells*
Items 13-16	Causal Diagnostic and Taxonomic Arguments *Infections*	Causal Predictive and Taxonomic Arguments *Infections*	Causal Diagnostic and Taxonomic Arguments *Cells*	Causal Predictive and Taxonomic Arguments *Cells*

The inductive arguments were presented on a laptop in red font. Responses were only accepted once the font turned green. In the delayed condition, the font turned to green after 10 seconds, whereas in the speeded condition, participants were encouraged to respond after 1 second. They entered their response on the keyboard by pressing any number between 1 (highly unlikely) and 9 (very likely).

Post-test

The post-test assessed people's beliefs about taxonomic and causal relatedness. For each of the 32 category pairs (8 predictive, 8 diagnostic, 16 taxonomic) about which people had made inferences in the main task, they were asked two questions, resulting in a total of 64 questions. One question asked them whether they believed that the two categories were from the same biological class. The second question prompted them to state whether the two categories were part of the same food chain. Participants could respond with YES, NO or DON'T KNOW, but were instructed to use the third option sparingly, as the emphasis was on their intuitions and beliefs rather than on factual correctness.

We calculated the mean proportion of positive responses to the two post-test questions about biological group membership and food chain relations across the two timing conditions and correlated these with both our web-based measures of co-occurrence (Spearman rho correlation coefficients ranged from -.1 to .06 all p's > .66), and our subjective measure of associative strength (Spearman rho correlation coefficients ranged from -.008 to .17 all p's > .17). With one exception (correlation between food chain beliefs in the speeded condition and subjective strength of association rating, r = .3, p = .016), none of the other correlations were significant, supporting the idea that the post-test is not measuring associative strength but beliefs based on more structured knowledge. However, results will be interpreted with this significant correlation in mind.

Procedure

Participants were told that we were interested in their beliefs about shared infections and cells. They received written instructions which included an example and enabled them to give informed consent. Apart from the response-input constraints, the experiment was self-paced. The order in which the problems were presented was completely randomized. Participants read the problems and once the font had changed from red to green, participants entered their response on the keyboard by pressing any number between 1 and 9.

Following the main experiment, participants read instructions for the post-test. The order in which participants received the post-test questions was completely randomized and there were no response-time constraints.

3.1.2 Results

ANOVA

For each participant, we calculated 6 mean inductive strength ratings, reflecting the 6 conditions created by having 3 types of relation for which people reasoned about cells and infections. Mean inductive strength ratings were analysed with a 3 (relation: predictive, diagnostic or taxonomic) by 2 (property: cells or infection) by 2 (timing: speeded or delayed) by 4 (list) mixed-design ANOVA, with timing and list as between-subjects variables. For the item analysis, inductive strength ratings were averaged across subjects rather than items and were analysed with a 2 (property) by 2 (timing) by 3 (relation) mixed-design ANOVA, with relation as a between-items variable. The continuous measure of strength of association between category pairs was included as a covariate in all the item analyses.

There was no significant main effect of list, $F_{(3, 52)} = .63$, $p = .6$, effect size d = .38. It did interact with property and relation, $F_{(6, 104)} = 3.11$, $p = .008$, effect size d = .85, so this will be commented upon later. The covariate of strength of association in the item analysis was highly significant, $F_{I\ (1,\ 60)} = 77.91$, $p < .0005$, effect size d = 2.28, but it did not interact with any of the other variables.

There was a significant main effect of relation, $F_{S\ (2,\ 104)} = 3.67$, $p = .029$, effect size d = .54, $F_{I\ (2,\ 60)} = 7.13$, $p = .002$, effect size d = .98. Post-hoc comparisons showed that there was no difference in inductive strength ratings for causal predictive (*M* = 4.27, *SE* = .16) and diagnostic inferences (*M* = 4.12, *SE* = .15, *p* = .66), nor was there a significant difference between causal predictive and taxonomic inductive strength ratings (*M* = 4.59, *SE* = .14, *p* = .39). The difference between diagnostic and taxonomic inferences was marginally significant (*p* = .051).

There was a significant main effect of property, $F_{S\ (1,\ 52)} = 5.73$, $p = .02$, effect size d = .66, such that inferences about infections (*M* = 4.50, *SE* = .12) were rated stronger than inferences about cells (*M* = 4.15, *SE* = .14). However, this was not significant across items, $F_{I\ (1,\ 60)} = 1.08$, $p = .3$, effect size d = .27.

Inferences drawn under speeded conditions (*M* = 4.36, *SE* = .15) were virtually identical to inferences made under delayed conditions (*M* = 4.29, *SE* = .15), $F_{S\ (1,\ 52)} = .11$, $p = .74$, effect size d = .04, $FI_{(1,\ 60)} = .49$, $p = .49$, effect size d = .18.

There was a significant interaction between property and relation, $F_{S\ (2,\ 104)} = 8.15$, $p = .001$, effect size d= .79, $F_{I\ (2,\ 60)} = 15.87$, $p < .0005$, effect size d = 1.46. Illustrated in Table 3.2 below and confirmed by Bonferroni post-hoc comparisons, taxonomic inferences about cells were rated significantly stronger than both causal predictive inferences (*p* = .012) and causal diagnostic inferences, (*p* < .0005), whereas there was no difference between the latter two types of relation (*p* = 1). In contrast, when reasoning about infections, there were no differences between the three types of inferences (all *p*’s > .93).

Furthermore, as mentioned earlier, there was a significant interaction between list, property and relation, however, this seemed to be driven predominantly by differences in the magnitude of the interaction. Across all lists, for inferences about cells, taxonomic inferences were higher than both causal predictive and causal diagnostic inferences, although many of the pairwise comparisons did not quite reach statistical significance. For inferences about infections, the differences between the three types of relations were fairly variable across the four lists, but none of the differences reached significance (p's > .1). As the interaction effect between property and relations was very robust across items, the significant three-way interaction between property, relation and list across participants is probably due to differences in people's knowledge about category relations across the different lists.

Table 3.2: Inductive Strength Ratings broken down by Property and Relation in Experiment 1

Property	Relation	Mean	Std. Error
Cell	predictive	3.93	.22
	diagnostic	3.81	.19
	taxonomic	4.72	.16
Infection	predictive	4.62	.20
	diagnostic	4.43	.15
	taxonomic	4.47	.16

None of the other higher-order interactions were significant (all p's > .11).

Causal Asymmetry Effect

We had made the a-priori prediction that context-sensitive reasoning might be compromised when people do not have sufficient time to recruit relevant structured knowledge. This ought to be especially evident when the application of such structured knowledge leads to reasoning phenomena such as the causal asymmetry effect, whereby people rate causal predictive inferences (from prey to predator or plant to consumer) stronger than causal diagnostic inferences (from predator to prey or consumer to plant). Thus, we explored how causal inferences were affected by timing manipulations with a 2 (causal relation: predictive versus diagnostic) by 2 (timing) mixed-design ANOVA. As the causal asymmetry effect should only hold in a relevant context, such as when reasoning about infections or diseases, we conducted separate ANOVAs for infections and cells.

Causal Asymmetry Infections

The covariate of strength of association in the item analysis was again highly significant, $F_{I\,(1, 29)} = 47.87$, $p < .0005$, effect size d = 2.52, but it did not interact with any of the other variables.

When reasoning about infections, there was no main effect of relation $F_{S(1,58)} = .87$, $p = .46$, effect size d = .25, $F_{I(1, 29)} = 1.86$, $p = .18$, effect size d = .5, nor a main effect of timing, $F_{S(1,58)} = .31$, $p = .58$, effect size d = .14, $F_{I(1, 29)} = .08$, $p = .87$, effect size d = .06.

However, the crucial result was a significant interaction between timing and relation, $F_{S\,(1, 58)} = 6.88$, $p = .011$, effect size d = .69, $F_{I\,(1, 29)} = 4.97$, $p = .034$, effect size d = .82. Further Bonferroni post-hoc analysis confirmed that in a relevant context, i.e. when reasoning about infections, people who were not under any time pressure showed a robust causal asymmetry effect. Thus, people made significantly stronger inferences about categories with a causal predictive relation (M = 4.78, SE = .27) than for categories with a diagnostic causal link (M =

4.10, *SE* = .22, *p*= .015, effect size d = .47). In contrast, the causal asymmetry effect when reasoning about infections completely vanished when people were under time pressure. Thus, their causal predictive inferences (*M* = 4.44, *SE* = .27) were the same as their causal diagnostic inferences (*M* = 4.77, *SE* = .22, *p* = .24, effect size d = .21). This interaction is illustrated in Figure 3.1 below.

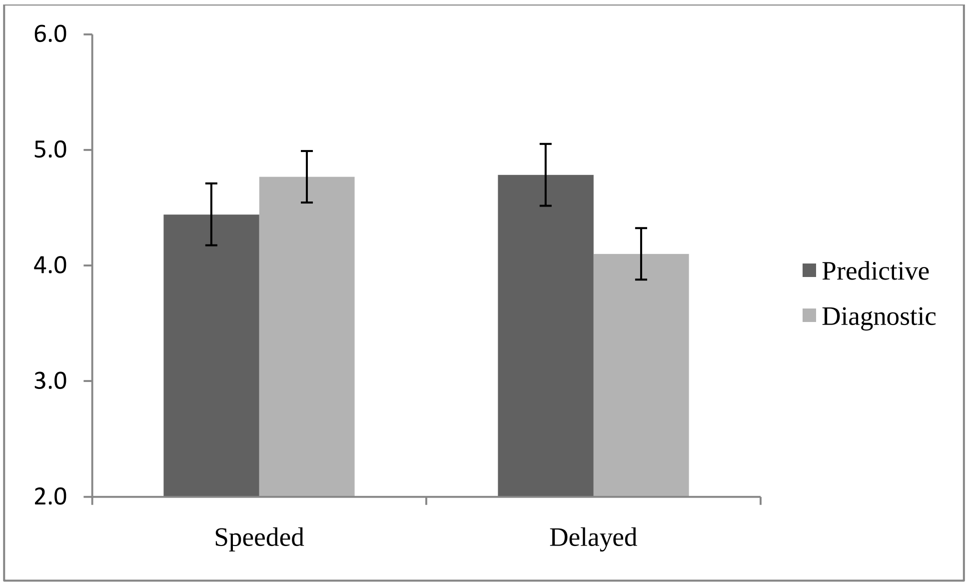

Figure 3.1: Causal Asymmetry Effect across Timing Conditions when reasoning about Infections

Causal Asymmetry Cells

As before, the covariate of strength of association in the item analysis was significant, $F_{I\ (1,\ 29)}$ = 5.96, *p* =. 021, effect size d = .91, but it did not interact with any of the other variables.

When reasoning about cells, causal relation had no effect on people's inductive strength ratings, $F_{S(1,58)}$ = .63, *p* = .43, effect size d = .21, $F_{I(1,\ 29)}$ = .43 , *p* = .52, effect size d = .24, nor did timing, $F_{S(1,58)}$ = .69, *p* = .41, effect size d = .22, $F_{I(1,\ 29)}$ = .007 , *p* = .94, effect size d = .001. There was also no significant interaction between timing and causal relation, $F_{S\ (1,\ 58)}$ = .41, *p* = .52, effect size d = .16, $F_{I\ (1,\ 29)}$ = .84, *p* = .37, effect size d = .34.

The fact that we observed a substantial causal asymmetry effect when reasoning about infections but not about cells confirms that people draw selectively on relevant knowledge. More importantly though, the absence of the causal asymmetry effect for inferences about infections when people were under time pressure strongly supports our hypothesis that drawing on structured causal knowledge as a guide to inductive strength appears to be a slower, more deliberate and analytic process. Under time pressure, people may simply rely on the existence of a causal association between the two categories, regardless of the directional nature of this link.

Regression Analyses

To explore how structured and unstructured types of knowledge influence category-based inductions under different timing conditions, we calculated mean inductive strength ratings for each item separately for the two types of property and timing conditions, resulting in 4 inductive strength scores for each item. Similarly, for each item we calculated the mean proportion of positive responses to the two post-test questions about biological group membership and food chain relations across the two timing conditions.

Unfortunately, beliefs about food chain relations in the speeded condition was significantly correlated with our strength of association measure ($r = .30$, $p = .016$), so this particular result needs to be interpreted with care. All 11 remaining correlations between our measures of structured knowledge and the measures of unstructured knowledge, i.e. associative strength and the two objective conditional co-occurrence indices, were non-significant (r's between -.008 and .17, all *p*'s > .17).

Hierarchical regression analyses were carried out on the mean inductive strength scores. As Shafto et al (2007) demonstrated that different knowledge is relevant in different contexts, we carried out the analyses separately for cells and infections. However, we make the theoretical assumption that people will be influenced by strength of association regardless of timing manipulations and property. Hence, we entered the associative strength variable in block 1. In a second block, we added proportion of positive responses to the biological group question and food chain question as the independent predictor variables. This enabled us to evaluate the degree to which adding variables reflecting structured knowledge accounted for additional variance above and beyond strength of association, our measure of unstructured knowledge.

Inductions about Infections

Reasoning about infections under no time pressure (R = .69) was significantly predicted by strength of association (beta = .53, t = 5.04, p <.0005). Of the structured knowledge variables entered in the second block, both knowledge about relevant causal food chain relations (beta = .45, t = 2.31, p = .025) and taxonomic beliefs (beta = .52, t = 2.69, p = .009) were significant predictors. Together, adding these two structured knowledge variables accounted for a significant amount of additional variance above strength of association alone (R^2 Change: 6.3%, $F_{(2, 60)} = 3.62$, $p = .033$).

Reasoning about infections under speeded conditions (R = .71) was also significantly predicted by associative strength (beta = .58, t = 5.29, p <.0005). The structured knowledge variables were not statistically significant predictors of inductive strength under speeded conditions (biological relatedness: beta = .33 t = 1.72, p = .091; relevant knowledge about food chain relations: beta = .33, t = 1.65, p = .10). Furthermore, these two structured knowledge predictors did not account for any additional variance compared to strength of association on its own (R^2 Change: 2.6%, $F_{(2, 60)} = 1.52$, $p = .23$). The beta coefficients for the two timing conditions are illustrated in Figure 3.2 below.

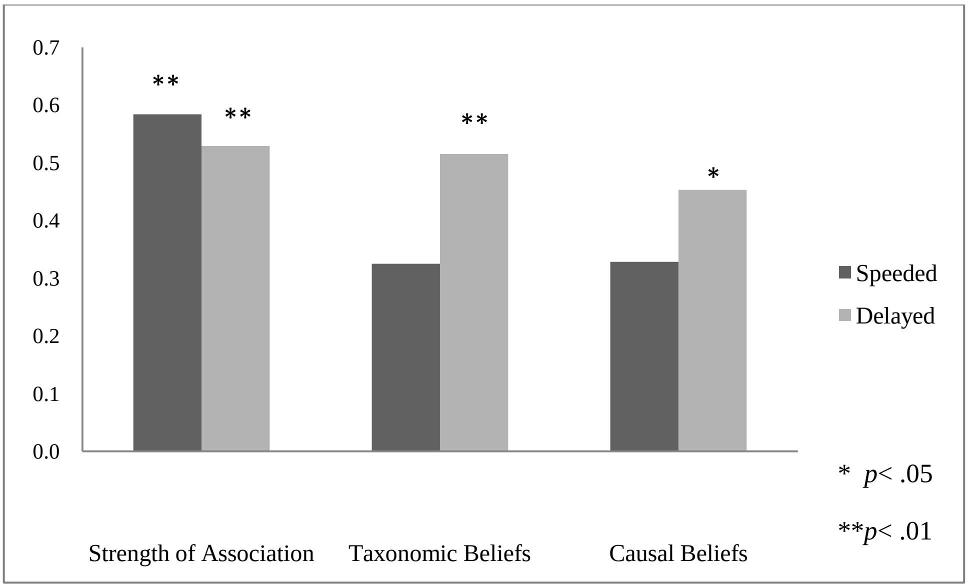

Figure 3.2: Standardized Regression Coefficients for Predictive Relations between Strength of Association, Taxonomic and Causal Beliefs and Inductive Strength Ratings for Infections

Inductions about Cells

A different pattern of variables predicted reasoning about cells, shown below in Figure 3.3. When people were forced to delay their responses, the multiple R was .82, meaning that 67.2% of the variance was accounted for by the three variables entered into the predictive equation. Strength of association (beta = .37, t = 4.43, p <.0005) and beliefs about biological relatedness (beta = .88, t = 2.69, p < .0005) were both significant predictors of inductive strength ratings across items, whereas the food chain belief variable was not a significant predictor (beta = .23, t = 1.49, p = .14). Adding the two structured knowledge variables added significant predictive power to the equation, accounting for an additional 48.2% of variance above strength of association on its own ($F_{(2, 60)} = 44.03$, $p < .0005$).

Inductions about cells when people had to give their responses under time pressure (R= .70) was predicted by all three variables. The most important predictor was taxonomic beliefs (beta = .93, t = 4.89, p < .0005), followed by beliefs about causal relatedness (beta = .6, t = 2.99, p = .004) and strength of association (beta = .28, t = 2.54, p = .014).

However, the simple correlation between causal beliefs and inferences about cells under speeded conditions was actually negative, r = -.1, p = .47. Given that each category in a causally related pair had been intentionally chosen from different taxonomic groups, this is not a surprising result. It also suggests that the reason the structured causal knowledge variable was a significant predictor above association on its own is via its correlation with strength of association (r = .3, p = .016) mentioned above. This would explain why knowledge that is not really relevant to reasoning about cells is a significant predictor under time pressure.

Although adding the structured knowledge variables accounted for additional variance above strength of association on its own (R^2 Change: 26.4%, $F_{(2, 60)} = 15.5$, $p < .0005$), this change in R^2 was not as substantial as when people were forced to delay their response (R^2 Change: 48.2%).

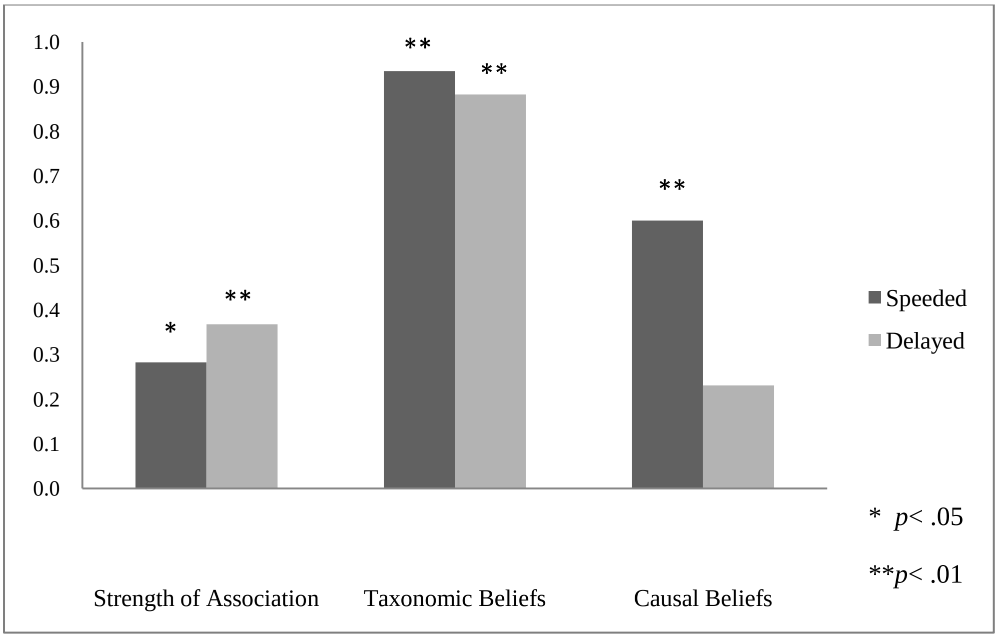

Figure 3.3: Standardized Regression Coefficients for Predictive Relations between Strength of Association, Taxonomic and Causal Beliefs and Inductive Strength Ratings for Cells

Availability of Domain-Specific Knowledge

To further scrutinize Shafto et al.'s (2007) claim that the taxonomic knowledge domain is fundamentally more accessible than ecological or causal knowledge domains, we looked at the effect of timing separately for taxonomically related categories and causally related categories, ignoring causal directionality. This enables a direct comparison with Shafto and colleagues' (2007) analysis, whilst controlling for strength of association.

Taxonomic Inferences

We carried out a 2 (property: cells versus infections) by 2 (timing: speeded or delayed) mixed design ANOVA, with timing as the between-subjects manipulation. For the item analysis, we included strength of association as a covariate and analyzed the data with a 2 (property) by 2 (timing) within-items ANOVA. The covariate of strength of association in the item analysis was highly significant, $F_{I\ (1,\ 30)} = 65.23$, $p < .0005$, effect size d = 2.95, but it did not interact with any of the other variables.

In line with Shafto et al.'s (2007) data, there was a main effect of property across participants, $F_{S\ (1,\ 58)} = 4.32$, $p = .042$, effect size d = .54. People tended to rate taxonomic inferences slightly higher when they were reasoning about cells ($M = 4.73$, $SE = .16$) than when they were reasoning about infections ($M = 4.39$, $SE = .16$). However, this effect was nonsignificant across items, $F_{I\ (1,\ 30)} = .01$, $p = .92$, effect size d = .001.

There was no effect of timing, $F_{S\ (1,\ 58)} = .13$, $p = .73$, effect size d = .09, $F_{I\ (1,\ 30)} = .61$, $p = .44$, effect size d = .28. Inferences were of similar strength in the speeded ($M = 4.51$, $SE = .20$) and unspeeded condition ($M = 4.61$, $SE = .20$).

Finally, there was no significant interaction between timing and property, $F_{S\ (1,\ 58)} = .85$, $p = .36$, effect size d = .24, $FI_{(1,\ 30)} = .017$, $p = .9$, effect size d = .06.

Causal Inferences

As for the taxonomic items, we performed a 2 (property: cells versus infections) by 2 (timing: speeded or delayed) mixed-design ANOVA, with timing as the between-subjects manipulation. The covariate of strength of association in the item analysis was again highly significant, $F_{I\ (1,\ 14)} = 12.71$, $p = .003$, effect size d = 1.91.

Unlike Shafto et al. (2007), who found that the availability of ecological knowledge was compromised when people were speeded up, we did not find an effect of timing, $F_{S(1,58)} = .78$, $p = .38$, effect size d = .22, $F_{I(1,\ 14)} = .62$, $p = .45$, effect size d = .42. Inferences were of similar strength in the speeded ($M = 4.32$, $SE = .18$) and unspeeded condition ($M = 4.09$, $SE = .18$).

There was also no interaction effect between timing and property, $F_{S\ (1,\ 58)} = .11$, $p = .74$, effect size d = .08, $FI_{(1,\ 14)} = .081$, $p = .78$, effect size d = .16.

There was however a main effect of property, $F_{S\ (1,\ 58)} = 10.61$, $p = .002$, effect size d = .48, $FI_{(1,\ 14)} = 5.32$, $p = .037$, effect size d = 1.23. Thus, people tended to rate inferences about causally related categories higher when they were reasoning about infections (*M* = 4.52, *SE* = .15) than when they were reasoning about cells (*M* = 3.89, *SE* = .18). These results suggest that strength of association, a measure of unstructured knowledge, has a pervasive influence, and once level of association is controlled for, no knowledge domain has fundamental superiority in its ability to inform category-based generalizations.

3.1.3 Discussion

The main goal of this first experiment was to show that knowledge effects in category-based reasoning can be distinguished with regards to two contrasting types of knowledge. On the one hand, inductive inferences can be influenced by effortlessly computable, unstructured knowledge such as strength of association (Rogers & McClelland, 2004) or similarity (Sloman, 1993b; Sloutsky & Fisher, 2004a). On the other hand, there are effects that can be ascribed to structured knowledge (Kemp & Tenenbaum, 2009; Osherson, et al., 1990; Rehder, 2009; Shafto, et al., 2008) which requires time and processing effort. Our results regarding the use of causal knowledge support this distinction. They suggest that the causal asymmetry effect depends upon considering the underlying causal structure and what implications it might have for the probability of sharing relevant properties such as infections or diseases (Shafto, et al., 2008). Thus, when people were not under any time pressure, they rated predictively-related categories as more likely to share an infection than diagnostically-related categories, whereas this causal asymmetry effect vanished when people had to respond rapidly.

Furthermore, the results showed that strength of association was a significant predictor of inductive strength ratings in both the speeded and unspeeded conditions, suggesting that unstructured knowledge influenced people’s inferences regardless of timing manipulations. In contrast, causal and biological beliefs accounted for more additional variance above strength of association on its own in the delayed condition compared to the speeded condition, suggesting that structured knowledge became more important when people had plenty of time to consider their responses. Thus, this supports our contention that drawing on structured knowledge is mediated by an effortful, time-consuming process, unlike the use of unstructured associative knowledge, which influences the reasoning process relatively automatically. For reasoning to be maximally effective, it is necessary to consider the deep structure of the domain with more scrutiny, which appears to be a slower, more elaborate and cognitively costly process.

Although the regression analyses broadly support this contention, as more additional variance was accounted for by the structured knowledge variables when people delayed their responses compared to when they had to respond rapidly, it is less clear how we should interpret some of the individual beta-weights.

To further substantiate our claim that the use structured knowledge is effortful, whereas the influence of unstructured knowledge is relatively automatic and hence have dissociable effects on category-based inferences, we carried out another experiment with the same design. Instead of a timing manipulation, we used a secondary task paradigm as a direct manipulation of available mental resources.

3.2 Experiment 2

The previous study suggests that certain qualitative phenomena which arise in category-based induction in knowledge-rich environments are a consequence of the recruitment of structured knowledge. This was especially clear when considering the causal asymmetry effect, in which inferences are stronger when they are predictive and go from prey to predator or from plant to consumer compared to diagnostic inferences in which the direction of the causal link is reversed. If our hypothesis is correct, and recruiting structured knowledge during inductive reasoning is a slower, more effortful process, which is separable from relying on unstructured knowledge such as strength of association between the two categories, then the former should be susceptible to manipulations of mental resources. Thus, if people do not have the available time or resources available, and fall back onto associative strength as a basis for making their inductive judgements, then causal asymmetry effects should not emerge.

Within the category-based induction literature, there is only one study (Feeney & Crisp, 2010) that has used manipulations of mental resources to assess whether some processes underlying category-based induction may require more effort than others. In that study, people rated arguments such as:

Grain has property X.
How likely is it that Mice have property X? (1)

Grain has property X.
How likely is it that Owls have property X? (2)

Grain has property X.
How likely is it that Mice and Owls have property X? (3)

People tended to rate the conjunctive category conclusion (3) higher than the single causally distant category conclusion (2), which is a violation of the conjunction law of probability. In contrast, only a minority of people tended to rate the conjunctive category conclusion (3) higher than the single causally near category conclusion (1). This tendency to commit the conjunction fallacy for causally distant conclusions was higher when people were under memory load, whereas it did not increase the fallacy rate for causally near conclusions. We attributed this to the fact that the causal relationship was effortlessly retrieved when the two categories were closely related, but that it took more mental resources to reconstruct the causal link between the premise and causally distant conclusion category. Thus, under memory load, people could not make use of more structured causal knowledge in order to overcome the fallacy.

We expected similar effects on people's ability to recruit structured knowledge for the current inductive evaluation task. This should be especially apparent for causally related categories, in which considerations of the mechanisms that relate categories to one another in the causal network seem to give rise to qualitative phenomena such as the causal asymmetry effect. Thus, based on the previous experiment, we expected that the emergence of the causal asymmetry effect would be tied to the availability of mental resources.

3.2.1 Overview Experiment 2

We used a secondary task paradigm in which people had to remember simple or complex dot patterns (de Neys, 2006a) whilst making their inductive strength evaluations. Remembering the simple patterns should only minimally affect working memory resources, whereas remembering the complex patterns places a much heavier burden on people's mental resources. In general, we expected people to exhibit a causal asymmetry effect when under minimal cognitive burden, rating predictive causal inferences (from *prey* to *predator*) higher than diagnostic causal inferences (from *predator* to *prey*). In contrast, this asymmetry effect should be attenuated or absent when people were burdened with a resource-engaging secondary task, in this case remembering complex dot configurations.

Furthermore, we hoped to replicate the findings from the two previous experiments, showing that associative strength, our measure of unstructured knowledge, predicts people's inductive strength ratings regardless of resource manipulations, whereas structured causal and taxonomic knowledge should account for more additional variance in the unloaded compared to the loaded condition.

Finally, we wanted to further substantiate our conclusion that Shafto et al's (2007) claim about an accessibility advantage for taxonomic knowledge cannot be fully upheld once level of association is controlled for. Thus, we expected the secondary task load to have a similar effect on taxonomic and causal inferences once the direction of the causal link is disregarded.

3.2.2 Method

Design

The experiment had a 2 (load: heavy or light) by 2 (property: cells or infections) by 3 (relation: taxonomic, causal predictive or causal diagnostic) by 4 (list) mixed-design, with list and load as between-subjects variables. Based on our findings from Experiment 1, we predicted an inductive selectivity effect, manifested as an interaction between property and relation. Furthermore, we predicted an effect of relation, and specifically, we expected a causal asymmetry effect manifested by causal predictive inferences being rated higher than causal diagnostic inferences. However, we expected this to be modulated by an interaction between load and relation, so that this causal asymmetry effect should be attenuated when people were under heavy memory load.

Materials

Inductive Reasoning Task and Load Manipulation

The inductive reasoning problems were the same ones used in Experiment 1. Before rating each inductive argument, participants were shown a 4 by 4 grid with 4 dots. In the heavy load condition, the dots were displayed in a random order with the constraint that they could never form a straight line or diagonal. In contrast, in the light load condition, the dots always formed a straight line or diagonal, placing minimal burden on working memory. The dot matrix was displayed for 2 seconds followed by the reasoning problem. Participants entered their response to the induction arguments on the keyboard by pressing any number between 1 (highly unlikely) and 9 (very likely). Once they had given their rating, an empty dot matrix appeared and participants tried to recall the location of the 4 dots by pressing the appropriate box with the mouse cursor.

Post-test

As in the previous experiments, we checked people's beliefs about category relations in a computerized post-test. We calculated the mean proportion of positive responses to the two post-test questions about biological group membership and food chain relations across the two timing conditions and correlated this with both our web-based measures of co-occurrence (Spearman rho correlation coefficients ranged from -.03 to .19, all p's > .13), and our subjective measure of associative strength (Spearman rho correlation coefficients ranged from -.2 to .19, all p's > .14). As none of the correlations were significant, we were confident that this test was assessing beliefs based on more structured knowledge.

Participants

40 participants took part in the study. They were volunteers from Queen's University Belfast and Durham University, who received either course credit for their participation or were paid £5 for their time. The mean age was 23.3 years (SD = 6.2 years).

Procedure

Participants were told that we were interested in their beliefs about shared infections and cells. They received written instructions which included an example. After giving informed consent, they began the self-paced experiment and completed two practice trials to familiarize themselves with the procedure.

For each trial, the dot matrix was presented on the computer screen for 2 seconds, followed by the inductive reasoning problem. Once participants had read the induction question, they entered their response on the keyboard by pressing a number between 1 and 9. They then saw an empty matrix and had to try and recall the location of the four dots by pressing on an appropriate box with the mouse cursor. Participants received no feedback on their recall performance. After the main experiment, participants completed the self-paced post-test. They were instructed to press key C if their response was YES, key M if their response was NO and key B if they didn't know.

3.2.3 Results

ANOVA

As in Experiment 1, we calculated 6 mean inductive strength ratings for each participant and analyzed these with a 3 (relation: predictive, diagnostic or taxonomic) by 2 (property: cells or infection) by 2 (memory load: heavy or light) by 4 (list) mixed-design ANOVA, with timing and list as between-subjects variables. For the analysis by items we used a 2 (property) by 2 (timing) by 3 (relation) mixed-design ANOVA, with relation as a between-items variable and strength of association as a covariate.

There was no significant main effect of list, $F_{(3, 32)} = .92$, $p = .44$, effect size d = .58, and none of its interactions with other variables were significant (all *p*'s > .33), so further analysis proceeded without this counterbalancing variable. For the item analysis, the covariate of strength of association was highly significant, $F_{I\ (1, 60)} = 44.5$, $p < .0005$, effect size d = 1.72, but it did not interact with any of the other variables.

There was no significant main effect of property, $F_{S\ (1, 33)} = .79$, $p = .38$, effect size d = .31, $F_{I\ (1, 60)} = 1.5$, $p = .23$, effect size d = .34. Inferences about infections (*M* = 3.93, *SE* = .21) were rated as strong as inferences about cells (*M* = 3.75, *SE* = .19).

Inferences made under heavy memory load (*M* = 3.7, *SE* = .25) were similar to inferences made under minimal memory load (*M* = 4.0, *SE* = .25), such that there was no significant effect of load $F_{S\ (1,\ 38)} = .75$, $p = .39$, effect size d = .3, $F_{I\ (1,\ 60)} = 1.3$, $p = .26$, effect size d = .29.

There was a significant main effect of relation, $F_{S\ (1.66,\ 53.03}$[5]$_{)} = 3.0$, $p = .007$, effect size d = .87, $F_{I\ (2,\ 60)} = 12.83$, $p < .0005$, effect size d = 1.38. Post-hoc comparisons showed that there was a difference in inductive strength ratings for causal predictive (*M* = 3.90, SE = .24) and diagnostic inferences (*M* = 3.33, *SE* = .23, *p* = .013, effect size d = .44), but no significant difference between causal predictive and taxonomic inductive strength ratings (*M* = 4.30, *SE* = .25, *p* = .23 effect size d = .2). The difference between diagnostic and taxonomic inferences was significant (*p* = .002, effect size d = .59).

As in the previous experiment, there was a significant interaction between property and relation, $F_{S\ (2,\ 76)} = 5.48$, $p = .006$, effect size d = .83, $FI_{(2,\ 60)} = 11.3$, $p < .0005$, effect size d = 1.22. As is shown below in Table 3.3, when people made inferences about cells, taxonomic inferences were significantly stronger than both causal predictive inferences (*p* =.01, effect size d = .62) and causal diagnostic inferences (*p* = .001, effect size d = .73), whereas there was no difference between the latter two types of relation (*p* = .93, effect size d = .15).

When reasoning about infections, people rated causal predictive inferences significantly higher than causal diagnostic inferences (*p* = .01, effect size d = .46). However, there was no difference between causal predictive and taxonomic inductive strength ratings (*p* = .99, effect size d = .1). Although taxonomic inferences were slightly higher than causal diagnostic inferences, this comparison did not meet statistical significance (*p* = .12, effect size d = .41).

[5] df adjusted for Non-Sphericity using Greenhouse-Geisser corrections

Table 3.3: Inductive Strength Ratings broken down by Property and Relation in Experiment 2

Property	Relation	Mean	Std. Error
Cell	predictive	3.50	.24
	diagnostic	3.26	.27
	taxonomic	4.50	.27
Infection	predictive	4.30	.33
	diagnostic	3.39	.30
	taxonomic	4.11	.25

None of the other two-way interactions was significant, however, there was a significant three-way interaction between property, load and relation, $F_{S\ (2,\ 76)} = 3.22$, $p = .047$, effect size d = .64; $F_{I\ (2,\ 60)} = 3.91$, $p = .025$, effect size d = .72. Figures 3.4 and 3.5 below show the mean inductive strength ratings for the three types of relations for each memory load condition in the two different property contexts.

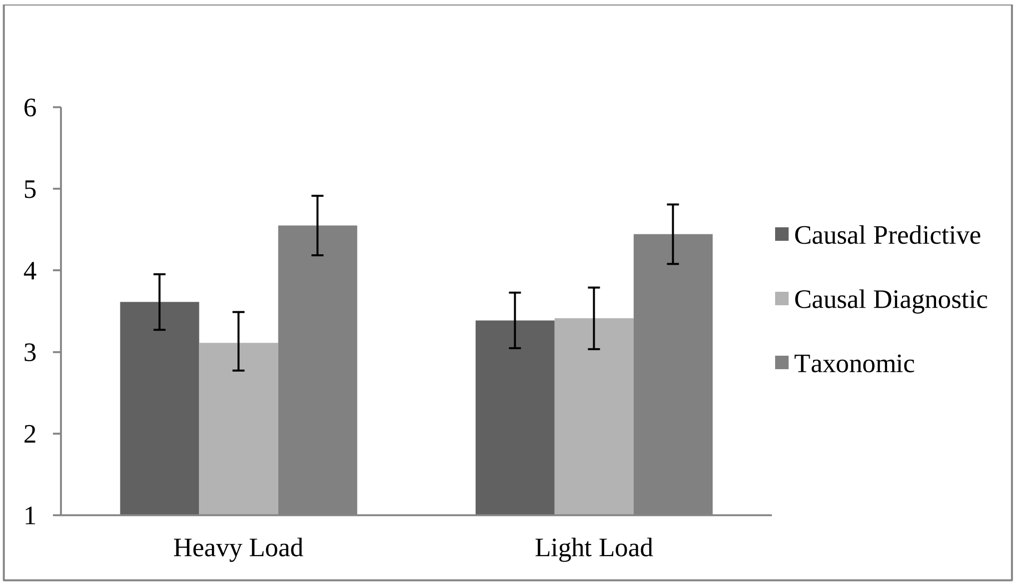

Figure 3.4: Mean Inductive Strength Ratings for Cells

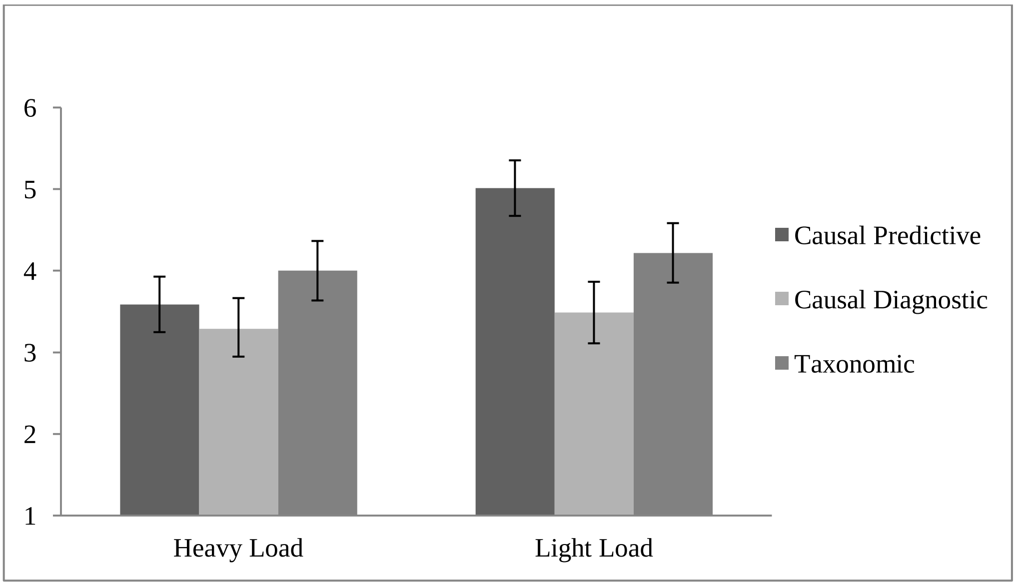

Figure 3.5: Mean Inductive Strength Ratings for Infections

Causal Asymmetry Effect

As we had made a-priori predictions about the effect of memory load on the emergence of the causal asymmetry effect, we carried out two separate 2 (causal relation: predictive versus diagnostic) by 2 (load: heavy versus light) mixed-model ANOVAs, one for cells and the other for infections.

Causal Asymmetry Infections

As in the previous experiment, strength of association was a highly significant covariate in the item analysis, $F_{I\,(1,\,29)} = 11.49$, $p = .002$, effect size d = 1.26.

Evaluating inferences about infections, there was a main effect of causal relation, $F_{S\,(1,\,38)} = 8.99$, $p = .005$, effect size d = .98, $F_{I\,(1,\,29)} = 10.24$, $p = .003$, effect size d = 1.18. Thus, across all conditions, there was a substantial causal asymmetry effect, with people rating inferences about category pairs with a predictive causal relation (*M* = 4.30, *SE* =.31) as significantly stronger than when the relation between the two categories was diagnostic (*M* = 3.39, *SE* = .29).

The main effect of load was nonsignificant, $F_{S\,(1,\,38)} = 2.41$, $p = .13$, effect size d = .5, $F_{I\,(1,\,29)} = 3.5$, $p = .07$, effect size d = .69.

Most importantly however, as in Experiment 1, the main effect of causal relation was modulated by the significant interaction between load and causal relation, $F_{S\,(1,\,38)} = 4.04$, $p = .051$, effect size d = .65, $F_{I\,(1,\,29)} = 7.76$, $p = .009$, effect size d = 1.04, pictured below in Figure 3.6. Bonferroni post-hoc tests showed that people who were not burdened by a heavy secondary memory load were able to take causal structure into account to maximize inductive potency. Indicative of a significant causal asymmetry effect, inferences with a causal predictive relation (*M* = 5.01, *SE* = .44) were rated much stronger than categories with a diagnostic causal link (*M* = 3.49, *SE* = .41, *p* = .001, effect size d = .73).

In contrast, the causal asymmetry effect whilst reasoning about infections was absent when people had to contend with a heavy secondary memory load. Thus, inductive strength ratings were almost identical for causal diagnostic inferences (*M* = 3.29, *SE* = .41) and causal predictive inferences (*M* = 3.49, *SE* = .41, *p* = .49, effect size d = .17).

This finding again suggests that the causal asymmetry effect arises because people consider how the underlying causal structure might influence the probability distribution of properties such as infections or diseases. When people are cognitively burdened, they may not be able to take the underlying causal structure into account, instead forcing them to rely on the existence of a non-directional association between the categories.

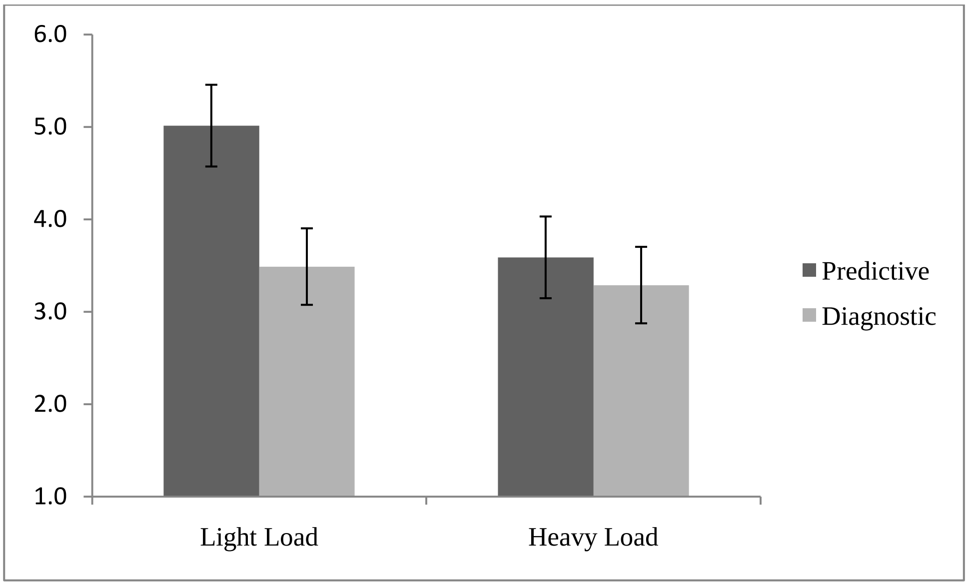

Figure 3.6: Causal Asymmetry Effect across Load Conditions

Causal Asymmetry Cells

As expected, for inferences about cells, there was no effect of causal relation, $F_{S\,(1,\,38)} = 1.05$, $p = .31$, effect size d = .34, $F_{I\,(1,\,29)} = 1.06$, $p = .31$, effect size d = .38. People rated inferences about category pairs with a predictive causal relation (*M* = 3.5, *SE* = .24) the same as when the relation between the two categories was diagnostic (*M* = 3.26, *SE* = .27).

The main effect of load was non-significant, $F_{S\ (1,\ 38)} = .007$, $p = .93$, effect size d = .001, $F_{I\ (1,\ 29)} = .55$, $p = .46$, effect size d = .28, as was the interaction between load and causal relation, $F_{S\ (1,\ 38)} = 1.29$, $p = .26$, effect size d = .36, $F_{I\ (1,\ 29)} = .97$, $p = .33$, effect size d = .36.

Surprisingly, the covariate of strength of association did not quite reach significance in the item analysis, $F_{I\ (1,\ 29)} = 3.65$, $p = .066$, effect size d = .71. However, the pattern of the other effects in the item analysis was identical whether or not this covariate was included.

Results Regression Analyses

As in the previous experiment, we ran hierarchical regression analyses to see how unstructured knowledge, represented by strength of association between categories, as well as how relevant domains of structured knowledge influence category-based inferences depending upon the availability of mental resources. For each item we averaged inductive strength scores across participants separately for the two types of property in the two memory load conditions. This meant that we had 4 inductive strength scores for each item. Furthermore, for each category pair, the mean proportions of positive responses to the two post-test questions about biological group membership and causal food chain relationships were calculated separately for the two load conditions. As there was neither a correlation between the structured knowledge variables and strength of association, nor between the structured knowledge variables and the objective conditional co-occurrence indices, we were confident that the structured knowledge variables were measuring a different type of knowledge than that assessed by our measure of unstructured knowledge, strength of association.

As before, strength of association was entered in the first block of the regression analysis, followed by taxonomic and causal beliefs in the second block. All four regression analyses were significant, but different predictors were of different importance across the two load conditions and across the two different properties.

Inferences about Infections

Illustrated in Figure 3.7, reasoning about infections when people were under minimal cognitive load (R = .69) was significantly predicted by strength of association (beta = .48, t = 4.62, p < .0005). However, the structured knowledge variables were far more important. Thus, in the second block, knowledge about relevant causal food chain relations was also a significant predictor (beta = .63, t = 2.65, p = .01), as was taxonomic knowledge (beta = .78, t = 3.39, p = .001). Together, adding these two structured knowledge variables accounted for a significant amount of additional variance above strength of association alone (R^2 Change: 11.1%, $F_{(2, 60)}$ = 6.35, p = .003).

Reasoning about infections when people were burdened with a heavy cognitive load (R = .48) was also significantly predicted by association (beta = .31, t = 2.58, p = .012). Surprisingly, whereas biological relatedness was a significant predictor of inductive strength (beta = .36 t = 2.45, p = .017), relevant knowledge about food chain relations (beta = .27, t = 1.77, p = .082) did not reach significance. Adding these two structured knowledge predictors in a second block accounted for a marginally significant amount of additional variance compared to strength of association on its own (R^2 Change: 7.8%, $F_{(2, 60)}$ = 3.02, p = .056).

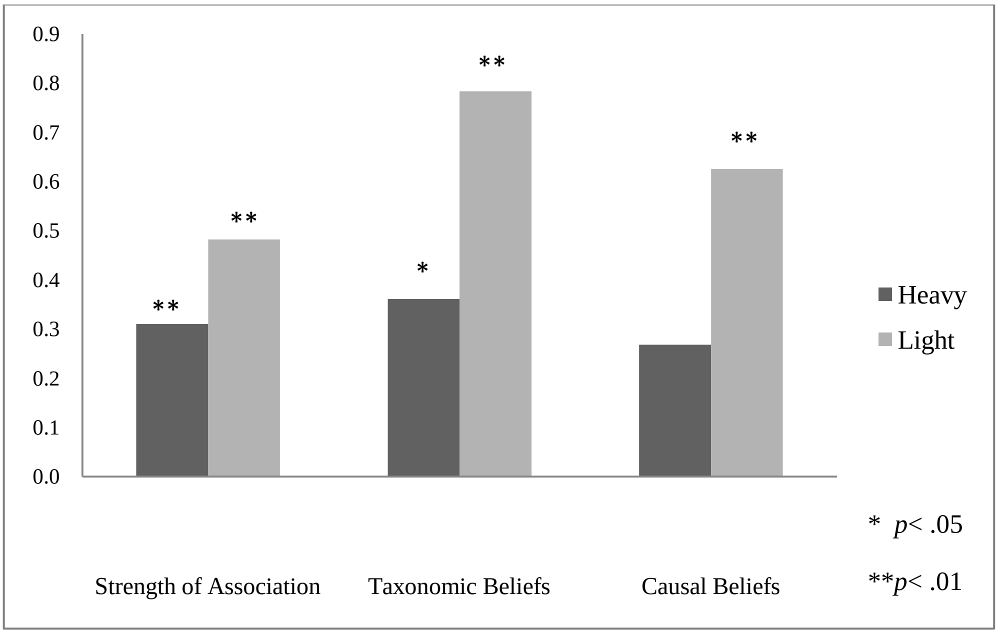

Figure 3.7: Standardized Regression Coefficients for Inductive Ratings about Infections in Experiment 2

Inferences about Cells

Reasoning about cells was predicted by a different constellation of variables, as shown in Figure 3.8 below. Overall, when people were not under a heavy memory load, 42.9% of the variance was accounted for by the 3 variables entered into the regression analysis (R = .66). Strength of association, which was forced into the regression equation in the first block, accounted for a significant amount of variance in inductive strength ratings (beta = .33, t = 3.07, *p* = .003).

Of the structured knowledge variables entered in the second block of the regression analysis, the most important predictor of inductive strength ratings were taxonomic beliefs (beta = .72, t = 2.99, *p* = .004), whereas causal beliefs were not predictive of people's inductive strength ratings about cells (beta = .21, t = .85, *p* = .4).

In summary, when people were not under time pressure, adding the structured knowledge predictors in a second block accounted for significantly more variance in inductive strength ratings about cells than strength of association on its own (R^2 Change: 29.1 %, $F_{(2,\ 60)}$ = 15.28, p < .0005).

Inductions about cells when people were cognitively burdened (R = .64) was predicted by strength of association (beta = .31, t = 2.95, p = .005). Of the structured knowledge variables entered in the second block, taxonomic beliefs were highly predictive of inductive strength ratings (beta = .51, t = 3.97, p < .0005), whereas structured knowledge about causal relatedness was not a significant predictor (beta = -.06, t = -.46, p = .65). Overall, as in the light load condition, adding the structured knowledge variables accounted for additional variance above strength of association on its own (R^2 Change: 30.4%, $F_{(2,\ 60)}$ = 15.51, p < .0005).

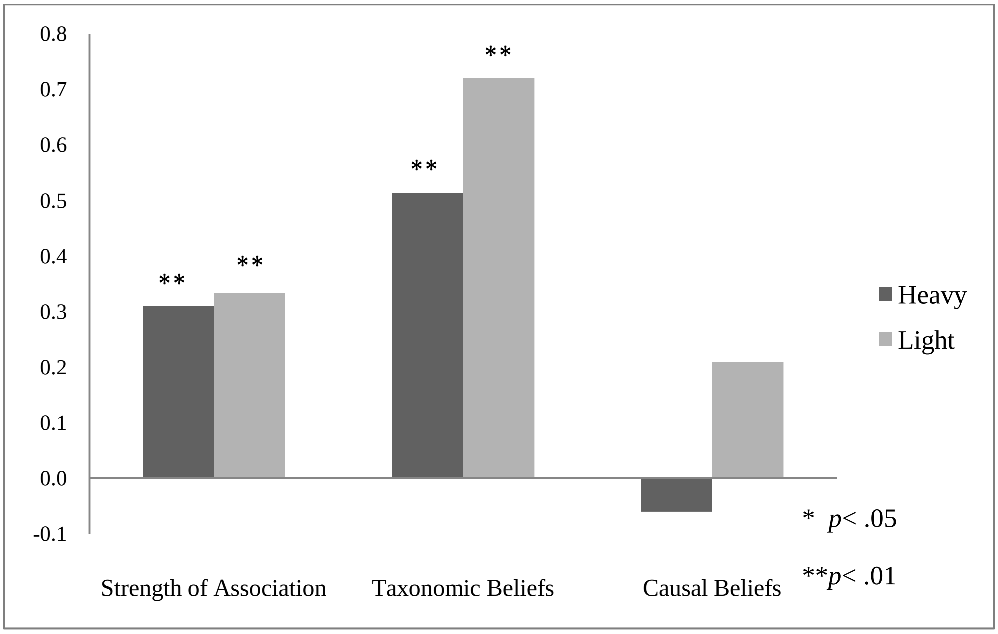

Figure 3.8: Standardized Regression Coefficients for Inductive Ratings for Cells in Experiment 2

Availability of Different Domain-Specific Knowledge

As in Experiment 1, we wanted to look at whether there was any evidence to support the claim made by other researchers (Shafto, Vitkin, & Coley, 2007) that access to ecological or causal knowledge is constrained by the availability of time and/or mental resources, whereas access to taxonomic knowledge is independent of domain-general cognitive factors. The evidence from the regression analyses in the first two experiments is somewhat ambivalent. For example, the fact that causal beliefs did not appear to predict people's inductive inferences about infections when they had to contend with a cognitive load could be used to support Shafto et al's (2008) position. However, once strength of association was controlled for, our first experiment suggested that timing manipulations did not have a dissociable effect upon people's causal and taxonomic inductive strength ratings. Thus, we repeated this analysis to explore how memory load affected the strength of people's inferences separately for taxonomically related categories and causally related categories, ignoring the directional nature of the causal link. As explained in the methods section, we controlled for strength of association and entered this variable as a covariate in the item analyses, providing an indication of the importance of unstructured knowledge.

Taxonomic Inferences

As expected, the item analyses indicated that the covariate of strength of association was highly significant $F_{I\,(1,\,30)} = 44.28$, $p < .0005$, effect size d = 2.42.

The 2 (property: cells versus infections) by 2 (load: heavy versus light) mixed-design ANOVA, with load as the between-subjects manipulation showed that across participants, there was a main effect of property, $F_{S(1,\,38)} = 5.34$, $p = .026$, effect size d = .75, People rated taxonomic inferences higher when they were reasoning about cells (*M* = 4.50, *SE* = .26) compared to the inferences they made about infections (*M* = 4.10, *SE* = .26). However, this effect was not significant across items, $F_{I\,(1,\,30)} = .83$, $p = .37$, effect size d = .33.

There was no effect of load, $F_{S\,(1,\,38)} = .013$ $p = .91$, effect size d = .001, $FI_{(1,\,30)} = .03$, $p = .86$, effect size d = .06. Inferences made under heavy load (*M* = 4.33, *SE* = .35) were almost identical to inferences drawn under light load (*M* = 4.28, *SE* = .35).

Finally, there was no significant interaction between load and property, $F_{S\,(1,\,38)} = .94$, $p = .34$, effect size d = .31, $FI_{(1,\,30)} = .97$, $p = .33$, effect size d = .36.

Causal Inferences

We then carried out the same analysis for causal inferences. Again, the covariate of strength of association in the item analyses was highly significant $F_{I\,(1,\,14)} = 10.47$, $p = .006$, effect size d = 1.73.

In line with our previous findings, and somewhat refuting the conclusions drawn by Shafto et al. (2007) about an accessibility disadvantage for ecological knowledge , we did not find any effect of load, $F_{S(1,38)} = 1.11$, $p = .3$, effect size d = .34, $F_{I(1,14)} = 4.0$, $p = .07$, effect size d = 1.06. Causal inferences received similar ratings under both loaded (*M* = 3.40, *SE* = .29) and unloaded conditions (*M* = 3.82, *SE* = .29).

Finally, there was no evidence that the effect of load on causal inferences might only arise in a relevant context. Thus, there was no interaction between load and property, $F_{S\,(1,\,38)} = 1.95$, $p = .17$, effect size d = .44, $F_{I\,(1,\,14)} = 1.91$, $p = .19$, effect size d = .74.

Analogous to Experiment 1, causal inferences were somewhat stronger when people reasoned about infections (*M* = 3.84, *SE* = .26) than when they reasoned about cells (*M* = 3.38, *SE* = .23), although this effect was not statistically significant, $F_{S\,(1,\,38)} = 2.78$, $p = .1$, effect size d = .54, $F_{I\,(1,\,14)} = 1.46$, $p = .25$, effect size d = .64.

Secondary Task Analysis

In a dual task paradigm, dissociable effects of the secondary task can reflect strategic trade-offs between primary and secondary tasks (Hegarty, Shah, & Miyake, 2000). To guard against the possibility that the dissociable effect of memory load on causal inferences reflected such a strategic trade-off, we calculated the number of dots correctly recalled separately for the trials preceding each of the unique property by load by relation conditions.

A 2 (relation: causal predictive, causal diagnostic or taxonomic) by 2 (property: cell of infection) by 2 (load: heavy versus light) by 4 (list) mixed-design ANOVA, with list and load as between-subjects variables showed that the only significant difference in the number of dots recalled was between the two load conditions, $F_{(1, 31)} = 17.71$, $p < .0005$, effect size d = 1.5. In the heavy load condition, participants recalled a mean of 3.2 dots (*SE* = .1), whereas they recalled on average 3.8 dots (*SE* = .1) when they were only under a light memory burden. This suggests that people were consistent in how they allocated their mental resources to the primary and secondary tasks across all problems and verified that the more complex patterns were harder to remember and more burdensome than the simple dot patterns.

3.2.4 Discussion

Experiment 2 corroborated the findings from the previous experiment. People's inductive strength ratings were sensitive to context, with inferences about cells being stronger when there was a taxonomic relationship between the categories than when there was a causal relation between them. In contrast, inferences about infections were equally strong when there was a taxonomic or causal predictive relation between the categories, but lower when the relation was diagnostic. Most importantly, this causal asymmetry effect disappeared when people were burdened with a secondary memory task, suggesting that the ability to consider how causal structure might modulate assessments of inductive strength is mediated by an effortful, analytical process.

The results from the regression analyses were less clear. Surprisingly, unlike in the previous experiment, adding the taxonomic structured knowledge variable when reasoning about cells explained similar amounts of additional variance above strength of association alone for both the heavy and light load condition. One explanation would be consistent with

Shafto et al.'s (2007) original suggestion that taxonomic knowledge may be more accessible than other knowledge structures. However, when causal directionality was ignored, load did not differentially affect people's taxonomic and causal inferences, suggesting that it is unlikely to be due to an accessibility advantage for the taxonomic knowledge domain. This is further supported by the finding that taxonomic knowledge explained much more variance when reasoning about infections in the light load condition (beta = .48) than in the heavy load condition (beta = .36).

Rather, a possible alternative explanation is that drawing on knowledge when reasoning about cells is relatively straightforward, as people only have to consider one relevant knowledge structure, taxonomic knowledge. In contrast, both taxonomic and causal knowledge is potentially relevant when reasoning about infections. This may make drawing on structured knowledge when reasoning about infections a cognitively more demanding task, especially given that the causal relations were complex, including directional aspects (Cobos, López, Cano, Almaraz, & Shanks, 2002; Fenker, Waldmann, & Holyoak, 2005; López, Cobos, & Cano, 2005).

Thus, in a final experiment we simplified the structure of causally related categories, including only causal predictive relations and their taxonomic counterparts. As it had been very difficult to find enough items for which strength of association level could be matched for causal and taxonomic category pairs, dropping the directional manipulation also allowed us to have enough individual causal and taxonomic arguments to manipulate strength of association in a factorial design by carrying out a median split on this variable.

3.3 Experiment 3

In the final experiment, the effects of putting people under time pressure was compared for causal predictive and taxonomically related category pairs which were either strongly or weakly associated. The primary goal was to confirm that depending upon available processing time, structured and unstructured knowledge could have dissociable effects upon people's inductive inferences. Whilst the previous two experiments strongly supported the idea that drawing on structured knowledge, such as causal relations, was subject to the availability of time and mental resources, the findings from the regression analyses were less clear-cut. This might have been because of the contrasting complexity of the knowledge structures underlying our causal and taxonomic arguments. We hoped to remedy this by dropping the directional manipulation for the causal arguments and looking at the effects of timing on weakly and strongly related category pairs in a factorial design.

The second goal was to advocate between two contrasting explanations for Shafto et al.'s (2007) results, regarding the availability of knowledge from different domains. These researchers suggest that knowledge from different domains diverge in their relative accessibility and thus the likelihood that this knowledge will come to bear on reasoning output. If different knowledge domains do indeed diverge in their accessibility, one would expect to find effects of timing on people's use of causal knowledge, but not on their use of taxonomic knowledge.

However, it might be that Shafto et al.'s (2007) findings were confounded by strength of association between the different categories. In this case one might expect timing to have similar effects on the use of causal and taxonomic knowledge once strength of association had been equated between domain-specific category pairs. Thus, highly associated categories should receive stronger inductive strength ratings than weakly associated categories regardless of whether they are causally or taxonomically related. Shafto et al. (2007) and Rehder (2006), make contradictory claims about the superiority of different knowledge domains, advocating the supremacy of taxonomic and causal knowledge respectively.

Arguing that neither of these positions is correct, we wanted to obtain further evidence that no specific knowledge domain was profoundly more privileged once level of association between the category pairs from different domains had been equated.

3.3.1 Method

Design

The experiment had a 2 (property: cells or diseases) by 2 (relation: causal or taxonomic) by 2 (level of association: high or low) by 2 (timing: speeded or delayed) by 4 (list) mixed design, with list and timing as between-subjects variables.

Participants

The 40 participants were from Durham University, and received course credit for their participation. The mean age was 24.2 years (*SD* = 5.8 years).

Materials

There were 20 items in total. Items consisted of a base category, a causally related target category and a taxonomically related target category. Causally related targets were always from different superordinate categories, for example, *plants* and *animals*, or *mammals* and *birds*. In contrast, taxonomically related pairs were always from the same superordinate taxonomic category. For example, the causal target for the base category *fly* would be *frog* and the taxonomic target would be *ant*. There were two exceptions in which the causally related categories were both mammals but belonged to different orders. For these items, the taxonomic alternative was from the same order as the base category. As in the previous two experiments, all our stimulus materials had been extensively pre-tested to ensure that people knew about the causal connection between the causally related category pairs. Also, it was ensured that strength of association between the base and the causal target was identical to the strength of association between the base and its taxonomic target.

For each of the 20 items, we created a taxonomic and a causal induction problem, resulting in a total of 40 inductive arguments. Participants were told that the base category had either a blank disease, such as *disease 3dfT*, or blank cells, such as *cells T78*. As before, the participants' task was to evaluate the likelihood that the cells/disease would be present in the conclusion category on a scale from 1 (*very unlikely*) to 9 (*very likely*).

To explore the role that level of association plays in the availability of knowledge from different domains, a median split based on level of association was carried out on the selected items. Thus, for 10 items the association between the base and its target categories was classed as strong and for the remaining 10 items this association was classed as weak. For half the strongly and weakly associated items participants made inductive inferences about diseases. For the other half, people evaluated inductive conclusions about cells. The participants reasoned about different items for cells and diseases, so whilst property was manipulated within-subjects, content was counterbalanced across participants in a Latin-square design, resulting in 4 different stimulus lists shown in Table 3.4 below.

Table 3.4: Design Stimulus Materials for Experiment 3

	List 1	List 2	List 3	List 4
Items 1-5 Strongly associated	Causal and Taxonomic Arguments *Cells*	Causal and Taxonomic Counterpart *Disease*	Causal and Taxonomic Arguments *Cells*	Causal and Taxonomic Arguments *Disease*
Items 6-10 Strongly associated	Causal and Taxonomic Arguments *Disease*	Causal and Taxonomic Counterpart *Cells*	Causal and Taxonomic Arguments *Disease*	Causal and Taxonomic Arguments *Cells*
Items 11-15 Weakly associated	Causal and Taxonomic Arguments *Cells*	Causal and Taxonomic Arguments *Disease*	Causal and Taxonomic Arguments *Disease*	Causal and Taxonomic Arguments *Cells*
Items 16-20 Weakly associated	Causal and Taxonomic Arguments *Disease*	Causal and Taxonomic Arguments *Cells*	Causal and Taxonomic Arguments *Cells*	Causal and Taxonomic Arguments *Disease*

Procedure

The induction problems were presented on a laptop. The premise and conclusions were presented simultaneously and appeared in a red font. Participants could only enter their response once the font changed from red to green. In the speeded condition, the font changed from red to green after one second and participants were instructed to read the problem and respond as fast as possible without sacrificing accuracy. In the delayed condition, the font only changed colour after 10 seconds and participants were instructed to carefully consider their responses. They entered their response on the key board by giving a rating between 1 and 9.

Post-Test

The post-test was identical to the one used in Experiments 1 and 2. Participants answered two questions about the 40 category pairs, assessing their beliefs about biological group membership, as well as their beliefs about causal relatedness.

The mean proportion of positive responses to the two post-test questions about biological group membership and food chain relations across the two timing conditions did not correlate with our two web-based measures of co-occurrence (Spearman rho correlation coefficients ranged from -.19 to .16, all p's > .27), nor did it correlate with our subjective measure of associative strength (Spearman rho correlation coefficients ranged from .10 to .22, all p's > .18) suggesting that this measure does not reflect associative strength but represents beliefs based on more structured knowledge.

3.3.2 Results

ANOVA

Mean inductive strength scores were calculated for each cluster of 5 inductive arguments representing the unique property by association by relation combination, resulting in 8 means for each participant. These were subjected to a 2 (property: disease or cell) by 2 (relation: causal or taxonomic) by association (high versus low) by 2 (timing: delayed or speeded) by 4 (list) mixed-design ANOVA, with list and timing as between-subject variables. For the analysis by items, inductive strength ratings were averaged across subjects rather than items and were analysed with a 2 (property) by 2 (timing) by 2 (relation) by 2 (association) mixed-design ANOVA, with association and relation as between-items variables.

There was no main effect of list, $F_{(3, 32)} = .15$, $p = .93$, effect size d = .24, however, as it interacted with strength of association and relation, $F_{(3, 32)} = 15.55$, $p < .0005$, this will be referred back to later on.

Inferences about diseases (*M* = 4.10, *SE* = .17) were weaker than inferences about cells (*M* = 4.40, *SE* = .17), however, this effect was only significant by items, $F_{I\ (1, 36)} = 4.5$, $p = .041$, effect size d = .71, and not by subjects, $F_{S\ (1, 32)} = 1.94$, $p = .17$, effect size d = .49.

The effect of relation was approaching significance by subjects, $F_{S\ (1, 32)} = 3.44$, $p = .07$, effect size d = .66, and was significant by items, $F_{I\ (1, 36)} = 4.12$, $p = .05$, effect size d = .67. Causal inferences received a mean strength rating of 4.05 (*SE* = .20) whereas taxonomic inferences were rated slightly higher at 4.45 (*SE* = .14).

As expected, there was a significant main effect of association, $F_{S\ (1, 32)} = 33.05$, $p < .0005$, effect size d = 2.0, $F_{I\ (1, 36)} = 7.7$, $p = .009$, effect size d = .92. Thus, inferences about highly associated categories (*M* = 4.52, *SE* = .14) were rated stronger than inferences about weakly associated categories (*M* = 3.98, *SE* = .14). However, as mentioned above, there was a significant interaction between list, association and relation. This showed that in list 2, there was no significant difference between strongly (*M* = 4.33, *SE* = .31) and weakly associated (*M* = 4.33, *SE* = .32, *p* = 1) taxonomically related categories. Similarly, the effect of association did not reach significance in list 3 for causally related categories, although the difference between weakly (*M* = 4.13, *SE* = .38) and strongly associated categories (*M* = 4.34, *SE* = .44, *p*= .36) was in the predicted direction. However, given that the effect of strength of association was so robust across items, this should not be of too much concern.

Furthermore, the effect of timing was significant by items $F_{I\ (1, 36)} = 18.1$, $p < .0005$, effect size d = 1.42, and was approaching significance for subjects, $F_{S\ (1, 32)} = 2.82$, $p = .1$, effect size d = .6. Inferences under speeded conditions (*M* = 4.03, *SE* = .19) were somewhat lower than inferences under delayed conditions (*M* = 4.48, *SE* = .19). However, timing did not interact with any of the other variables. Thus, once strength of association had been equated between knowledge domains, there was no evidence to support Shafto et al.'s (2007) contention that accessing taxonomic and causal knowledge domains are subject to contrasting processing constraints.

The only significant two-way interaction was between property and relation, $F_{S\ (1,\ 32)} = 23.17$, $p < .0005$, effect size d = 1.7, $F_{I\ (1,\ 36)} = 35.68$, $p < .0005$, effect size d = 1.99, suggesting that people showed some context-sensitive reasoning. Bonferroni post hoc tests showed that when reasoning about cells, people rated taxonomic inferences (*M* = 5.01, *SE* =.2) significantly stronger than causal inferences (*M* = 3.79, *SE* = .22, $p < .0005$, effect size d = .9). When reasoning about diseases, people rated causal inferences slightly higher (*M* = 4.32, *SE* = .26) than taxonomic inferences (*M* = 3.89, *SE* = .17) although this difference was not significant ($p = .16$, effect size d = .3). None of the other higher-order interactions were significant (all *p*'s > .08).

Regression Analyses

As in Experiments 1 and 2, we ran hierarchical regression analyses to see how different types of knowledge influence category-based inferences under different timing conditions. Inductive strength scores were averaged across participants for each item, separately for the two types of property and the two timing conditions, resulting in 4 inductive strength scores for each item. Structured knowledge was indexed by beliefs about taxonomic and causal relatedness. Thus, the mean proportion of positive responses to the two post-test questions about biological group membership and causal food chain relationship was calculated for each category pair, separately for the two timing conditions. Unstructured knowledge was instantiated by the strength of association between the category pairs. We confirmed that the structured knowledge ratings were not correlated with any of our measures of association (see Method), suggesting that they were indeed tapping different types of knowledge.

All four regression analyses were significant, but different relevant knowledge influenced inductive strength under different conditions. Firstly, as suggested by the larger multiple correlation coefficients in the delayed condition, people used different types of knowledge to inform their inferences when they had time to do so, whereas under time pressure, the ability to recruit relevant structured knowledge seemed to be attenuated.

Inferences about Diseases

As Figure 3.9 below shows, reasoning about diseases under speeded conditions (R= .59) was significantly predicted by strength of association (beta = .45, t = 3.13, p = .003). Regarding variables entered in the second block of the regression analysis, knowledge about relevant causal food chain relations was also a significant predictor (beta = .35, t = 2.04, p = .049), whereas taxonomic knowledge was not a significant predictor (beta = .08, t = .44, p = .67). Together, adding these two structured knowledge variables accounted for a non-significant amount of additional variance (R^2Change: 9.6%, $F_{(2, 36)} = 2.64$, $p = .085$).

In contrast, reasoning about diseases under delayed conditions (R= .68) was no longer significantly predicted by association (beta = .24, t = 1.8, p = .08). However, inductive strength was strongly predicted by relevant knowledge about food chain relations (beta = .61, t = 4.34, $p < .0005$), but also by beliefs about biological relatedness (beta = .34, t = 2.33, p = .026). Adding the structured knowledge predictors in a second block did account for significantly more variance in inductive strength ratings than strength of association on its own (R^2Change: 25.8%, $F_{(2, 36)} = 9.46$, $p < .0005$).

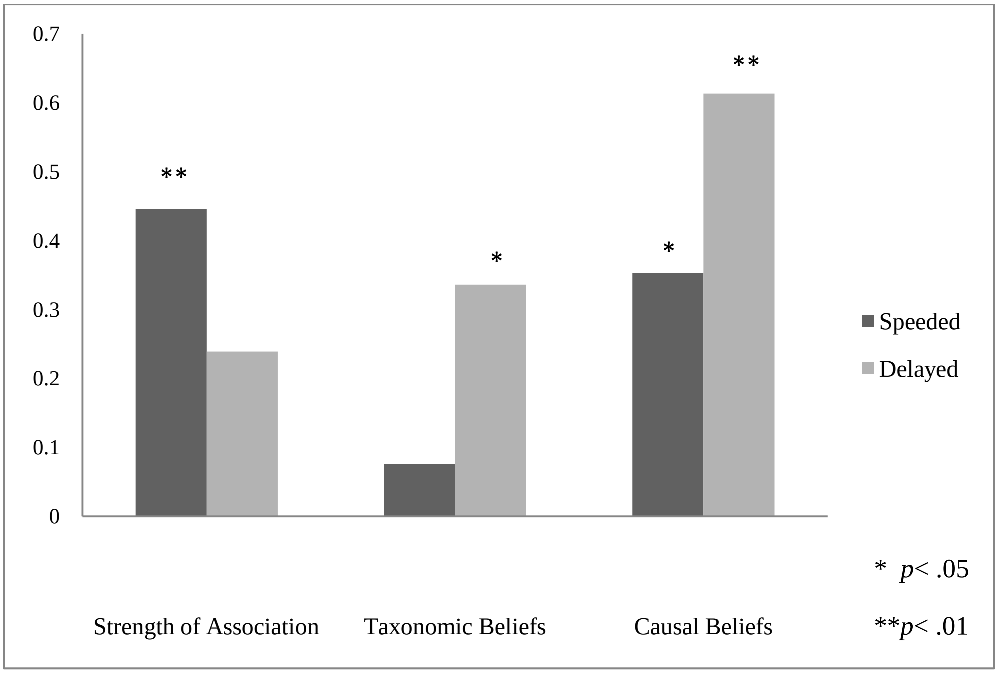

Figure 3.9: Standardized Regression Coefficients for Predictive Relations between Taxonomic and Causal Beliefs, Strength of Association and Inductive Strength Ratings for Diseases

Inferences about Cells

Reasoning about cells showed a different pattern, as depicted in Figure 3.10 below. Under delayed conditions, strength of association was not a significant predictor of inductive strength (beta = .19, t = .15, p = .14). Inductive inferences were however predicted by beliefs about biological relatedness (R = .72) (beta = .48, t = 3.48, p = .001), and were negatively predicted by beliefs about causal relatedness (beta = -.31, t = -2.29, p = .028). Given that we had selected causal targets that were always from different superordinate categories (i.e. taxonomically unrelated) it is not surprising that believing in the existence of a causal link was a negative predictor of inferences about cells. As when reasoning about diseases, adding the structured knowledge predictors in a second block of the regression analysis accounted for significantly more variance in inductive strength ratings than strength of association on its own when people were not under time pressure (R^2Change: 44.2%, $F_{(2, 36)} = 16.33$, $p < .0005$)

Inductions about cells under speeded conditions (R = .64) were predicted by strength of association (beta = .51, t = 3.76, *p* = .001) and were negatively predicted by beliefs about causal relatedness (beta = -.34, t = -2.5, *p* = .048). Taxonomic beliefs were not a significant predictor of inductive strength under speeded conditions (beta = .11, t = .65, *p* = .52). However, adding the structured knowledge coefficients did explain some additional variance above strength of association on its own (R^2Change: 16.7%, $F_{(2, 36)} = 5.09$, $p= .011$).

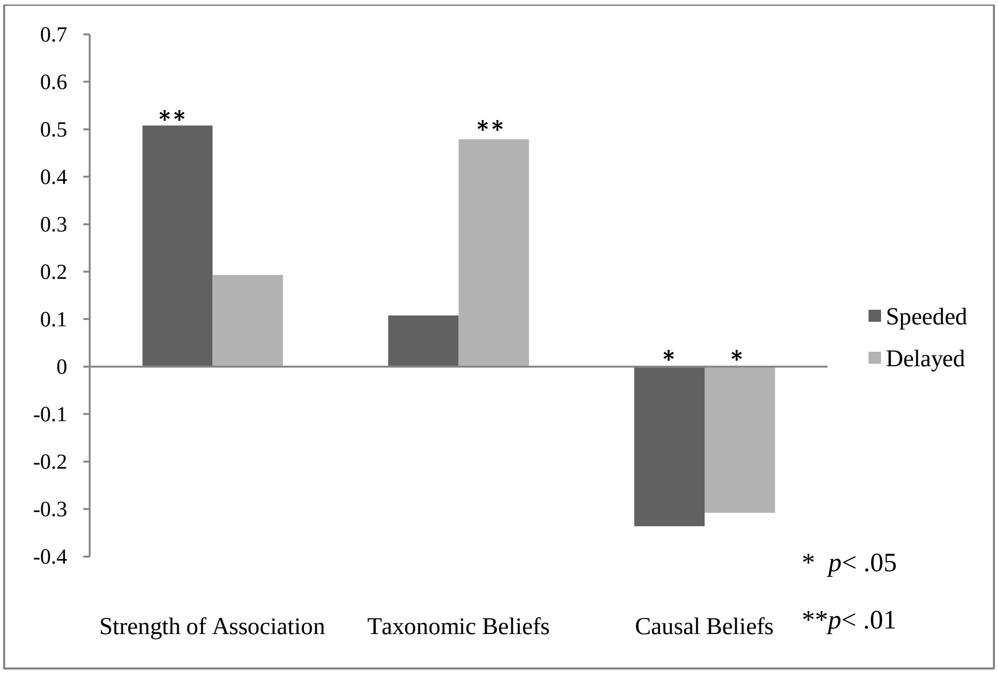

Figure 3.10: Standardized Regression Coefficients for Predictive Relations between Strength of Association, Taxonomic and Causal Beliefs and Inductive Strength Ratings for Cells

3.3.3 Discussion

The primary goal of this experiment was to corroborate the finding that structured and unstructured types of knowledge, which differ in their respective processing characteristics, have dissociable effects on peoples' inferences. To this end, we changed our experimental design in order to give us enough items to compare inferences about strongly and weakly associated category pairs. As expected, strongly associated category pairs were given higher inductive strength ratings than weakly related category pairs, irrespective of the structured relation between the categories. Furthermore, whereas our measure of unstructured knowledge, strength of association, was a strong predictor of inductive strength ratings when people were put under time pressure, structured causal and taxonomic knowledge was more influential when people were forced to delay their responses. This suggests that when people had enough time, they could appraise the nature of the relationship between the categories, and consider what impact this might have on inductive strength ratings.

The second goal was to examine whether differences in the accessibility of knowledge from different domains arise when strength of association was controlled for. Based on their experimental findings, Shafto and colleagues (2007) concluded that taxonomic knowledge is privileged. In contrast, ecological knowledge was relatively less accessible and did not come to bear on the reasoning process under conditions of time pressure. However, some of Shafto et al.'s (2007) findings seem to be the result of their material selection rather than due to differences in accessibility to different knowledge structures per se. As the current experiment demonstrates, once level of association was equated between causally and taxonomically related category pairs, the advantage for taxonomic knowledge disappeared. This suggests that no domain of knowledge is fundamentally more privileged than any other.

3.4 General Discussion

We began this series of experiments with the proposition that selective category-based induction is influenced by two contrasting types of knowledge, effortlessly computable unstructured knowledge, such as strength of association (Rogers & McClelland, 2004; Sloutsky & Fisher, 2008) or similarity (Sloman, 1993b; Sloutsky & Fisher, 2004a) on the one hand, and on the other hand, structured knowledge (Kemp & Tenenbaum, 2009; Shafto, Coley, et al., 2007; Shafto, et al., 2008; Tenenbaum, et al., 2007), which requires more time and processing effort. Overall, our results strongly support the idea that different types of knowledge which differ in their processing characteristics can have dissociable effects on people's category-based inferences. The findings also suggest that when examining the effects of knowledge on reasoning output the distinction of knowledge by *type* is perhaps more fruitful than the distinction by *domain*, as no knowledge domain was consistently more available once level of association was controlled for.

Experiments 1 and 2 used speeded response and secondary task paradigms to show that an inductive phenomenon arising from the use of structured knowledge, the causal asymmetry effect, was subject to time and mental resource constraints. Thus, when people were forced to respond rapidly or were cognitively burdened, they failed to show a causal asymmetry effect. This suggests that people could only consider what implications the nature of the causal relation had for their inductive inferences when they had sufficient time and cognitive capacity, but would fall back onto the use of unstructured knowledge under time pressure or cognitive load.

Despite this important role for structured causal knowledge, the regression analyses from these two experiments suggested that there was a pervasive influence of unstructured knowledge on people's inductive evaluations, as measured by strength of association between the categories. This is all the more impressive given that the association ratings were obtained from a different set of participants.

However, whilst the regression results, summarized in Table 3.5 below, suggested that structured knowledge accounted for more additional variance above strength of association alone when people had enough time and free cognitive resources, the pattern of beta-weights for the structured knowledge variables were less clear.

Table 3.5: Summary Regression Results Experiments 1 to 3

	Beta-weights	Strength of Association	Causal Knowledge	Taxonomic Knowledge	R^2 Change
Expt 1	Speeded Infection	.58 *	.33 ns	.33 ns	2.6%ns
	Delayed Infection	.53 *	.45*	.52 *	6.3%*
	Speeded Cells	.28 *	.60 *	.91 *	26.4%*
	Delayed Cells	.37 *	.23 ns	.88 *	44%*
Expt 2	Loaded Infection	.31*	.27 ns	.36 *	7.8%ns
	Unloaded Infection	.48 *	.63*	.78 *	11.1%*
	Loaded Cells	.33 *	-.06 ns	.51 *	30.4%*
	Unloaded Cells	.31 *	.21ns	.72 *	29.1%*
Expt3	Speeded Disease	.45 *	.35 *	.08 ns	9.6 %ns
	Delayed Disease	.24 ns	.61 *	.34 *	25.8%*
	Speeded Cells	.51 *	-.34 *	.11 ns	16.7 %ns
	Delayed Cells	.15 ns	-.31 *	.72 *	44.2%*

Some of the individual beta weights from the first two experiments might be construed as supporting Shafto et al's (2007) suggestion that taxonomic knowledge is more accessible than causal knowledge. However, this could be attributable to the directional manipulation for our causally related categories in Experiments 1 and 2. Thus, people may simply not have spotted the causal connection for diagnostically related categories, especially under speeded or loaded conditions. This might be analogous to Fenker et al's (2005) results which showed that people are slower to verify the existence of a diagnostic compared to a predictive causal link. In the third experiment we therefore dropped the directional manipulation and also included a manipulation of strength of association. Experiment 3 demonstrated that inferences were stronger for strongly associated categories than for weakly associated categories, and that this was not affected by time pressure. The regression analyses for this experiment were also very clear: strength of association predicted people's inferences under time pressure, whereas under delayed conditions, they flexibly drew on different knowledge structures depending upon the property they were reasoning about.

We did observe some interesting contextual effects, suggested both by the pattern of beta-weights across all three experiments when reasoning about infections and diseases, and by the robust interaction between property and relation. Thus, across all three experiments, causal and taxonomic beliefs predicted people's inferences about diseases and infections when they had time and available mental resources. However, unlike previous studies (Coley, et al., 2005; Heit & Rubinstein, 1994; Shafto et al., 2007), we did not find a clear-cut crossover interaction between property and category relations. Rather, whereas taxonomic arguments were rated stronger when reasoning about cells, taxonomic and causal arguments were of similar strength when reasoning about infections. This suggests that people might be thinking in detail about possible other mechanisms by which categories come to share a disease or infection (Keil, Levin, Richman, & Gutheil, 1999; Springer & Keil, 1989), apart from the obvious food chain relation. For example, it is feasible that categories that have a close taxonomic connection might share a genetic vulnerability for a disease. In this case, taxonomic relatedness acts as another causal explanation as to why categories might come to share a property (Rehder, 2006; Rehder & Hastie, 2001).

Whilst theories which emphasize unstructured knowledge (e.g. Rogers & McClelland, 2004) could in principle deal with contextual effects by assuming different contexts selectively 'prime' different features, they do not have any means of explaining the subtle differences in how available time and mental resources affected people's use of different knowledge structures to inform their inferences. Thus, in contrast to the unloaded and delayed timing conditions, people's inferences made under time pressure or cognitive load rarely seemed to be informed by more than one knowledge domain.

Similarly, it is difficult to see how connectionist models, whose hallmark processes are associative and nondirectional, could explain the asymmetries that arise when considering causal predictive inferences compared to diagnostic inferences. Based on a rational Bayesian analysis, inferences about disease transmission ought to be stronger from prey to predator than from predator to prey (Shafto, et al., 2008), as categories higher up the food chain may have contracted diseases for reasons outside the food chain. However, in order to appraise this, it is necessary to both assign structural roles to each category and bring to mind alternative mechanisms by which an organism may contract the disease. Whilst the current experiments do not pinpoint precisely which of these steps is mediated by an effortful analytical process, one or both of these aspects are likely to require mental resources and processing which goes beyond simply evaluating the strength of association between the categories. Thus, the current experiments suggest that distinguishing knowledge by type (i.e. *structured* and *unstructured* knowledge) rather than by domain (e.g. *causal* versus *taxonomic*) could be crucial for furthering our understanding of the mental processes that underlie category-based inductive reasoning.

In general, the current findings extend our understanding of the mental processes mediating category-based induction, providing a link to other domains of reasoning. For example, Verschueren, Schaeken and d'Ydewalle (2005) contend that causal conditional reasoning can be mediated by a heuristic process, in this case by a fairly effortless assessment of the probability with which the cause and effect co-vary. However, this can be overruled by an analytical process responsible for retrieving counterexamples. Their dual-process explanation parallels our current work, in which we suggest that unstructured knowledge is based on the extent to which both categories are associated or co-occur in the environment and which appears to influence inductive reasoning in a heuristic manner. In contrast, in order to appraise the importance of structural relations between categories, it is necessary to assign structural roles and/or perhaps retrieve alternative mechanisms by which categories come to share properties. Thus, the effects of structured knowledge seem to be mediated by more effortful mental processes, underscoring Evans' (Evans, 2006; 2007; 2008) repeated observation that analytical reasoning can be contextualized.

Our claim that the use of unstructured knowledge such as feature similarity or strength of association is mediated by an effortless heuristic process is reflected by the tendency to rely more on unstructured, associative knowledge when under time pressure or when burdened with a memory task. This is consistent with findings from other causal reasoning paradigms. For example, Evans, Handley, Neilens & Over (2008) showed that people lower in cognitive ability seem to be more prone to interpreting a causal conditional statement such as *'if primary school class sizes are reduced, then national literacy improves'* to also imply its converse, i.e. that *'if national literature improves, then class sizes were reduced'*. Apparently, they fail to take the directionality of the causal conditional into account, relying more on simple associations between events. This might suggest that inductive and deductive reasoning are mediated by similar mental processes, especially when people use quick heuristic estimates based on probabilistic information encountered in the environment (Chater & Oaksford, 2007; Oaksford & Chater, 2001, 2009; Oaksford & Hahn, 2007). What sets these two types of reasoning apart are the knowledge structures that more effortful analytical reasoning processes are based upon. Thus, whereas inductive reasoning is more likely to be informed by domain-specific information such as causal interactions, taxonomic relations and similar knowledge structures, accurate deductive reasoning is more likely to be based on abstract logical structures (Rotello & Heit, 2009).

Conclusions

We provide support for the claim that structured and unstructured knowledge, which are mediated by two contrasting mental processes (Rehder, 2009), can have dissociable effects on category-based inductive reasoning. Use of unstructured knowledge, such as nondirectional associative strength (Rogers & McClelland, 2004; Smith & DeCoster, 2000), seems to reflect a relatively effortless process, thus having a pervasive influence on people's inductive inferences. However, this can be supplemented by the use of more elaborate structured knowledge (Kemp & Tenenbaum, 2009; Shafto, et al., 2008), mediated by a cognitively more demanding analytical process. Structured knowledge can encode intuitive theories about the structural relationships between categories and can give rise to qualitative reasoning phenomena, such as the causal asymmetry effect when reasoning about disease transmission. Use of this type of knowledge is constrained by time and cognitive resources but can maximize inductive potency of inferences beyond the use of unstructured knowledge, such as simple associative strength between categories.

Chapter IV

Generative Category-Based Induction

The experiments described in the previous chapter suggest that manipulations designed to compromise slow analytical processes, such as encouraging people to respond quickly or burdening their working memory, decrease people's ability to use structured sources of knowledge as a guide to their inductive inferences. When analytical processes are compromised, people seem to rely more on unstructured knowledge, such as degree of association (Dickinson, 2001; Rogers & McClelland, 2004; Shanks, 2007) between categories, as a cue to inductive strength. This acts as a useful heuristic when more effortful evaluations are not possible. In contrast, people may prefer to draw on structured sources of knowledge to guide their inductive generalizations when they have the time and mental resources to do so. For example, if people do not have time or the necessary mental resources, they may infer that it is highly likely that *carrots* have a disease if they learn that *rabbits* have it, given that they are strongly associated. In contrast, they may be less likely to infer that *rabbits* have certain cells given that *camels* have them, as these two categories are only weakly associated. However, upon careful reflection, people may realize that rabbits and camels are taxonomically related. Consequently, when people draw on structured taxonomic knowledge rather than relying solely upon unstructured knowledge, they may be more confident that *rabbits* would also have the cells.

However, Coley et al. (2005) have argued that some phenomena may simply be task-specific artifacts, especially if people are unaware of the nature of the relation between categories. For example, studies such as Lopez et al. (1997) and Proffitt et al. (2000) suggest that whereas experts are able to draw on their extensive background knowledge to let thematic relations guide their inductive inferences, novices tend to use taxonomic similarity as a default reasoning strategy. Such findings are predominantly based on experimental paradigms in which people evaluate the strength of an inductive argument (*Rabbits have property X, therefore, Foxes have property X*), evaluate a series of conclusions (*Rabbits have property X. How likely is that Foxes have property X? Eagles? Hares?*), or are forced to choose between two alternative conclusion categories (*Rabbits have property X. Is it more likely that Hares or Foxes have property X?*). Yet when people make inferences in tasks that are less rigid than the aforementioned paradigms, even novices show extensive use of structured knowledge about causal and ecological relations. For example, Baker and Coley (Baker & Coley, 2005; Coley & Baker, 2004) gave their participants two related category pairs and asked them to make inference about which other categories might also have a novel property. Overall, people tended to make more ecological than taxonomic inferences, suggesting that the open-ended methodology allowed them to use relevant knowledge more flexibly.

Vitkin, Coley and Hu (2005) used a similar paradigm with children, in which the young children were asked to generate categories that might share a disease or internal substance known to be present in two base categories. They found that older rural children, who have more experience with the natural world, were inclined to use the relation between the base categories as an inductive guide. Thus, they made more similarity-based inferences when the base pairs were taxonomically related (e.g. *newt* and *box turtle*) and more interaction-based inferences when the base category pairs were ecologically related (e.g. *salmon* and *black bear*). In contrast, urban children and younger rural children tended to be guided by similarity regardless of the nature of the relation between the two categories.

Both of the studies described above focus on the flexible use of domain-specific knowledge (such as *taxonomic* versus *ecological*) rather than looking at the use of different types of knowledge that diverge in their processing properties. Yet using such an alternative open-ended induction paradigm would also be another way of exploring whether people who are cognitively compromised do indeed rely more strongly on unstructured knowledge, such as simple associations, rather than drawing on highly structured knowledge sources which involves effortful processing. We would have a stronger claim for the importance of considering the influence of unstructured types of knowledge alongside more structured sources of knowledge if we could show that the dissociable effects of structured and unstructured knowledge demonstrated in the previous experiments emerge for *generative inferences* as well as *evaluative inferences* of the kind used in Experiments 1 to 3. In this context, *evaluative inferences* are those which ask people to evaluate the likelihood that two (or more) categories share a property, whereas for *generative inferences* people need to produce their own conclusion category upon learning that a base category has a certain property.

4.1 Experiment 4

In this experiment we used a generative inference paradigm where people were told that a base category had a novel infection or cells and were asked to infer which other category was most likely to also have the infection or cells. We predicted that under cognitive load, people who were asked to generate a conclusion category would fall back onto unstructured knowledge, such as simple associations. Consequently, category pairs generated under conditions of memory load should be rated as more strongly associated than categories generated under uncompromised circumstances. For example, if people learn that *rabbits* have a novel disease, they might respond that *hares* are likely to also have the disease, based on the strong association between these two categories. In contrast, it might be that people who have time and available mental resources maximize the potency of their inductive inferences by drawing on several sources of structured knowledge. In the above example, people might reason that the disease found in *rabbits* is most likely to also be present in *foxes*, by virtue of their shared habitat, food chain relation, as well as belonging to the same biological group of *mammals*. Thus, whilst it could be said that "*rabbits* and *foxes*" are related in more ways than "*rabbits* and *hares*", the mean association rating from our pre-test showed that the latter category pair was nonetheless rated as being more strongly associated (M = 8.28) than the former category pair (M = 6.00).

4.1.1 Method

Design and Materials

The experiment had two phases, the induction generation phase and an associative rating phase.

Induction Generation Phase

The first phase of the experiment had a 2 (load: heavy or light) by 2 (property: infection or cells) between-subjects design.

Thirty six people (the "reasoners") completed the generative inductive task. They were presented with 10 base categories and were told that each category had a novel property, either a disease (e.g. *has disease 5y5u*) or cells (e.g. *has 3-yu-cells*). Participants were then asked to generate ONE other category that they believed was most likely to also have the property. Half of the participants generated their inferences whilst carrying out a resource-demanding secondary task. The remaining participants in the light load condition carried out a secondary task designed to place minimal burden on cognitive resources. Of the 18 participants in each load condition, 9 reasoned about infections, whereas the other 9 reasoned about cells. For example, people would read the following generative induction problem:

Weasels have 4Ou-cells / infection 4Ou.

Which other category is most likely to also have 4Ou-cells/ infection 4OU?

Once people had written down their response, they rated how likely they thought it was that the two categories shared the property on a scale from 1 (very unlikely) to 9 (highly likely).

Preceding each of the induction trials was a secondary memory task. People were presented with a 4*4 dot matrix with 4 randomly placed black dots. Participants had to remember the location of the dots, complete the induction task and then recall the location of the dots in an empty matrix. The configuration of the dots was different for each of the 10 trials.

In the heavy load condition, the dots were completely randomly placed, with the restriction that they could never appear in a straight or diagonal line. In the light load control condition, the dots always appeared in a straight or diagonal line, placing minimal burden on working memory.

Association Rating Phase

Following this generative task, in the second phase of the study each individual's 10 category pairs were transcribed onto an association rating sheet and interspersed with 10 weakly associated distracter items. Category pairs from 2 participants in the loaded condition could not be transcribed, as more than 4 generated categories were unreadable or hadn't been completed. Thus, we had 34 different rating sheets.

A group of 136 participants (the "raters") who had not taken part in the first phase received one of 34 different sheets (approximately 4 participants per sheet) and were asked to rate the strength of association on a scale from 1 (unrelated) to 9 (very highly associated) between 20 pairs of categories. These included the 10 idiosyncratic category pairs generated by participants in phase 1 and 10 identical weakly related distracter pairs. The format for the association ratings was identical to the one used in the pre-test reported in Chapter 2.

We then compared the mean strength of association ratings assigned to the categories the "reasoners" had generated when reasoning about cells or infections under loaded or unloaded conditions. It might be the case that reasoners under loaded conditions rely more heavily upon easily available unstructured knowledge. Consequently we expected that the categories generated under a heavy memory load would be rated as more strongly associated than category-pairs generated under a minimal load.

Structured Relation Ratings

To look at the nature of the relationship between the base and the generated target category, the experimenter and a second blind coder rated whether there was a taxonomic and/or interaction-based relationship between the two categories for each the 10 category pairs generated by the "reasoners" in phase 1. We tried to keep this very simple as it can be difficult to judge what exact relation people were thinking about. Thus, categories were classed as taxonomically related if they belonged to the same biological group. Categories could be coded as causally related if they had a similar diet, lived in a similar habitat, or if they exhibited behavioural or food chain interactions. The detailed template can be found in Appendix A. Coders attached a score of 0 if there was no discernable relation between the two categories (e.g. *fox* → *hedgerow*). They attached a score of 1 if they were exclusively interaction-based (e.g. *fox* → *chicken*) or taxonomically related (e.g. *fox* → *dingo*). As the two relations were not mutually exclusive, coders attached a score of 2 if they were related in both ways (e.g. *fox* → *rabbit*).

Participants

The 36 "reasoners" who took part in phase 1 were recruited via the Durham Psychology participant pool and received course credit for their participation. There were 5 males and 31 females with a mean age of 21.3 years (*SD*= 3.1 years).

The 136 "raters" in phase 2 completed the association rating task at the beginning of one of their Psychology lectures at Durham University. There were three groups, one group of 58 first-year Psychology students, a group of 54 second-year Psychology students and a group of 24 Psychology Masters students.

Procedure

The inductive generation task was run individually on a laptop. Instructions stated that we were interested in people's beliefs about shared infections and cells whilst remembering dot patterns. They were asked to think of ONE other category which was most likely to share the disease or cells with a base category. They were given the following example not used in the main experiment:

Squirrels have infection Rtt5/ t3e-cells.

Which other category do you think might have infection Rtt5/ t3e-cells?

The instructions emphasized that the category they generated could be anything, such as another animal, a plant and that there were no right or wrong answers.

Participants were randomly assigned to either the heavy load or control condition and reasoned either about cells or infections. To familiarize them with the experimental sequence, participants completed two practice trials which had an identical format as the main trials. On each trial, participants saw a matrix with 4 randomly placed dots. This was displayed for 2 seconds. Following this, people read the premise which stated that a specific plant or animal category had a novel infection or cells. They were then asked to generate one other category that they thought would most likely share the property with the premise category. Once they had written down their response on a score sheet, participants were instructed to rate how confident they were that that this category also has the infection/cell by pressing any number between 1 (very unlikely to share the infection/cells) to 9 (highly likely) on the keyboard. This was followed by an empty dot matrix and people recalled the location of the dots by holding and pressing down the mouse cursor over the appropriate box. Following this, a newly configured dot matrix was presented, followed by the next premise sentence.

In phase two of the study, groups of “raters” who had not taken part in the first phase randomly received one of 34 different association rating sheets. The format and instructions were identical to the pre-test association rating task reported in Chapter 2. To summarize, participants were informed that we were interested in their beliefs about the strength of association between category pairs. They were asked to think about all kinds of possible associations, such as causal, functional or similarity-based, and although these examples included structured relations, we emphasized that we wanted them to give the answer that came to mind first, as fast as possible. They were asked to rate the association between the 20 pairs of categories (10 category pairs generated by one of the participants in phase 1, interspersed with 10 weakly associated distracter pairs) on scale from 1 (unrelated) to 9 (very strongly associated).

4.1.2 Results

Two “reasoners” who had taken part in phase 1 in the heavy load condition had to be eliminated, one because they gave ‘don’t know’ responses on more than 25% of the trials in the generative task, and another because more than 25% of their generated categories could not be deciphered due to bad handwriting.

Strength of Association Ratings

First and foremost, we were interested whether people whose resources are restricted rely more heavily on unstructured knowledge, such as simple associations, when making inductive inferences. Association ratings made by the 136 “raters” were averaged across the 10 idiosyncratic category pairs generated by the “reasoners” and were analysed with a 2 (load: heavy or light) by 2 (property: cells or infection) between-subjects ANOVA.

There was a main effect of property, $F_{(1, 132)} = 6.34$, $p = .013$, effect size d = .44. Thus, on average, people rated category pairs which had been generated for shared cells as more highly associated ($M = 6.69$, $SE = .14$) than categories generated for shared infections ($M = 6.21$, $SE = .13$). Likewise, there was a main effect of load, $F_{(1, 132)} = 8.08$, $p = .005$, effect size d = .50, such that categories generated under conditions of heavy load ($M = 6.72$, $SE = .14$) were rated as more strongly associated than categories generated by people whose resources had been minimally taxed ($M = 6.16$, $SE = .13$). This supports our hypothesis that unstructured knowledge such as strong associations between categories, act as a useful heuristic for inductive reasoning when cognitive resources are sparse.

Finally, there was no interaction between property and load condition $F_{(1, 132)} = .17$, $p = .68$, effect size d = .06.

Generative Inductive Strength Ratings

We examined "reasoners" actual inductive strength ratings for the categories they had generated. The results from a 2 (load: heavy versus light) by 2 (property: cells or infections) between-subjects ANOVA showed that inductive strength ratings did not differ between the load conditions, $F_{(1,\ 30)} = 1.22$, $p = .28$, effect size d = .4. People under heavy load gave a mean inductive strength rating of 5.79 (*SE* = .38) whereas people under minimal load rated the strength of their inductions at 5.22 (SE = .35). There was also no main effect of property, $F_{(1,\ 30)} = .53$, $p = .47$, effect size d = .26. Inferences about cells (*M* = 5.70, *SE* = .37) were rated as strong as inferences about infections (*M* = 5.32, *SE* = .37). Finally, there was no interaction between property and load, $F_{(1,\ 30)} = .088$, $p = .77$, effect size d = .11.

Relation between Inductive Strength Ratings, Strength of Association and Structured Relations between Categories

Table 4.1 below shows descriptive statistics for inductive strength, strength of association and the measure of structured knowledge across the two load conditions for the 34 "reasoners".

Table 4.1: Descriptive Statistics for Inductive Strength, Strength of Association and Structured Relations across the 34 "Reasoners"

Measure		Mean	SD	Range
Inductive Strength	Heavy	5.79	1.07	3.60
	Light	5.22	1.61	5.90
Strength of Association	Heavy	6.67	.69	2.35
	Light	6.22	.86	2.97
Structured Relation	Heavy	1.05	.17	.60
	Light	1.17	.15	.50

We wanted to explore whether there were systematic differences across the two load conditions with regards to the types of knowledge (i.e. structured versus unstructured knowledge) that influence people's inductive strength ratings,. For each of the 34 "reasoners" who had generated a category and rated the inductive strength of this argument, we calculated an *association strength* measure and a *structured relation* measure for each of their 10 individual category pairs. For each "reasoner", we then used these as predictor variables in a linear regression analysis to predict their inductive strength ratings.

The *associative strength* measure represented the mean strength of association scores for a "reasoner's" 10 category pairs across the four "raters" from phase 2. For each of the 34 "reasoners", we calculated Cronbach's Alpha across the association ratings. The mean Cronbach's Alpha across all "reasoners" was .68 (SD = .18), which indicated that the association ratings had an acceptable inter-rater reliability.

The *structured relation* measure represented an index of the possible underlying structural relations between the 10 base categories and the categories generated by a "reasoner". The experimenter and a second coder who was unaware of the conditions under which the participants had generated their targets evaluated in how many ways the generated target could be related to the base, assigning 0 if there was no discernable link, 1 if they were either exclusively taxonomically related or shared an interaction-based relation, and 2 if they were related in several ways. Concordance rate between the two coders was 82%. For the category pairs where there was disagreement, this was resolved through discussion with a third independent coder.

For each of the 34 "reasoners", we then ran individual linear regression analyses. We used the *associative strength* and *structured relation* measures to predict each "reasoner's" inductive strength ratings across the 10 category pairs. We then used the beta weights assigned to these two predictors in a 2 (load: heavy or light) by 2 (type of beta-weight:

associative or *structured relation* beta weight) mixed-design ANOVA, with type of beta-weight as the repeated-measures variable.

Although the *associative strength* beta-weight (*M* beta = .19, *SE* = .07) was of larger magnitude than the *structured relation* beta-weight (*M* beta = .02, *SE* = .07), this difference was not statistically significant, $F_{(1, 32)} = 3.07$, *p* = .089, effect size d = .6.

There was no significant main effect of load, $F_{(1, 32)} = .16$, *p* = .69, effect size d = .14. Thus, under heavy load conditions the mean beta-weight was .12 (*SE* = .07), whereas in the light load condition, it was .09 (*SE* = .06).

However, there was a significant interaction between beta-weight type and load, $F_{(1, 32)} = 4.18$, *p* = .049, effect size d = .72. Bonferroni posthoc tests showed that when people were under heavy memory load, the *associative* beta-weight (*M* beta = .3, *SE* = .09) was significantly larger than *structured relation* beta-weight (*M* beta = -.06, *SE* = .1, *p* = .014, effect size d = .7). In contrast, this difference was not significant when people were not under a heavy memory load, where the *associative* beta weight (*M* beta = .07, *SE* = .09) was of a similar magnitude as the *structured relation* beta-weight (*M* beta = .1, *SE* = .09), *p* = .83, effect size d = .05). This interaction is illustrated in Figure 4.1 below.

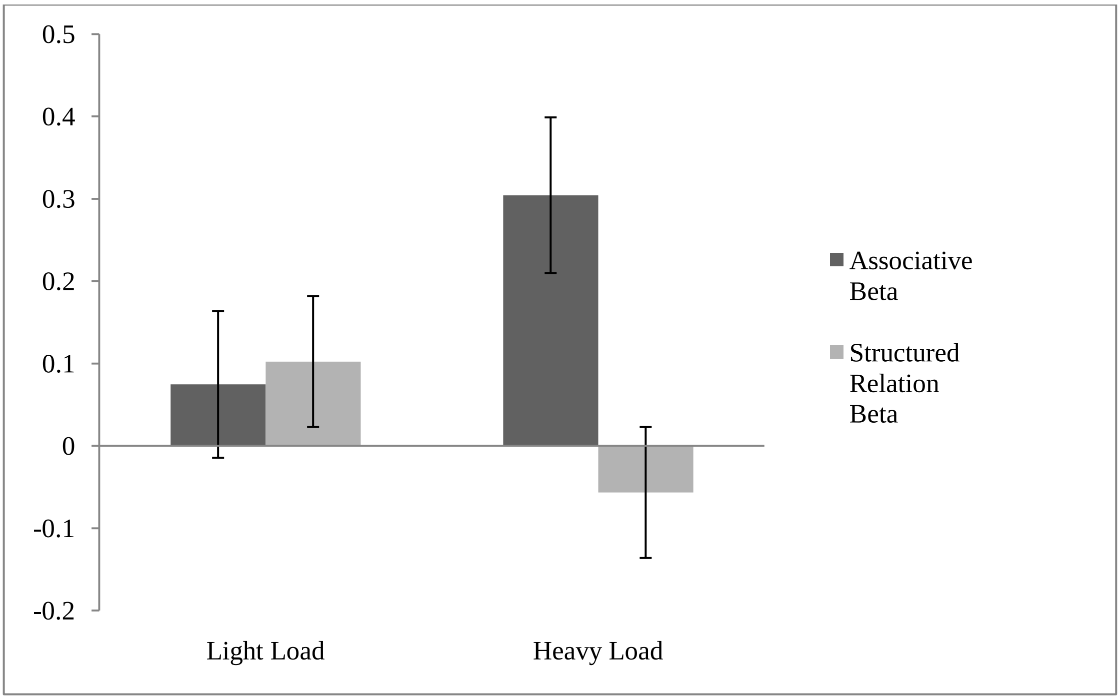

Figure 4.1: Mean Beta-Weights across the two Load Conditions in Experiment 4

Secondary Task Manipulation Check

To check the effectiveness of our secondary task manipulation, we compared the mean number of dots people correctly recalled across the two load conditions. This confirmed that the complex dot patterns were more difficult than the simple dot patterns. Reasoners were significantly worse at remembering the location of the complex dot patterns in the heavy load condition (*M* correct recall = 2.8, *SD* = .73) than the people who had to remember the correct location for the simple dot patterns used in the light load condition (*M* correct recall = 3.6, *SD* = .50), $t_{(32)} = 3.69$, $p = .001$, effect size $d = 1.3$.

Relations between Association Ratings, Index of Structured Relations and Web-Based Co-occurrence

If structured and unstructured knowledge are indeed dissociable, then mean associative strength ratings should not be correlated with the index of structured relations for the category pairs. However, they should be correlated with a measure of conditional co-occurrence. To check this, we calculated the correlation between the mean associative strength rating and the index of structured relations for each individual across the 10 category pairs they had generated. We then calculated the conditional World Wide Web co-occurrence as described in Chapter 2 for each of the individual "reasoner's" 10 category pairs, and correlated this with the mean associate strength score. For both correlations, we used Kendall's Tau-b rather than a parametric test of correlation, as each measure was on a different scale, and the index of structured knowledge could only take on 3 different values.

These two correlation coefficients were then compared across the 34 reasoners with a paired samples t-test, $t_{(32)} = 3.29$, $p = .002$. This showed that the mean correlation coefficient between association ratings and co-occurrence (*M* Kendall's $\tau = .19$, *SD* = .26) was significantly larger than the mean correlation coefficient between association ratings and the index of structured knowledge (*M* Kendall's $\tau = -.02$, *SD* = .28). This was further supported by a one-sample t-test, which showed that the mean correlation coefficient between association ratings and co-occurrence was significantly different from zero, $t_{(33)} = 3.87$, $p < .0005$. In contrast, the mean correlation coefficient between association ratings and the index of structured knowledge was not significantly different from zero, $t_{(33)} = -.36$, $p = 72$.

4.1.3 Discussion

The current study looked at whether the influence of different types of knowledge, unstructured and as well as structured sources of knowledge, is also evident in a paradigm where people were asked to generate rather than simply evaluate a category-based inductive inference. We showed that as well as structured sources of knowledge, such as taxonomic group membership or causal relations, strength of association had a crucial influence on people's generative inferences. Categories produced under heavy load were rated as more strongly associated than categories generated under a light load. Furthermore, under heavy load conditions, those association ratings were better predictors of inductive strength than an index of structured relations. In contrast, in the light load condition there was no difference in how well association ratings and the index of structured relations predicted inductive strength ratings. Thus, we have further evidence that inductive reasoning is influenced by different and dissociable types of knowledge. Unstructured knowledge, such as associative strength, is especially powerful when people can dedicate relatively little cognitive resources to recruiting more structured knowledge. Rather than being paradigm-specific, unstructured or associative knowledge appears to be a useful heuristic for guiding induction across a range of paradigms, especially when more analytical processes are compromised.

Interestingly, there were no differences in peoples' inductive confidence across the two load conditions. The reason may be that people based their inductive evaluations on different types of knowledge in the two conditions. As suggested by the systematic differences in the *associative strength* and *structured relation* beta weights across the two load conditions, people in the heavy load condition seemed to base their ratings more on strength of association, whereas people in the unloaded condition appeared to rely equally on strength of association and structured information about possible diverse relations between the base and generated target categories. This suggests that inductive inferences can be informed by dissociable types of knowledge, which appear to be subject to diverging cognitive processing constraints.

4.2 Experiment 5

The goal of the second experiment was to replicate the finding that people seem to rely more heavily on unstructured knowledge when making generative inductions under cognitive load, but draw on structured sources of knowledge when they are not cognitively compromised. As drawing on different kinds of structured knowledge is especially crucial when people make inductions about a variety of different properties, we manipulated property (cells or infections) about which people made generative inductive inferences in a within-subjects design. This might amplify the need to draw on different sources of structured knowledge, especially when people are not burdened with a heavy memory load.

4.2.1 Method

Participants

The "reasoners" who took part in the first phase of the study were 23 Undergraduate and Graduate students from Durham University who received either course credit or £5 for their involvement. There were 4 males and 19 females with a mean age of 23.5 years (SD = 5.2 years).

The "raters" recruited for the second phase were 92 Durham Undergraduate Psychology students who completed the association rating sheets at the beginning of their group seminars or practical lectures.

Materials and Design

Inductive Generation Task

The first phase of the experiment had a 2 (load: heavy or light) by 2 (property: infection or cells) mixed design, with load as the between-subjects manipulation. The materials had the same format as the previous experiment, with some minor changes so that property could be manipulated within participants. Thus, each "reasoner" was presented with 20 base categories and asked to generate ONE other category that they believed was most likely to also have the property. For half of the base categories people reasoned about infections, whereas they reasoned about cells for the other half. 11 "reasoners" generated their inference under heavy cognitive load and 12 generated their inferences under minimal cognitive load.

As before, depending on the condition "reasoners" were randomly allocated to, each trial was preceded by either a complex or simple 4*4 dot matrix. Once people generated their responses, they rated how likely they thought it is that the two categories shared the property on a scale from 1 (very unlikely) to 9 (highly likely). Following this, participants recalled the location of the dots as best as they could.

Association Ratings

Following the generative task, in the second phase of the study each "reasoner's" 20 category pairs were transcribed onto an association rating sheet and interspersed with 15 weakly associated distracter items. Each of these sheets was randomly given to one of the 92 "raters" (approximately 4 "raters" per sheet), who rated the strength of association between the two category pairs on a scale from 1 (unrelated) to 9 (very highly associated). People were instructed to include all kinds of associations, such as causal, functional or similarity-based and were asked to give the first response that came to mind, as fast as possible. Based on Experiment 4, we predicted that categories generated by "reasoners" under the condition of memory load would be rated as more strongly associated than category pairs generated under minimal cognitive load.

Structured Relation Ratings

As in the previous experiment, in order to determine the underlying structural relations between the base categories and the categories generated by "reasoners" in phase 1, the experimenter and a second blind coder rated whether there was a taxonomic and/or interaction-based relationship between the 20 category pairs. The criteria for judging a response as taxonomic and/or causal were the same as in the previous experiment. To recapitulate, participants were awarded 0 if there was no discernable link between the base and the generated category (e.g. *alligator* → *soil*), 1 if they were taxonomically related (e.g. *zebra* → *horse*), 1 if they were related through a causal link or ecological interaction (e.g.

hawk → *mouse*) and 2 points if there was both a taxonomic and interaction-based relation between the categories (e.g. *cod* → *shark*).

Procedure

The inductive generation task was run individually on a laptop. “Reasoners” were given verbal and written instructions which included an example not used in the main task. They were randomly assigned to either the heavy or light load condition. As before, 2 practise trials familiarized participants with the experimental sequence.

On each trial, participants saw a matrix with 4 randomly placed dots for 2000 milliseconds. This was followed by the premise stating that a specific plant or animal category had a novel infection or cells. Participants generated one other category that they believed would most likely share the property with the premise category and rated how strong they thought this inference was on a scale from 1 (very unlikely to share the infection/cells) to 9 (highly likely) by pressing any number between 1 and 9 on the keyboard. This was followed by an empty dot matrix. People had to recall the location of the dots in the empty dot matrix by pressing down the mouse cursor over the appropriate box. Following this, a newly configured dot matrix was presented, followed by the next premise sentence.

In the second part of the study, a group of 92 “raters” who had not taken part in the first phase received one of 23 different sheets (4 “raters” per sheet) and were asked to rate the strength of association between the 35 pairs of categories (20 idiosyncratic pairs generated by “reasoners” in phase 1 and 15 weakly related distracter pairs). They were instructed to consider all kinds of associations but to give the answer that came to mind first.

4.2.2 Results

Association Ratings

As in the first generative induction experiment, we averaged association ratings made by "raters" in phase 2 across all 20 category pairs generated by "reasoners" in phase 1. One "rater" in phase 2 failed to complete more than 50% of the association ratings and was excluded from the analysis.

The mean association scores were then analyzed with a 2 (load: heavy or light) by 2 (property: cells or infection) mixed-design ANOVA, with load as the between-subjects variable.

There was no main effect of property, $F_{(1, 89)} = 1.08$, $p = .30$, effect size d = .22. People gave a mean association rating of 6.23 (*SE* = .12) for category pairs which had been generated about shared cells, and a mean association rating of 6.15 (*SE* = .13) for category pairs generated about infections.

As predicted though, there was a main effect of load, $F_{(1, 89)} = 4.03$, $p = .048$, effect size d = .42, such that categories generated under conditions of heavy load (*M* = 6.42, *SE* =.16) were rated as more strongly associated than categories generated by "reasoners" whose resources were minimally taxed (*M* = 5.96, *SE* = .16).

Finally, there was no interaction between property and load condition $F_{(1, 89)} = .55$, $p = .46$, effect size d = .16.

Generative Inductive Strength Ratings

We examined "reasoners'" actual inductive strength ratings for the categories they had generated. The results from a 2 (load) by 2 (property) mixed-design ANOVA with load as a between-subjects variable paralleled those from Experiment 4. Inductive strength ratings did not differ between the load conditions, $F_{(1, 21)} < .001$, $p = .99$, effect size d < .01. "Reasoners" under heavy load gave a mean inductive strength rating of 5.55 (*SE* = .40) whereas those under minimal load rated the strength of their induction at 5.56 (*SE* = .42).

There was also no main effect of property, $F_{(1,\ 21)} = 2.1$, $p = .16$, effect size d = .63. Inferences about cells (*M* =5.68, *SE* = .32) were rated as strong as inferences about infections (*M* =5.42, *SE* = .29).

The interaction between load and property was not statistically significant, $F_{(1,\ 21)} = 2.68$, $p = .12$, effect size d = .71. Thus, these results mirror those from the previous experiment.

Relations between Inductive Strength Ratings, Structured Relations and Associative Strength

Means, standard deviations and ranges for inductive strength, structured relations and associative strength across the 23 "reasoners" are shown in table 4.2 below. As in Experiment 4, in order to explore whether "reasoners'" inductive strength ratings might be influenced by different types of knowledge in the two load conditions we used an *associative strength measure* and an index of *structured relations* to predict their inductive strength ratings.

Table 4.2: Descriptive Statistics for Inductive Strength, Strength of Association and Structured Relations across the 23 "Reasoners"

Measure	Load	Mean	SD	Range
Inductive Strength	Heavy	5.56	1.31	4.45
	Light	5.55	1.43	4.90
Strength of Association	Heavy	6.42	0.84	2.85
	Light	5.92	0.80	2.39
Structured Relation	Heavy	0.92	0.17	0.65
	Light	1.03	0.14	0.50

To create *associative strength measure* we averaged the mean strength of association scores attached to each "reasoner's" 20 category pairs across the four "raters" from phase 2. We then calculated Cronbach's Alpha for each of the 23 "reasoners" across the association ratings. The mean Cronbach's Alpha across all "reasoners" was .71 (*SD* = .13), showing that the association ratings had good inter-rater reliability.

To create the *structured relation measure*, the experimenter and a second blind coder then assessed in how many ways the generated target could be related to the base. 0 was attached if there was no obvious structured link, 1 if there was either a taxonomic or interaction-based connection, and 2 if they were related in more than one way. Disagreement was resolved through discussion with a third independent coder. Concordance rate was somewhat lower than in the first experiment at 67%, probably due to the additional number of fairly unusual base categories (e.g. *Eucalyptus*). Disagreements between the two primary coders were resolved through discussion with two further colleagues.

For each "reasoner" who had taken part in phase 1, we used the *associative strength* and *structured relation* measures to predict his/her inductive strength ratings. The beta weights were then subjected to a 2 (load: heavy or light) by 2 (type of beta weight: *associative* versus *structured relation* beta weight) mixed-design ANOVA, with type of beta weight as the repeated-measures variable.

There was no significant main effect of type of beta weight, $F_{(1, 21)} = .068$, $p = .80$, effect size d = .11. The *associative* beta weight (*M* beta = .16, *SE* = .05) was similar to the *structured relation* beta weight (*M* beta = .15, *SE* = .03).

Although the average beta weight in the light load condition was larger (*M* beta= .20, *SE* = .04) than in the heavy load condition (*M* beta= .11, *SE* = .04), this main effect did not reach statistical significance, $F_{(1, 21)} = 3.22$, $p = .09$, effect size d = .78. However, there was a significant interaction between beta weight type and load, $F_{(1, 21)} = 6.53$, $p = .018$, effect size d = 1.1. This is illustrated below in Figure 4.2

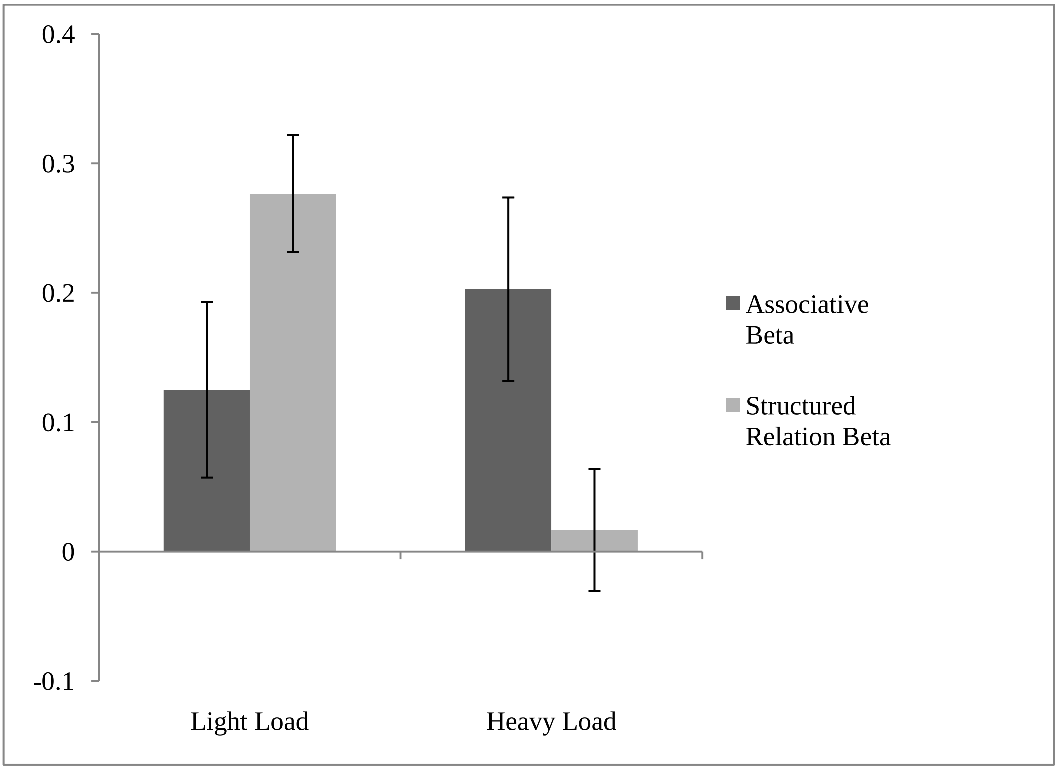

Figure 4.2: Beta Weights across the two Load Conditions in Experiment 5

Bonferroni posthoc tests showed that when "reasoners" were under heavy memory load, the *associative strength* beta weight (*M* beta= .20, *SE* = .07) was larger than the *structured relation* beta weight (*M* beta = .02, *SE* = .02), although this difference was not quite statistically significant due to the small number of participants in this condition ($p = .065$, effect size d = 1.2). The pattern was reversed when "reasoners" were not under a heavy memory load. Thus, the *structured relation* beta weight (*M* beta = .28, *SE* = .05) was slightly but not significantly larger in magnitude than *associative strength* beta weight (*M* beta = .11, *SE* = .07, p=.11, effect size d = .62).

Across the two load conditions, the *associative strength* beta weight was slightly but not significantly larger for “reasoners” who generated their inferences under load compared to those who were not cognitively compromised (p = .44, effect size d = .32). In contrast, the mean *structured relation* beta weights were significantly larger for “reasoners” who generated their inferences under minimal cognitive load compared to “reasoners” who were cognitively burdened by the complex dot matrix task (p = .001, effect size d = 1.6).

This suggests that “reasoners” seemed to rely on different types of knowledge depending upon the availability of mental resources. It seems that both structured and unstructured types of knowledge were influencing inductive strength ratings when “reasoners” were only under minimal cognitive load. In contrast, it appears that “reasoners’” inductions in the heavy load condition were predominantly influenced by unstructured knowledge.

Relations between Association Ratings, Index of Structured Relations and Web-Based Co-occurrence

As in the previous experiment, it is important to confirm that association ratings and the index of structured relations are measuring different types of knowledge. If this is the case, then there should be a stronger correlation between mean associative strength ratings and conditional web-based co-occurrence (two measures of unstructured knowledge) than between associative strength and the measure of structured relations between category-pairs. We calculated the Kendall’s Tau-b correlation coefficient between the mean associative strength rating and the index of structured relations for each of the 23 individuals across the 20 category pairs they had generated. For each of the 23 individual “reasoner’s” 20 category pairs, we then calculated the conditional World Wide Web co-occurrence as described in Chapter 2. Again, for each individual we obtained a Kendall’s Tau-b correlation coefficient between web-based co-occurrence and the mean association ratings attached to each of the 20 category pairs.

We then compared these two correlation coefficients across the 23 reasoners using a paired-samples t-test, $t_{(21)} = 5.97$, $p < .0005$. The mean correlation coefficient between association ratings and co-occurrence (*M* Kendall's $\tau = .41$, SD = .12) was significantly larger than the mean correlation coefficient between association ratings and the index of structured knowledge (*M* Kendall's $\tau = .14$, SD = .20). These correlations were somewhat larger than in the previous experiment. The one-sample t-test showed that both the mean correlation coefficient between association ratings and co-occurrence, $t_{(22)} = 16.62$, $p < .0005$, and the mean correlation coefficient between association ratings and the index of structured knowledge, $t_{(22)} = 3.37$, $p = .002$ were significantly different from zero. However, given that a correlation of .14 is very small and the correlation coefficient between the two indices of unstructured knowledge was three times larger, this does support our contention that association ratings and the index of structured relations are measuring dissociable types of knowledge.

Secondary Task Manipulation Check

In order to verify our manipulation of compromising "reasoners'" available mental resources with a secondary memory task, we compared the mean number of dots people correctly recalled across the two load conditions. As expected, remembering the complex dot patterns was more difficult than memorizing the simple dot patterns. People made significantly more errors recalling the location of the 4 dots in the heavy load condition (*M* correct recall = 2.7, *SD* = .65) compared to the light load condition (*M* correct recall = 3.4, *SD* = .45), $t_{(18)} = 2.9$, $p = .009$, effect size $d = 1.3$.

4.2.3 Discussion

The current findings largely replicated those from the previous Experiment 4. In phase 1, "reasoners" generated their own inference categories either under a heavy or light cognitive load and rated their confidence in their inference. In a second phase, another group of “raters” rated the strength of association between the category pairs generated in phase 1. Crucially, “raters” in phase 2 rated the base and conclusion categories as more strongly associated when the inferences had been generated by “reasoners” who were under a heavy cognitive burden than when the inferences had been made by “reasoners” under minimal cognitive load. Furthermore, on average, strength of association between the category pairs was slightly more predictive of “reasoners’” evaluations of the strength of their inferences in the heavy load condition compared to the light load condition. In contrast, a measure of structured relations between the category pairs was predictive of “reasoners’” inference strength in the minimal load condition but did not predict evaluations of inductive inference strength in the heavy load condition.

One interesting difference was the fact that the average *structured relation* beta weight was much larger than the *associative strength* beta weight in the light load condition, whereas in the previous experiment, there was no difference between the two types of beta weights in the light load condition. Unlike Experiment 4, in this experiment the property “reasoners” generated their inferences about was manipulated within- rather than between-subjects. It is feasible that having to make inferences about two different properties might have highlighted the need to draw on different domains of structured knowledge.

The current findings also raise some interesting questions about the role of inhibitory control in inductive reasoning. Thus, whilst a strong association between categories is often a good guide for inductive reasoning, there seem to be occasions where inductions can be more potent when a strong association is inhibited. For example, consider the case in which oranges are known to have a novel infection, certain kinds of cells or another specific property. The primary associate of *orange* is *apple* (Moss & Older, 1996). Although both categories belong to the superordinate category of fruits, *apples* are pomaceous fruits whereas *oranges* are citrus fruits, which differ in important characteristics such as ideal growing conditions and botanical anatomy. Thus, an inductively more potent inference might be from *oranges* to *lemons*, *grapefruits*, *limes* or *kumquats*. However, to make such an inference, people have to inhibit the strong association they have between *apples* and *oranges* and recruit structured knowledge about *oranges*, such as facts about their physical, anatomical and botanical attributes, how and where they are grown, typical culinary uses or any other knowledge that might be relevant for generalizing a certain property (Barsalou, 1982; Lin & Murphy, 2001; Ross & Murphy, 1999). However, inhibiting unstructured associative knowledge is likely to involve effortful executive processing.

Additionally, recruiting structured knowledge presumably requires considerable time, effort and available mental resources (Fenker, et al., 2005; Satpute, et al., 2005). Hence, these two sources of difficulty, inhibiting unstructured knowledge on the one hand and recruiting structured knowledge on the other, might explain why people who are cognitively compromised rely more heavily on the useful heuristic of generalizing to strongly associated categories.

4.3 General Discussion

The goal of the two studies was to try and replicate the finding that category-based induction is shaped by two contrasting types of knowledge, structured and unstructured, with a less artificial paradigm. In the spirit of the philosophical principle of diversity (Bacon, 1620/1898; Hempel, 1966; Myrvold, 1996; Nagel, 1939; Steel, 1996, Heit, Hahn & Feeney, 2005)[6], using a variety of methods should deliver more convincing evidence for our hypothesis that

[6] For a critical evaluation, see Wayne (1995)

inductive reasoning is shaped by unstructured as well as structured knowledge, each of which has different processing characteristics.

As suggested by Coley et al. (2005), some inductive inference phenomena may simply be artifacts arising from the use of specific paradigms, rather than reflecting the fundamental principles governing category-based inductive reasoning. The evaluation task has been a popular method for assessing features of category-based induction since its introduction by Rips (1975). In this task, people evaluate arguments pre-selected by the experimenter. Some of these inferences may appear unnatural to participants if they lack the relevant knowledge. In normal everyday circumstances we are unlikely to ever juxtapose two unrelated categories and then evaluate the '*unlikelihood*' that they share common features. Thus, if people fail to show a particular phenomenon, such as the use of causal knowledge, it may simply reflect the fact that they had no specific factual knowledge about the relationship between two categories. Our post-test from the previous experiments for example showed that there were vast differences in people's knowledge about food chain relations. Whereas only a minority knew that *cats* hunt *squirrels*, or that *eagles* prey on *meerkats*, almost everyone knew that *snails* feed on *cabbage* or that *frogs* eat *flies*. Thus, using evaluative paradigms can make it problematic to disentangle lack of knowledge from the effects of more domain-general manipulations on the use of knowledge, such as putting people under time pressure (e.g. Shafto et al. 2007).

The risk of masking or distorting the use of domain-general reasoning strategies because of lack of domain-specific knowledge is akin to a problem that arises in developmental research on the diversity effect. Early studies (Gutheil & Gelman, 1997; Lopez, et al., 1992) suggested that children under the age of 9 years did not use diversity as a guide to their inductive inferences, leading them to conclude that adult and child inductive reasoning is mediated by fundamentally different processes. However, Heit and Hahn (2001) demonstrated that the results were largely an artifact of the experimental design and stimulus materials. Thus, when they used familiar categories (e.g. *dolls* or *balls*), transparent (e.g. *belongs to*) rather than hidden properties (e.g. *has leukocytes inside*) and did not force the children to switch strategies between adjacent trials, they showed that even 5-year-old children used diversity as a category-based inductive reasoning strategy. If the reasoner lacks relevant knowledge, there will be an inevitable discrepancy between performance and competence. Such discrepancies suggest that the differences observed between adults' and children's reasoning patterns do not lie in the nature of the processes per se, but are determined by crucial differences in content-specific knowledge.

The situation could be similar in evaluative induction paradigms, whereby the failure to observe a specific reasoning phenomenon might be attributed to domain-general factors, when it is in fact driven by a lack of domain-specific knowledge. The use of the generative induction paradigm enables people to recruit whatever relevant knowledge they have, enabling us to be more certain that differences which emerge under divergent levels of cognitive pressure are attributable to processing differences rather than divergences in content-specific knowledge.

As knowledge is central to inductive reasoning (Coley, et al., 2005; Heit, 2000), it is crucial to allow people to draw on whatever knowledge structures they deem relevant. Coley, Hayes, Lawson and Moloney (2004) suggest that people's inductive inferences are often driven by relatively abstract conceptual structures about interrelationships between organisms. Although these knowledge structures may not be factually accurate or have much explanatory depth (Keil, 2003; Rozenblit & Keil, 2002), they can nonetheless generate expectations about features of specific category members that go beyond what is known about specific individual instances, forming a foundation for category-based inductive generalizations. Thus, classifying *zebra* as *mammals* is likely to generate expectations about properties this category instance may have. Asked to generalize a property from *zebra* to another instance, people might reasonably assume that *horses* are most likely to share a feature by virtue of their close taxonomic link. Incidentally, these two instances are also fairly closely associated. Thus, regardless of whether or not people explicitly think of the structured relationship between the two instances, they may fall back on this unstructured associative knowledge when they are under time pressure or cognitive burden, offering a sound basis for making a confident generalization. However, if people have the time and mental resources to do so, they may carry out a more fine-grained analysis of the interrelationships between members of a specific category, drawing on other relevant relations between categories (Medin, et al., 2003). For example, because *zebras* are *African mammals,* they may prefer to make inferences to *hyenas* or *lions*. In this case, not only are the categories related by virtue of a taxonomic link, they also share an ecological (same habitat) and causal connection (part of the same food chain). Thus, in this case, structured knowledge is the foundation for a sound inductive inference.

Conclusions

Using a naturalistic paradigm, we obtained further evidence that people's inductive inferences can be influenced by two dissociable types of knowledge. Unstructured knowledge seems to encode statistical regularities within the environment, such as temporal or spatial co-occurrence, and can be psychologically captured by strength of association. People appear to rely more heavily on this type of knowledge when they do not have the available mental resources, suggesting that unstructured knowledge influences reasoning in a heuristic, effortless manner. In contrast, structured knowledge, which captures the underlying abstract interrelationships between categories within a particular domain, appears to be of more importance to the inductive reasoning process when cognitive processing capacity is not compromised. Thus, the involvement of this type of knowledge seems to require a more analytical, effortful mode of reasoning. However, a weakness is that the current experiments cannot precisely identify which processes involved in the recruitment and use of structured knowledge are effortful. This weakness will be addressed in the next chapter.

Chapter V

When Two Types of Knowledge conflict: Association versus Structured Knowledge

The previous experiments suggest that two separable types of knowledge drive category-based inductive reasoning. On the one hand, the degree to which people believe the conclusion of an inductive argument can be influenced by a fairly automatic and effortless process that is based largely on associative strength without any necessary reference to the precise nature of the relationship between the categories. This might be especially true when time and cognitive resources are sparse. Such unstructured associative knowledge is likely to reflect the frequency of repeated co-occurrence between two categories, perhaps resulting in a proportionately weighted connective link in long-term memory. As reviewed in the introduction in Chapter 1, theories which emphasize such unstructured knowledge include Rogers and McClelland’s (2004) connectionist model of category-based semantic induction, Sloman’s (1993b) associative feature-overlap model and Sloutsky and Fisher’s (2004) SINC (Similarity, Induction and Categorization) model.

On the other hand, people are also able to draw on more structured sources of knowledge, taking into account the relationship between the categories. Such structured knowledge can be based on the relation that is deemed most relevant in a specific context, such as taxonomic knowledge when people are reasoning about anatomical or genetic properties (Tenenbaum et al., 2006; 2007), similarity relations between sub- and superordinate categories (Sloman, 1998) or causal relations between categories when reasoning about disease transmission (Medin et al., 2003; Shafto et al., 2008). Drawing on such structured knowledge during the inference process seems to be a slower, cognitively more demanding task. To derive maximum inductive accuracy the person must not only appraise the nature of the relationship between the categories, but must also evaluate whether this relationship provides a mechanism by which a specific property might be common to both categories.

This apparent dissociation between the two types of knowledge leads to some interesting questions regarding the interaction between effortless processes based on unstructured knowledge and the more effortful use of structured knowledge when reasoning selectively about different properties. Of particular interest here is what happens when use of different types of knowledge leads to opposing inferences. For example, sometimes two categories might be very highly associated, but depending on the property, the actual nature of the relation between them might not warrant a strong inference. To illustrate, owing to the strong association between *carrots* and *rabbits*, people might infer that these two categories are likely to share cells. However, if they apply structured knowledge, they may realize that the causal relation between *carrots* and *rabbits* does not warrant a strong inference about shared cells. Similarly, people may initially be unwilling to make a strong inference from *carrots* to *bamboo*, as these two categories are not strongly associated. Yet if people base their inference on the structural taxonomic relation, they should be more confident that the two plants, *carrots* and *bamboo*, might share a property such as cells. In the former case, people would have to inhibit the automatic activation of a strong association in order to make an accurate inference. In the latter example, an accurate inference might depend more on the effortful retrieval of the relational link and hence mechanisms by which the two are related and might share cells.

Putting different types of knowledge into conflict with one another bears some resemblance to research strategies employed by dual process theorists in order to disentangle the nature of the processes underlying human reasoning (De Neys, 2009; Sloman, 1996). It can also help illuminate how contrasting processes might interact with one another (Sloman, 1996; De Neys, in press). To explore the processes underlying the use of structured and unstructured knowledge, we used a triad task in which people learnt that a base category had a novel property and had to choose which of two target categories was more likely to share the novel property with the base. People could either make an inference based on unstructured knowledge, represented by a strong association between the base and one of the targets (e.g. *carrot* has cells X → *rabbit*?), or based on structured knowledge about an appropriate taxonomic link between the base and the alternative target (e.g. *carrot* has cells X → *bamboo*?). By manipulating the nature of the property, we ensured that the inference based on structured knowledge ought to be more accurate than an inference based on unstructured knowledge. Thus, when reasoning about cells, only an inference based on structured taxonomic knowledge (between *carrots* and *bamboo*) is appropriate. In the above example, it might be the case that the use of unstructured knowledge is mediated by an effortless, heuristic process. In contrast, inhibiting such unstructured knowledge, or retrieving structured knowledge, might be a more time-consuming and effortful process. When structured and unstructured knowledge compete, people may need more available mental resources in order to select appropriate structured knowledge. Selecting appropriate structured knowledge ought not to require additional resources or inhibitory control if such competition is absent.

To substantiate our suggestion that there may be a link between degree of inhibitory control and accurate selective reasoning, we also included a correlational aspect in the current study. If accurate selective induction is related to inhibitory control, we might expect people lower in inhibitory control to be less accurate in their selective inductions than people with a high level of inhibitory control. Thus, people high in inhibitory control should display more accurate inductive selectivity and make more inferences based on structured knowledge. In contrast, people lower in inhibitory control might be more inaccurate and swayed by strong but irrelevant associations between categories as a basis for making their inferences. If they cannot inhibit a prepotent response based on unstructured knowledge, such as associative strength, they may fail to thoroughly assess how context and the actual structural relation between the two categories impacts upon inductive strength.

There are several caveats to bear in mind when considering correlational studies which have implicated inhibitory control in the ability to resist belief-based answers (Handley, Handley, Capon, Beveridge, Dennis, & Evans, 2004; Markovits & Doyon, 2005), or in our case, in the hypothesized need to inhibit different types of knowledge. One problem with the construct of inhibitory control is that there seems little consensus about whether it is a unidimensional construct, or whether there are different kinds of inhibitory control (Nigg, 2000). For example, Logan (1994) defines inhibitory control as an executive function that is necessary in order to dynamically pursue goals in an ever-changing environment, in which new information may change one’s course of action. Thus, when new information becomes available it may at times be necessary to inhibit a prepared or overlearnt response. Logan’s (1994) definition suggests that people’s level of inhibitory control can be indexed by their ability to stop a pre-programmed response. One typical task used for this is the Stop Signal task in which people have to respond as fast as possible by pressing either X or O, depending upon which symbol appears on the computer screen. After a series of such response trials, people are instructed to withhold their response when they hear a tone. As this only occurs on a quarter of all trials (25%), the dominant pre-programmed X and O response has to be inhibited.

Others such as Hasher, Zacks and May (1999) view inhibitory control as consisting of a) the ability to prevent irrelevant knowledge from entering working memory, b) the ability to remove outdated information from working memory, and c) the ability to refrain from responding to an incorrect prepotent response. These can be summarized as different aspects of dealing with interference from irrelevant stimuli. This parallels several authors' notion of intentional inhibition (Collette, Germain, Hogge, & Van der Linden, 2009; Harnishfeger, 1995). Whereas unintentional inhibition occurs automatically and relatively early on during processing and is thus unlikely to reach conscious awareness, intentional inhibition is a consciously controlled process used to manage interference from irrelevant information.

If inhibition is primarily defined as the ability to block out interference from irrelevant stimuli, then in order to measure an individual's level of inhibitory control, it is necessary to create a conflict between a to-be-inhibited prepotent response in favour of an alternative answer. One of the most widely used tasks is the Stroop task (Macleod, 1991). Participants see colour words in which the physical colour of the word is incongruent with the linguistic colour. Participants are asked to name the physical colour of the word whilst ignoring the semantically implied colour. For example, they may see the word green printed in blue ink, to which they have to respond with 'blue'. In contrast to trials in which the word is linguistically neutral (e.g. xxxx), people are slower when there is conflict between the physical and linguistic colour. As the linguistically implied colour is processed automatically, people have to inhibit this prepotent response in order to name the physical colour of the stimulus, which is a slower, more effortful process.

Another frequently used task is the Hayling test (Burgess & Shallice, 1996). People have to complete sentences in which the last word has been omitted. The sentences are designed so that they strongly prime the automatic retrieval of a particular word (e.g. *The captain wanted to stay with the sinking* _______. → *Ship*). On the test trials, people are permitted to complete the sentence with the primed word. However, on the inhibitory trials, people are instructed to withhold the primed answer and complete the sentence with an unrelated word (e.g. *The captain wanted to stay with the sinking* _______. → *Thunder*). The difference in reaction time between the test and inhibitory trials is taken as an index of inhibitory efficiency, with smaller differences indicating a more proficient inhibitory process. People with frontal lobe lesions perform poorly on the inhibitory trials (Burgess & Shallice, 1996) and there is increased left prefrontal activation in healthy subjects during the inhibitory trials (Collette, et al., 2001), lending credence to the use of this task as an instrument for measuring the construct of inhibitory control. As this task is primarily about semantic inhibitory control, we used a variant of this measure to explore whether there was a relationship between use of structured knowledge and level of inhibitory control.

5.1 Overview Experiment 6

To summarize, we used a triad task in which we created a conflict between responses based on unstructured and structured knowledge. People had to make an inductive choice between two target categories, one which was strongly associated, and a competing target which was weakly associated but structurally related to the base category. Compared to our control triads, where people had to choose between a weakly associated but structurally related target category and an unrelated category, we expected people to make more reasoning errors when there was competition between inferences based on a structured relation or a strong association between the base and target categories. The primary goal was to demonstrate that the ability to inhibit inappropriate unstructured knowledge in favour of more appropriate structured knowledge is an effortful process that draws on working memory resources. Thus, we compared the effects of a resource-demanding secondary task on people's ability to inhibit irrelevant unstructured knowledge and choose the more appropriate target category based on structured knowledge.

Furthermore, we expected that people higher in inhibitory control might be more likely to reason based on structured rather than unstructured knowledge. That is, we expected them to be more successful at inhibiting the strongly associated choice in favour of the weakly associated but structurally related choice, resulting in a significant positive correlation between level of inhibitory control and successful selective reasoning.

5.1.1 Method

Participants

34 participants from Durham either earned course credit or were paid £6 for their participation in the experiment. The mean age was 20.9 years (*SD* = 2.5 years).

Triad Task with Memory Load

For the triad task, people were given a base category which they were told had novel cells, and two target categories. We chose 14 conflict triads in which the base had a causal relationship with the strongly associated target and a taxonomic relation with the other weakly associated target. Whereas the causal relationship provides a route for the transmission of infection or disease, when reasoning about cells, structured taxonomic knowledge is likely to be more relevant and previous studies (Shafto et al., 2007) have shown that people are more likely to draw on knowledge about taxonomic relations when reasoning about cells than when reasoning about infections or diseases. For 8 conflict triads the causal relation between the base and target was predictive (i.e. from prey to predator), whereas for the remaining 6 conflict triads, this relation was diagnostic (i.e. from predator to prey). The causal target was always from a different superordinate category than the base, whereas the taxonomically related target was always from the same superordinate category. Based on our pre-test described in Chapter 2, in which we had asked people to rate the strength of association between two categories on a scale from 1 (unrelated) to 9 (very highly associated), we selected *conflict triads* in which the strength of association between the target and causal choice was always significantly higher (all paired-samples t-tests had p's $< .05$) than the association between the base and the taxonomic target. Thus, we expected that if people did choose the causal target, this would be attributable to the strong association, and hence unstructured knowledge, rather than due to the underlying causal relation. This is because the causal relation ought to be irrelevant when reasoning about cells. We also added some new items which were pre-tested by asking 15 individuals to rate the strength of association for 41 new category pairs, using the procedure described in Chapter 2. A list of all triads and the strength of association between each base and its different target categories can be found in Appendix 4A.

Participants were told that the base category has novel cells and were asked to make an inductive choice between two conflicting targets. For example, a cell conflict triad would be *Butterflies have dF4-cells → Flowers or Locusts?* For these *conflict triads* the accurate inductive choice should be based on structured knowledge about the taxonomic relation between the base and the target, in which case people should choose *Locusts* in the above example. In contrast, the erroneous choice would be based on unstructured knowledge represented by a strong association between the base and the alternative target, in which case people might choose *Flowers*.

Control triads consisted of the base, the weakly associated taxonomic target and a completely unrelated alternative target (different superordinate category, no causal or ecological relation). For the above example, the corresponding cell control triad would be *Butterflies have 5tR-cells → Seaweed or Locusts?*

In order to ensure that people were paying attention to the nature of the property, and to stop them developing the default strategy of always selecting the taxonomic choice, we included 28 filler triads in which people were asked to reason about diseases (e.g. *has disease wQ2*). Depending upon the mechanism people were basing their reasoning on, either the weakly associated taxonomic or strongly associated causal target could be an appropriate inductive choice. Consequently, unlike when reasoning about cells, there would not necessarily be a strong need to inhibit the strongly associated choice. As we were primarily interested in pitching structured knowledge against unstructured associative knowledge, we did not include these in our analysis. Thus, each participant reasoned about 14 conflict cell triads (8 in which the strongly associated causal choice was predictively related to the base and 6 in which the strongly associated causal target was diagnostically related to the base), 14 control cell triads and 28 disease filler triads (i.e. 56 triads in total).

The 14 triads consisted of real picture photographs of the category. Below the base pictures, participants read the statement:

[Category name] have C4x-cells/ disease sA3

Which category is more likely to also have these cells/this disease?

This was followed by two arrows, one pointing to the left target picture and one pointing to the right target picture. The category label was printed below the target images. For an example of a conflict triad in which people were requested to reason about cells, see Figure 5.1. Figure 5.2 illustrates the corresponding cell control triad.

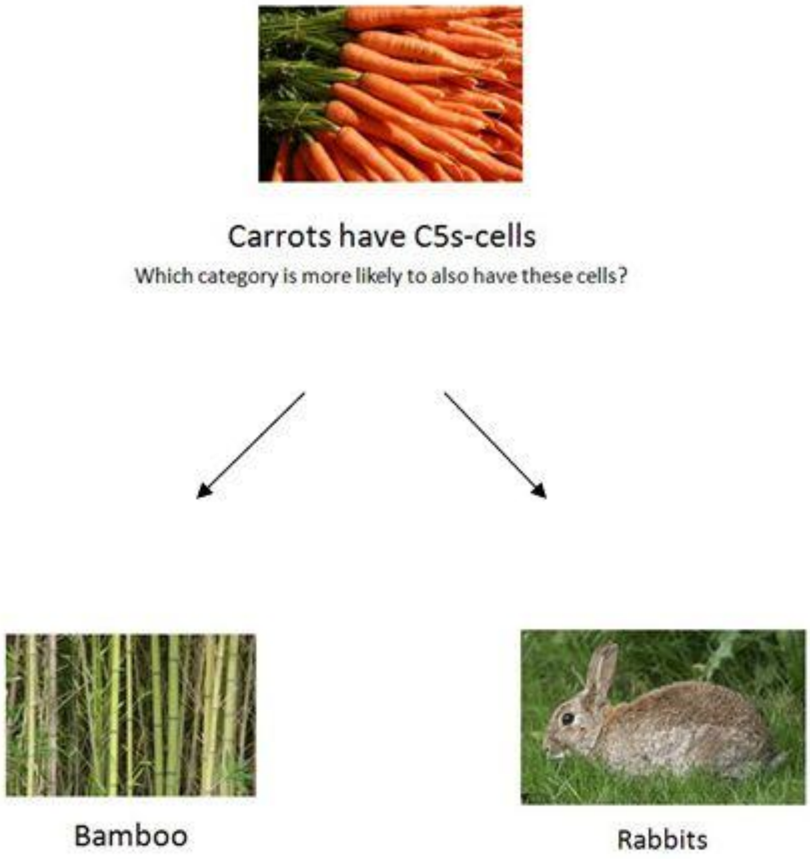

Figure 5.1: Example Cell Conflict Triad

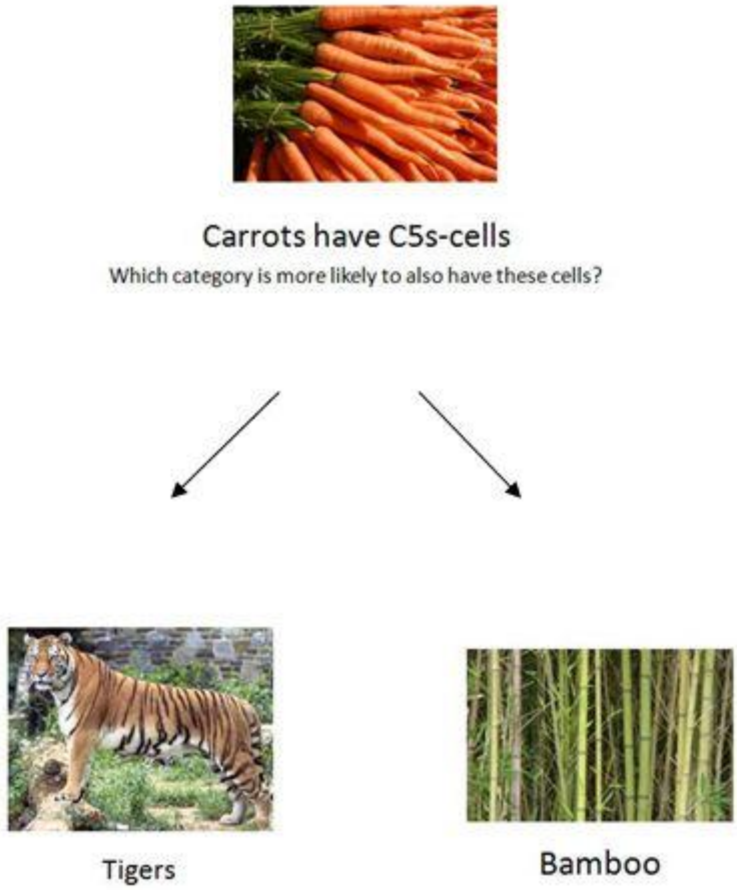

Figure 5.2: Example Cell Control Triad

The triads were presented on a laptop. Before each of the triad trials, people saw a 4 by 4 dot matrix, filled with four randomly placed black dots. These were shown for 2 seconds and participants had to try and memorize the location of the dots whilst carrying out the reasoning task. In the heavy load condition, the dots were randomly distributed, with the constraint that they could never form a straight or diagonal line. In contrast, in the light load condition, the dots always had an easily memorizable straight or diagonal line configuration.

The base image was then displayed for 2 seconds, followed by the two target pictures. Participants made their choice by pressing 1 if they thought the category in the left picture was more likely to share the property with the base, and 9 if they thought the category on the right was more likely to also have the property.

Upon choosing a conclusion category, participants saw an empty dot matrix and were asked to recall the location of the dots by clicking on an empty matrix box with the mouse cursor. The dots appeared in the clicked box, regardless of whether or not this was the correct location. There was a 500 ms delay before the appearance of the next filled dot matrix. The order of presentation of the triads was fully randomized.

Design

The experiment had a 2 by 2 mixed-design. The independent within-subjects variable was the type of triad (control or conflict) and the between-subjects independent variable was memory load (heavy or light). The dependent variable was the number of appropriate taxonomic choices.

We expected a main effect of type of triad, with people making more inferences based on structured knowledge in the control compared to the conflict condition where there was the lure from the strongly associated but inappropriate causal target. We also expected an interaction between load and triad type. People in the in both load conditions should make more inferences based on structured knowledge in the control compared to the conflict condition. However, people in the heavy load condition might not be able to effortfully recruit structured knowledge, amplifying this difference between the control and conflict triads compared to the light load condition.

In a further analysis, we compared the proportion of taxonomic choices for the conflict triads in which the causal link was predictive to the proportion of taxonomic choices for the conflict triads in which the causal link was diagnostic. We did not expect an effect of causal direction (i.e. predictive verses diagnostic). This is because we assume that people may be lured by the contextually inappropriate causal alternative because it is strongly associated with the base, in which case the underlying structural relation should not have an effect on the appeal of the strongly associated target (Fenker, et al., 2005).

Finally, we assessed inhibitory control by means of a task that assesses inhibition of automatically activated semantic knowledge. If people have to inhibit a response based on strong semantic associations between categories in favour of a response that is based on a structured relation between the categories that is more effortful to retrieve, we might expect correlations between people's use of more appropriate structured knowledge and their ability to withhold a prepotent response.

Procedure

Participants were randomly allocated to either the heavy or light load condition. They received detailed instructions emphasizing that there were no right or wrong answers. If they understood the instructions, they completed two practice trials to familiarize them with the alternating sequence of the dot pattern and induction task.

Post-test

In the post-test, participants were asked two questions about the 42 different category pairs (14 base and causal target pairs, 14 base and taxonomic target pairs and 14 base and unrelated target pairs). The questions were presented in a random order on a laptop PC. The first question assessed their belief about biological group membership. If people thought that the two categories belonged to the same biological group, they pressed C (corresponding to YES), whereas if they didn't believe they were taxonomically related they pressed M (corresponding to NO). If they didn't know, they pressed B (corresponding to DON'T KNOW), although participants were asked to use this option sparingly.

The second question asked participants to indicate whether there was a causal relationship between the two categories and in particular, whether the two categories were part of the same food chain. As with the first question, participants pressed C (YES), B (DON'T KNOW) or M (NO). People were given as much time as they wanted to make their responses. The order of presentation was fully randomized.

Semantic Inhibitory Control Task

To assess people's level of inhibitory control which includes a semantic component we used a variant of the Hayling test (Burgess & Shallice, 1996) which was adapted by Markovits and Doyon (2004). This task requires the inhibition of a strongly associated response without explicit instructions to do so. Inhibition in this task occurs at the semantic level and is designed so that people cannot adopt a low-level inhibitory strategy.

People were given incomplete sentences for 2000 milliseconds, such as "His favourite sport is _______". This was followed by a letter string which was a contextually appropriate real word (e.g. "cricket"), a contextually inappropriate real word (e.g. "daylight"), a contextually appropriate non-word that strongly resembled a real word (e.g. "footbalf") and a contextually inappropriate non-word (e.g. "sinema"). People were instructed to decide whether the letter string was a real word which would appropriately complete the previous sentence. If they thought it was a contextually appropriate real word, they were instructed to respond with YES by pressing C on the keyboard. If they thought the letter string was a non-word, or a contextually inappropriate real word, they responded with NO by pressing M.

We adapted a total of 30 cloze sentences from Bloom and Fischler's (1980) database, which strongly prime the missing word. Each of these 30 sentences was combined with the four types of letter strings, resulting in a total of 120 trials. All sentences and filler words can be found in Appendix 4B. The order in which the sentences and letter strings were presented was fully randomized, with the constraint that identical sentences paired with a different letter string could never appear adjacent to one another.

Participants completed the experiments in one individual session of around 40 minutes. The tasks were completed in a fixed order, starting with the triad task, followed by the post-test belief assessment and finishing with the semantic inhibitory control task.

5.1.2 Results

For each individual, we calculated the number of times they made an accurate taxonomic choice for the cell conflict triads, and likewise for the cell control triads. These were analysed with a 2 (triad type: conflict versus control) by 2 (load: heavy versus light) mixed-design ANOVA, with load as the between-participants manipulation. For the item analysis, the proportion of people who chose the taxonomic category was calculated and analysed with a 2 (triad type) by 2 (load) repeated-measures ANOVA.

The ANOVA showed that there was a marginally significant effect of load across subjects, $F_{S\ (1,\ 32)} = 3.97$, $p = .055$, effect size d = .35. This effect was highly significant across items, $F_{I\ (1,\ 13)} = 40.2$, $p < .0005$, effect size d = 1.76. People under heavy load made fewer appropriate taxonomic choices ($M = 9.2$, $SE = .61$) compared to people under a light cognitive load ($M = 10.9$, $SE = .61$).

Furthermore, there was a main effect of triad type, $F_{S\ (1,\ 32)} = 66.71$, $p < .0005$, effect size d = 1.44, $F_{I\ (1,\ 13)} = 35.07$, $p < .0005$, effect size d = 1.64. As predicted, people made more accurate taxonomic choices for cell control triads ($M = 12.3$, $SE = .28$) than for cell conflict triads ($M = 7.8$, $SE = .66$).

However, these two main effects were modulated by a significant interaction between triad type and load, $F_{S\ (1,\ 32)} = 4.96$, $p = .033$, effect size d = .39, $FI_{(1,\ 13)} = 27.3$, $p < .0005$, effect size d = 1.45. Illustrated in Figure 5.3, and confirmed by Bonferroni post-hoc comparisons, load only had an effect on people's number of accurate taxonomic choices for cell conflict triads (p < .0005, effect size d = .76), but not on the number of accurate taxonomic choices for the cell control triads (p = .41, effect size d = .29). Thus, whilst people more often failed to make an accurate taxonomic choice for cell conflict triads than for cell control triads across both load conditions, this effect was amplified when people were under cognitive load. This suggests that people under load struggled to inhibit the lure from the strongly associated but inappropriate target category.

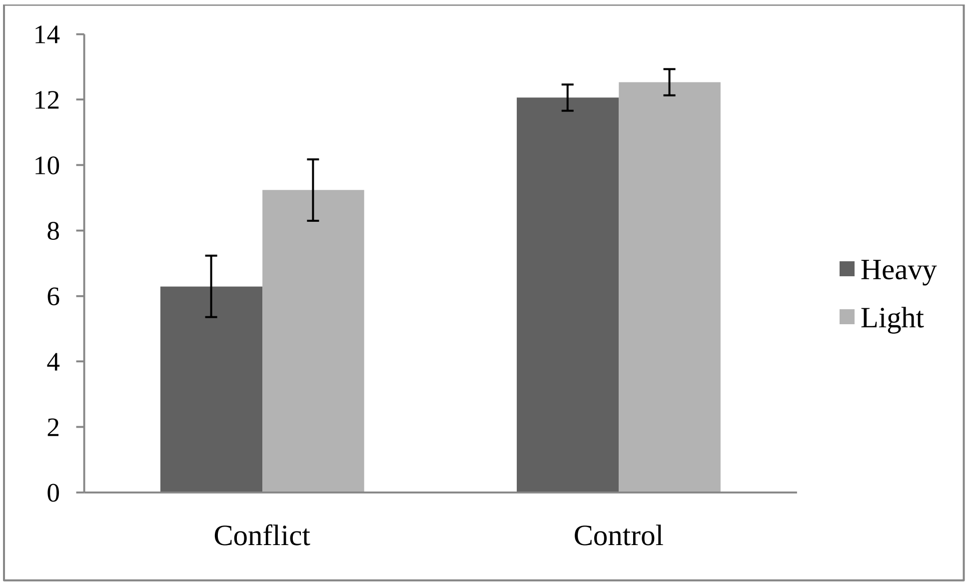

Figure 5.3: Number of Relevant Taxonomic Choices for the Two Types of Cell Triads across the Two Load Conditions

Category lures in the conflict trial shared two relations with the base categories. The categories had a strong associative relation, but they were also part of the same food chain. Our hypothesis was that people would have to inhibit a response based on the associate relation in order to choose the taxonomically related target. However, it is possible that participants who chose the lures did so on the basis of the food chain relation. To rule out this possibility, we also looked at whether there was a causal asymmetry effect. As demonstrated in Experiments 1 and 2 reported in Chapter 3, inferences based on structured causal knowledge should lead to a causal asymmetry effect, with more causal choices in the predictive (e.g. *salmon* → *grizzly bears* or goldfish?) compared to the diagnostic case (e.g. *monkeys* → *peanuts* or seals?). If we are dealing predominantly with competition from unstructured associative knowledge then we would not expect to find effects that depend upon people using structured causal knowledge, such as the aforementioned causal asymmetry effect. As we had a different number of conflict triads in which the strongly associated but irrelevant causal target was predictively (8 conflict triads) or diagnostically (6 conflict triads) related to the base category, we compared the effect of causal direction on the *proportion* of taxonomic choices for the conflict triads. The proportion of taxonomic choices for the conflict triads in which the causal choice was predictively related to the base was .55 (*SD* = .34), which was not significantly different than the proportion of taxonomic choices when the causal link was diagnostic (*M* = .52, *SD* = .27), $t_{S\ (33)}$ = .99, *p* = .33, effect size d = .17. An independent samples t-test with direction as the between-item variable verified the above result across items, $t_{I\ (12)}$ = .51, *p* = .62, effect size d = .26. The absence of a causal asymmetry effect supports our contention that it is the unstructured associative aspect of the strongly associated but inappropriate causal target which is swaying people from making a more appropriate inference based on structured knowledge.

Post Test Check

We checked that people knew about the purported relationships between the categories. In the post-test, people stated whether they believed that the two categories were part of the same food chain (*causal relatedness*) and whether the category pairs belonged to the same biological group (*taxonomic relatedness*). For the causally related categories, the mean causal relatedness endorsement proportion was .85 (*SD* = .13) and the mean endorsement proportion for being taxonomically related was .11 (*SD* =.10). For the taxonomically related categories, the taxonomic relatedness endorsement proportion was .71 (*SD* = .19) and .11 (*SD* =.10) for being causally related. The causal relatedness endorsement rate for unrelated category pairs was .19 (*SD* =.11), and the taxonomic relatedness endorsement proportion was .07 (*SD* =.08).

A one-way ANOVA on causal relatedness endorsement proportions confirmed that there was a significant main effect of type of relation between the categories, $F_{(2, 66)} = 406.27$, $p < .0005$. Bonferroni post-hoc tests verified that the mean causal relatedness endorsement was significantly higher for the causally related categories than for both the taxonomically and unrelated categories (p's< .0005). The causal relatedness endorsement for the unrelated categories was significantly higher than for the taxonomically related categories ($p = .003$).

Likewise, a one-way ANOVA on taxonomic relatedness endorsement proportions showed that there was a main effect of relation between the categories, $F_{(1.08, 39.12}$[7]$_{)} = 266.96$, $p < .0005$. Bonferroni post-tests showed that as expected, the taxonomic endorsement proportion for the taxonomically related categories was significantly higher than for both the causally ($p < .0005$) and unrelated categories ($p < .0005$). The taxonomic relatedness endorsement proportion was marginally higher for the causally related categories than for the unrelated categories ($p = .053$).

One criticism may be that any effects we find in which people are swayed by a strong but inappropriate association rather than choosing the weakly associated category with the appropriate structural relation might be because people simply did not have the knowledge that the two weakly associated categories were in fact related. As the causally related categories were always strongly associated and taxonomically related pairs were always weakly associated, if it was indeed lack of knowledge driving an effect, we would expect to see significant differences in post-test endorsement rates for causal and biological beliefs. We compared the proportion of positive responses to the question whether the (strongly

[7] df adjusted for Non-sphericity using Greenhouse-Geisser corrections

associated) causally related pairs were part of the same food chain with the proportion of positive responses to the question whether the (weakly associated) taxonomically related category pairs belonged to the same biological group. A paired-samples t-test was significant, $t_{(33)} = 3.69$, $p = .001$, effect size d = .63. However, this seemed to be driven by a minority of participants whose taxonomic knowledge was significantly below their knowledge about the causal food chain links. We calculated the difference between the taxonomic and causal endorsement ratings and excluded participants whose difference scores were further than 1.5 standard deviations away from the mean.

Another way of checking that people do have the relevant knowledge but are still swayed by the strong but irrelevant association for the conflict triads is to compare the proportion of biological relatedness endorsements for the taxonomically related targets where people chose the taxonomic target with the proportion of biological endorsements where they chose the causal target. This is a very stringent test, because in cases where people did not make any causal choices, the proportion of taxonomic endorsement when choosing the causal target would be zero, even though this does not reflect their actual lack of knowledge, but merely a lack of causal choices. When all participants were included (3 people made no causal choices), the mean biological endorsement proportion for conflict triads where people made a taxonomic choice was .77 (*SD* = 17), whereas it was .58 (*SD* = .36) for conflict triads in which people chose the causal target. A paired samples t-test showed that this difference was significant, $t_{(26)} = 2.92$, $p = .007$, effect size d = .56. As with the previous method of assessing people's level of structured knowledge, this seemed to be driven by a small group of individuals who lacked relevant taxonomic knowledge, therefore choosing the strongly associated causal alternative. Hence, we calculated the difference between the proportion of taxonomic endorsements when people chose the weakly associated but appropriate taxonomic target and the proportion of taxonomic endorsements when they chose the strongly associated but irrelevant causal target and excluded participants whose difference score was larger than 1.5 standard deviations than the mean difference.

We then re-ran the comparison of people's level of structured knowledge excluding those individuals whose knowledge level difference scores were 1.5 standard deviations larger than the mean difference on either of aforementioned knowledge difference measures (10 participants). Two paired-samples t-tests showed that for the remaining participants, there was no difference between their taxonomic endorsement proportions for taxonomically related categories (M = .77, SD = .33) and their causal endorsement for causally related category pairs (M = .84, SD = .15) ($t_{(23)}$ = 1.81, p = .084, effect size d = .37), nor for their taxonomic endorsement proportions for the cell conflict triads for which they made a taxonomic (M = .85, SD = .17) compared to a causal choice (M = .76, SD = .20) ($t_{(23)}$ = -1.78, p = .088, effect size d = .36).

To confirm that these people were lured by the appeal of the strong association between the base and the inappropriate target despite having knowledge about the appropriate structural relation between the base and the alternative target, we re-ran the 2-way ANOVA on the number of accurate taxonomic choices on the subgroup of 24 individuals who had relevant taxonomic knowledge. This corroborated the findings from the full analysis, with a significant main effect of memory load $F_{(1, 22)}$ = 4.29, p = .05, effect size d = .44, a significant main effect of triad type, $F_{(1, 22)}$ = 43.33, p < .0005, effect size d = 1.4, and a significant interaction between triad type and memory load, $F_{(1, 22)}$ = 6.12, p = .002, effect size d = .53. As with the full sample, the interaction arose because the load manipulation had an effect on the conflict cell triads but not on cell control triads. Thus, compared to those under light load, people under heavy load were less able to choose the weakly related but appropriate taxonomically related target when there was a highly associated but irrelevant alternative target (p = .026). In contrast, when there was no competition between structured taxonomic and unstructured associative knowledge, people in both load conditions made a similar number of relevant taxonomic choices (p = .56).

Relation between Selective Inductive Reasoning and Semantic Inhibition

Next we explored whether people's tendency to inhibit a response based on a strong but contextually inappropriate association between two categories and instead to choose the weakly associated but structurally related category was correlated with their general ability to withstand interference by semantic knowledge on our semantic inhibition task. In this task, people had been presented with an incomplete sentence (e.g. *The sailor wanted to stay with the sinking* _____) which was followed by one of four letter strings, a contextually appropriate non-word (*shifp*), a contextually appropriate real word (*boat*), a contextually inappropriate non-word (*basana*) or a contextually inappropriate real word (*ketchup*) and were asked to judge whether the letter string would suitably complete the sentence. In line with Markovits and Doyon (2004), we argue that correctly rejecting contextually appropriate non-words as suitable sentence fillers necessitates the inhibition of semantic content activated by both the sentence and the similarity between the non-word and a real cloze sentence filler word.

To guard against the possibility that people simply adopt a low-level processing strategy in which they ignore the sentence and base their judgements purely on whether the letter string is a real or non-word, we also looked at their performance on the real words. Whilst this low-level strategy would result in correct judgements for the non-word letter strings regardless of the sentence context, people would be at chance performance for the real words, where half of the words are in fact suitable fillers. The probability of making 50 out of 60 correct judgements for real words if people were merely guessing is below .000001. As none of our participants made more than 10 mistakes when judging real words, we can safely assume that they were processing the sentence context in order to make their decision. Table 5.1 summarizes the results from the inhibitory control task.

Table 5.1: Means and Standard Deviations for the Semantic Inhibition Task Experiment 6

Measure	Mean	SD	Min	Max
Correct Judgement Contextually Appropriate Non-word	28.5	1.7	23	30
Correct Judgement Contextually Appropriate Word	27.5	1.6	23	30
Correct Judgement Contextually Inappropriate Non-word	29.9	0.4	28	30
Correct Judgement Contextually Inappropriate Word	29.7	0.6	28	30

If using structured knowledge requires the inhibition of inappropriate unstructured knowledge, we would expect to find correlations between individuals' ability to choose the weakly associated but appropriate taxonomic choice and the ability to correctly reject contextually appropriate non-words. When all participants were included there was only a small and non-significant correlation, $r_{(34)} = .29$, $p = .1$. However, the correlation between inhibitory control and use of structured knowledge may only hold for individuals with appropriate taxonomic knowledge. When the 10 individuals who lacked this structured knowledge were excluded, the correlation rose to $r_{(24)} = .37$, $p = .08$, which was still non-significant. As semantic inhibition requires the availability of cognitive resources, we were interested to see whether there might be differences between the two load conditions. We therefore looked at the correlations separately for the two memory load conditions.

Light Load Condition

For the 12 knowledgeable people who only had to remember very simple straight line dot patterns, inhibitory control was significantly negatively related to the difference in number of taxonomic choices between cell conflict and cell control triads, $r_{(12)} = -.72$, $p = .008$. This means that people high in semantic inhibitory control exhibited more consistent reasoning patterns, making a similar number of relevant taxonomic choices in the conflict and control triad trials. Inhibitory control was also positively related to the number of taxonomic choices in the conflict triads, $r_{(12)} = .55$, $p = .06$. However, due to the small sample size this correlation did not quite reach statistical significance. Nonetheless, it does provide some preliminary evidence for the idea that people higher in inhibitory control were better at inhibiting the lure of a strongly associated irrelevant target.

Heavy Load Condition

In contrast, for the 12 knowledgeable participants who had to remember complex configurations of dots whilst carrying out the inductive reasoning task, scores on the semantic inhibitory control task were neither correlated with the difference in number of taxonomic choices between cell conflict and cell control triads, $r_{(12)} = -.14$, $p = .67$, nor with the overall number of taxonomic choices they made for conflict triads, $r_{(12)} = .2$, $p = .54$. This might suggest that semantic inhibitory control is effortful and requires available working memory resources. Thus, load might only be detrimental in people who are high in inhibitory control, but have less effect on people who are lower in inhibitory control and are already performing poorly. This hypothesized divergent effect of load upon high and low performers might be similar to studies (Kane & Engle, 2000) which show that secondary tasks only leads to a decline in performance for individuals with high working memory spans but not for people with low working memory spans. However, the contention that secondary task differentially affects inductive reasoning in people who are high or low in inhibitory control would have to be followed up with a factorial design (*load*: heavy versus light by *semantic inhibitory control*: high versus low) in a larger sample.

Secondary Task Analysis

In a dual task paradigm, there is a risk that the dissociable effects of memory load on the number of taxonomic choices for conflict and control cell triads could reflect a strategic trade-off between primary and secondary tasks (Hegarty, et al., 2000). That is, participants may simply allocate more mental resources to the conflict cell triads than to the control cell triads, neglecting the secondary task.

To check for such a trade-off, we calculated the number of dots correctly recalled for the trials preceding the two crucial types of triads separately for the two load conditions. A 2 (triad type: conflict or control) by 2 (load: heavy versus light) mixed-design ANOVA, with load as the between-subjects variable showed that the only significant difference in the number of dots recalled was between the two load conditions, $F_{(1, 32)} = 27.76$, $p < .0005$, effect size d = .93. In the heavy load condition, participants recalled a mean of 3.2 dots (*SE* = .07), whereas they recalled on average 3.7 dots (*SE* = .07) when they were only under a light memory burden. There was no main effect of triad type $F_{(1, 32)} = .003$, $p = .96$, nor an interaction between load and triad type, $F_{(1, 32)} = .13$, $p = .72$.

This suggests that people were consistent in how they allocated their mental resources to the primary and secondary tasks across all problems and also verified that the more complex patterns were harder to remember and more burdensome than the simple dot patterns.

5.1.3 Discussion

In the current experiment we looked at people's inductive choices in a triad task when there was a conflict between the two types of knowledge, i.e. unstructured versus structured knowledge. Whilst reasoning about cells, we pitched a weakly associated but appropriate taxonomic choice against a strongly associated but inappropriate causal choice. Compared to the control triads, in which the alternative to the weakly related taxonomic choice was an unrelated category, people were more frequently swayed by the strongly associated but inappropriate target. The results from the current study also suggest that the ability to inhibit unstructured knowledge in favour of contextually more appropriate structured knowledge is a resource-demanding process. Thus, the failure to make the inductively most potent choice by inhibiting the lure from a highly associated target was amplified when people had to contend with a heavy secondary memory burden.

It seems that reasoning based on the contextually more appropriate source of structured knowledge requires people to inhibit the competing strong association activated in semantic memory. For people who had relevant structured knowledge and who were not cognitively compromised, the ability to draw on this in favour of inappropriate unstructured knowledge was correlated with performance on a semantic inhibitory control task. By demonstrating such a correlation, the current experiment links the use of different types of knowledge in category-based inductive reasoning to more domain-general reasoning processes. We focused on two major aspects, the processing characteristics of the two types of knowledge and the nature of the inhibitory process required for successful selective induction.

Regarding these two points, our findings suggest that activation of unstructured knowledge is automatic, fast and effortless, whereas use of structured knowledge might require more deliberate and effortful processing. This resonates with the characteristics ascribed to heuristic and analytical processes in dual-process theories of reasoning (Evans, 2007, 2008; Evans & Over, 1996; Sloman, 1996). These theories suggest that when output from two processes conflict, people have to withhold a response based on effortless, automatic processing in favour of more effortful reasoning. Applying this to our findings, people may only be able to engage in slow, deliberate reasoning based on structured knowledge if they have sufficient cognitive resources and can inhibit an appealing inductive choice based on unstructured associative knowledge.

This implies that the processes which mediate accurate selective induction when there is conflict between the two types of knowledge are domain-general. For example, similar results ascribing an important role to inhibitory control in ensuring a correct response have been obtained in deductive reasoning paradigms. In a typical task, people have to inhibit belief-based responses in order to reason in line with the logical structure of a problem. Studies have shown that the ability to resist belief-based responses in syllogistic reasoning follows a curvilinear trend across development, thus tracing the same developmental trajectory as inhibitory capacities (De Neys & Van Gelder, 2009). Similarly, people who resist belief-based reasoning are higher in semantic inhibitory control (Markovits & Doyon, 2004). Resisting a belief-based response activates brain regions in the lateral prefrontal cortex, which are implicated in general inhibitory functions (De Neys, et al., 2008).

The current experiment is the first demonstration that it might be inhibitory functions that are crucially involved in maximizing the potency of category-based inductive reasoning. However, the number of participants was fairly small and the correlations which support our hypothesis about the role of inhibitory control in category-based induction were based on only 12 knowledgeable participants in the light load condition. Thus, to try and gain more support for our position, and to try and elucidate the nature of the inhibitory control process, we carried out a further study in which we increased number of participants and dropped the cognitive resource manipulation.

5.2 Overview Experiment 7

In the next experiment, we wanted to firstly replicate the finding that in order to use structured knowledge, people may have to inhibit conflicting unstructured knowledge.

Secondly, we wanted to explore the nature of inhibitory control in more detail. The range of theoretical frameworks and resultant inhibitory control tasks make it challenging to evaluate the level at which inhibition may come into play in inductive reasoning and to find possible correlations between measures of inhibitory control and use of different types of knowledge in category-based induction. Also, even within the literature primarily concerned with the construct of inhibitory control there is little consensus as to whether the processes that underlie inhibition of simple motor tasks are identical to those that are needed to inhibit higher-level competing semantic knowledge (Friedman & Miyake, 2004). For example, the Stop Signal task measures inhibition at the motor level, in which people have to inhibit a prepotent motor response. Unlike some of the other inhibitory measures, there is no direct conflict between two responses. The Stroop task on the other hand involves more semantic processing and a direct conflict between two alternative responses, although the nature of the pre-potent and target response are somewhat different. Thus, the pre-potent response involves word processing, whereas the latter involves perceptual colour naming. In contrast, the Hayling task, and our variant thereof, requires the inhibition of a pre-potent semantic response in favour of a less automatic answer that is also semantic in nature. As the knowledge recruited in category-based induction is semantic, we might expect the strongest correlations between use of different types of knowledge and tasks that assess inhibitory control at the semantic rather than at a lower perceptual or the motor-response level of processing. If there is a dissociation in the degree to which the two inhibitory control measures correlate with patterns of inductive reasoning, we can get a better idea of the nature of the prepotent response and the processes needed to inhibit this unstructured knowledge when it is inappropriate. In the following experiment we explore this possibility by contrasting the strength of correlation between reasoning performance and scores on the Hayling task (our measure of *semantic* inhibitory control) with the strength of the correlation between reasoning performance and the stop signal task (our measure of *low-level* inhibitory control).

5.2.1 Method

Participants

28 participants from Durham were paid £5 for their participation. Their mean age was 26.4 years (*SD* = 5.6 years).

Materials

Triad Task

The materials were similar to the ones used in the previous experiment, but we dropped the secondary memory load component. To recapitulate, there were 14 triads in which unstructured knowledge was pitched against structured knowledge. Thus, when people were reasoning about cells, the appropriate choice (taxonomically related target) was weakly associated, whereas the inappropriate choice (causally related target) was strongly associated. For the cell control triads, we replaced the inappropriate but strongly associated target with a completely unrelated alternative target. To try and ensure that people paid attention to the property they were reasoning about, we included 28 filler triads in which they made inductive inferences about diseases (e.g. *has disease wQ2*). When reasoning about diseases, both the taxonomic or causal target could be an appropriate inductive choice, obviating the necessity to inhibit unstructured knowledge. In total, each participant reasoned about 56 triads, 14 conflict cell, 14 control cell triads and 28 disease filler triads.

Stop Signal Task

The Stop Signal task (Aman, Roberts, & Pennington, 1998; Handley, et al., 2004) assesses inhibitory control by means of a computerized reaction time experiment. There were two kinds of trials, primary go trials and stop signal trials. For the primary go trials, participants first saw a fixation point for 500 ms (a small smiley face), followed by either an “X” or “O” stimulus. Participants were instructed to respond as fast as possible by pressing the corresponding letter on the keyboard without sacrificing accuracy. On the stop signal trials, participants were instructed not to press anything if they heard a beep before seeing the “X” or “O”. The tones were presented randomly 150 ms or 250 ms before the participant’s mean reaction time to primary trials. Adjusting the tone delay to each individual’s mean reaction time ensured that the inhibitory requirements were roughly alike across participants and independent from primary response times. As the stop signal trials constituted the minority of trials (25%) the dominant response was to press “X” or “O”, requiring people to inhibit this action when hearing the tone.

There were 4 blocks of trials. In the first block, participants encountered 30 primary trials (equal number of “X”s and “O”s), which were used to measure a participant’s individual mean reaction time. Thus, this reaction time was used in later trials to determine the magnitude of the delay between the tone and the letter stimulus.

The next block consisted of 23 practice trials, of which 8 were stop signal trials. The next two experimental blocks consisted of 48 trials each, with 32 primary trials and 16 stop signal trials (8 trials at the two different delay settings). The presentation of the primary and stop signal trials was completely randomized in all blocks.

Post-test and Semantic Inhibitory Control Task

The materials for the post-test and the semantic inhibitory control task were identical to the previous experiment. To reiterate, for the post-test, people answered two questions about each category pair, firstly, whether they belonged to the same biological group and secondly, whether they were part of the same food chain.

For the semantic inhibitory control task, people verified whether a letter string constituted a suitable sentence filler.

Procedure

Participants completed the four experimental components in one individual session of around 50 minutes. They started with the triad task. This was followed by the post-test belief assessment and the stop signal task. Participants finished with the semantic inhibitory task.

5.2.2 Results

Triad Task

For each individual, we counted the number of times a participant chose the weakly associated taxonomic category separately for the cell conflict triads and cell control triads and analysed these with a paired-samples t-test, with type of triad as the independent variable. For the analysis by items, we calculated the proportion of participants which chose the weakly associated category for the cell conflict and control triads and analysed these with a paired-samples t-test.

As expected, there was an effect of type of triad, $t_{S\ (27)} = -6.78$, $p < .0005$, effect size d = 1.28; $t_{I\ (13)} = -3.67$, $p = .003$, effect size d = .98. Thus, as Figure 5.4 below illustrates, people chose the weakly associated taxonomic target more frequently for the control triads (*M* number of taxonomic choices = 12.86, *SD* = 1.51) than for the conflict triads where there was competition from the strongly associated but irrelevant causal target (*M* = 10.18, *SD* = 2.84).

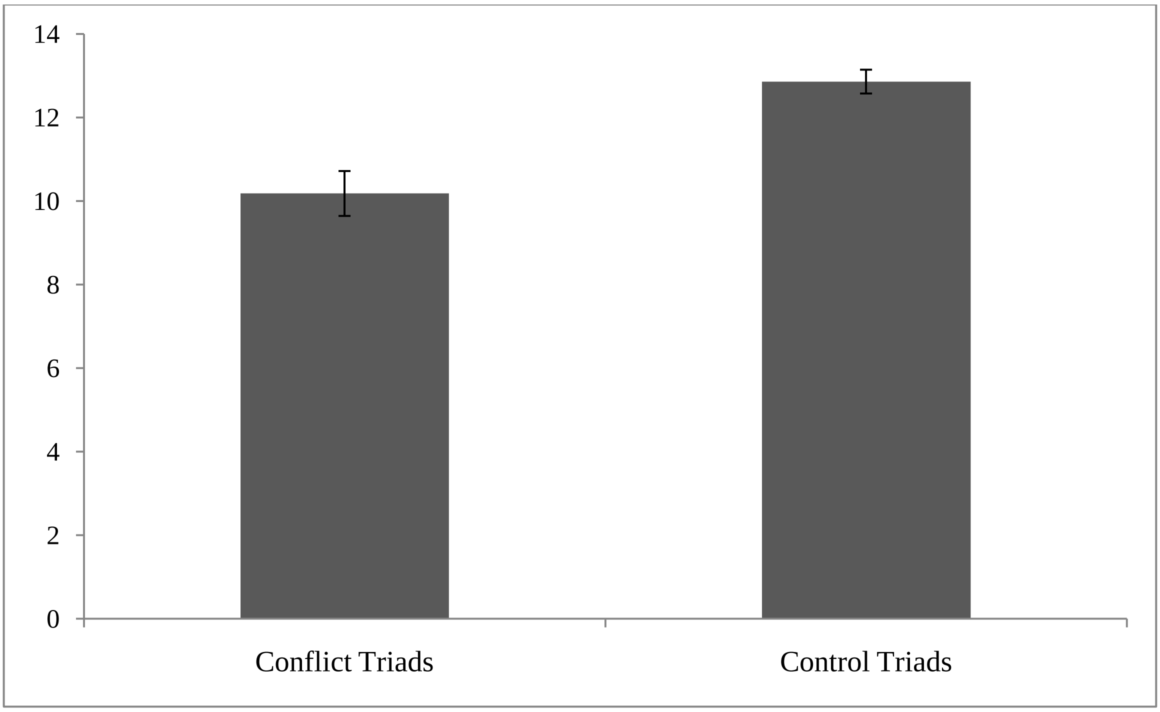

Figure 5.4: Mean Number of Taxonomic Choices across the two Types of Triads when reasoning about Cells

This suggests that people were indeed swayed by the strong but contextually inappropriate association between the base and the causally related target.

As in the previous experiment, we wanted to show that it was unstructured knowledge in form of a strong association drawing people towards an inappropriate inference rather than the structured causal relation between the strongly associated base and target. If this is the case, there should be no causal asymmetry effect. We had 8 conflict triads in which the strongly associated but inappropriate causal target was predictively related to the base and 6 conflict triads in which the direction of the causal link was diagnostic, so we compared the effect of causal direction on the proportion of taxonomic choices for the conflict triads with a paired-samples t-test. The proportion of taxonomic choices for the conflict triads in which the causal choice was predictively related to the base was .75 (*SD* = .23), which was not significantly different from the proportion of taxonomic choices when the causal link was diagnostic (*M* taxonomic proportion = .70, *SD* = .21), $t_{S\ (27)}$ = 1.51, *p* = .14, effect size d = .28. An independent samples t-test with direction as the between-item variable verified the above result across items, $t_{I\ (12)}$ = .38, *p* = .71, effect size d = .21.

In fact, the direction of the difference is opposite to the one that would be predicted if people were using structured causal knowledge, in which case they should be more likely to make a causal choice (and hence less likely to make a taxonomic choices) when the causal link was predictive (see Shafto et al., 2008, for a Bayesian justification of this assumption). This supports the claim that the above effect of triad type was indeed due to the strong association between the causal target and the base rather than being based on the structural relation.

Post-test check

To check that people knew about the purported relationships between the categories, we checked the beliefs about food chain relations (*causal relatedness*) and biological group membership (*taxonomic relatedness*) in a post-test. Data from one participant was lost due to equipment failure.

For the causally related categories, the mean causal relatedness endorsement proportion was .86 (*SD* = .21) and the mean endorsement proportion for being taxonomically related was .17 (*SD* = .20). For the taxonomically related categories, the taxonomic relatedness endorsement proportion was .73 (*SD* = .17) and .13 (*SD* =.12) for being causally related. Finally, for the unrelated categories, the causal relatedness endorsement rate was .22 (*SD* = .19) and the taxonomic relatedness endorsement proportion was .07 (*SD* = .10).

A one-way ANOVA on causal relatedness endorsement proportions showed that there was a significant main effect of type of relation between the categories, $F_{(1.52,\ 39.39^{8})}$ = 102.12,

[8] Adjusted for Nonsphericity using Greenhouse-Geisser

$p < .0005$. Thus, Bonferroni post-hoc tests confirmed that as expected, the mean causal relatedness endorsement was significantly higher for the causally related categories than for the taxonomically and unrelated categories (p's $< .0005$), whereas there was no difference between the latter two ($p = .98$).

Similarly, a one-way ANOVA on taxonomic relatedness endorsement proportions showed that there was a main effect of relation between the categories, $F_{(1.27,33.03)}$[9] $= 209.84$, $p < .0005$, with endorsement proportions significantly higher for the taxonomically related categories than for both the causally and unrelated categories (p's $< .0005$). Also, taxonomic endorsement proportion was higher for causally related categories than for unrelated categories ($p = .051$), probably owing to the fact that one causally related category pair (*cat* and *squirrel*) was indeed from the same biological group.

As in the previous experiment, in order to rule out the possibility that our effects were confounded by lack of appropriate structured taxonomic knowledge, we carried out two comparisons on people's causal and taxonomic relatedness endorsements. Firstly, we compared the proportion of positive responses to the question whether the (strongly associated) causally related pairs were part of the same food chain with the proportion of positive responses to the question whether the (weakly associated) taxonomically related category pairs belonged to the same biological group. A paired-samples t-test was significant, $t_{(26)} = 2.83$, $p = .009$, effect size $d = .54$. As in the previous experiment, this seemed to be driven by a small minority of participants whose taxonomic knowledge was significantly below their knowledge about the causal food chain links. We calculated the difference between the taxonomic and causal endorsement ratings and excluded participants whose difference scores were more than 1.5 standard deviations away from the mean difference.

[9] Adjusted for Nonsphericity using Greenhouse-Geisser

Secondly, we compared the proportion of biological relatedness endorsements for the taxonomically related targets for the conflict triads where people chose the taxonomic target with the proportion of biological endorsements where they chose the causal target. The mean biological endorsement proportion for conflict triads where people made a taxonomic choice was .77 (*SD* = 17), which was significantly higher than the biological endorsement proportion (*M* = .58, *SD* = .36) for conflict triads in which participants chose the causal target, $t_{(26)} = 2.92$, *p* = .007, effect size d = .56. Again, this seemed to be due to a minority of individuals who did not have enough taxonomic knowledge, therefore choosing the strongly associated causal alternative. Thus, we calculated the difference between the proportion of taxonomic endorsements when people chose the appropriate taxonomic target and the proportion of taxonomic endorsements when they chose the strongly associated but inappropriate causal target. We excluded participants whose difference score was larger than 1.5 standard deviations than the mean difference.

People whose knowledge level difference scores were 1.5 standard deviations larger than the mean difference on either of aforementioned knowledge difference measures (8 participants) were excluded. Two paired-samples t-tests showed that for the remaining participants, there was no longer a significant difference between their taxonomic endorsement proportions for taxonomically related categories (*M* = .79, *SD* = .15) and their causal endorsement for causally related category pairs (*M* = .84, *SD* = .23), $t_{(19)} = 1.03$, *p* = .32, effect size d = .23. There was also no longer a significant difference in biological endorsement proportions for taxonomic (*M* taxonomic endorsement = .79, *SD* = .17) versus causal inductive choices (*M* taxonomic endorsement = .71, *SD* = .31), $t_{(19)} = 1.18$, *p* = .25, effect size d = .24.

For thoroughness, the comparison on the mean number of taxonomic choices for the cell conflict triads and cell control triads was also checked for this subgroup. This confirmed that the difference was still highly significant, $t_{(19)} = 6.78$, $p < .0005$, effect size d = 1.52. Taxonomic choices for the cell conflict triads were less frequent (*M* taxonomic choice cell conflict = 10.15, *SD* = 2.68) than for the cell control triads (*M* taxonomic choice cell control = 12.85, *SD* = 1.60). This supports our contention that participants chose the lure on conflict trials because they were influenced by the appeal of the strongly associated but inappropriately related causal target, and not because they lacked the structured knowledge required to select the taxonomically related target.

Semantic Inhibition

We suggested that in order to reason based on appropriate structured knowledge, people have to inhibit contextually inappropriate unstructured knowledge. If this is the case then people who are better able to withhold a prepotent response should show more appropriate selective inductive reasoning and be less likely to be swayed by the inappropriate but strongly associated alternative target. As in the previous experiment, we assessed whether people's tendency to inhibit a response based on a strong but inappropriate association between two categories and to choose the weakly associated but appropriately related category was correlated with their performance on our semantic inhibition task. Table 5.3 below shows the results from the Semantic Inhibition task. Again, to ensure that people were not employing a low-level processing strategy by ignoring the prime sentence and simply judging whether the letter string was a real or non-word, we checked their performance on the real words. The probability of making 50 out of 60 correct judgements for real words by guessing is below .000001. As can be seen from the descriptive statistics shown in table 5.2 below, none of our participants made more than 10 mistakes when judging real words. We take this as confirmation that people processed the prime sentence context before making their judgement.

Table 5.2: Means and Standard Deviations for the Semantic Inhibition Task in Experiment 7

Measure	Mean	SD	Min	Max
Correct Judgement Contextually Appropriate Non-word	27.6	4.5	9	30
Correct Judgement Contextually Appropriate Word	27.8	1.5	24	30
Correct Judgement Contextually Inappropriate Non-word	29.7	0.5	29	30
Correct Judgement Contextually Inappropriate Word	29.5	1.2	25	30

As predicted, there was a significant correlation between people's tendency to choose the weakly associated taxonomic category when reasoning about cells and their level of semantic inhibitory control, i.e. their ability to make a correct judgement for the related non-word trials. In particular, there was a negative correlation between inhibitory control and the difference in number of taxonomic choices between the conflict and control trials when reasoning about cells, $r_{(28)} = -.47$, $p = .012$. Thus, people higher in inhibitory control showed more consistency in their reasoning, making a similar number of relevant taxonomic choices in the conflict and control triad trials. Furthermore, inhibitory control was positively correlated with people's tendency to make a taxonomic choice in the conflict trials, thus inhibiting the strongly associated alternative, $r_{(28)} = .64$, $p < .0005$.

Whilst this suggests that people high in semantic inhibitory control can withstand the interference of strongly associated but irrelevant unstructured knowledge in order to reason on the basis of more structured relevant knowledge, this is only true for people who actually have the relevant knowledge and for whom there is a conflict between irrelevant unstructured and contextually more appropriate structured knowledge. As mentioned in the previous section, there was a minority of 8 individuals who did not have the necessary structured knowledge. We re-analyzed the correlation between taxonomic choices and inhibitory control for the 20 participants who had the relevant taxonomic knowledge. The correlation between inhibitory control and the difference in number of taxonomic choices between cell conflict and control triads was $r_{(20)} = -.46$, $p = .041$. Similarly, the correlation between semantic inhibitory control and total number of taxonomic choices for the cell conflict triads was $r_{(20)} = .65$, $p = .002$. The results of this re-analysis rule out the possibility that our results are due to a correlation between semantic inhibition and possession of structured semantic knowledge. When only those participants with high levels of structured semantic knowledge are included, we continue to observe a correlation between semantic inhibition and reasoning performance. This suggests that people who are high in inhibitory control are most capable of maximizing inductive potency by withstanding interference from strongly associated but contextually irrelevant unstructured knowledge.

Low-Level Motor Inhibitory Control

Whereas the semantic inhibition task taps people's ability to inhibit semantic information that has been automatically activated by context, inhibitory control can be tapped at a lower level. Perhaps the relation between semantic inhibition and people's ability to resist the lure of the strong but irrelevant association and reason based on a relevant relation between base and target premise merely reflects their ability to withhold an appealing prepotent response in order to give a more carefully considered answer. To explore the possibility that the relation between inhibitory control and use of structured knowledge is not constrained to the semantic level of processing, we also looked at the relation between reasoning patterns on the conflict triads and performance on the stop signal task. Descriptive statistics for the stop signal task are presented in Table 5.3 below.

Table 5.3: Descriptive Statistics for the Stop Signal Task in Experiment 7

	Mean	SD	Min	Max
Stop signal accuracy	27.04	4.65	7	32
Primary accuracy	60.11	8.76	24	64
MRT primary trials in ms	559	131	352	971
Stop accuracy 250 ms delay	12.54	2.65	3	16
Stop accuracy 150 ms delay	14.50	2.35	4	16

Total accuracy on the stop signal trials of the inhibitory control measure neither correlated with the mean number of taxonomic choices for the conflict triads when reasoning about cells, $r_{(28)} = .05$, $p = .80$, nor with the difference in number of taxonomic choices between the conflict and control trials when reasoning about cells, $r_{(28)} = -.17$, $p = .39$. Even when people who lacked appropriate taxonomic knowledge were excluded, the correlation between total stop signal accuracy and the mean number of taxonomic choices remained non-significant $r_{(20)} = -.27$, $p = .28$, as did the correlation between stop signal task performance and the difference in number of taxonomic choices between the conflict and control trials , $r_{(20)} = .06$, $p = .08$. It thus seems that the lack of a significant relation is because the stop signal task measures inhibitory control at a simple motor response level. This suggests that in order to resist the lure of strong associative knowledge in order to recruit more structured knowledge requires the ability to inhibit the intrusion of semantic knowledge rather than the ability to simply withhold a response. Interestingly, the measure of semantic inhibition (i.e. rejecting contextually appropriate non-words as suitable sentence fillers) was not correlated with people's performance on the stop signal task, suggesting that they are not tapping into the same domain-general construct of inhibition, $r_{(28)} = -.1$, $p = .61$.

5.2.3 Discussion

In the current experiment we confirmed that recruiting contextually appropriate structured knowledge is harder when there is competition from highly available but inappropriate unstructured knowledge. Thus, people in our triad task more frequently chose the weakly associated but contextually more appropriate taxonomic target category when there was no competition from the inappropriate but strongly associated causal target. This suggests that the use of contextually more appropriate source of structured knowledge can at times necessitate the inhibition of a strong but inappropriate association activated in semantic memory. This was supported by our findings that accurate context-sensitive reasoning and use of structured knowledge on the triad task was related to performance on a semantic inhibitory control task.

The nature of this inhibitory process deserves a more fine-grained analysis, as not all inhibitory measures show equal levels of association with reasoning performance (Miyake, et al., 2000). We found that people's ability to resist the inappropriate choice based on unstructured associative knowledge and reason based on a more suitable structural relation between the base and the alternative target category was correlated with their ability to resist interference by automatic semantic knowledge on a lexical decision-making task. In contrast, there was no relationship between the measure of inhibition obtained from the stop-signal task and inductive reasoning performance. Furthermore, there was a near-zero correlation between the semantic inhibition measure and the stop-signal task. This concurs with other research suggesting that different measures of inhibitory control are poorly correlated (Fan, Flombaum, McCandliss, Thomas, & Posner, 2003; Kramer, Humphrey, Larish, Logan, & Strayer, 1994; Salthouse & Meinz, 1995; Shilling, Chetwynd, & Rabbitt, 2002; Wager, et al., 2005).

A recent study by Bissett, Nee and Jonides (2009) suggests that dissociable inhibitory control processes operate at different processing levels, explaining the lack of correlations between tasks that selectively tap into the independent inhibitory processes. This might explain why the nature of the inhibitory process required to reason based on contextually more appropriate knowledge appears to be circumscribed and highly selective. Concerning our current findings, the most interesting dissociation found by Bissett et al. (2009) was between interference control and conflict resolution at the response selection stage and prepotent inhibition at the response output stage. Research has also confirmed this distinction by demonstrating that these two kinds of inhibitory processes activate common as well as dissociable neural regions (Nee, Wager, & Jonides, 2007). In our reasoning task, automatic activation of strongly associated but inappropriate unstructured knowledge is likely to create a conflict at the response selection stage. To maximize the potency of their selective inference, people have to decide whether to reason based on this strong associative but inappropriate relation or allocate additional resources to processing the more pertinent structural relation. Once this conflict has been resolved, it is unlikely to require further inhibitory processes to provide the correct answer.

Our semantic task also creates a conflict at the response selection stage. Thus, the response cued by the semantic resemblance of the non-word with the primed cloze word conflicts with the response required by the non-word itself. The idea that response selection is prolonged in tasks that require conflict resolution at this processing stage is supported by a study demonstrating increased ERP amplitude arising from activity in the lateral frontal and anterior cingulate cortex (ACC) (West, 2003; West, Jakubek, Wymbs, Perry, & Moore, 2005). In contrast, this conflict at the response selection stage is supposedly absent in the stop-signal task (Rubia, Russell, et al., 2001; Rubia, Smith, et al., 2001), instead only requiring the inhibition of a prepotent response at the output stage. This suggests that inhibitory functions operating in category-based induction are fairly subtle and specific rather than reflecting a global ability to control prepotent responses.

5.3 Overview Experiment 8

In the next experiment, we wished to replicate the finding that in order to maximize selective inductive potency, people must at times inhibit inappropriate unstructured knowledge if this conflicts with more appropriate structured knowledge. Furthermore, we also wanted to explore the possibility that in addition to inhibitory control, the recruitment of structured knowledge might pose a secondary source of difficulty. For example, in order to exhibit the diversity effect with specific conclusions, it is necessary to generate a superordinate category that includes both the premise and conclusion categories. Feeney (2007) has shown that the diversity effect for specific conclusion categories is correlated with people's cognitive ability, suggesting that only people high in cognitive resources might be able to engage in an effortful assessment of taxonomic interrelationships. Thus, in addition to failures of inhibition, failing to effortfully recruit structured knowledge might further explain why people are sometimes swayed by inappropriate unstructured relations between categories, such as a strong association. In this case, people high in working memory resources might be better able to effortfully recruit contextually appropriate structured knowledge. Besides the relationship we found between inhibitory control and selective inference, we might expect a correlation between working memory capacity and appropriate selective inductive reasoning.

Based on the previous experiments, we expected to replicate the finding that people make selectivity errors when they reasoned about cells and there was a conflict between a highly associated but inappropriate causal target and a more appropriate, but only weakly related taxonomic target. We also hoped to repeat the observation that people higher in inhibitory control are more successful at inhibiting unstructured knowledge in favour of more appropriate structured knowledge. Finally, we examined whether people who were more successful at using appropriate structured knowledge outperformed people who were swayed by the inappropriate unstructured knowledge on the Operation Span task, a measure of working memory capacity.

5.3.1 Method

Design

The experiment had a repeated-measures design. The independent variable was the type of triad (type of triad: control or conflict) and the dependent variable was the number of appropriate taxonomic choices. We expected a main effect of type of triad, with people making more contextually appropriate inferences in the control compared to the conflict condition where there was the lure from the strongly associated but inappropriate causal target.

As before, we expected correlations between people's use of more appropriate structured knowledge and their ability to withhold a prepotent semantic response, measured by the variant of the Hayling Task. Thus, there should be a positive correlation between level of semantic inhibitory control and successful selective reasoning, and a negative correlation between inhibitory control and the difference between conflict and control trials.

In addition, we expected to find correlations between people's appropriate selective inductive reasoning on the cell conflict triads and their working memory capacity as measured by the Operation Span task.

Participants

50 participants from Durham University earned course credit for their participation in the experiment. Their mean age was 23.1 years (*SD* = 4.2 years).

Triad Task

The materials were identical to the ones used in the previous experiment. To briefly summarize, we chose 14 triads in which we pitched unstructured knowledge in form of a strongly associated target category against structured knowledge in form of a more appropriately related, but weakly associated alternative target. Thus, when people were reasoning about cells, the appropriate choice (taxonomically related target) was weakly associated, whereas the inappropriate choice (causally related target) was strongly associated. For the cell control triads, the inappropriate but strongly associated target was substituted for a completely unrelated alternative target. As before, we included 28 filler triads in which people were asked to reason about diseases (e.g. *has disease wQ2*) for which both the taxonomic or causal target could be appropriate, in order to ensure that people had to pay attention to the property they were reasoning about. In total, each participant reasoned about 56 triads, 14 conflict cell, 14 control cell triads and 28 disease filler triads.

Materials and Procedure Post-Test and Semantic Inhibitory Task

The materials for the post-test and the semantic inhibitory control task were identical to the previous experiment.

Automated Operation Span

Materials

People's working memory was assessed by means of the automated operation task, AOSPAN (Unsworth, Heitz, Schrock, & Engle, 2005). In this task, to-be-remembered items are interspersed with processing items. The memory items consist of 12 letters (F, H, J, K, L, N, P, Q, R, S, T and Y). The processing items consist of mathematical operations (e.g. (1*3) + 1 = ?), followed by a digit (e.g. 4) and a "true" and "false" box. Subjects had to solve the mathematical operation and click the appropriate box, verifying whether the presented digit was a correct or incorrect solution. For half of the mathematical operations, the subsequent digit represented a correct solution, whereas for the other half, the digit was not a valid result.

Procedure

The task was presented on a laptop and was mouse-driven. There were three practise elements. In the first section, participants completed the memory task. They saw a letter on-screen for 800 milliseconds. After a varying number of letters, participants were presented with 4*3 letter matrix and they had to try and recall the previously presented letter sequence in the correct order by clicking a box next to the chosen letter. Participants were then provided with feedback on how many letters they had correctly remembered. Figure 5.5 below, taken directly from Unsworth et al (2005), shows the basic task procedure.

The second practice session familiarized participants with the mathematical operations and also served as the basis for determining an individual's unique response deadline for the main trials. Participants were asked to respond to the maths problems as fast as possible without sacrificing accuracy. Following each of the 15 trials, participants received feedback on their performance in form of the percent of maths problems solved correctly.

For the final practice session, participants completed the letter and maths problems simultaneously. They were first presented with the mathematical operation. The response deadline was adjusted to each individual based on his/her performance on 15 pre-test maths problems by adding 2.5 standard deviations to their response mean. Once they had solved the maths operation, they were presented with a letter after 200 milliseconds. In total, there were three practice trials consisting of a letter set size of two.

Next followed the main experimental trials. In total, there were five different set sizes, ranging from three to seven letters. There were three sets of each set size, totaling 75 letters and 75 mathematical operations. The order of presentation of a given set size was completely random for each participant.

In order to ensure that people did not develop a strategy of rehearsing the letters and disregarding the mathematical operations, participants were asked to try to be at least 85% accurate on the maths operations. Accuracy was fed back to participants during the recall stage, where a cumulative percentage in red was shown in the top right-hand corner.

If recruitment of relevant structured knowledge does indeed require cognitive effort, we might expect to see correlations between people's number of accurate taxonomic choices in the conflict triads and their working memory capacity, operationalized as the total number of letters that they could correctly recall.

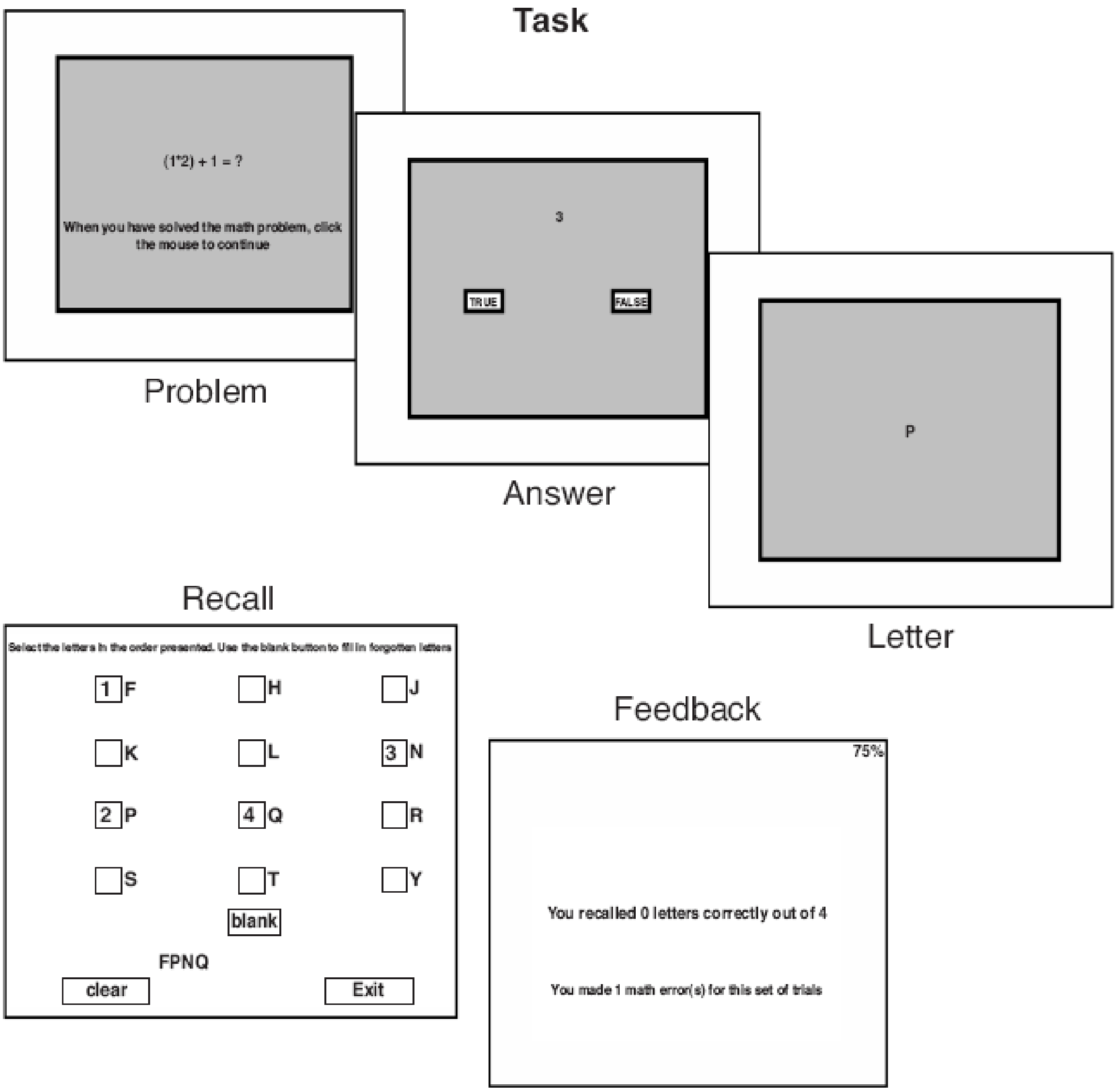

Figure 5.5: Illustration of the AOSPAN. Presentation of a math operation followed by a digit. Participants chose whether the digit is a correct or incorrect solution to the math operation. Followed by a letter for 800 ms. At recall, participants choose the presented letters in the correct order from an array. Followed by feedback for 2000 ms.

Participants carried out the four tasks in one individual session of around 50 minutes. The experimental components were completed in fixed sequence, starting with the triad task, followed by the post-test belief assessment, the semantic inhibitory control task, and finishing with the AOSPAN task.

5.3.2 Results

Triad Task

We summed the number of times a participant chose the weakly associated taxonomic category separately for the cell conflict triads and cell control triads and analysed these with a paired-samples t-test.

As in the previous two experiments, there was an effect of type of triad, $t_{S\ (49)} = -7.68$, $p < .0005$, effect size d = 1.1; $t_{I\ (13)} = -3.79$, $p = .002$, effect size d = 1.0. Illustrated in Figure 5.6, people chose the weakly associated taxonomic target more frequently for the control triads (*M* number of taxonomic choices = 12.7, *SD* = 1.3) than for the conflict triads where there was competition from the strongly associated but inappropriate causal target (*M* = 9.7, *SD* = 3.0).

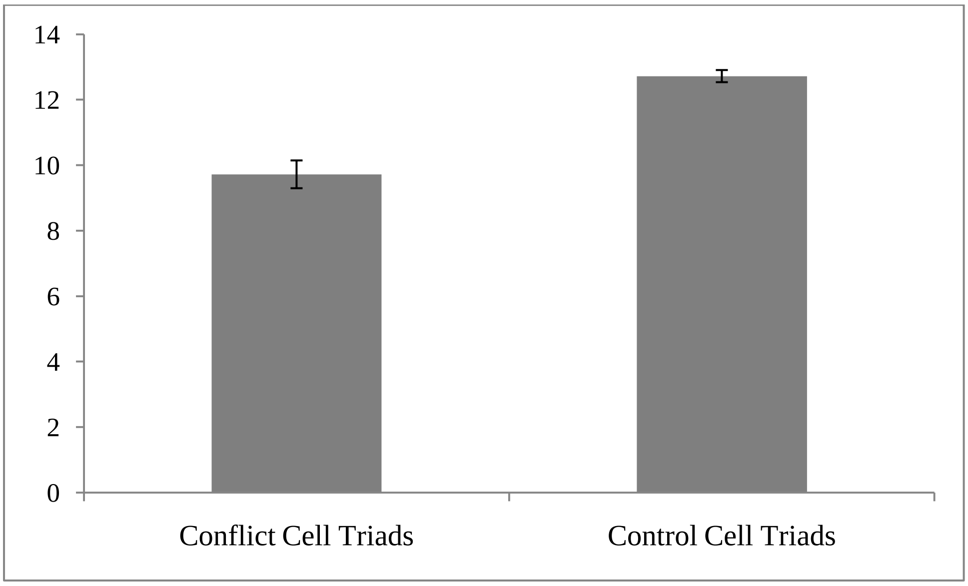

Figure 5.6: Mean Number of Taxonomic Choices across the two Types of Triads when reasoning about Cells

This suggests that inappropriate unstructured knowledge was detrimentally affecting people's ability to base their inductive choices on more appropriate structured taxonomic knowledge.

In order to show that people were swayed by unstructured knowledge in form of a strong association between the base and the causal target we compared the proportion of taxonomic choices for the conflict triads in which the causal link was predictive (8 triads) to conflict triads in which this link was diagnostic (6 triads). If people were lured by the strong association rather than by the structured relation, we would not expect to observe a causal asymmetry effect.

The proportion of taxonomic choices for the conflict triads in which the causal choice was predictively related to the base (e.g. *ants* → *anteaters* or dragonflies?) (*M* taxonomic choice = .73, *SD* = .22) was significantly higher ($t_{S\,(49)}$ = 2.95, *p* = .005, effect size = .37) than the proportion of taxonomic choices for the triads in which the causal link was diagnostic (e.g. *butterflies* → *flowers* or locusts?) (*M* taxonomic proportion = .64, *SD* = .26). However, the effect of direction was only significant across participants, not across items ($t_{I\,(12)}$ = .6, *p* = .56, effect size d = .32). As the direction of this difference was opposite to the one that would be predicted if people were drawing on structured causal knowledge, we believe that the above effect of triad type can be attributed to the strong unstructured association between the causal target and the base rather than being based on the underlying structural relation.

Post Test Check

To check that people knew about the structured relations between the categories, we checked beliefs about food chain relations (*causal relatedness*) and biological group membership (*taxonomic relatedness*) in a post-test. For the causally related categories, the mean causal relatedness endorsement proportion was .84 (*SD* = .13) and the mean endorsement proportion for being taxonomically related was .15 (*SD* = .16). For the taxonomically related categories, the taxonomic relatedness endorsement proportion was .73 (*SD* = .19) and .12 (*SD* = .15) for being causally related. Finally, for the unrelated categories, the causal relatedness endorsement rate was .17 (*SD* = .09) and the taxonomic relatedness endorsement proportion was .07 (*SD* = .1).

A one-way ANOVA on causal relatedness endorsement proportions showed that there was a significant main effect of type of relation between the categories, $F_{(1.62,\ 79.39[10])} = 567.66$, $p < .0005$. Bonferroni post-hoc tests confirmed the expected pattern of results. Thus, the mean causal relatedness endorsement was significantly higher for the causally related categories than for both the taxonomically and unrelated categories (p's $< .0005$). There was no difference in causal relatedness endorsement for the unrelated categories compared to the taxonomically related categories ($p = .06$).

Similarly, a one-way ANOVA on taxonomic relatedness endorsement proportions showed that there was a main effect of relation between the categories, $F_{(1.6,\ 78.57[11])} = 329.23$, $p < .0005$. Bonferroni post-tests showed that the taxonomic relatedness endorsement proportion was significantly higher for the causally related categories than for the unrelated categories ($p = .002$). However, the important finding was that the taxonomic endorsement proportion for the taxonomically related categories was significantly higher than for both the causally and unrelated categories (p's $< .0005$).

To check that people weren't simply choosing the strongly associated but inappropriate target because they lacked structured knowledge about the taxonomic relation between the two weakly associated categories, we compared the proportion of positive responses to the question whether the (strongly associated) causally related pairs were part of the same food chain with the proportion of positive responses to the question whether the (weakly

[10]df adjusted for non-sphericity using Greenhouse-Geisser
[11] df adjusted for non-sphericity using Greenhouse-Geisser

associated) taxonomically related category pairs belonged to the same biological group. As in the previous two experiments, a paired-samples t-test was significant ($t_{(49)} = 3.56$, $p < .0005$, effect size d = .55), showing that the endorsement rates about biological relatedness for the taxonomically related pairs (*M* proportion = .73, *SD* = .19) was significantly lower than the endorsement rate of being part of the same food chain for the causally related pairs (*M* proportion = .84, *SD* = .13). We worked out the difference between the taxonomic and causal endorsement ratings and eliminated participants whose difference scores were larger than 1.5 standard deviations away from the mean difference.

As in the previous experiments, we also compared the proportion of biological relatedness endorsements for the conflict triads where people chose the taxonomic target with the proportion of biological endorsements where they chose the causal target. When all participants were included, the mean biological endorsement proportion where people made a taxonomic choice was .80 (*SD* = .19), whereas it was only .60 (*SD* = .35) for conflict triads in which people chose the causal target. A paired samples t-test showed that this difference was statistically significant ($t_{(49)} = 4.07$, $p < .0005$, effect size d = 1). Again, we computed the difference between the proportion of taxonomic endorsements when people chose the appropriate taxonomic target and the proportion of taxonomic endorsements when they chose the inappropriate causal target and excluded participants whose difference score was larger than 1.5 standard deviations than the mean difference

To remedy the problem of lack of appropriate taxonomic knowledge, we excluded 18 participants whose difference scores were further than 1.5 standard deviations away from the mean difference on either of the two aforementioned structured knowledge indices. For the remaining participants, the results showed that there was neither a difference in the mean biological endorsement proportion (*M* = .82, *SD* = .16) and their causal relatedness endorsement proportion (*M* = .83, *SD* = .14), $t_{(31)}$ = .84, *p* = .41, effect size d = .09, nor a significant difference ($t_{(31)}$ = .42, *p* = .68, effect size d = .08) in biological endorsement proportions for taxonomic (*M* taxonomic endorsement = .81, *SD* = .21) versus causal inductive choices (*M* taxonomic endorsement = .79, *SD* = .23).

To be sure, we repeated the comparison on the mean number of taxonomic choices for the cell conflict triads and cell control triads for people who had relevant structured knowledge. This confirmed that the difference was still highly significant ($t_{(31)}$ = 5.5, $p <$ .0005, effect size d = .95). Participants made fewer appropriate taxonomic choices for the cell conflict triads (*M* = 10.2, *SD* = 3.0) than for the cell control triads (*M* = 12.9, *SD* = 1.0). This supports our suggestion that they were swayed by unstructured knowledge instantiated by the strongly associated but inappropriately related causal target.

Relation between Selective Inductive Reasoning and Semantic Inhibition

The first correlations we looked at were between people's successful use of structured knowledge and their performance on the inhibitory control task, which assesses inhibition at a semantic level. Table 5.4 below gives descriptive statistics for the semantic inhibition task.

Table 5.4: Means and Standard Deviations for the Semantic Inhibition Task Experiment 8

Measure	Mean	SD	Min	Max
Correct Judgement Contextually Appropriate Non-word	28.3	2.1	21	30
Correct Judgement Contextually Appropriate Word	27.5	2.1	20	30
Correct Judgement Contextually Inappropriate Non-word	29.8	0.6	27	30
Correct Judgement Contextually Inappropriate Word	29.8	0.5	28	30

In line with the findings from Experiments 7, there was a significant correlation between choosing the weakly associated taxonomic category when reasoning about cells and level of semantic inhibitory control, i.e. people's ability to make a correct judgement for the related non-word trials. Across the whole sample, inhibitory control was negatively related to the difference in number of taxonomic choices between cell conflict and cell control triads, $r_{(50)} = -.33$, $p = .019$. This means that people high in semantic inhibitory control exhibited more consistent reasoning patterns, making a similar number of appropriate taxonomic choices in the conflict and control triad trials. Inhibitory control was also positively related to the number of taxonomic choices in the conflict triads, $r_{(50)} = .30$, $p = .034$, supporting the idea that they were better at inhibiting unstructured knowledge in favour of more appropriate structured knowledge.

However, these correlations were substantially smaller than those obtained in the previous experiment. This decreased magnitude of the relationship between semantic inhibitory control and appropriate context-sensitive inductive reasoning might be driven by the people who did not have relevant taxonomic knowledge as suggested by the post-test. To explore this possibility, we divided our sample into two groups of individuals, those with and those lacking structured taxonomic knowledge. The group lacking appropriate structured knowledge (N = 18) was made up of people whose difference scores between their taxonomic and causal endorsements was further than 1.5 standard deviations away from the mean. It also included people whose taxonomic endorsement proportions were more than 1.5 standard deviations away from the mean difference between conflict triads where they chose the causal versus the taxonomic target.

We then re-ran the correlational analyses separately for the two groups of people who differed in whether or not they had structured taxonomic knowledge. Figure 5.7 below shows a visual illustration of the divergent relation between semantic inhibitory control and appropriate context-sensitive reasoning depending upon structured knowledge level.

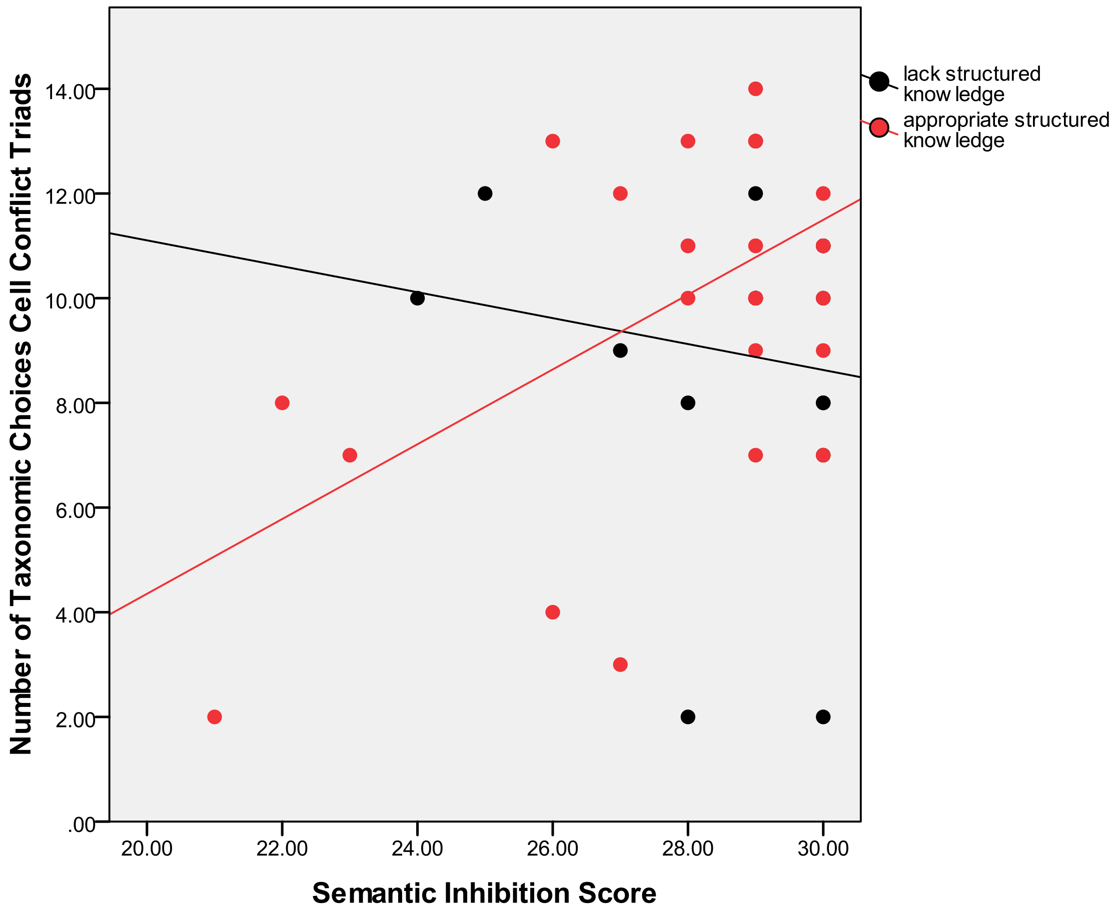

Figure 5.7: Relation between Semantic Inhibition and Number of Taxonomic Choices depending upon Presence of Structured Knowledge

For the 32 individuals with taxonomic knowledge the correlation between inhibitory control and the difference in number of taxonomic choices between cell conflict and control triads increased to $r_{(32)} = -.57$, $p = .001$. Similarly, the correlation between semantic inhibitory control and total number of taxonomic choices for the cell conflict triads increased to $r_{(32)} = .54$, $p = .001$. This suggests that amongst those who have the appropriate background knowledge, appropriate context-sensitive reasoning is best when people are high in inhibitory control, as they can better withstand the interference of strongly associated but inappropriate unstructured knowledge.

Furthermore, when we ran the same correlation analyses on the individuals who did not have the relevant taxonomic knowledge, hence eliminating the potential conflict between structured and unstructured associative knowledge, the correlation between taxonomic choices and semantic inhibitory control was substantially reduced and no longer significant. Thus, for individuals lacking structured knowledge the correlation between semantic inhibitory control and the difference in number of taxonomic choices between cell conflict and control triads decreased to $r_{(18)} = -.15$, $p = .55$. The association between semantic inhibitory control and total number of taxonomic choices for the cell conflict triads decreased to $r_{(18)} = .22$, $p = .39$.

Relation between Working Memory Capacity and Appropriate Selective Induction

People with greater working memory capacity might show superior performance on the inductive reasoning task, as they are better able to recruit appropriate structured knowledge whilst facing distraction/competition from inappropriate unstructured knowledge. The AOSPAN task assesses people's ability to store and recall letters from memory whilst simultaneously carrying out mathematical operations (Unsworth, et al., 2005). We used the total number of letters participants could correctly recall, thus giving them credit for partially correct sets. This method is superior to all-or-nothing scoring, showing the highest internal consistencies across different span tasks (Conway, et al., 2005). Table 5.5 below shows the mean and standard deviations for the correctly recalled letters, as well as mean number of speed and accuracy errors for the mathematical operations. Data from 2 people was missing due to equipment failure.

Table 5.5: Means and Standard Deviations for the AOSPAN in Experiment 8

Measure	Mean	SD	Min	Max
Total Number Correct Recall	55.8	14.3	13	73
Math Speed Error	1.9	3.6	0	21
Math Accuracy Error	6.2	6.3	0	42

Overall, scores on the AOSPAN were significantly correlated with the difference between number of taxonomic choices for cell conflict compared to cell control triads ($r_{(48)} = -.35$, $p = .015$), and correlated, but not significantly, with the overall number of taxonomic choices for cell conflict triads ($r_{(48)} = .27$, $p = .067$). These relationships appear to be fairly modest, however, as with the relation between selective inductive reasoning and semantic inhibition, this may be attenuated by people who did not have relevant taxonomic knowledge. To explore this possibility, the correlations were re-run separately for the two subgroups differing in their level of taxonomic knowledge. Figure 5.8 below shows this relation for the two subgroups.

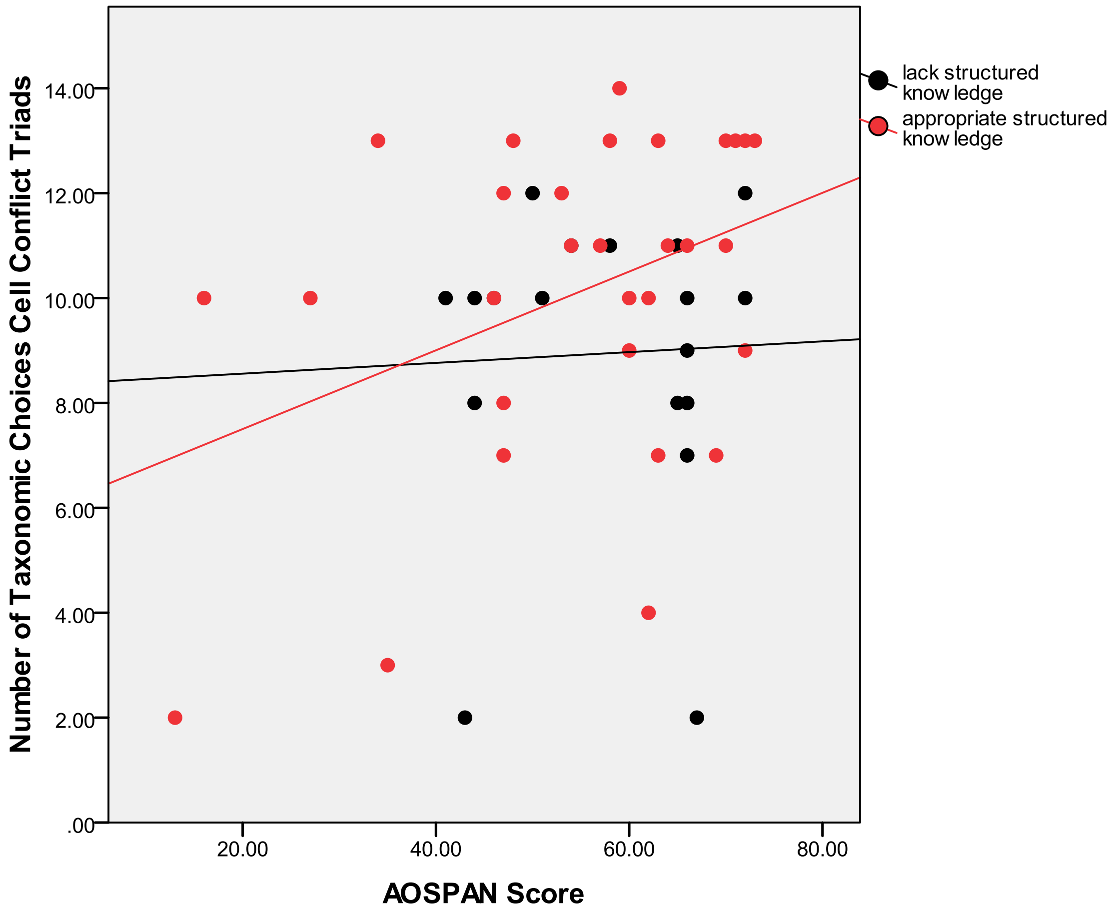

Figure 5.8: Relation between Working Memory AOPSAN Score and Number of Taxonomic Choices depending upon Presence of Structured Knowledge

As can be seen from Tables 5.6 and 5.7 below, there were robust correlations between working memory and the ability to choose the appropriate taxonomic target category in favour of the strongly associated but inappropriate competing category for people who had taxonomic knowledge, but not for those individuals lacking structured knowledge.

Table 5.6: Correlations between Taxonomic Choices, Semantic Inhibition and Working Memory Span for Individuals *with* Structured Knowledge (N=30)

	Number Taxonomic Choices	Semantic Inhibition Score	AOPSAN Score
Number Taxonomic Choices	1	**.54****	**.39***
Semantic Inhibition Score		1	**.36***
AOSPAN Score			1

* $p < .05$ ** $p < .001$

Table 5.7: Correlations between Taxonomic Choices, Semantic Inhibition and Working Memory Span for Individuals *lacking* Structured Knowledge (N=18)

	Number Taxonomic Choices	Semantic Inhibition Score	AOPSAN Score
Number Taxonomic Choices	1	.04	-.15
Semantic Inhibition Score		1	-.18
AOSPAN Score			1

Partial Correlations between Appropriate Context-Sensitive Induction, Semantic Inhibition and Working Memory

According to Engle (2002), one important component of the working memory construct can be conceptualized as executive attention. Thus, we explore to what extent the AOSPAN and our measure of semantic inhibition each explains unique variance in the total number of appropriate taxonomic choices on the conflict triads. As we assume that inhibition is only required when there is a conflict between structured and unstructured knowledge, and working memory is involved with the recruitment of weakly associated but appropriate structured knowledge, we limited this analysis to the people who had structured knowledge.

Firstly, there was a significant correlation between semantic inhibitory control and AOSPAN scores, $r_{(28)} = .36$, $p = .05$. Furthermore, the significant zero-order correlation between people's number of appropriate taxonomic choices for the cell conflict triads and their AOSPAN score, $r_{(28)} = .39$, $p = .034$, decreased and became non-significant when controlling for people's semantic inhibitory control score, $r_{(27)} = .25$, $p = .2$. However, the substantial zero-order correlation, $r_{(28)} = .54$, $p = .002$, between number of taxonomic choices and semantic inhibitory control score was only slightly decreased when controlling for AOPSAN score, $r_{(27)} = .46$, $p = .011$. This suggests that what makes the selection of appropriate knowledge difficult in our task is the ability to block out interference from inappropriate unstructured knowledge and focus attention upon more appropriate structured relations. This supports several authors' views (McCabe, Roediger, McDaniel, Balota & Hambrick, 2010; Engle, 2002; Barrett, Tugade & Engle, 2004; Kane & Engle, 2003) that working memory tasks and inhibitory control tasks share an underlying executive attention component. It seems that it is this component of working memory that drives its relation to inductive reasoning performance in our task.

5.3.3 Discussion

The results from the current study corroborate those from Experiment 6 and 7. When people had to choose which target was more likely to share novel cells with a base category, they were often swayed by an inappropriate but strongly associated target rather than choosing the appropriate taxonomically related category. The ability to choose the weakly associated but appropriate target was strongly related to people's performance on a measure of semantic inhibitory control, but only when people had the relevant background knowledge. Furthermore, in people with relevant structured knowledge, performance on the inductive reasoning triads was also related to scores on the AOSPAN, a measure of working memory, although this correlation became non-significant when variance explained by semantic inhibitory control was partialled out. Most interestingly, when we split our participants according to their level of structured knowledge, the correlations between inductive reasoning performance, semantic inhibitory control and working memory were only significant for those participants with structured knowledge, whereas the relation was absent for the group who lacked this knowledge. This would be expected, as the conflict between structured knowledge and unstructured associative knowledge would only be present in those who actually had the relevant background knowledge, strengthening our claim that category-based induction can be influenced by two types of knowledge that differ in their processing characteristics.

5.4 General Discussion

The three experiments described in this chapter provide further evidence that structured and unstructured knowledge influence selective category-based induction, and that the ability to reason based on structured knowledge is related to an individual's level of inhibitory control. Although these two types of knowledge will often coincide and afford similar inferences, there are occasions when structured and unstructured knowledge may conflict. In the current experiments, we created a conflict between inferences based on unstructured knowledge (strong association between base and target) and structured knowledge (weakly associated, but appropriate structural relation between base and alternative target).

The results suggest that when the to-be-generalized property of a base category afforded a generalization to an appropriately related but weakly associated target, people were often swayed by the very strongly associated but inappropriate alternative target. The findings also showed that the tendency to inhibit the strong association in favour of the target with an appropriate structural relation was highly correlated with a specific type of inhibitory control, namely semantic inhibitory control, but not with more basic-level motor-level inhibitory control.

The notion that inhibitory control forms a crucial component of working memory is borne out in Engle's (2002; 2010) conceptualization of working memory as attention control. He claims that working memory reflects the interface between attention and memory, that is, the capability to either maintain or inhibit memory representations depending upon their relevance to the current task. In the first experiment, semantic inhibitory control only correlated with reasoning performance in the light load condition, not in the heavy load condition. To some extent, this dissociable effect of memory load mirrors the findings from Kane and Engle (2000), who showed that only people with superior working memory showed a decline in performance on a recall task under secondary load, whereas people with low working memory spans were unaffected by the secondary task. As secondary tasks are thought to interfere with the efficiency of working memory, it would appear that the predominant resource-demanding challenge in our category-based inductive reasoning task lies in inhibiting irrelevant memory representations, rather than in the subsequent ability to activate and maintain more task-relevant knowledge structures. This idea is further reinforced by the finding that although superior selective inductive reasoning performance was related to a measure of working memory in the third experiment, this was largely via the shared variance between the working memory and semantic inhibitory control measure. Once the variance accounted for by the semantic inhibitory control measure was partialled out, the association between working memory score and performance on the selective inductive reasoning task was no longer significant.

The findings also help our understanding of the more domain-general processes involved in category-based induction and reasoning more generally. Thus, as already demonstrated in the previous two chapters, reasoning based on unstructured knowledge seems to be mediated by an effortless, heuristic process, whereas the ability to draw on structured knowledge requires the involvement of more effortful, analytical reasoning processes. The experiments reported in the current chapter go some way in illuminating the nature of the interaction between the two contrasting processes, echoing similar findings in social cognition research (Macrae, Bodenhausen, Schloerscheidt, & Milne, 1999; von Hippel, Silver, & Lynch, 2000), as well as in the deductive reasoning and decision-making literature. For example, De Neys (2006b) demonstrated that reasoning performance was negatively affected by a secondary cognitive load, regardless of people's working memory capacity. De Neys (2006b) findings suggests that differences in reasoning performance do not arise because of qualitative differences in the reasoning architecture, but are due to the efficiency with which people may be able to inhibit an automatic response in favour of applying a more effortful process. Our results suggest a slightly more elaborate interpretation. Whilst a secondary memory load led to an overall decline in selection of appropriate knowledge, it also eliminated the relation between semantic inhibitory control and reasoning performance. One could speculate that the secondary task was more detrimental to people who are better at inhibiting irrelevant information compared to people who are less proficient at inhibiting appealing but inappropriate knowledge. The latter group might already be already performing poorly, so the secondary task would simply level the playing ground between people who are more or less proficient at inhibitory control, analogous to Kane & Engle's (2000) findings that a secondary task was more detrimental for people with higher working memory scores. The fact that the correlation between the measure of working memory control and reasoning performance was largely mediated by scores on the semantic inhibitory control task suggests that it is the shared component of executive attentional control that is crucial for selecting appropriate structured knowledge, and which is disrupted by a secondary memory task.

Chapter VI

Summary, Discussion and Final Conclusions

A series of experiments examined the role of knowledge in category-based induction and how different types of knowledge, which are subject to contrasting processing constraints, can have dissociable effects on people's inferences. Experiments 1-3 (Chapter 3) used timing manipulations and a secondary task paradigm to explore whether it is possible to dissociate the effects of different types of knowledge in category-based inductive reasoning. It also examined whether knowledge from different structural domains differ in their relative accessibility. In Chapter 4, Experiments 4 and 5 used a more naturalistic paradigm to substantiate the claim that both structured and unstructured types of knowledge determine inductive reasoning output, and that these different types of knowledge are mediated by contrasting cognitive processes. The final series of experiments (Chapter 5) explored the role of mental processing capacities and inhibitory control in mediating the use of structured and unstructured knowledge. This chapter summarizes how the experimental findings answer the questions posed at the outset of this thesis. It then critically reviews what implications these answers have for theories of category-based inductive reasoning. With reference to more general theories of reasoning, in particular dual process frameworks, this chapter also evaluates what the findings can tell us about the fundamental nature of cognitive processes involved in reasoning. The chapter concludes with an outlook onto future challenges and opportunities.

6.1 Structured and Unstructured Types of Knowledge

We began with the hypothesis that theories of category-based inductive reasoning form a dichotomy in how they conceptualize the nature of knowledge driving people's inferences. On the one hand, there are frameworks which call attention to knowledge that cannot be captured by a higher-order theoretical structure. We summarize these approaches under theories that emphasise the role of unstructured knowledge, such as featural similarity (Sloman, 1993b), perceptual similarity (Sloutsky & Fisher, 2004a) and associative strength (Rogers & McClelland, 2004). On the other hand, there are frameworks in which theoretical interrelationships between categories are central to understanding people's inferences. We subsume these under theories that favour structured knowledge, such as taxonomic hierarchies (Osherson, et al., 1990), ecological relationships (Coley, et al., 2005) and causal relations (Rehder, 2006; 2009). One way to resolve this apparent theoretical dichotomy is to allow for both types of knowledge to influence people's inductions, but that different types of knowledge are subject to contrasting processing constraints. We re-interpreted the findings from several experiments (Barraff & Coley, 2003; Coley & Barraff, 2003; Shafto, Coley, et al., 2007) to argue that the application of unstructured knowledge may be a heuristic process, which is effortless and automatic, whereas the use of structured knowledge may require more analytical processing, requiring considerable time and cognitive processing resources (see Rehder, 2009). Our experiments demonstrate the tenability of this proposition, allowing us to answer the more specific questions raised in the introduction.

Question 1:

Are unstructured and structured knowledge dissociable, and if so, do they have differential impact on the reasoning output?

The first step to answering this question was to devise measures of each type of knowledge and show that they did not correlate with one another. We defined unstructured knowledge as the degree of simultaneous activation of two or more category instances, which may arise from spatial and/or temporal contiguity, or from having similar features. One way of psychologically representing unstructured knowledge seems to be strength of association between categories. Two categories ought to be more strongly associated the more frequently they co-occur, thus encoding statistical properties of our environment. One feature of unstructured knowledge is that people do not need to reason explicitly about how and why two categories are related/unrelated. In fact, there need not even be a deep underlying structure that relates one or more categories. This has been used extensively in advertising, where a brand (e.g. bathroom tissue paper) is associated with a structurally unrelated category which has favourable or positive properties (e.g. a cute, soft and cuddly puppy) (Knobil, 2003).

In a first step, we asked people for an evaluation of associative strength between categories, providing us with a subjective measure of unstructured knowledge. However, it was important to verify this subjective measure against a more objective criterion of associative strength, such as frequency of co-occurrence. Thus, in a second step we carried out a co-occurrence search on the World Wide Web using two different search engines, Google and Exalead. Using a formula suggested by Heylighen (2001), the conditional co-occurrence was computed, which takes the frequency of each individual world in a pair into account. The correlations between the subjective associative strength measure and the two objective co-occurrence indices were highly correlated, suggesting that we succeeded in creating an index of associative strength between categories.

In contrast to unstructured knowledge, structured knowledge was defined as abstract and underlying theories which describe lawful relations between categories within a domain. We determined people's structured knowledge by asking them to explicitly assess causal and taxonomic relatedness between category pairs. When we correlated these two structured knowledge variables with the three different measures of unstructured knowledge in Experiments 1 to 3, only one out of the 36 individual correlation coefficients was significant. This supports our contention that we managed to create dissociated measures of structured and unstructured knowledge.

To answer the second part of the question, whether structured and unstructured knowledge have dissociable effects on the reasoning output, it is necessary to look at whether inferences diverge depending upon which type of knowledge has most influence on the reasoning output. Several characteristics follow from the aforementioned conceptualization of unstructured and structured knowledge which permit a distinction between the influence of these two contrasting types of knowledge. For example, an important characteristic of unstructured knowledge that allows it to be dissociated from structured knowledge is that it should be nondirectional, so that inferences from A to B ought to be of similar strength as inferences from B to A. In contrast, structured knowledge has an underlying abstract theoretical organization which defines interrelationships between category instances (Tenenbaum, et al., 2007). Causal knowledge is one such type of knowledge, characterized by its complex underlying structure (Sloman, 2005). Categories might be related by a 'common causal category' (e.g. *pigeons* and *rats* are treated as vermin by *humans*), or they may be related by a 'common effect category' (e.g. *hawks* and *cats* both eat *mice*). However, the simplest relation that can be exploited for maximizing the potency of inductive inferences is the asymmetry of causal knowledge, with causes always preceding, or at least coinciding, with their effects. For example, if people are reasoning about the transmission of diseases, an inference from prey to predator (e.g. from *rabbits* to *foxes*) should be stronger than from predators to prey (e.g. from *foxes* to *rabbits*) (Medin, et al., 2003; Rehder, 2009; Shafto, et al., 2008). Such causal asymmetry effects would not be expected if people were treating causal relations as simple symmetrical associations, a strategy employed by a substantial minority in Rehder's (2009) experiment. However, the challenge is to combine descriptive and computational-level models, such as the theory-based Bayesian approaches which predict phenomena such as the causal asymmetry effect, with concrete theories about online processing in order to explain when and how different types of knowledge affect the reasoning output.

In the first study (Experiments 1 and 2), we used the standard inductive inference evaluation paradigm to explore whether the causal asymmetry effect is mediated by considerations of the underlying causal structure rather than relying on unstructured associative knowledge. The findings strongly suggested that when people had the available mental resources, and were not under any time pressure or heavy cognitive burden, they were able to apply structured causal knowledge. In a relevant context, such as when reasoning about the likelihood of sharing diseases, these individuals thought that causal predictive inferences (i.e. from *prey* to *predator* or from *plant* to *consumer*) were stronger than causal diagnostic inferences (i.e. from *predator* to *prey* or from *consumer* to *plant*). Crucially though, this difference disappeared when people were not granted enough time or working memory resources to draw on structured causal knowledge. Instead, they seemed to use a non-directional associative strength heuristic as the basis for making their inductive evaluations. This also parallels the finding that directional asymmetries in reaction times between causally related concepts only emerge when people assessed causal relatedness but not when considering degree of association (Fenker, et al., 2005; Satpute, et al., 2005). Thus, it appears that the use of structured knowledge can lead to a qualitatively different reasoning output than the use of unstructured knowledge.

The second study (Experiments 4 and 5) used a completely different paradigm to explore whether people can draw on both structured and unstructured sources of knowledge to make category-based inductive inferences. Rather than being asked to evaluate pre-determined inferences between category pairs, participants were asked to generate their own inference in response to learning that a category had a specific property. This circumvents concerns that people may not always have the relevant structured knowledge that is necessary to spot the relation between base and target categories (Coley, et al., 2005). For example, in Shafto et al.'s (2007) post-test, participants were asked to state whether category pairs were ecologically and/or taxonomically related. Participants were less certain about their ecological knowledge than about their taxonomic knowledge, giving twice as many 'don't know' answers in response to the ecological relatedness question compared to the taxonomic relatedness question. Similarly, in our last series of experiments (Experiments 6, 7 and 8), there was a substantial minority of individuals whose taxonomic knowledge was substantially worse than their causal knowledge.

When people are given the option to generate their own inferences, they may draw on several domains of structured background knowledge. For example, when reasoning about the transmission of diseases, people may apply causal knowledge (e.g. being part of the same food chain), ecological knowledge (e.g. living in the same habitat) as well as taxonomic knowledge (e.g. closely related species may share genetic vulnerabilities to certain diseases). This would almost certainly be more inductively potent than merely relying on the first category instance that sprung to mind. To test the assumption that people can draw on unstructured and structured knowledge when generating an inference, we manipulated the amount of mental resources people could dedicate to the inductive generation task, and thus presumably the likelihood that they could make use of structured knowledge. The results from both Experiments 4 and 5 demonstrated that when another group of people were asked to evaluate the strength of association between the base and generative target category, the pairs people had generated under a heavy cognitive burden were judged as being more strongly associated than the category pairs generated under minimal cognitive load. Furthermore, the inductive strength people assigned to their own inferences was better predicted by associative strength when people were under cognitive strain, but by an index of structured knowledge when they were not cognitively taxed. This is similar to the findings of Baraff and Coley (Baraff & Coley, 2003; Coley & Barraff, 2003), who asked musical experts and novices to make inductive inferences about musicians and/or composers. They found that music experts seemed to draw on structured context-dependent relational knowledge when they had plenty of time, but reverted to a similarity-based strategy akin to that used by novices when they were under time pressure. Such a strategy presumably reflects the effortless and heuristic use of more unstructured types of knowledge.

Together the findings from the first and second set of experiments suggest that structured and unstructured knowledge can be dissociated, and demonstrate how these two types of knowledge can have a differential effect on the reasoning output. This provides a parsimonious way of uniting apparently dichotomous theories of inductive reasoning. It also links our work on category-based inductive reasoning to more general theories of reasoning, and specifically, dual process frameworks (Epstein, Pacini, DenesRaj, & Heier, 1996; Evans, 2006, 2007; Evans & Over, 1996; Sloman, 1996; Stanovich & West, 1998), whereby unstructured knowledge has a relatively automatic effect upon reasoning, whereas the use of structured knowledge will be constrained by contextual and cognitive factors.

If we accept the assertion that the use of structured knowledge is facilitated by more effortful, time-consuming, analytical reasoning, this still leaves open the question of whether different domain-specific knowledge structures are equivalent in terms of their processing demands and availability. For example, Shafto et al. (2007) suggested that taxonomic knowledge is privileged and more available than ecological knowledge. They based this claim on the finding that timing had a detrimental effect on the use of ecological but not on taxonomic knowledge. However, the researchers did not control for strength of association between the categories. This could potentially confound any conclusion drawn about domain-specific differences, an addressed by the next question.

Question 2:
Are apparent differences regarding the importance and availability of knowledge from different domains (Rehder, 2006; Shafto, Coley, et al., 2007) still evident when level of association between the categories has been controlled for?

Different researchers have taken contrasting positions on which domain of knowledge might dominate the reasoning output. To ensure that we were dealing with the characteristics of a domain-specific knowledge structure rather than with a confounding variable such as unstructured strength of association between categories, we ensured that the level of association between causally and taxonomically category pairs was identical. Across all three experiments, there was neither a fundamental advantage for taxonomic knowledge, as suggested by Shafto et al (2007), nor for causal knowledge, as proposed by Rehder (2006). Some of the regression analyses did suggest that taxonomic knowledge might be less vulnerable to manipulations of time and mental resources. However, this was largely the case when people were reasoning about cells. When they were reasoning about infections and diseases, it appears that the use of both taxonomic and causal structured knowledge was somewhat compromised when people were forced to respond fast or had to contend with a secondary memory load task. It could be speculated that when people reason about cells, they only need to consider one appropriate knowledge structure, taxonomic relations, whereas when they reason about diseases and infections, they might have to consider both causal and taxonomic relatedness, making the latter a more effortful and complex task. Thus, rather than being due to a fundamental availability advantage for taxonomic knowledge, the apparent advantage may be a by-product of only having to consider one knowledge structure. This would also be supported by the finding that when causal directionality was ignored, no apparent domain differences emerged depending upon the availability of time and mental resources. Rather, when level of association was controlled for, people seemed to use the property they were reasoning about as a guide for selecting relevant knowledge domains.

This may also explain some previous differences between experts and novices in their use of knowledge from different domains. It is feasible that based on their extensive background knowledge and frequent involvement with relevant categories, experts would build stronger associations between various categories than novices. Furthermore, differences in associative strength might also emerge between experts depending upon the focus of their expertise. Consider the experiment carried out by Proffitt et al. (2000), who looked at inductive reasoning patterns amongst three different types of US tree experts, landscapers, taxonomists and maintenance workers. In one of their experiments (Experiment 1), they showed that whereas taxonomists exhibited a diversity effect, this effect was completely absent in maintenance workers. It is possible that such differences found in experts' use of reasoning strategies might be explained by group differences in the degree of association between various tree species. For example, maintenance workers may have a stronger association between trees they encounter in a similar context on a daily basis, even if they are taxonomically diverse. In the long run, this frequent co-occurrence may lead to a stronger association between certain tree species in maintenance workers compared to taxonomists. Maintenance workers may thus be less likely to generalize a property shared by two diverse but nonetheless highly associated tree species to all other tree species compared to taxonomists.

6.2 Executive Functions and Mental Resources

To summarize thus far, our findings suggest that two types of knowledge influence category-based induction, unstructured knowledge mediated by an effortless, automatic mental process, and structured types of knowledge mediated by effortful mental processes which have a longer time course. If knowledge effects in category-based induction depend crucially upon the interaction between easily available unstructured knowledge and more elaborate but less easily available structured knowledge, we can ask more specific questions about the nature of how these two contrasting types of knowledge interact, and what domain-general processes might be implicated. This leads to the next cluster of questions regarding the role of executive functions and mental resources in use of knowledge in category-based reasoning.

Question 3:
Do people have to inhibit one type of knowledge in order for the other to dominate the inference process?

Having established that structured knowledge and unstructured knowledge differ in their processing characteristics we are able to link our research to broader frameworks of reasoning which explore how different processes might interact with one another. Characterizing the nature and interaction of different mental processes has been pursued by dual-process researchers such as Evans and Curtis-Holmes (2005) and De Neys (De Neys, 2006a, 2006b; De Neys & Everaerts, 2008; De Neys & Franssens, 2009; De Neys & Glumicic, 2008; De Neys, et al., 2008), using manipulations like restricting response times, loading working memory as well as drawing on correlational (De Neys, & Everaerts, 2008; Kokis, MacPherson, Toplak, West, & Stanovich, 2002; Stanovich & West, 1998) and neuro-imaging (De Neys, Vartanian & Goel, 2008) techniques. However, despite the range of reasoning domains covered (e.g. conditional reasoning, decision-making, probabilistic reasoning), the stimulus materials all have one feature in common, that is, there is always a clear conflict between a correct normative response and a dichotomous belief-based response. In contrast, responses to our inductive reasoning problems in Experiments 1 to 3 are of a more continuous nature. Thus, it is not possible to assign a response strictly to the output from either a heuristic or an analytical process. Rather, it suggests that the extent to which people's reasoning is determined by processes that are either relatively automatic or effortless, or controlled and effortful is a proportionate rather than a discrete matter (Crisp & Feeney, 2009).

The fact that it is not possible to assign a response strictly to one process or the other elucidates the interactive nature (Elqayam, 2009) of the two purported processes, and fits parsimoniously with Evans's (2006, 2007) default-interventionist architecture. That is, a heuristic and effortless process might deliver an initial response based on unstructured knowledge arising from temporal or spatial contingencies, associations, or frequent co-occurrences. If people then have time and available mental resources, they may be able to recruit more structured theoretical knowledge that may or may not lead to modifications to the response suggested by the heuristic process. This is seen most clearly in the case of the causal asymmetry effect about disease transmission in food chains. When people do not have available time and mental resources, they may rely strongly on the existence of an association between two categories. Both structured and unstructured knowledge might concur on the fact that there is a strong relation between two categories (e.g. between *carrots* and *rabbits*). However, unstructured knowledge stops at the fact that there is a relation. In contrast, structured knowledge might go beyond this to explicitly account for knowledge as to **why and how** they are related (*rabbits* eat *carrots, carrots* don't eat *rabbits*), resulting in a robust causal asymmetry effect that follows the predictions of a rational Bayesian analysis of property distribution (Kemp & Tenenbaum, 2009; Shafto, et al., 2008). This raises questions about exactly how mental resources are utilized, and how these two types of knowledge may interact. One possibility is that people have to exert inhibitory control over unstructured knowledge if they want to reason based on structured knowledge, a question addressed in the following section.

The regression analyses from Experiments 1 to 3 demonstrated that when people have available time and mental resources, structured knowledge may augment unstructured knowledge. However, whilst there is a high probability that unstructured and structured knowledge coincide to deliver the same adaptive response in everyday life, there are instances where actions based on structured knowledge may oppose the intuitive response suggested by unstructured knowledge. For example, upon encountering a dangerous bear, most people's intuitive reaction would be to run away. However, running would trigger a bear's instinctual chase response, most likely resulting in a full-blown attack. In contrast, taking a non-threatening position is more likely to appease the bear. Using such elaborate causal knowledge would then afford the inference that lying flat on the ground is the better course of action. This is a prime example where people have to inhibit unstructured knowledge in favour of more elaborate structured knowledge, resulting in a very different and more adaptive inference.

In Experiments 6 to 8 we looked at how people deal with such a conflict between structured and unstructured knowledge. The results showed that people found it harder to draw on weakly associated but appropriate structured knowledge when they were lured by unstructured associative knowledge even if it was contextually inappropriate. As mentioned, this supports the idea that unstructured knowledge has to be inhibited if inferences are to be based on less available structured knowledge. The fact that this effect was amplified when people were put under cognitive load suggests that the inhibitory process required considerable mental effort. This is consistent with evidence from other reasoning domains, which suggests that inhibiting background knowledge relies on availability of central executive resources such as attention (Barton, Fugelsang, & Smilek, 2009) and inhibitory control. For example, the ability to reason counterfactually, such as imagining an alternative outcome had a preceding event not occurred, or imagining the outcome if a detail inconsistent with one's world knowledge was in fact true, is correlated with 3- and 4-year-old children's performance on inhibitory control tasks (Beck, Riggs, & Gorniak, 2009). Similarly, older children who are higher in inhibitory control perform better on belief-laden reasoning tasks, in which a response based on background beliefs conflicts with the correct logical response (Handley, et al., 2004). Finally, Markovits, Saelen and Forgues (2009) demonstrated that people have to inhibit contradictory background knowledge if it conflicts with the logical demands of a reasoning task and that this ability is related to a measure of semantic inhibitory control (Markovits & Doyon, 2004). Interestingly, the ability to inhibit inappropriate unstructured knowledge in our adult task was strongly related to a measure of high-level semantic inhibitory control, but not with lower-level motor inhibitory control, a matter considered in more detail below.

Question 4:

Does the ability to withhold a response based on unstructured knowledge correlate with general measures of inhibitory control? If so, are there differences depending upon the level at which an instrument measures the construct of inhibition?

Another crucial finding from the final study was that for people who had relevant structured background knowledge, individual differences in the tendency to base their inductions on weakly associated but appropriate structured knowledge were related to semantic inhibitory control (Experiments 6 to 8) and to an index of working memory capacity (Experiment 8). Several researchers (Engle, 2002; Engle, et al., 1995; Lustig, Hasher & Tonev, 2001) posit that working memory reflects the ability to control attention. In this conceptualization, inhibitory control is a sub-process which ensures that conflicting memory representations or response tendencies do not interfere with the current goals of a task or action. The finding from our final study (Experiment 8) that scores on the working memory capacity task did not account for additional variance in people's performance on once semantic inhibitory control was accounted for suggests that it was the inhibitory control sub-process of working memory that played the greatest role in the inductive reasoning task. This coincides with evidence from Barton, Fugelsang and Smilek (2009) who have shown that inhibiting inappropriate background beliefs during a reasoning task requires working memory resources, leading to a decline in performance on a secondary task. Engle's (2010) conceptualization of inhibitory control as being at the interface between attention and memory raises further questions about the nature of the inhibitory process that is crucial in allowing people to draw on structured knowledge.

Whereas we found strong correlations between the semantic inhibitory control task and selective reasoning performance, there was an absence of a correlation between reasoning performance and a low-level motor inhibitory control task. This goes some way to elucidating the nature of the inhibitory control process. Firstly, it appears that inhibitory control is not a unitary construct, and that only certain aspects of inhibitory control are implicated in people's ability to withhold a response based on unstructured knowledge in favour of structured knowledge. Research on the construct of inhibitory control suggests that inhibition-related processes that operate on response selection are dissociable from those that operate on response output (Bissett, et al., 2009). The stop-signal task mainly taps inhibition at the response output stage, whereas our measure of semantic inhibition measured people's ability to rejct an erronous response in order to select the correct response. Whilst there was a strong correlation between inductive reasoning performance and our measure of semantic inhibitory control, there was no correlation with the stop-signal task This strongly suggests that successful selective reasoning requires effortful inhibition at the response selection stage, where there is a potential conflict between memory representations based on unstructured associative knowledge and the need to activate structured knowledge representations. Thus, people may have to inhibit the selection of a response based on unstructured knowledge and select a response based on an analysis of the structural relations between categories. Once they have resolved the potential conflict by selecting an appropriate response, there is no longer any need to exert inhibitory control over the subsequent motor response.

To summarize, we suggest that one effortful process involved in basing inductive inferences on structured rather than on automatic and easily available unstructured knowledge is inhibitory control. This inhibitory control process seems to occur at the semantic level, where there is a conflict between memory representations at the response selection stage. Thus, people may have to resist the lure from a response based on easily available unstructured knowledge in order to select a response based on more elaborate structured knowledge.

6.3 Future Directions

In the final section we discuss how widening the application of the concept of structured and unstructured knowledge can further elucidate the nature of the processes underlying selective category-based inductive reasoning, and help inform and refine theories trying to account for inductive reasoning.

Category-Based Inductive Reasoning Phenomena

An important issue is to extend the notion of different types of knowledge to a wider scope of inductive reasoning phenomena. The thesis focused on one-category premises and conclusions. In future it will be important to extend the notion of different types of knowledge to arguments with multiple premise and conclusion categories. Consider the following multi-premise arguments:

Cows have property X
Horses have property X

Therefore, Bears have property X (Argument 1)

Cows have property X
Rats have property X

Therefore, Bears have property X (Argument 2)

People tend to believe that argument 2, which includes more diverse premise categories than argument 1, provides stronger evidence that the property will be present in the conclusion category. Known as the diversity effect, this finding has been explained by the Feature-Based Induction Model (FBM) (Sloman, 1993) with reference to the extent to which the premise arguments cover the featural space of the conclusion category. In contrast, the Similarity-Coverage Model (SCM) (Osherson et al., 1990) explains the effect in terms of the degree of similarity between the premise and conclusion categories and the extent to which the premise categories are similar to members of the lowest superordinate category which includes both the premise and conclusion categories.

These competing explanations could be reconceptualized in terms of the roles played by structured and unstructured knowledge. Whereas the SCM would assume that people have to apply structured knowledge about taxonomic relationships in order to be able to generate an inclusive superordinate category, the FBM only requires people to draw on unstructured knowledge encoding the degree to which each category activates overlapping features. This leads to some straightforward predictions: if people do indeed have to apply structured knowledge in order to manifest the diversity phenomenon, then the diversity effect should be subject to cognitive processing constraints such as general cognitive ability, and availability of time and working memory resources. In contrast, if the diversity effect is only dependent upon featural overlap, factors related to cognitive resources should not affect the reasoning output.

Experimental findings by Feeney (2007) suggest that in adults the diversity effect might best be explained with reference to the use of structured knowledge: he found that people higher in cognitive ability were more sensitive to diversity than people lower in cognitive ability. However, to provide further support for the contention that the diversity effect depends on the use of structured knowledge, future work would be necessary to show that manipulations of cognitive resources, such as restricting inference time or adding a cognitive burden, leads to a decrease in the rate with which people exhibit the diversity effect (Evans, 2010).

Considering the role played by structured and unstructured types of knowledge would also provide a parsimonious explanation for phenomena that contradict those predicted by the similarity-based models. For example, Medin et al (2003) find that people will often violate the diversity principle and rate an argument with less diverse premise categories (e.g. *sparrows* & *dogs* → *all living things*) as being stronger than an argument with a more diverse set of premise categories if there is a causal relation between the two premise categories (e.g. *sparrows* & *seeds* → *all living things*). Similarly, people may not show the diversity phenomenon if the more diverse premise categories share a salient property (e.g. *penguins* & *polar bears* → *camels*) compared to the less diverse premise categories (e.g. *penguins* & *eagles* → *camels*). These researchers explain such causal non-diversity and non-diversity via property reinforcement effects in terms of relevance, whereby relations (e.g. a salient property, causal link or thematic relation) between the premise categories can supply information that is potentially more informative than similarity or category membership. If such a salient relation is not shared by the conclusion category, it may decrease people's belief in the inductive potency of an argument, even if the premise categories are more diverse.

An alternative way of thinking about this is in terms of structured versus unstructured knowledge. Thus, causally related categories might become more strongly associated because they more frequently co-occur than unrelated categories. Similarly, categories sharing a salient property could be more strongly associated via activation of the shared property in semantic memory, akin to spreading activation mechanisms in connectionist models. To explore this possibility and provider a starting point for further research, we asked 23 participants to rate the strength of association between Medin et al.'s (2003) category pairs, as we had done in the pre-test described in Chapter 2. This showed that Medin et al's (2003) diverse, but thematically related premise pairs were rated as being more strongly associated (M = 5.4, SD = 1.2) than the non-diverse premise category pairs (M = 3.4, SD = 0.8, $t_{(30)}$ = 5.82, p < .0005, effect size d= 2).

In order to show a diversity effect with Medin et al.'s (2003) materials, people would have to inhibit easily available unstructured knowledge, instantiated by the strong association between the premise categories, and dedicate additional resources to activating structured knowledge about taxonomic relations between the premise categories and conclusion. However, given that people only tend to entertain one hypothesis at a time[12], such strong associations between the premise categories may reduce the likelihood that people would see a need to dedicate cognitive resources to assessing diversity. This alleged role of unstructured associative knowledge and structured knowledge in the manifestation of phenomena related to the diversity principle could be tested in future work by asking people to generalize a property that is contrary to the one suggested by the relation between the premise categories. Consider if people were asked to judge which of the two arguments is stronger:

Sparrows have gene X
Seeds have gene X

Therefore, all living things have gene X (Argument 3)

Sparrows have gene X
Dogs have gene X

Therefore, all living things have gene X (Argument 4)

People would have to inhibit the strong association between the premise categories in Argument 3 in order to effortfully draw on structured knowledge about taxonomic hierarchies and hence spot that these premise categories actually provide better coverage of the superordinate category *all living things* than the unrelated and less diverse categories in Argument 4. Furthermore, if we were able to show that any tendency to pick the more diverse

[12] see Evans's (2006) three principles of hypothetical thought

premise in the above example is compromised by manipulations of cognitive resources or time constraints, we would have even more robust evidence that the diversity effect depends upon the inhibition of unstructured knowledge and effortful application of structured knowledge.

Theories of Category-Based Induction

Another important challenge will be to refine and extend existing theories of category-based induction in order to deal with the diverging processing aspects underlying the use of different types of knowledge. Several theories seem to offer a good account of either structured or unstructured knowledge. For example, Sloman's (1993b) Feature-Based Induction model and Roger and McClelland's (2004) Parallel Distributed Processing model can both easily explain why people instantly reject inferences in which the base and conclusion categories are only weakly associated or unrelated. As such categories are unlikely to have many features in common, there will be little common activation between the two categories, resulting in a weak inference. Sloman's (1993b) model has somewhat more difficulty explaining how two very dissimilar but strongly associated categories (e.g. *honey* and *bees*) result in strong inferences. Roger and McClelland's (2004) model fares somewhat better, as one could assume that the two concepts might be simultaneously activated via relational 'nodes' in a spreading activation network.

One limitation of most models that deal with unstructured types of knowledge is that they cannot easily accommodate context-sensitive inferences, that is, they make no predictions with regards to what domain of knowledge people will rely on. However, this is also a weakness of one of the models which includes a structured knowledge element, Osherson et al.'s (1990) Similarity-Coverage model. Thus, whilst it parsimoniously explains inferences that draw on structured taxonomic knowledge, it cannot easily account for patterns of inductions driven by other domain-specific types of knowledge, such as causal relations.

Bayesian frameworks can more easily explain such selectivity effects by assuming that people draw on different theoretical knowledge structures depending upon context (Kemp & Tenenbaum, 2009). They also arguably provide an interesting possibility of integrating both structured and unstructured types of knowledge. For example, it may be that the individual probability components, such as likelihoods and prior probabilities, can instantiate unstructured sources of knowledge, as well as be based on more rigid structured sources of knowledge. Although there is great debate regarding the question where such priors come from and what would constitute reasonable prior assumptions (Oaksford & Chater, 2007, p. 81), one possibility that would fit with our current notion of different types of knowledge is that priors can either be based upon unstructured knowledge acquired through simple statistical principles such as co-occurrences, or be derived from the application of highly structured background knowledge (Gopnik & Tenenbaum, 2007). The ability to integrate prior probabilities and likelihoods in Bayes' theorem in contrast may depend upon structured knowledge about how the different probability components relate to one another. Consider how people update their beliefs in the conditional probability when encountering a counterexample (Chater & Oaksford, 2007; Oaksford & Hahn, 2007). For example, based upon repeated experience and without having to have any specific structured knowledge, people may have a high degree of belief in the conditional that *if x is a bird, then x flies*. However, if they encounter a bird that does not fly, for example a *penguin*, they may apply structured knowledge that birds will only fly if their wings are sufficiently large relative to their bodyweight. This in turn might lead people to revise their initial belief in the conditional downwards to incorporate this counterexample via Bayesian updating.

In this context, an important task would be to pinpoint the mental processes that mediate the Bayesian inferential process. Take for example the case of the causal asymmetry effect, whereby people assign higher inductive strength ratings to predictive compared to diagnostic inferences when they have plenty of time and mental resources available. People are likely to hold beliefs about the prior probability with which a property is transferred between two species A and B. For example, people may believe that the probability with which *carrots* transfer a disease to *rabbits* can be assigned an approximate probability value of .4. In Shafto et al's (2008) model, this would be formalized as the transmission rate parameter. If people were asked to evaluate the probability that *rabbits* have the disease given that *carrots* have it, they might assess the conditional probability, which in this case is simply the probability of the conditional (i.e. the probability with which *carrots* transfer diseases to *rabbits*). However, Shafto et al's (2008) disease transmission model includes a second parameter, a background rate, specifying the probability that the disease was acquired from an alternative source outside the food web. Whilst people often neglect such base rates when reasoning causally, they are less inclined to do so when reasoning diagnostically (Fernbach & Darlow, 2009). Thus, if people are asked to make the reverse diagnostic inference, i.e. to evaluate the probability that *carrots* have a disease given that *rabbits* have it, they might apply their qualitative knowledge that diseases are more likely to be passed on from prey to predator rather than vice versa. This might induce people to think about alternative sources by which rabbits acquired the disease, for example contracting the disease from *other rabbits*, or eating contaminated *grass*. Such counterexamples might then lead to a downward revision in people's beliefs about the conditional probability that *carrots* have the disease given that *rabbits* have it, and hence in the probability of the conditional, explaining the emergence of the causal asymmetry effect. This would be akin to Oaksford, Chater and Larkin's (2000) probabilistic approach to human conditional inference, where the activation of counterexamples explains why valid modus tollens inferences are less frequently endorsed than valid modus ponens inferences.

Another important undertaking will be to pinpoint precisely which aspect of the Bayesian inductive inference process is compromised by burdening mental resources or restricting inference time. One possibility is that certain stages of a Bayesian inference process cannot operate optimally when cognitive resources are sparse. Thus, not only does a person have to activate relevant domain-specific representations, they also have to integrate these with Bayesian computations to arrive at the inductively most potent inference, a process which is likely to take time and considerable processing effort (Copeland & Radvansky, 2004; Straubinger, Cokely, & Stevens, 2009). Regarding the causal asymmetry effect, people's ability to integrate transmission probability with the probability that a disease might have been acquired from an alternative source may be the process that is compromised when people are under time pressure or cognitive load. Thus, when people are burdened by a cognitive load, they may rely exclusively on easily activated unstructured knowledge about co-occurrence, which would be a very useful heuristic guide to transmission probabilities. Consequently, one would not expect asymmetries to arise based on the direction of a potential causal link. In contrast, if time and mental resources are available, people may apply more elaborate structured knowledge about disease transmission mechanisms, leading to a downward revision of inductive strength for diagnostic compared to predictive causal inferences.

Another possibility is that people under heavy cognitive load cannot keep track of the probabilities assigned to different hypotheses. For example, in Kemp and Tenenbaum's (2009) theory-based Bayesian model the probabilities assigned to all possible hypotheses is exhaustive, that is, the probabilities have to add up to 1. However, there is ample evidence suggesting that people frequently fail to be coherent (Ayton, 1999; Pfeifer & Kleiter, 2009; Tversky & Koehler, 1994). For example, even highly-trained physicians are prone to sub-additive probability estimations when there is abundant evidence to support exhaustive hypotheses (Redelmeier, Koehler, Liberman & Tversky, 1995) or superadditive estimations when people have little supporting evidence for two mutually exhaustive hypotheses (Macchi, Krantz & Osherson, 1999). Indirect evidence that the ability to reason coherently requires mental effort comes from Feeney and Crisp (in 2010), who showed that inductive conjunction fallacies increase under cognitive load.

Thus, it will be fruitful for future work to answer the question how the mind might instantiate processes that approximate theory-based Bayesian strategies. By exploring the interaction between types of knowledge and domain-general processing characteristics, it would also be possible to specify which of these Bayesian processes require mental resources. As a result, we would not only be able to understand the output of knowledge-based inductive reasoning, but we would also be able to explain the mental processes by which this feat is accomplished.

References

Aman, C. J., Roberts, R. J., & Pennington, B. F. (1998). A neuropsychological examination of the underlying deficit in attention deficit hyperactivity disorder: Frontal lobe versus right parietal lobe theories. *Developmental Psychology, 34*(5), 956-969.

Anaki, D., & Henik, A. (2003). Is there a "strength effect" in automatic semantic priming? *Memory & Cognition, 31*(2), 262-272.

Atran, S., Medin, D., Lynch, E., Vapnarsky, V., Eka, E. U., & Sousa, P. (2001). Folkbiology doesn't come from folkpsychology: Evidence from Yukatak Maya in cross-cultural perspective. *Journal of Cognition and Culture 1*, 3-42.

Ayton, P. (1997). How to be incoherent and seductive: Bookmakers' odds and support theory. *Organizational Behavior and Human Decision Processes, 72*(1), 99-115.

Bacon, F. (1620/1898). *Novum Organum*. London: George Bell and Sons.

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56*(11), 851-864.

Baddeley, A. D., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and Retrieval from Long-Term Memory. *Journal of Experimental Psychology-General, 113*(4), 518-540.

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: biological categorization and reasoning within and across cultures. *Cognition, 84*(1), 1-53.

Baker, A., & Coley, J. D. (2005). *Taxonomic and ecological relations in open-ended induction*. Paper presented at the 27th Annual Conference of the Cognitive Science Society.

Baraff, L., & Coley, J. D. (2003, Jul 31-Aug 02). *Thinking about music: Novice and expert inductive reasoning.* Paper presented at the 25th Annual Conference of the Cognitive-Science-Society, Boston, MA.

Barraff, E., & Coley, J. D. (2003). *Thinking About Music: Novice and Expert Inductive Reasoning*. Paper presented at the Annual Conference of the Cognitive Science Society.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*(4), 553-573.

Barsalou, L. W. (1982). Context-Independent and Context-Dependent Information in Concepts. *Memory & Cognition, 10*(1), 82-93.

Barton, K., Fugelsang, J., & Smilek, D. (2009). Inhibiting beliefs demands attention. *Thinking & Reasoning, 15*(3), 250-267.

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking & Reasoning, 15*(4), 337-354.

Bissett, P. G., Nee, D. E., & Jonides, J. (2009). Dissociating Interference-Control Processes Between Memory and Response. *Journal of Experimental Psychology-Learning Memory and Cognition, 35*(5), 1306-1316.

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory and Cognition, 8*, 631-642.

Brown, A. S. (1976). Catalog of scaled verbal material. *Memory & Cognition, 4*, 1S-45S.

Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia, 34*(4), 263-272.

Canas, J. J. (1990). Associative Strength Effects in the Lexical Decision Task. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 42*(1), 121-145.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, Mass. ; London: MIT Press.

Chater, N., & Oaksford, M. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Chen, S., Duckworth, K., & Chaiken, S. (1999). Motivated heuristic and systematic processing. *Psychological Inquiry, 10*(1), 44-49.

Church, K. W., & Hanks, P. (1990). *Word Association Norms, Mutual Information, and Lexicography*. Morristown: Assoc Computational Linguistics.

Cobos, P. L., López, F. J., Cano, A., Almaraz, J., & Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology-Animal Behavior Processes, 28*(4), 331-346.

Coley, J. D., & Baker, A. (2004). *Taxonomic and Ecological Relations in Open-Ended Folk Biological Induction.* Paper presented at the Fifth International Conference on Thinking.

Coley, J. D., & Barraff, E. (2003). *Effects of Time Pressure on Expert and Novice Category-Based Induction*. Paper presented at the Annual Meeting of the Psychonomic Society.

Coley, J. D., Hayes, B., Lawson, C., & Moloney, M. (2004). Knowledge, expectations, and inductive reasoning within conceptual hierarchies. *Cognition, 90*(3), 217-253.

Coley, J. D., Medin, D. L., & James, L. B. (1999). *Folk biological induction among Native American children*. Paper presented at the Biennial Meeting of the Society for Research in Child Development.

Coley, J. D., Shafto, P., Stepanova, O., & Barraff, E. (2005). Knowledge and Category-Based Induction. In W. K. Ahn, Goldstone, R. L., Love, B. C., Markman, A. B. and Wolf, P. (Ed.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin.* Washington, D.C.: American Psychological Association.

Collette, F., Germain, S., Hogge, M., & Van der Linden, M. (2009). Inhibitory control of memory in normal ageing: Dissociation between impaired intentional and preserved unintentional processes. *Memory, 17*(1), 104-122.

Collette, F., Van der Linden, M., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2001). The functional anatomy of inhibition processes investigated with the Hayling task. *Neuroimage, 14*(2), 258-267.

Collins, A. M., & Loftus, E. F. (1975). Spreading Activation Theory of Semantic Processing. *Psychological Review, 82*(6), 407-428.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112*(2), 347-382.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Conway, A. R. A., Tuholski, S. W., Shisler, R. J., & Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory & Cognition, 27*(6), 1042-1050.

Copeland, D., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 57*(8), 1437 - 1457.

Cramer, P. (1968). *Word association*. New York: Academic Press.

Crisp, A. K., & Feeney, A. (2009). Causal conjunction fallacies: The roles of causal strength and mental resources. *Quarterly Journal of Experimental Psychology, 62*(12), 2320-2337.

De Deyne, S., & Storms, G. (2008a). Word associations: Network and semantic properties. *Behavior Research Methods, 40*(1), 213-231.

De Deyne, S., & Storms, G. (2008b). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40*(1), 198-205.

De Groot, A. M. B. (1989). Representational Aspects of Word Imageability and Word Frequency as assessed through Word Association. *Journal of Experimental Psychology-Learning Memory and Cognition, 15*(5), 824-845.

De Neys, W. (2006a). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology, 59*(6), 1070-1100.

De Neys, W. (2006b). Dual processing in reasoning - Two systems but one reasoner. *Psychological Science, 17*(5), 428-433.

De Neys, W. (2009). Beyond response output: More logical than we think. *Behavioral and Brain Sciences, 32*, 87-88.

De Neys, W. (in press). Counterexample retrieval and inhibition during conditional reasoning: Direct evidence from memory probing. In M. Oaksford (Ed.), *The Psychology of conditionals*. Oxford: Oxford University Press.

De Neys, W., & Everaerts, D. (2008). Developmental trends in everyday conditional reasoning: The retrieval and inhibition interplay. *Journal of Experimental Child Psychology, 100*(4), 252-263.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*(1), 45-61.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*(3), 1248-1299.

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the lifespan: the rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science, 12*(1), 123-130.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science, 19*(5), 483-489.

Dickinson, A. (2001). Causal learning: Association versus computation. *Current Directions in Psychological Science, 10*(4), 127-132.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4*(1), 3-23.

Elman, J. L. (1996). *Rethinking innateness : a connectionist perspective on development*. Cambridge, Mass. ; London: MIT.

Elqayam, S. (2009). Models of dependence and independence: A two-dimensional architecture of dual processing. *Thinking & Reasoning, 15*(4), 377-387.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23.

Engle, R. W. (2010). Role of Working-Memory Capacity in Cognitive Control. *Current Anthropology, 51*(1).

Engle, R. W., Conway, A. R. A., Tuholski, S. W., & Shisler, R. J. (1995). A Resource Account of Inhibition. *Psychological Science, 6*(2), 122-125.

Epstein, S., Pacini, R., DenesRaj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71*(2), 390-405.

Estes, Z. (2003). A tale of two similarities: comparison and integration in conceptual combination. *Cognitive Science, 27*(6), 911-921.

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*(4), 321-339.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J. S. B. T. (2010). On the Nature of Inductive Inference. *The American Journal of Psychology, 123*(1), 119-122.

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*(4), 382-389.

Evans, J. S. B. T., & Frankish, K. (Eds.). (2009). *In two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2008). Understanding causal conditionals: A study of individual differences. *Quarterly Journal of Experimental Psychology, 61*(9), 1291-1297.

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology.

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *Neuroimage, 18*(1), 42-57.

Feeney, A. (2007). How many processes underline category-based induction? Effect of conclusion specificity and cognitive ability. *Memory & Cognition, 35*, 1830-1839.

Feeney, A. and Crisp, A.K. (2010). A conjunction of fallacies: What different types of causal conjunction error reveal about dual processes for thinking. In *The Science of Reason: A Festschrift for Johnathan St. B.E. Evans.* Manktelow, K., Over, D. and Elquayam, S. (eds). Psychology Press: Hove, UK.

Feeney, A., Coley, J., & Crisp, A. K. (2010). The Relevance Framework for Category-Based Induction: Evidence from Garden-Path Arguments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(4), 906-919.

Feeney, A., Crisp, A. K., & Wilburn, C. J. (2008). Inductive reasoning and semantic cognition: More than just different names for the same thing? *Behavioral and Brain Sciences, 31*(6), 715.

Fenker, D. B., Waldmann, M. R., & Holyoak, K. J. (2005). Accessing causal relations in semantic memory. *Memory & Cognition, 33*(6), 1036-1046.

Fernbach, P. M., & Darlow, A. (2009). *Causal Asymmetry in Inductive Judgement.* Paper presented at the Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society, Amsterdam, Netherlands.

Fisher, A. V., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development, 76*(3), 583-597.

Foss, D. J., & Ross, R. J. (1983). Great expectations: Context effects during sentence processing. In G. D. Flores d'Arcais & R. J. Jarvella (Eds.). *The process of language understanding* (pp. 169-191). Chichester: Wiley.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3-to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology-General*, *133*(3), 382-397.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology-General*, *133*(1), 101-135.

Gelman, S. A., & Markman, E. M. (1986). Categories and Induction in Young Children. *Cognition*, *23*(3), 183-209.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working Memory and Strategies in Syllogistic Reasoning Tasks *Memory & Cognition, 21*(1), 115-124.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*(1), B11-B22.

Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science, 10*(3), 281-287.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*(4), 334-384.

Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64*(2), 159-174.

Hadley, R. F. (1999). Connectionism and novel combinations of skills: Implications for cognitive architecture. *Minds and Machines, 9*(2), 197-221.

Hadley, R. F. (2000). Cognition and the computational power of connectionist networks. *Connection Science*, *12*(2), 95-110.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87*(1), 1-32.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning, 10*(2), 175-195.

Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster & C. J. Brainerd (Eds.). *Interference and inhibition in cognition* (pp. 175-204). San Diego, CA: Academic Press.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and Performance Xvii - Cognitive Regulation of Performance: Interaction of Theor and Application* (Vol. 17, pp. 653-675). Cambridge: M I T Press.

Hatano, G., & Inagaki, K. (1997). Qualitative changes in intuitive biology. *European Journal of Psychology of Education, 12*(2), 111-130.

Hatano, G., & Inagaki, K. (2000). Domain-specific constraints of conceptual development. *International Journal of Behavioral Development*, *24*(3), 267-275.

Hayes, B. K., & Thompson, S. P. (2007). Causal relations and feature similarity in children's inductive reasoning. *Journal of Experimental Psychology-General, 136*(3), 470-484.

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition, 28*(3), 376-385.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7*(4), 569-592.

Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin & Review, 12*(2), 340-344.

Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology, 43*(4), 243-273.

Heit, E., Hahn, U. & Feeney, A (2005). Defending diversity. In W. Ahn, Goldstone, B.C., Love, A.B. & Wolff, P. (Eds.), *Categorization inside and outside of the lab: Festschrift in Honor of Douglas L. Medin* (pp. 87-100). Washington, DC: American Psychological Association.

Heit, E., & Rubinstein, J. (1994). Similarity and Property Effects in Inductive Reasoning *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(2), 411-422.

Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, New Jersey: Prentice Hall.

Heylighen, F. (2001). Mining Associative Meanings from the Web: from Word Disambiguation to the Global Brain: Standaard Publishers.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network- investigations of acquired dyslexia. *Psychological Review, 98*(1), 74-95.

Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. *Cognition, 113*(1), 62-79.

Howson, C. (2000). *Hume's problem: induction and the justification of belie.* Oxford: Clarendon.

Ide, N., & Veronis, J. (1998). Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics, 24*(1), 1-40.

Inagaki, K., & Hatano, G. (1993). Young Children's Understanding of the Mind-Body Distinction. *Child Development, 64*(5), 1534-1549.

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development, 67*(6), 2823-2840.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science, 5*(2), 219-232.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267-294). Cambridge: Cambridge University Press.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology-Learning Memory and Cognition, 26*(2), 336-358.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637-671.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology-General, 132*(1), 47-70.

Keil, F. C. (2003). Folkscience: coarse interpretations of a complex reality. *Trends in Cognitive Sciences, 7*(8), 368-373.

Keil, F. C., Levin, D. T., Richman, B. A., & Gutheil, G. (1999). Mechanism and explanation in the development of biological thought: The case of disease. In D. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233-284). Cambridge, MA: MIT Press.

Kemp, C., & Tenenbaum, J. B. (2003). *Theory-based induction.* Paper presented at the Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society, Pts 1 and 2.

Kemp, C., & Tenenbaum, J. B. (2009). Structured Statistical Models of Inductive Reasoning. *Psychological Review, 116*(1), 20-58.

Kemp, C., Tenenbaum, J. B., Niyogi, S., & Griffiths, T. L. (2010). A probabilistic model of theory formation. *Cognition, 114*(2), 165-196.

Klaczynski, P. A. (2004). A dual-process model of adolescent development: Implications for decision making, reasoning, and identity. *Advances in Child Development and Behavior, 32*, 73-123.

Klaczynski, P. A., & Lavallee, K. L. (2005). Domain-specific identity, epistemic regulation, and intellectual ability as predictors of belief-biased reasoning: A dual-process perspective. *Journal of Experimental Child Psychology, 92*(1), 1-24.

Knobil, M. (2003). *Consumer Superbrands 2003: An Insight into 70 of Britain's Strongest Brands: An Insight into 100 of Britain's Strongest Brands.* Superbrands Ltd.

Kokis, J. V., MacPherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83*(1), 26-52.

Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and Inhibition- Beyond a Unitary View of Inhibitory Processing in Attention. *Psychology and Aging, 9*(4), 491-512.

Kucera, H., & Francis, W. N. (1967). *Computational Analysis of Present Day American English*. Providence: Brown University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology-General, 130*(1), 3-28.

Logan, G. D. (1994). On the ability to inhibit thought and action: A user's guide to the stop signal paradigm. In D. Dagenbach & D. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189-239).

Lopez, A. (1995). The Diversity Principle in the Testing of Arguments. *Memory & Cognition, 23*(3), 374-382.

Lopez, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*(3), 251-295.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The Development of Category-Based Induction. *Child Development, 63*(5), 1070-1090.

López, F. J., Cobos, P. L., & Cano, A. (2005). Associative and causal reasoning accounts of causal induction: Symmetries and asymmetries in predictive and diagnostic inferences. *Memory & Cognition, 33*(8), 1388-1398.

Lupker, S. J. (1984). Semantic Priming without Association- A Second Look *Journal of Verbal Learning and Verbal Behavior, 23*(6), 709-733.

Lustig, C., Hasher, L., & Tonev, S. T. (2001). Inhibitory control over the present and the past. *European Journal of Cognitive Psychology, 13*(1-2), 107-122.

Macchi, L., Osherson, D., & Krantz, D. H. (1999). A note on superadditive probability judgment. *Psychological Review, 106*(1), 210-214.

Macleod, C. M. (1991). Half a Century of Research on the Stroop Effect- An Integrative Review. *Psychological Bulletin, 109*(2), 163-203.

Macrae, C. N., Bodenhausen, G. V., Schloerscheidt, A. M., & Milne, A. B. (1999). Tales of the unexpected: Executive function and person perception. *Journal of Personality and Social Psychology, 76*(2), 200-213.

Markman, A. B., & Gentner, D. (1993). Structural Alignment during Similarity Comparisons. *Cognitive Psychology, 25*(4), 431-467.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24*(2), 235-249.

Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory & Cognition, 32*(4), 592-601.

Markovits, H., Saelen, C., & Forgues, H. L. (2009). An inverse belief-bias effect: more evidence for the role of inhibitory processes in logical reasoning. *Experimental Psychology, 56*(2), 112-120.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The Relationship Between Working Memory Capacity and Executive Functioning: Evidence for a Common Executive Attention Construct. *Neuropsychology, 24*(2), 222-243.

McEvoy, C. L., & Nelson, D. L. (1982). Category Name and Instance Norms for 106 Categories of Various Sizes. *American Journal of Psychology, 95*(4), 581-634.

McNamara, T. P. (1992a). Priming and Constraints it places on Theories of Memory and Retrieval. *Psychological Review, 99*(4), 650-662.

McNamara, T. P. (1992b). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(6), 1173-1190.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*(3), 517-532.

Medin, D. L., & Waxman, S. (2007). Interpreting asymmetries of projection in children's inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive Reasoning* (pp. 55-80). Cambridge: Cambridge University Press.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology, 41*(1), 49-100.

Morton, J. B., & Munakata, Y. (2005). What's the difference? Contrasting modular and neural network approaches to understanding developmental variability. *Journal of Developmental and Behavioral Pediatrics, 26*(2), 128-139.

Moscovitch, M. (1995). Models of consciousness and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1341-1356). Cambridge: MIT Press.

Moss, H. E., & Older, L. (1996). *Birkbeck word association norms*. Hove: Psychology Press.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing Different Types of Lexical Semantic Information- Evidence from Priming. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*(4), 863-883.

Myrvold, W. C. (1996). Bayesianism and diverse evidence: A reply to Andrew Wayne. *Philosophy of Science, 63*(4), 661-665.

Nagel, E. (1939). *Principles of the theory of probability*. Chicago: University of Chicago Press.

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognitive Affective & Behavioral Neuroscience, 7*(1), 1-17.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. Hillsdale: New Jersey: Erlbaum.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition, 28*(6), 887-899.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. From http://www.usf.edu/FreeAssociation

Nelson, D. L., & Schreiber, T. A. (1992). Word Concreteness And Word Structure as Independent Determinants of Recall. *Journal of Memory and Language, 31*(2), 237-260.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126*(2), 220-246.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The Use of Statistical Heuristics in Everyday Inductive Reasoning. *Psychological Review, 90*(4), 339-363.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*(8), 349-357.

Oaksford, M., & Chater, N. (2009). The uncertain reasoner: Bayes, logic, and rationality. *Behavioral and Brain Sciences, 32*(1), 105-120.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology-Learning Memory and Cognition, 26*(4), 883-899.

Oaksford, M., & Hahn, U. (2007). Induction, Deduction, and Argument Strength in Human Reasoning and Argumentation. In A. Feeney & E. Heit (Eds.), *Inductive Reasoning* (pp. 269-301). Cambridge: Cambridge University Press.

Opfer, J. E., & Bulloch, M. J. (2007). Causal relations drive young children's induction, naming, and categorization. *Cognition, 105*(1), 206-217.

Opfer, J. E., & Doumas, L. A. A. (2008). Analogy and conceptual change in childhood. *Behavioral and Brain Sciences, 31*(6), 723-+.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185-200.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*(6), 972-987.

Palermo, D. S., & Jenkins, J. J. (1964). *Word association norms: Fourth grade through college*. Minneapolis: University of Minnesota Press.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976-987.

Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge, U.K. ; New York: Cambridge University Press.

Pfeifer, N., & Kleiter, G. D. (2009). Mental probability logic. *Behavioral and Brain Sciences, 32*(1), 98.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology-Learning Memory and Cognition, 26*(4), 811-828.

Rabbitt, P. (1997). Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1-38). Hove, UK: Psychology Press.

Rapp, R., & Wettler, M. (1991). *Prediction of Free Word Associations based on Hebbian Learning.* Paper presented at the Proceedings of the International Joint Conference on Neural Networks, Singapore.

Redelmeier, D. A., Koehler, D. J., Liberman, V., & Tversky, A. (1995). Probability Judgment in Medicine- Discounting Unspecified Possibilities. *Medical Decision Making, 15*(3), 227-230.

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition, 34*(1), 3-16.

Rehder, B. (2009). Causal-Based Property Generalization. *Cognitive Science, 33*(3), 301-344.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*(3), 264-314.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology-General, 130*(3), 323-360.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 665-681.

Roberts, M. J., & Newton, E. J. (2001). Inspection times, the change task, and the rapid-response selection task. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 54*(4), 1031-1048.

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

Rogers, T. T., & McClelland, J. L. (2008a). Precis of Semantic Cognition: A Parallel Distributed Processing Approach. *Behavioral and Brain Sciences, 31*(6), 689-+.

Rogers, T. T., & McClelland, J. L. (2008b). A simple model from a powerful framework that spans levels of analysis. *Behavioral and Brain Sciences, 31*(6), 729-749.

Rosen, E., & Russell, W. A. (1957). Frequency Characteristics of Successive Word Association. *American Journal of Psychology, 70*(1), 120-122.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology-General, 126*(3), 211-227.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*(4), 495-553.

Rotello, C. M., & Heit, E. (2009). Modeling the Effects of Argument Length and Validity on Inductive and Deductive Reasoning. *Journal of Experimental Psychology-Learning Memory and Cognition, 35*(5), 1317-1330.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science, 26*(5), 521-562.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., et al. (2001). Mapping motor inhibition: Conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage, 13*(2), 250-261.

Rubia, K., Smith, A., Lidzba, K., Toone, B., Simmons, A., Williams, S. C. R., et al. (2001). Neural substrates of successful versus unsuccessful stopping in a cognitively challenging event related stop task. *Neuroimage, 13*(6), S351-S351.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing : explorations in the microstructure of cognition*. Cambridge, Mass: MIT Press.

Salthouse, T. A., & Meinz, E. J. (1995). Aging, Inhibition, Working memory, and Speed. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 50*(6), 297-306.

Satpute, A. B., Fenker, D. B., Waldmann, M. R., Tabibnia, G., Holyoak, K. J., & Lieberman, M. D. (2005). An fMRI study of causal judgments. *European Journal of Neuroscience, 22*(5), 1233-1238.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Prelexical and Postlexical Loci of Contextual Effects on Word Recognition. *Memory & Cognition, 12*(4), 315-328.

Shafto, P., & Coley, J. D. (2003a). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology-Learning Memory and Cognition, 29*(4), 641-649.

Shafto, P., & Coley, J. D. (2003b). Development of Categorization and Reasoning in the Natural World: Novices to Experts, Naive Similarity to Ecological Knowledge. *Journal of Experimental Psychology / Learning, Memory & Cognition, 29*(4), 641-649.

Shafto, P., Coley, J. D., & Baldwin, D. (2007). Effects of time pressure on context-sensitive property induction. *Psychonomic Bulletin & Review, 14*(5), 890-894.

Shafto, P., Kemp, C., Bonawitz, E. B., Coley, J. D., & Tenenbaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition, 109*(2), 175-192.

Shafto, P., Vitkin, A. Z., & Coley, J. D. (2007). Availability in category-based induction. In A. Feeney & E. Heit (Eds.), *Induction*. Cambridge: Cambridge University Press.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology, 60*(3), 291-309.

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology-Animal Behavior Processes, 24*(4), 405-415.

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: an investigation of the construct validity of inhibition in older adults. *Neuropsychologia, 40*(6), 605-619.

Simons, D. J., & Keil, F. C. (1995). An Abstract to Concrete Shift in the Development of Biological Thought- The Insides Story. *Cognition, 56*(2), 129-163.

Sloman, S. A. (1993a). Do Simple Associations Lead to Systematic Reasoning. *Behavioral and Brain Sciences, 16*(3), 471-472.

Sloman, S. A. (1993b). Feature-Based Induction. *Cognitive Psychology, 25*(2), 231-280.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35*(1), 1-33.

Sloman, S. A. (2005). *Causal Models: How People Think about the World and Its Alternatives*. New York: Oxford University Press.

Sloutsky, V. M., & Fisher, A. V. (2004a). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology-General, 133*(2), 166-188.

Sloutsky, V. M., & Fisher, A. V. (2004b). When development and learning decrease memory - Evidence against category-based induction in children. *Psychological Science, 15*(8), 553-558.

Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development, 79*(3), 639-651.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*(2), 179-185.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108-131.

Spellman, B. A., Holyoak, K. J., & Morrison, R. G. (2001). Analogical priming via semantic relations. *Memory & Cognition, 29*, 383-393.

Spence, D. P., & Owens, K. C. (1990). Lexical Co-occurrence and Association Strength. *Journal of Psycholinguistic Research, 19*(5), 317-330.

Sperber, D., & Wilson, D. E. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford, UK: Blackwell.

Springer, K. (1992). Children's Awareness of the Biological Implications of Kinship. *Child Development, 63*(4), 950-959.

Springer, K., & Keil, F. C. (1989). On The Development of Biologically Specific Beliefs - The Case of Inheritance. *Child Development, 60*(3), 637-648.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology-General, 127*(2), 161-188.

Steel, D. (1996). Bayesianism and the value of diverse evidence. *Philosophy of Science, 63*(4), 666-674.

Stepanova, O., & Coley, J. D. (2003). *Animals And Alcohol: The Role of Experience in Inductive Reasoning Among College Students*. Paper presented at the Annual Meeting of the Psychonomic Society.

Straubinger, N., Cokely, E. T., & Stevens, J. R. (2009). The dynamics of development: Challenges for Bayesian rationality. *Behavioral and Brain Sciences, 32*(1), 103.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. *Advances in Neural Information Processing Systems 13, 13*, 59-65.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*(7), 309-318.

Tenenbaum, J. B., Kemp, C., & Shafto, P. (2007). Theory-based Bayesian models of inductive reasoning. In A. Feeney & E. Heit (Eds.), *Induction*. Cambridge: Cambridge University Press.

Thompson, V. (2009). Dual process-theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond* (pp. 171-196). Oxford: Oxford University Press.

Toates, F. (2006). A model of the hierarchy of behaviour, cognition, and consciousness. *Consciousness and Cognition, 15*(1), 75-118.

Tulving, E., & Donaldson, W. (1972). *Organization and memory*. New York: Academic Press.

Tversky, A., & Koehler, D. J. (1994). Support Theory- A Nonexistential Representation of Subjective Probability. *Psychological Review, 101*(4), 547-567.

Tversky, A. (1977). Features of Similarity. *Psychological Review, 84*(4), 327-352.

Unsworth, N., Heitz, R. R., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning, 11*(3), 239-278.

Vitkin, A. Z., Coley, J. D., & Hu, R. (2005). *Children's Use of Relevance in Open-Ended Induction in the Domain of Biology.* Paper presented at the 27th Annual Conference of the Cognitive Science Society.

Vitkin, A. Z., Coley, J. D., & Kane, R. (2005). *Salience of Taxonomic and Ecological Relations in Children's Biological Categorization*. Paper presented at the Biennial Meetings of the Society for Research in Child Development.

von Hippel, W., Silver, L. A., & Lynch, M. E. (2000). Stereotyping Against Your Will: The Role of Inhibitory Ability in Stereotyping and Prejudice among the Elderly. *Pers Soc Psychol Bull, 26*(5), 523-532.

Wager, T. D., Sylvester, C. Y. C., Lacey, S. C., Nee, D. E., Franklin, M., & Jonides, J. (2005). Common and unique components of response inhibition revealed by fMRI. *Neuroimage, 27*(2), 323-340.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology*. Harmondsworth: Penguin Books.

Wayne, A. (1995). Bayesianism and diverse evidence. *Philosophy of Science, 62*, 111-121.

West, R. (2003). Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. *Neuropsychologia, 41*(8), 1122-1135.

West, R., Jakubek, K., Wymbs, N., Perry, M., & Moore, K. (2005). Neural correlates of conflict processing. *Experimental Brain Research, 167*(1), 38-48.

Wisniewski, E. J., & Bassok, M. (1999). What Makes a Man Similar to a Tie? Stimulus Compatibility with Comparison and Integration. *Cognitive Psychology, 39*(3-4), 208-238.

Appendices

Appendix 1A: Causal Strength Ratings

Category Pairs	Mean Causal Strength	Std. Deviation	One-sample t-test *p*- value
Acorn & Cat	1.1	0.3	0.343
Acorn & Squirrel	6.1	2.2	0.000
Butterfly & Swallow	2.8	1.9	0.016
Cabbage & Hedgehog	1.9	1.3	0.054
Cabbage & Snail	5.0	2.9	0.002
Carrot & Fox	2.7	1.3	0.003
Carrot & Rabbit	7.4	1.6	0.000
Flour & Bread	7.3	1.7	0.000
Flower & Butterfly	6.3	2.4	0.000
Flower & Swallow	2.9	1.7	0.007
Fly & Frog	5.7	3.4	0.002
Fly & Heron	3.3	2.7	0.023
Frog & Heron	3.4	2.3	0.010
Grass & Sheep	7.6	1.5	0.000
Grass & Sweater	1.9	1.5	0.095
Grasshopper & Snake	4.9	1.7	0.000
Grasshopper & Toad	4.7	1.8	0.000
Krill & Orca	5.9	2.3	0.000
Krill & Penguin	5.7	1.4	0.000
Lead & Pipe	6.2	1.6	0.000
Lead & Plumber	5.3	2.4	0.000

Meerkat & Eagle	4.4	1.6	0.000
Mercury & Fisherman	5.2	3.6	0.005
Mercury & Tuna Fish	4.1	3.0	0.010
Mouse & Buzzard	4.5	2.3	0.001
Paper & Book	7.9	1.5	0.000
Penguin & Orca	5.2	2.7	0.001
Pesticide & Frog	3.6	1.6	0.001
Pesticide & Snail	4.6	2.0	0.000
Pipe & Plumber	6.4	3.0	0.000
Rabbit & Fox	6.2	1.5	0.000
Salmon & Grizzly Bear	6.1	3.0	0.000
Saucepan & Cook	7.4	1.9	0.000
Scorpion & Eagle	3.4	2.0	0.004
Scorpion & Meerkat	5.0	2.9	0.002
Sheep & Sweater	5.6	1.6	0.000
Shrimp & Grizzly Bear	2.1	1.6	0.057
Shrimp & Salmon	6.1	2.5	0.000
Snail & Frog	4.0	1.6	0.000
Snail & Hedgehog	3.4	2.1	0.005
Squirrel & Cat	2.9	1.7	0.006
Steel & Cook	6.9	1.9	0.000
Steel & Saucepan	4.8	1.5	0.000
Toad & Snake	4.5	2.8	0.003
Tree & Book	6.3	2.2	0.000
Tree & Paper	8.2	1.3	0.000
Tuna Fish & Fisherman	6.8	1.3	0.000
Wheat & Bread	8.5	1.0	0.000
Wheat & Buzzard	2.6	1.8	0.019
Wheat & Flour	8.0	1.3	0.000
Wheat & Mouse	2.1	1.4	0.040

Appendix 1B: Mean Subjective Association Ratings and Co-Occurrence Z-Scores

Category Pair	Mean Association Rating	Exalead Co-Occurrence Z-Score	Google Co-Occurrence Z-Score
Acorn & Cat	1.33	-.24	-.41
Acorn & Coconut	5.46	-.25	-.43
Acorn & Lychee	3.86	-.26	-.44
Acorn & Pecan	5.78	-.09	-.38
Acorn & Squirrel	6.06	-.24	-.26
Acorn & Walnut	6.33	-.22	-.27
Ant & Aardvark	5.72	-.25	-.02
Ant & Cockroach	5.39	-.25	-.01
Ant & Praying Mantis	4.33	-.23	-.35
Antelope & Crocodile	3.76	-.25	-.35
Antelope & Goat	5.59	-.24	-.21
Antelope & Leopard	4.83	-.25	-.35
Book & Newspaper	6.33	-.25	-.18
Book & Radio	4.33	-.24	-.34
Book & Television	5.17	-.18	-.28
Book & World Wide Web	4.94	-.25	-.13
Bread & Polenta	4.67	-.23	-.22
Bread & Rice Cracker	5.28	-.26	.39
Bread & Tortilla	6.44	-.18	.18
Butterfly & Ant	4.67	-.24	-.39
Butterfly & Bee	6.17	-.24	-.06
Butterfly & Beetle	4.89	-.25	-.33
Butterfly & Moth	7.61	-.21	.13
Butterfly & Silverfish	2.11	-.26	-.41
Butterfly & Swallow	4.72	-.25	-.40
Buzzard & Penguin	3.22	-.26	-.43

Buzzard & Seagull	5.72	-.26	.35
Buzzard & Vulture	7.39	-.23	-.09
Cabbage & Cauliflower	7.33	-.06	3.52
Cabbage & Hedgehog	2.39	-.26	-.44
Cabbage & Horseradish	4.72	-.23	-.19
Cabbage & Rapeseed	3.72	-.25	-.40
Cabbage & Snail	4.11	-.25	-.43
Cabbage & Turnip	6.67	-.18	.30
Carrot & Fennel	4.65	-.21	-.06
Carrot & Fox	1.50	-.26	-.43
Carrot & Parsley	5.20	-.21	-.27
Carrot & Potato	7.28	-.07	.31
Carrot & Rabbit	6.06	-.20	-.05
Carrot & Radish	6.39	-.18	.36
Carrot & Rapeseed	1.61	-.23	-.42
Cat & Badger	3.50	-.25	-.38
Cat & Cheetah	7.44	-.22	-.13
Cat & Ferret	4.35	-.20	.88
Cat & Lynx	7.56	-.22	-.24
Cat & Puma	7.28	-.20	.94
Cook & Builder	4.39	-.25	-.43
Cook & Hairdresser	4.29	-.25	-.31
Cook & Shop Assistant	3.94	-.26	-.41
Eagle & Buzzard	7.11	-.20	.63
Eagle & Canary	5.50	-.26	-.43
Eagle & Penguin	3.80	-.25	-.41
Feathers & Fur	5.88	-.24	.15
Feathers & Leather	3.94	-.25	-.23
Feathers & Pillow	6.61	-.25	-.23
Fisherman & Captain	5.83	-.24	-.37

Fisherman & Farmer	5.71	-.23	-.31
Fisherman & Waitress	3.33	-.26	-.43
Flour & Bread	8.11	-.21	.90
Flour & Ground Nuts	4.47	-.26	-.03
Flour & Sugar	5.39	-.09	2.78
Flower & Bamboo	4.09	-.15	-.32
Flower & Butterfly	6.39	-.13	.48
Flower & Conifer	5.11	-.25	-.40
Flower & Fern	4.82	-.25	-.27
Flower & Moss	4.61	-.24	-.39
Flower & Seaweed	3.91	-.26	-.38
Flower & Swallow	1.78	-.25	-.42
Fly & Ant	5.17	-.22	-.37
Fly & Beetle	5.44	-.25	-.29
Fly & Cockroach	5.33	-.21	-.11
Fly & Earwig	4.80	-.22	-.12
Fly & Frog	4.83	-.20	-.33
Fly & Heron	2.89	-.25	-.40
Fly & Locust	5.94	-.25	-.41
Fly & Mosquito	6.61	-.23	-.05
Fly & Spider	6.17	-.15	-.28
Fly & Swallow	3.67	-.25	-.40
Fly & Waterbug	5.61	-.13	11.37
Fox & Black Bear	3.83	-.26	-.36
Fox & Dingo	5.65	-.25	-.39
Fox & Dog	6.06	-.26	-.25
Fox & Jackal	6.56	-.23	.01
Fox & Kangaroo	1.89	-.24	-.36
Fox & Wolf	7.22	-.21	-.24
Frog & Heron	4.29	-.24	-.40

Frog & Newt	6.65	-.23	-.09
Frog & Salamander	5.83	-.25	.09
Goose & Feathers	7.29	-.25	-.11
Goose & Finch	4.65	-.25	-.41
Goose & Kingfisher	3.17	-.26	-.36
Goose & Pillow	4.56	-.23	-.12
Grass & Bamboo	6.67	-.20	-.25
Grass & Dandelion	6.24	-.24	-.25
Grass & Eucalyptus	5.61	-.24	-.36
Grass & Maize	5.94	-.23	-.24
Grass & Palm Tree	4.68	-.25	-.37
Grass & Poison Ivy	5.75	-.26	-.37
Grass & Sheep	6.94	-.25	-.30
Grass & Shepherd	4.56	-.25	-.40
Grasshopper & Lice	3.72	-.26	-.43
Grasshopper & Locust	6.83	-.25	-.18
Grasshopper & Snake	3.17	-.25	-.40
Grasshopper & Termite	5.11	-.23	-.42
Grasshopper & Toad	3.78	-.25	-.42
Grizzly Bear & Panda	5.83	-.16	-.32
Grizzly Bear & Polar Bear	7.83	-.24	.13
Grizzly Bear & Weasel	2.83	-.26	-.43
Grizzly Bear & Wolf	4.28	-.22	-.21
Hedgehog & Mole	5.28	-.26	-.40
Hedgehog & Shrew	4.33	-.25	-.36
Hedgehog & Sloth	3.33	-.26	-.43
Heron & Condor	4.67	-.25	-.40
Heron & Dove	4.76	-.19	-.39
Heron & Hawk	5.06	-.21	-.25
Heron & Parrot	5.15	-.26	-.42

Heron & Penguin	4.00	-.26	-.42
Heron & Stork	7.28	-.23	-.13
Krill & Barnacle	5.33	-.25	-.43
Krill & Crayfish	5.94	-.19	-.37
Krill & Orca Whale	4.61	-.26	-.43
Krill & Penguin	3.56	-.25	-.36
Krill & Woodlice	2.83	-.26	-.34
Lead & Cadmium	6.59	.27	4.26
Lead & Gold	6.76	-.20	-.27
Lead & Mercury	6.72	-.19	.88
Lead & Pipe	5.94	-.20	-.33
Lead & Plumber	4.22	-.14	-.42
Meerkat & Badger	4.39	-.25	-.41
Meerkat & Eagle	3.41	-.26	-.42
Meerkat & Lizard	2.61	-.26	-.39
Meerkat & Mouse	3.83	-.25	-.33
Meerkat & Polar Bear	3.27	-.26	-.41
Meerkat & Skunk	4.56	-.26	-.42
Meerkat & Weasel	5.89	-.26	-.42
Mercury & Cadmium	6.94	-.16	2.28
Mercury & Fisherman	2.35	-.26	-.43
Mercury & Silver	6.39	-.23	-.25
Mercury & Tuna Fish	2.72	-.23	-.21
Mouse & Buzzard	3.78	-.25	-.42
Mouse & Cat	6.33	-.25	.05
Mouse & Gerbil	6.88	-.13	.45
Mouse & Goat	2.71	-.19	-.33
Mouse & Hamster	7.18	.00	-.23
Mouse & Hawk	5.22	-.24	-.41
Mouse & Porcupine	3.72	-.25	-.35

Mouse & Porpoise	3.04	-.25	-.42
Mouse & Squirrel	4.41	-.25	-.26
Orca Whale & Cod	5.28	-.25	-.43
Orca Whale & Cow	2.24	-.25	-.43
Orca Whale & Dolphin	7.29	-.17	.76
Orca Whale & Hippopotamus	3.00	-.26	-.43
Orca Whale & Seal	6.28	-.24	-.28
Paper & Book	7.72	-.15	-.12
Paper & Papyrus	6.06	-.21	.23
Paper & Plastic	4.06	-.13	.57
Penguin & Buzzard	3.71	-.26	-.43
Penguin & Chicken	3.11	-.25	-.41
Penguin & Duck	4.29	-.26	-.38
Penguin & Orca Whale	4.83	-.25	-.36
Pesticide & Fertilizer	5.88	-.14	.74
Pesticide & Frog	1.88	-.25	-.43
Pesticide & Ink	1.89	-.25	-.43
Pesticide & Perfume	3.00	-.25	-.43
Pesticide & Snail	3.28	-.26	-.43
Pillow & Sheet	7.72	-.15	.01
Pillow & Tablecloth	4.00	-.25	-.29
Pipe & Cable	5.67	-.21	-.24
Pipe & Plumber	7.39	.00	-.11
Pipe & Wire	4.44	-.17	-.24
Plumber & Electrician	6.65	.08	1.44
Plumber & Gardener	5.12	-.21	-.38
Plumber & Pilot	2.89	-.26	-.41
Rabbit & Beaver	5.11	-.25	-.37
Rabbit & Camel	2.88	-.24	-.41
Rabbit & Dolphin	2.86	-.22	-.37

Rabbit & Falcon	4.06	-.26	-.43
Rabbit & Fox	6.00	-.18	-.27
Rabbit & Hare	8.28	-.20	-.09
Rabbit & Mouse	4.78	-.25	-.20
Rabbit & Porcupine	3.61	-.25	-.27
Rabbit & Rat	3.89	-.19	.04
Robin & Rat	2.22	-.23	-.42
Salmon & Eel	4.82	-.25	.04
Salmon & Frogfish	4.84	-.26	-.42
Salmon & Goldfish	5.00	-.25	-.41
Salmon & Grizzly Bear	5.94	-.20	.20
Salmon & Herring	6.56	-.24	.42
Salmon & Trout	8.11	-.08	4.37
Saucepan & Cake Tin	6.61	-.26	-.42
Saucepan & Cook	7.44	-.08	1.63
Saucepan & Plate	6.11	-.26	-.39
Saucepan & Wineglass	3.44	-.26	-.41
Scorpion & Eagle	4.47	-.26	-.41
Scorpion & Meerkat	3.39	-.25	-.41
Scorpion & Spider	5.72	-.17	-.19
Sheep & Antelope	4.17	-.09	.00
Sheep & Goat	6.35	-.18	1.17
Sheep & Ox	4.44	-.21	-.21
Sheep & Shepherd	8.22	-.21	.25
Sheep & Sweater	5.33	-.25	-.40
Sheep & Wool	8.11	-.05	.51
Shepherd & Bus Driver	3.28	-.26	-.44
Shepherd & Cowboy	6.17	-.25	-.42
Shepherd & Zookeeper	5.94	-.26	-.44
Shrimp & Barnacle	4.67	-.22	-.38

Shrimp & Crab	6.88	.00	2.18
Shrimp & Grizzly Bear	2.41	-.26	-.44
Shrimp & Haddock	4.29	-.25	-.18
Shrimp & Lice	2.89	-.26	-.43
Shrimp & Lobster	7.72	-.14	1.56
Shrimp & Salmon	5.33	-.16	1.16
Shrimp & Woodlice	2.66	-.24	.04
Snail & Frog	4.47	-.23	-.33
Snail & Hedgehog	4.28	-.25	-.41
Snail & Limpet	5.56	-.25	-.19
Snail & Octopus	2.83	-.25	-.38
Snail & Slug	7.72	-.22	.18
Snail & Squid	3.22	-.26	-.39
Snake & Crocodile	4.41	-.20	-.12
Snake & Gecko	5.41	-.25	-.33
Snake & Iguana	5.72	-.23	-.26
Snake & Lizard	6.78	-.18	.17
Sparrow & Hawk	5.89	-.24	.15
Squirrel & Cat	3.89	-.23	.43
Squirrel & Chipmunk	6.56	-.21	.22
Squirrel & Goat	3.14	-.24	-.39
Squirrel & Llama	2.97	-.26	-.43
Squirrel & Prairie Dog	4.33	-.26	-.30
Squirrel & Sheep	2.67	-.25	-.25
Steel & Brass	7.06	-.08	.99
Steel & Bronze	7.11	-.18	.21
Steel & Cook	3.94	-.25	-.36
Steel & Saucepan	6.11	-.24	.30
Swallow & Nightingale	6.56	-.25	-.42
Swallow & Ostrich	3.56	-.26	-.43

Swallow & Ostrich	5.00	-.26	-.43
Swallow & Parrot	5.56	-.26	-.43
Swallow & Pigeon	6.11	-.25	-.42
Swallow & Robin	6.56	-.26	-.41
Sweater & Blazer	6.17	-.25	-.07
Sweater & Hat	5.28	-.19	-.14
Sweater & T-Shirt	6.44	-.24	.02
Toad & Newt	6.83	-.25	-.16
Toad & Salamander	5.00	-.24	-.08
Toad & Snake	3.78	-.20	-.32
Tree & Book	6.00	-.20	-.31
Tree & Bush	7.06	-.25	-.31
Tree & Flower	6.35	-.19	.07
Tree & Grass	5.94	-.18	-.20
Tree & Paper	6.94	-.21	-.34
Tuna Fish & Fisherman	5.17	-.26	-.33
Tuna Fish & Goldfish	5.47	-.23	-.42
Tuna Fish & Mackerel	7.33	-.16	1.89
Tuna Fish & Swordfish	8.06	-.15	1.87
Weasel & Earthworm	3.71	-.26	-.43
Weasel & Otter	6.39	-.26	-.33
Weasel & Vole	6.44	-.25	-.42
Wheat & Bamboo	3.83	-.24	-.38
Wheat & Barley	7.83	.11	3.32
Wheat & Bread	7.56	-.17	1.69
Wheat & Buzzard	1.94	-.26	-.44
Wheat & Corn	7.72	-.07	2.13
Wheat & Flour	7.94	.21	3.01
Wheat & Maize	7.76	-.16	1.30
Wheat & Mouse	4.39	-.26	-.43

Wheat & Rice	7.17	-.18	1.08
Wool & Cotton	6.82	-.12	1.99
Wool & Silk	5.78	-.12	2.27
Wool & Sweater	7.17	-.15	1.28

Appendix 1C: Co-Occurrence Indexing using Google and Exalead Search Engines

By the middle of 2008, the World Wide Web indexed over just below 571 million domains, representing a growth factor of around 148 over a 14-year period (Carpenter, 2009). As such, it probably represents the world's largest freely accessible source of printed word, making it ideal for Psycholinguistic research (Heylighen, 2001). We chose this as our data source for creating a co-occurrence index against which to verify subjective association ratings made by participants.

One potential problem when using internet search engines for creating an index of co-occurrence between two categories is that the search algorithms used to rank web pages, as well as the techniques used to 'crawl the web' are closely guarded trade secrets. We believe that this is less of a problem in our case, as we are trying to get a measure of the absolute frequency with which the two categories co-occur. However, one way of circumventing possible confounds that are specific to one search engine, we used two very different search engines to carry out co-occurrence searches, the popular Google and the less well-known Exalead which supports explicit proximity searches. The extent to which the two indices correlate would indicate the validity of using the World Wide Web and search engines in order to estimate frequency of co-occurrence.

Exalead (www.exalead.com) is a French search engine founded in 2000. It has an index of over 8 billion web pages (the biggest after Google, Yahoo! and Microsoft). One of the features which sets this search engine aside from others is that it specifically supports proximity functions with a NEAR operator. So for example, to search for the co-occurrence of dog and mouse within 6 characters, one would specify the following search term:

dog NEAR/6 *mouse*

In Google, unordered proximity searches have to be specified by combining a string of ordered searches, for example:

"*dog****mouse*" OR "*mouse****dog*"

The fact that the search engines use slightly different search algorithms, which tend to be a close-guarded trade secret, the finding that there was such a significant correlation between the two contrasting search engines lends credence to this method of measuring co-occurrence.

Appendix 2

Appendix 2A: Mean Inductive Strength Ratings Experiment 1

Property	Infection		Cell	
Timing	**speeded**	**delayed**	**speeded**	**delayed**
Krill & Penguin	4.00	4.00	4.00	3.50
Squirrel & Cat	3.00	3.70	4.43	4.00
Fly & Frog	4.29	4.80	3.57	3.38
Carrot & Rabbit	5.57	6.20	3.71	4.38
Snail & Hedgehog	3.71	4.63	3.57	3.10
Wheat & Mouse	4.71	3.63	3.43	4.30
Flower & Butterfly	4.71	4.50	2.57	4.00
Tree & Bark Beetle	5.71	6.75	4.00	4.60
Grasshopper & Toad	4.20	3.80	2.83	3.29
Meerkat & Eagle	3.00	3.80	2.83	3.00
Butterfly & Swallow	3.70	4.00	4.33	3.29
Shrimp & Salmon	5.70	5.20	6.17	4.86
Cabbage & Snail	5.50	4.71	4.70	5.00
Frog & Heron	4.00	4.71	3.60	3.00
Acorn & Squirrel	4.33	5.57	5.10	3.80
Grass & Sheep	5.67	6.00	5.00	4.40
Penguin & Krill	5.50	4.20	6.17	2.86
Cat & Squirrel	4.10	4.80	3.50	4.71
Frog & Fly	5.90	5.60	5.33	3.00
Rabbit & Carrot	5.80	3.40	5.00	2.00
Hedgehog & Snail	4.00	3.29	3.80	3.40
Mouse & Wheat	2.83	3.86	3.10	2.80
Butterfly & Flower	4.00	4.71	4.30	3.80
Bark Beetle & Tree	5.00	5.86	4.80	4.00
Toad & Grasshopper	4.00	2.80	3.43	3.13
Eagle & Meerkat	3.43	2.60	2.71	2.00

Swallow & Butterfly	3.86	4.20	3.86	3.38
Salmon & Shrimp	5.14	5.10	4.71	5.13
Snail & Cabbage	3.71	4.25	3.71	4.90
Heron & Frog	5.00	3.25	3.71	3.90
Squirrel & Acorn	6.14	5.00	3.71	3.60
Sheep & Grass	6.43	3.88	3.14	4.40
Krill & Woodlice	4.43	3.20	3.71	2.63
Squirrel & Prairie Dog	3.71	3.70	4.00	3.38
Fly & Ant	4.86	3.20	3.71	5.25
Carrot & Radish	6.57	6.00	5.14	6.50
Snail & Squid	3.14	2.00	3.00	2.70
Wheat & Bamboo	3.71	3.25	4.00	4.20
Flower & Grass	5.71	4.75	6.29	5.50
Tree & Bush	5.43	5.63	6.14	6.30
Grasshopper & Lice	3.40	4.60	4.00	3.86
Meerkat & Skunk	5.00	6.20	5.00	4.57
Butterfly & Ant	2.70	4.80	2.33	4.43
Shrimp & Barnacle	5.10	4.40	4.17	5.00
Cabbage & Horseradish	4.67	4.86	5.50	5.20
Frog & Newt	5.33	7.14	6.00	7.80
Acorn & Pecan	3.00	4.57	5.20	5.20
Grass & Dandelion	6.50	6.29	6.50	7.00
Penguin & Chicken	3.50	3.60	2.00	4.00
Cat & Badger	3.60	4.80	5.17	4.43
Frog & Salamander	5.00	5.40	4.83	4.57
Rabbit & Beaver	5.00	4.80	5.17	4.43
Hedgehog & Shrew	4.33	5.57	5.30	5.00
Mouse & Squirrel	4.33	4.43	4.30	5.20
Butterfly & Bee	4.67	6.14	4.80	4.60
Bark Beetle & Praying Mantis	3.00	4.43	4.60	5.20
Toad & Salamander	4.71	3.70	3.57	5.25
Eagle & Penguin	4.00	2.30	3.57	4.00
Swallow & Parrot	4.29	3.40	5.29	5.13

Salmon & Goldfish	5.14	3.40	4.29	6.25
Snail & Limpet	3.29	3.75	3.71	5.40
Heron & Penguin	4.29	3.63	4.57	3.80
Squirrel & Prairie Dog	3.71	3.38	3.29	4.10
Sheep & Goat	6.14	6.25	5.43	6.60

Appendix 2B: Post-Test Endorsements Experiment 1

	Proportion who believe Categories are Causally Related		Proportion who believe Categories are from same Taxonomic Group	
Timing	**speeded**	**delayed**	**speeded**	**delayed**
Krill & Penguin	0.77	0.78	0.17	0.11
Squirrel & Cat	0.15	0.22	0.75	0.50
Fly & Frog	0.92	1.00	0.00	0.11
Carrot & Rabbit	0.92	1.00	0.08	0.00
Snail & Hedgehog	0.92	0.67	0.08	0.06
Wheat & Mouse	0.92	0.94	0.00	0.06
Flower & Butterfly	0.62	0.83	0.00	0.06
Tree & Bark Beetle	0.92	0.94	0.00	0.06
Grasshopper & Toad	0.63	0.55	0.19	0.27
Meerkat & Eagle	0.44	0.55	0.31	0.09
Butterfly & Swallow	0.63	0.64	0.25	0.00
Shrimp & Salmon	0.63	0.45	0.63	0.64
Cabbage & Snail	1.00	1.00	0.19	0.00
Frog & Heron	0.67	0.82	0.19	0.18
Acorn & Squirrel	0.94	1.00	0.00	0.00
Grass & Sheep	1.00	1.00	0.00	0.00
Penguin & Krill	0.81	0.82	0.31	0.09
Cat & Squirrel	0.13	0.27	0.69	0.73
Frog & Fly	1.00	0.91	0.06	0.00
Rabbit & Carrot	0.88	1.00	0.13	0.00
Hedgehog & Snail	0.75	0.82	0.06	0.09
Mouse & Wheat	0.88	0.91	0.06	0.00
Butterfly & Flower	0.88	0.82	0.00	0.09
Bark Beetle & Tree	1.00	0.82	0.06	0.00
Toad & Grasshopper	0.54	0.61	0.25	0.11
Eagle & Meerkat	0.23	0.33	0.00	0.00
Swallow & Butterfly	0.54	0.67	0.08	0.06

Salmon & Shrimp	0.38	0.50	0.67	0.56
Snail & Cabbage	1.00	0.94	0.08	0.22
Heron & Frog	0.69	0.61	0.00	0.00
Squirrel & Acorn	0.92	1.00	0.00	0.17
Sheep & Grass	0.92	0.89	0.08	0.11
Krill & Woodlice	0.08	0.11	0.31	0.44
Squirrel & Prairie Dog	0.08	0.12	0.69	0.59
Fly & Ant	0.17	0.17	0.85	0.94
Carrot & Radish	0.08	0.11	1.00	1.00
Snail & Squid	0.27	0.18	0.08	0.50
Wheat & Bamboo	0.00	0.11	1.00	0.89
Flower & Grass	0.17	0.00	0.92	0.94
Tree & Bush	0.17	0.11	0.92	1.00
Grasshopper & Lice	0.25	0.64	0.63	0.45
Meerkat & Skunk	0.13	0.20	0.88	1.00
Butterfly & Ant	0.13	0.36	0.44	0.73
Shrimp & Barnacle	0.50	0.36	0.75	0.45
Cabbage & Horseradish	0.19	0.18	0.56	1.00
Frog & Newt	0.25	0.36	0.93	1.00
Acorn & Pecan	0.13	0.27	0.80	0.73
Grass & Dandelion	0.31	0.09	0.94	1.00
Penguin & Chicken	0.00	0.00	0.63	0.45
Cat & Badger	0.06	0.09	0.81	0.91
Frog & Salamander	0.07	0.10	0.93	0.50
Rabbit & Beaver	0.00	0.18	0.81	0.73
Hedgehog & Shrew	0.19	0.09	0.94	0.64
Mouse & Squirrel	0.19	0.18	0.88	0.91
Butterfly & Bee	0.19	0.18	0.94	1.00
Bark Beetle & Praying Mantis	0.50	0.60	0.69	0.70
Toad & Salamander	0.17	0.11	0.92	0.76
Eagle & Penguin	0.00	0.11	0.62	0.83
Swallow & Parrot	0.00	0.00	1.00	1.00
Salmon & Goldfish	0.17	0.11	1.00	0.94

Snail & Limpet	0.27	0.12	0.75	0.71
Heron & Penguin	0.08	0.22	0.69	0.83
Squirrel & Prairie Dog	0.08	0.00	0.77	0.53
Sheep & Goat	0.17	0.06	0.92	0.94

Appendix 2C: Mean Inductive Strength Ratings Experiment 2

Property	**Infection**		**Cell**	
Load	**heavy**	**light**	**heavy**	**light**
Krill & Penguin	3.80	3.20	1.40	4.20
Squirrel & Cat	4.20	3.00	3.40	4.00
Fly & Frog	5.80	5.60	3.80	4.00
Carrot & Rabbit	6.60	5.40	2.20	3.80
Snail & Hedgehog	3.60	2.20	1.60	3.20
Wheat & Mouse	3.60	2.20	3.40	2.20
Flower & Butterfly	5.00	3.20	3.00	2.20
Tree & Bark Beetle	5.20	4.60	3.80	5.00
Grasshopper & Toad	5.20	3.80	4.40	2.60
Meerkat & Eagle	5.80	1.40	4.00	3.80
Butterfly & Swallow	5.20	4.00	3.00	2.60
Shrimp & Salmon	6.00	4.60	4.80	3.00
Cabbage & Snail	6.60	3.80	4.00	3.40
Frog & Heron	3.80	4.00	3.60	3.80
Acorn & Squirrel	4.40	4.20	4.60	4.00
Grass & Sheep	5.40	5.40	3.20	6.00
Penguin & Krill	4.60	3.40	5.40	1.80
Cat & Squirrel	2.60	2.40	4.00	4.80
Frog & Fly	4.60	3.60	2.80	2.80
Rabbit & Carrot	4.60	3.00	3.60	1.20

Hedgehog & Snail	3.80	4.80	4.80	2.80
Mouse & Wheat	2.80	2.20	4.00	2.40
Butterfly & Flower	5.20	3.80	2.80	3.20
Bark Beetle & Tree	5.20	5.00	5.20	3.60
Toad & Grasshopper	2.00	3.20	2.80	2.60
Eagle & Meerkat	2.20	1.40	1.80	3.00
Swallow & Butterfly	2.00	2.00	3.00	2.00
Salmon & Shrimp	5.80	4.80	3.40	4.20
Snail & Cabbage	2.80	3.20	2.80	3.40
Heron & Frog	1.60	2.60	2.00	3.20
Squirrel & Acorn	3.60	3.80	2.80	4.40
Sheep & Grass	2.40	3.40	3.40	4.40
Krill & Woodlice	4.00	3.40	3.20	3.40
Squirrel & Prairie Dog	3.40	5.00	3.80	4.00
Fly & Ant	3.80	3.40	2.40	4.20
Carrot & Radish	3.80	5.60	5.40	6.20
Snail & Squid	1.60	2.60	2.20	2.40
Wheat & Bamboo	1.00	2.60	3.20	4.80
Flower & Grass	4.00	5.00	5.80	5.80
Tree & Bush	5.20	5.00	5.80	7.20
Grasshopper & Lice	4.20	2.20	5.60	3.40
Meerkat & Skunk	4.00	4.20	5.20	4.20
Butterfly & Ant	5.60	3.80	3.60	2.60
Shrimp & Barnacle	5.00	4.40	5.40	4.20
Cabbage & Horseradish	4.60	3.80	5.60	4.20

Frog & Newt	7.60	5.40	7.00	6.20
Acorn & Pecan	4.40	3.60	4.60	3.60
Grass & Dandelion	7.00	4.80	6.20	5.40
Penguin & Chicken	3.80	1.60	4.40	5.00
Cat & Badger	4.80	2.40	3.80	4.20
Frog & Salamander	4.80	4.20	6.60	4.00
Rabbit & Beaver	6.40	1.40	4.80	3.60
Hedgehog & Shrew	5.80	3.80	4.20	4.40
Mouse & Squirrel	4.80	3.20	4.60	5.00
Butterfly & Bee	5.60	5.40	4.00	4.00
Bark Beetle & Praying Mantis	3.60	5.20	5.60	3.60
Toad & Salamander	3.80	6.60	3.40	5.80
Eagle & Penguin	2.00	3.40	4.00	4.40
Swallow & Parrot	4.60	4.00	4.60	4.80
Salmon & Goldfish	4.80	3.40	4.00	5.00
Snail & Limpet	2.40	4.20	3.60	3.60
Heron & Penguin	2.00	3.60	2.00	3.60
Squirrel & Prairie Dog	3.60	4.00	3.20	5.40
Sheep & Goat	4.25	5.00	4.80	7.40

Appendix 2D: Post-Test Endorsements Experiment 2

	Proportion who believe Categories are Causally Related		Proportion who believe Categories are from same Taxonomic Group	
Load	**heavy**	**light**	**heavy**	**light**
Krill & Penguin	0.71	0.65	0.11	0.10
Squirrel & Cat	0.26	0.15	0.47	0.60
Fly & Frog	0.89	0.95	0.05	0.05
Carrot & Rabbit	0.95	0.95	0.05	0.05
Snail & Hedgehog	0.63	0.70	0.05	0.00
Wheat & Mouse	0.79	0.90	0.05	0.05
Flower & Butterfly	0.68	0.74	0.00	0.00
Tree & Bark Beetle	0.89	0.90	0.05	0.00
Grasshopper & Toad	0.68	0.75	0.11	0.20
Meerkat & Eagle	0.68	0.60	0.11	0.00
Butterfly & Swallow	0.68	0.45	0.05	0.10
Shrimp & Salmon	0.63	0.25	0.68	0.45
Cabbage & Snail	0.84	1.00	0.00	0.00
Frog & Heron	0.63	0.80	0.00	0.05
Acorn & Squirrel	0.95	1.00	0.00	0.00
Grass & Sheep	0.95	1.00	0.00	0.00
Penguin & Krill	0.71	0.65	0.11	0.10
Cat & Squirrel	0.26	0.15	0.47	0.60
Frog & Fly	0.89	0.95	0.05	0.05
Rabbit & Carrot	0.95	0.95	0.05	0.05
Hedgehog & Snail	0.63	0.70	0.05	0.00

Mouse & Wheat	0.79	0.90	0.05	0.05
Butterfly & Flower	0.68	0.74	0.00	0.00
Bark Beetle & Tree	0.89	0.90	0.05	0.00
Toad & Grasshopper	0.68	0.75	0.11	0.20
Eagle & Meerkat	0.68	0.60	0.11	0.00
Swallow & Butterfly	0.68	0.45	0.05	0.10
Salmon & Shrimp	0.63	0.25	0.68	0.45
Snail & Cabbage	0.84	1.00	0.00	0.00
Heron & Frog	0.63	0.80	0.00	0.05
Squirrel & Acorn	0.95	1.00	0.00	0.00
Sheep & Grass	0.95	1.00	0.00	0.00
Krill & Woodlice	0.22	0.30	0.25	0.60
Squirrel & Prairie Dog	0.20	0.30	0.50	0.80
Fly & Ant	0.30	0.20	0.70	0.60
Carrot & Radish	0.50	0.30	1.00	1.00
Snail & Squid	0.30	0.10	0.50	0.60
Wheat & Bamboo	0.10	0.20	0.80	0.90
Flower & Grass	0.20	0.20	0.90	0.90
Tree & Bush	0.10	0.20	1.00	0.90
Grasshopper & Lice	0.33	0.30	0.78	0.60
Meerkat & Skunk	0.11	0.00	0.89	0.90
Butterfly & Ant	0.11	0.20	0.78	0.80
Shrimp & Barnacle	0.56	0.20	1.00	0.70
Cabbage & Horseradish	0.44	0.00	1.00	1.00
Frog & Newt	0.33	0.30	1.00	0.90

Acorn & Pecan	0.33	0.20	0.89	1.00
Grass & Dandelion	0.56	0.00	1.00	0.80
Penguin & Chicken	0.00	0.00	0.44	0.60
Cat & Badger	0.22	0.00	0.89	0.90
Frog & Salamander	0.33	0.50	0.89	0.60
Rabbit & Beaver	0.11	0.00	0.44	0.70
Hedgehog & Shrew	0.44	0.30	0.89	0.70
Mouse & Squirrel	0.44	0.10	0.89	0.80
Butterfly & Bee	0.33	0.10	1.00	0.90
Bark Beetle & Praying Mantis	0.33	0.30	0.63	0.70
Toad & Salamander	0.20	0.30	0.20	0.80
Eagle & Penguin	0.30	0.00	0.30	0.90
Swallow & Parrot	0.10	0.10	0.10	0.70
Salmon & Goldfish	0.30	0.00	0.30	1.00
Snail & Limpet	0.30	0.20	0.30	0.80
Heron & Penguin	0.40	0.20	0.40	0.90
Squirrel & Prairie Dog	0.30	0.30	0.30	0.60
Sheep & Goat	0.10	0.20	0.10	1.00

Appendix 2E: Mean Inductive Strength Ratings Experiment 3

Property	**Disease**		**Cell**	
Timing	**speeded**	**delayed**	**speeded**	**delayed**
Acorn & Squirrel	4.80	5.90	3.30	3.40
Acorn & Walnut	4.10	4.50	5.50	6.20
Butterfly & Ant	3.40	3.40	4.10	5.90
Butterfly & Swallow	3.60	4.60	3.50	4.00
Cabbage & Rapeseed	2.90	4.60	4.20	4.40
Cabbage & Snail	4.20	5.00	3.50	3.50
Carrot & Rabbit	3.70	3.60	4.30	2.80
Carrot & Radish	5.20	4.60	4.60	5.90
Fly & Ant	3.70	3.80	4.60	4.40
Fly & Frog	4.30	5.00	4.30	4.70
Grass & Bamboo	4.10	3.80	6.10	6.60
Grass & Sheep	4.90	5.00	3.50	3.00
Grasshopper & Lice	2.90	3.90	3.80	4.80
Grasshopper & Toad	3.10	3.50	3.40	4.10
Krill & Penguin	4.80	4.60	3.90	3.40
Krill & Woodlice	3.00	2.50	2.60	3.50
Meerkat & Badger	3.90	4.00	4.00	5.70
Meerkat & Eagle	3.50	4.30	2.00	4.50
Mouse & Buzzard	4.80	6.00	3.00	3.80
Mouse & Squirrel	4.70	5.60	5.20	6.00
Penguin & Duck	3.60	3.20	4.90	6.90
Penguin & Orca	4.80	5.30	4.20	4.90
Rabbit & Fox	6.30	5.30	6.00	4.90
Rabbit & Squirrel	4.60	4.10	5.50	5.10
Salmon & Grizzly	3.70	6.50	3.10	4.50
Salmon & Herring	4.30	5.00	6.50	6.50

Scorpion & Eagle	2.70	3.00	1.80	2.80
Scorpion & Spider	3.70	5.20	4.10	4.80
Shrimp & Barnacle	4.90	4.50	4.60	4.70
Shrimp & Salmon	4.80	5.10	5.00	3.80
Snail & Frog	3.00	2.90	3.90	4.20
Snail & Hedgehog	3.90	4.40	2.60	3.60
Snail & Limpet	3.00	4.50	4.50	6.10
Snail & Squid	2.90	2.50	2.80	4.00
Squirrel & Cat	3.60	3.40	4.80	5.30
Squirrel & Prairie Dog	2.30	3.40	6.00	4.60
Toad & Salamander	4.60	4.10	5.40	5.60
Toad & Snake	3.30	3.30	4.00	3.70
Wheat & Bamboo	2.80	3.90	4.40	5.20
Wheat & Mouse	3.30	4.80	2.80	3.60

Appendix 2F: Post-Test Endorsement Proportions Experiment 3

	Proportion who believe Categories are Causally Related		Proportion who believe Categories are from same Taxonomic Group	
Timing	**speeded**	**delayed**	**speeded**	**delayed**
Acorn & Squirrel	0.90	1.00	0.70	0.80
Acorn & Walnut	0.25	0.16	0.95	1.00
Butterfly & Ant	0.50	0.35	1.00	0.65
Butterfly & Swallow	0.55	0.61	0.00	0.00
Cabbage & Rapeseed	0.10	0.05	0.80	0.74
Cabbage & Snail	1.00	1.00	0.05	0.05
Carrot & Rabbit	0.95	0.85	0.05	0.05
Carrot & Radish	0.20	0.20	1.00	0.95
Fly & Ant	0.00	0.10	0.85	0.65
Fly & Frog	0.90	0.80	0.15	0.35
Grass & Bamboo	0.25	0.35	0.95	0.95
Grass & Sheep	0.90	0.90	0.00	0.00
Grasshopper & Lice	0.30	0.24	0.75	0.50
Grasshopper & Toad	0.70	0.63	0.40	0.50
Krill & Penguin	0.90	0.82	0.05	0.10
Krill & Woodlice	0.05	0.06	0.37	0.16
Meerkat & Badger	0.11	0.16	0.80	0.94
Meerkat & Eagle	0.74	0.63	0.11	0.11
Mouse & Buzzard	0.85	0.83	0.85	0.89
Mouse & Squirrel	0.15	0.22	0.95	0.70
Penguin & Duck	0.45	0.50	0.85	0.75
Penguin & Orca	0.79	0.60	0.05	0.11
Rabbit & Fox	0.90	0.95	0.00	0.05
Rabbit & Squirrel	0.25	0.26	0.95	0.84
Salmon & Grizzly	0.90	0.95	0.40	0.60
Salmon & Herring	0.35	0.33	0.90	0.90

Scorpion & Eagle	0.45	0.25	0.10	0.11
Scorpion & Spider	0.30	0.15	0.65	0.55
Shrimp & Barnacle	0.35	0.30	0.65	0.72
Shrimp & Salmon	0.55	0.55	0.00	0.10
Snail & Frog	0.40	0.20	0.05	0.15
Snail & Hedgehog	0.70	0.53	0.50	0.35
Snail & Limpet	0.05	0.20	0.80	0.76
Snail & Squid	0.20	0.37	0.35	0.35
Squirrel & Cat	0.10	0.15	0.15	0.06
Squirrel & Prairie Dog	0.25	0.30	0.65	0.74
Toad & Salamander	0.25	0.15	0.60	0.67
Toad & Snake	0.55	0.47	0.05	0.05
Wheat & Bamboo	0.06	0.95	0.90	0.90
Wheat & Mouse	0.85	0.15	0.35	0.35

Appendix 3

Appendix 3A: Base Categories Experiment 4

Property (between subjects)	**Cell**	**Disease**
	Carrot	Carrot
	Tuna	Tuna
	Worm	Worm
	Trout	Trout
	Hawk	Hawk
	Mosquito	Mosquito
	Dog	Dog
	Zebra	Zebra
	Crocodile	Crocodile
	Weasel	Weasel

Appendix 3B: Base Categories Experiment 5

Property (within-subjects)	**Cell**	**Disease**
	Bat	Kangaroo
	Tuna	Dog
	Bee	Weasel
	Trout	Elephant
	Skunk	Seagull
	Mosquito	Rose
	Dog	Flea
	Snake	Pigeon
	Pig	Chimpanzee
	Heron	Eucalyptus

Appendix 3C: Classification Criteria for Categories generated by Participants

Taxonomic Relationship	
Category Membership	Both Categories belong to the same class or category. *Example: Zebra & Antelope→ Mammals*
Causal Relationship	
Similar Diet	Both Categories are similar with respect to diet or eating the same kind of thing. *Example: Mosquito & Vampire Bat→ feed on blood*
Similar Habitat	Both Categories share similar or the same habitat. *Example: Zebra & Antelope→ live in African Savannah*
Behavioural Interaction	Both Categories interact via some aspect of behaviour. *Example: Seagull &Fisherman→ Gulls trail fishing boats*
Food Chain Interaction	Both Categories interact with respect to diet or eating, i.e. one category eats or is eaten by the other. *Example: Worm & Blackbird→ dietary connection*

Appendix 4

Appendix 4A: Triads and Strength of Association

Table 4.1: Category Pairs for the Triad Items and Mean Strength of Association Ratings

	Target Category (and Strength of Association with Base)		
Base Category	Irrelevant but Strong Association	Taxonomic Weak Association	Unrelated
Orca Whales	Cod (5.3)	Cows (2.2)	Pigeons
Snails	Hedgehogs (4.3)	Octopuses (2.8)	Sloths
Butterflies	Flowers (7.6)	Locusts (5.5)	Seaweed
Salmon	Grizzly Bears (7.5)	Goldfish (5.0)	Hedgehogs
Monkeys	Peanuts (6.8)	Seals (2.5)	Almonds
Acorns	Squirrels (7.9)	Lychees (3.7)	Seals
Bananas	Monkeys (7.5)	Tulips (2.4)	Sea Lions
Shepherds	Sheep (8.2)	Bus Drivers (3.3)	Porpoises
Mice	Wheat (4.4)	Goats (2.7)	Bamboo
Grass	Sheep (7.5)	Palm Trees (5.0)	Hyenas
Penguins	Orca Whales (5.9)	Chickens (3.0)	Dogs
Ants	Anteaters (7.6)	Dragonflies (4.5)	Moose
Carrots	Rabbits (7.7)	Bamboo (3.1)	Tigers
Dolphins	Cod (7.3)	Llamas (2.0)	Parrots

Appendix 4B: Semantic Inhibitory Control Task

Prime Sentence	Ap. Non-Word	Ap. Word	In-ap. Non-Word	Ap. Non-Word
He was tired and went to ____	sleap	bed	efent	goat
John felt sorry, but the accident was not his ___	foult	responsibility	lettor	billboard
The teacher wrote the problem on the ____	backboad	paper	craud	talent
Shuffle the cards before you ____	deel	play	klothes	college
The pigs wallowed in the ____	mudde	dirt	cource	wrist
Karen awoke after a bad ____	nitemair	dream	troubel	ankle
The movers put the sofa down on the bare ___	fhlore	concrete	rotton	father
The child was born with a rare ____	deseese	illness	delite	planet
The squirrel stored some nuts in the ____	threa	nest	grount	blond
Betty had no sense of ____	humah	direction	parens	steak
The captain wanted to stay with the sinking ___	shibp	boat	werde	marathon
John poured himself a glass of ____	millkh	wine	strewp	duck
He liked lemon and sugar in his ____	cofea	tea	peetal	rooster
They went as far as they ____	ccoud	wanted	lucke	voice
The derelict house will be torn____	dauwne	apart	proplem	book
Most cats see well at ____	nhite	dusk	crums	moustache
Sharon dried the dishes with a ____	taull	cloth	rowen	safety
Jean was glad that the affair was____	ovar	finished	traine	parents
The whole town came to hear the mayor ____	speek	talk	bibel	newspaper
The game was called off when it began to ___	rayn	snow	shing	sing
He scraped the cold food from his ____	plade	dish	vally	hint
The pizza was too hot to ____	eet	handle	trak	sofa
She called her husband at his ____	ofise	work	mohn	horse
The wealthy child attended a private ____	scholl	nursery	jocke	tomorrow
The crime rate has gone up this ____	yeer	month	jellow	garlic
Her new shoes were the wrong ____	sise	colour	croud	hormone

The lawyer knew that his client was ____	inosant	guilty	cemical	circus
The dog chased the cat up the ____	trea	street	magasin	health
The doctor said that his leg was ____	proken	injured	naighbor	talked
John must keep his dog on a ____	leache	chain	paiper	money

Appendix 4C: Mean Proportion of Taxonomic Choices for Items Experiment 6

	Triads	Mean Proportion Taxonomic Choice	
		Heavy Load	Light Load
Conflict Triads Cells	Orca Whales → Cod or Cows?	0.12	0.35
	Snails → Hedgehogs or Octopuses?	0.59	0.76
	Butterflies → Flowers or Locusts?	0.65	0.88
	Salmon → Grizzly Bears or Goldfish?	0.59	0.82
	Monkeys → Peanuts or Seals?	0.47	0.59
	Acorns → Squirrels or Lychees?	0.59	0.82
	Bananas → Monkeys or Tulips?	0.53	0.65
	Shepherds → Sheep or Bus Drivers?	0.82	0.88
	Mice → Wheat or Goats?	0.35	0.76
	Grass → Sheep or Palm Trees?	0.65	0.71
	Penguins → Orca Whales or Chickens?	0.06	0.29
	Ants → Anteaters or Dragonflies?	0.53	0.76
	Carrots → Rabbits or Bamboo?	0.29	0.65
	Dolphins → Cod or Llamas?	0.06	0.29
Control Triads Cells	Orca Whales → Pigeons or Cows?	0.88	0.88
	Snails → Sloths or Octopuses?	0.71	0.82
	Butterflies → Seaweed or Locusts?	0.94	0.94
	Salmon → Hedgehogs or Goldfish?	1.00	1.00
	Monkeys → Almonds or Seals?	0.59	0.53
	Acorns → Sea Lions or Lychees?	0.94	1.00
	Bananas → Seals or Tulips?	0.82	1.00

Shepherds → Porpoises or Bus Drivers?	0.94	1.00
Mice → Bamboo or Goats?	0.88	0.94
Grass → Hyenas or Palm Trees?	0.94	0.88
Penguins → Dogs or Chickens?	0.71	0.82
Ants → Moose or Dragonflies?	0.88	0.94
Carrots → Tigers or Bamboo?	0.88	1.00
Dolphins → Parrots or Llamas?	0.79	0.84

Appendix 4D: Post-Test Endorsements Experiment 6

	Proportion who believe Categories are Causally Related		Proportion who believe Categories are from same Taxonomic Group	
Timing	**Light load**	**Heavy load**	**Light load**	**Heavy load**
Orca & Cod	0.56	1	0.29	0.35
Snail & Hedgehog	0.76	0.76	0.24	0
Flower & Butterfly	0.94	0.82	0.12	0
Salmon & Grizzly Bear	0.94	0.94	0	0
Monkey & Peanut	0.94	0.94	0.06	0.12
Acorn & Squirrel	1	0.94	0	0.12
Banana & Monkey	0.88	1	0	0
Sheep & Shepherd	0.65	0.71	0.24	0.06
Mouse & Wheat	0.82	1	0	0
Grass & Sheep	0.94	0.94	0.06	0
Penguin & Orca Whale	0.47	0.59	0. 25	0.29
Ant & Anteater	0.82	1	0.18	0.06
Carrot & Rabbit	1	0.94	0.12	0.06
Dolphin & Cod	0.82	0.88	0.29	0.47
Orca Whale & Cow	0.12	0.06	0.71	0.41
Snail & Octopus	0.12	0	0.76	0.25
Butterfly & Locust	0.35	0.29	0.82	0.88
Salmon & Goldfish	0	0.18	0.94	1
Monkey & Seal	0.12	0	0.41	0.29
Acorn & Lychee	0.24	0.06	0.82	0.94
Banana & Tulip	0.12	0	0.53	0.69
Shepherd & Bus Driver	0	0.29	0.94	1
Mouse & Goat	0.12	0.12	0.53	0.47
Grass & Palm Tree	0.12	0.06	1	1
Penguin & Chicken	0.12	0	0.59	0.59
Ant & Dragonfly	0.47	0.24	0.94	0.94
Carrot & Bamboo	0.06	0.12	0.82	0.82

Dolphin & Llama	0	0	0.47	0.41
Orca Whale & Pigeon	0	0	0	0.06
Snail & Sloth	0.24	0.24	0.2	0.06
Butterfly & Seaweed	0.12	0	0.12	0.06
Salmon & Hedgehog	0	0	0	0
Monkey & Almond	1	0.71	0.06	0
Acorn & Sea Lion	0.12	0	0.12	0
Banana & Seal	0	0	0.12	0
Shepherd & Porpoise	0.12	0.06	0.56	0.19
Mouse & Bamboo	0.29	0.12	0	0
Grass & Hyena	0.53	0.63	0.12	0.12
Penguin & Dog	0	0	0.06	0.12
Ant & Moose	0.29	0.24	0	0
Carrot & Tiger	0.06	0.2	0.12	0
Dolphin & Parrot	0.12	0	0	0.06

Appendix 4E: Mean Proportion of Taxonomic Choices across Individual Items

	Triads	Mean Proportion Taxonomic Choice	
		Experiment 7	Experiment 8
Conflict Triads Cells	Orca Whales → Cod or Cows?	0.32	0.34
	Snails → Hedgehogs or Octopuses?	0.79	0.66
	Butterflies → Flowers or Locusts?	0.82	0.76
	Salmon → Grizzly Bears or Goldfish?	0.86	0.92
	Monkeys → Peanuts or Seals?	0.86	0.78
	Acorns → Squirrels or Lychees?	0.86	0.86
	Bananas → Monkeys or Tulips?	0.89	0.80
	Shepherds → Sheep or Bus Drivers?	0.89	0.86
	Mice → Wheat or Goats?	0.79	0.88
	Grass → Sheep or Palm Trees?	0.75	0.84
	Penguins → Orca Whales or Chickens?	0.21	0.08
	Ants → Anteaters or Dragonflies?	0.93	0.86
	Carrots → Rabbits or Bamboo?	0.86	0.84
	Dolphins → Cod or Llamas?	0.36	0.24
Control Triads Cells	Orca Whales → Pigeons or Cows?	0.82	0.82
	Snails → Sloths or Octopuses?	0.96	0.68
	Butterflies → Seaweed or Locusts?	0.82	0.98
	Salmon → Hedgehogs or Goldfish?	1.00	1.00
	Monkeys → Almonds or Seals?	1.00	0.88
	Acorns → Sea Lions or Lychees?	0.89	1.00
	Bananas → Seals or Tulips?	0.93	0.98
	Shepherds → Porpoises or Bus Drivers?	1.00	0.98

Mice → Bamboo or Goats?	0.89	0.92
Grass → Hyenas or Palm Trees?	0.96	0.96
Penguins → Dogs or Chickens?	0.86	0.76
Ants → Moose or Dragonflies?	0.96	0.96
Carrots → Tigers or Bamboo?	0.96	0.96
Dolphins → Parrots or Llamas?	0.79	0.84

Appendix 4F: Post-Test Endorsements Experiment 7 and 8

	Proportion who believe Categories are Causally Related		Proportion who believe Categories are from same Taxonomic Group	
Timing	**Expt 7**	**Expt 8**	**Expt 7**	**Expt 8**
Orca & Cod	0.81	0.76	0.33	0.50
Snail & Hedgehog	0.74	0.66	0.07	.0.20
Flower & Butterfly	0.81	0.80	0.15	0.04
Salmon & Grizzly Bear	0.81	0.84	0	0.06
Monkey & Peanut	0.93	0.94	0	0
Acorn & Squirrel	0.89	0.96	0.04	0.04
Banana & Monkey	0.93	0.94	0.04	0
Sheep & Shepherd	0.85	0.70	0.30	0.18
Mouse & Wheat	0.89	0.92	0	0.02
Grass & Sheep	0.93	1	0	0.02
Penguin & Orca Whale	0.67	0.58	0.56	0.46
Ant & Anteater	0.93	0.94	0.04	0.14
Carrot & Rabbit	0.93	0.92	0	0.02
Dolphin & Cod	0.74	0.76	0.33	0.52
Orca Whale & Cow	0.11	0	0.67	0.54
Snail & Octopus	0.11	0.1	0.54	0.53
Butterfly & Locust	0.41	0.20	0.93	0.90
Salmon & Goldfish	0.26	0.28	1	0.96
Monkey & Seal	0.07	0.02	0.89	0.50
Acorn & Lychee	0.30	0.26	0.74	0.90
Banana & Tulip	0.11	0.02	0.67	0.76
Shepherd & Bus Driver	0.19	0.22	1	0.98
Mouse & Goat	0.12	0.12	0.44	0.70
Grass & Palm Tree	0.19	0.06	0.89	0.94
Penguin & Chicken	0.04	0.02	0.78	0.62
Ant & Dragonfly	0.30	0.29	0.81	0.82
Carrot & Bamboo	0.15	0.12	0.74	0.86

Dolphin & Llama	0.04	0	0.52	0.46
Orca Whale & Pigeon	0.08	0.04	0..07	0.08
Snail & Sloth	0.35	0.27	0.15	0.20
Butterfly & Seaweed	0.07	0.04	0	0
Salmon & Hedgehog	0.07	0	0.04	0.02
Monkey & Almond	0.74	0.76	0	0
Acorn & Sea Lion	0.07	0.04	0	0
Banana & Seal	0	0	0	0.02
Shepherd & Porpoise	0.12	0.06	0.36	0.27
Mouse & Bamboo	0.27	0.18	0	0
Grass & Hyena	0.50	0.56	0	0
Penguin & Dog	0.15	0	0.22	0.24
Ant & Moose	0.22	0.14	0.07	0.04
Carrot & Tiger	0.04	0.04	0	0
Dolphin & Parrot	0.04	0	0.04	0.08

Appendix 5

Appendix 5A: Association Ratings Premise Categories in Medin et al. (2003)

Relation	Category Pair	Strength of Association
Property Reinforcement Diverse	Skunks & Stink Bugs	5.1
	Chimpanzees & Dolphins	4.1
	Bats & Robins	4.9
	Camels & Desert Rats	5.9
	Pigs & Chickens	4.7
	Penguins & Polar Bears	5.9
	Kangaroos & Frogs	2.4
Non-diverse	Penguins & Eagles	5.3
	Robins & Iguanas	2.3
	Pigs & Whales	3.2
	Cats & Rhinos	3.0
	Penguins & Frogs	3.1
	Koalas & Wolves	3.6
	Polar Bears & Antelopes	3.6
	Horses & Ants	2.5
	Kangaroos & Elephants	4.1
	Rabbits & Zebras	3.0
	Sparrows & Dogs	3.1
	Bats & Elephants	2.4
	Fleas & Butterflies	4.5
	Camels & Rhinos	4.3
	Chimpanzees & Cows	3.1
	Skunks & Deer	3.6
Causal Diverse	Rabbits & Carrots	6.8
	Robins & Worms	6.1
	Fleas & Dogs	5.8
	Horses & Grass	6.5
	Cats & Sparrows	4.9

Rabbits & Lettuce	6.9
Koalas & Gum Trees	5.6
Sparrows & Seeds	6.0