Factors Shaping Process and Representation in Multiple-Cue Judgment

DOCTORAL DISSERTATION
To be publicly examined in BT102, Betendevetarhuse, Umeå University, on Friday May 28th, 2004, at 10.15 a.m.
For the degree of Doctor of Philosophy

by

Anna-Carin Olsson

Department of Psychology
Umeå University, Umeå, Sweden
2004
ABSTRACT


This thesis investigates the cognitive processes and representations underlying human judgment in a multiple-cue judgment task. Several recent models assume that people have several qualitatively distinct and competing levels of knowledge representations (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994; Sloman, 1996). The most successful cognitive models in categorization and multiple-cue judgment are, respectively, exemplar-based models and cue abstraction models. The models are different in the computations and processes implied, but the structure of the task is similar. Study 1 investigated if the different theoretical conclusions in categorization and multiple-cue judgment derive from genuine differences in the processes, or are accidental to the different research methods. The results revealed large individual differences and a shift from exemplar memory to cue abstraction when the criterion is changed from a binary to a continuous variable, and especially for a probabilistic criterion. People appear to switch between qualitatively distinct processes in the two tasks. In Study 2, we expected learning in dyads to promote explicit cue abstraction as a consequence of verbalization (a social abstraction effect) and performance to improve due to the larger joint exemplar knowledge base (an exemplar pooling effect). Study 2 suggests that dyads make better judgments than participants working alone, but we failed to detect any difference in the representation of knowledge. Dyads can store more exemplars in memory allowing more efficient exploitation of memory and exemplar retrieval dominates the judgments. In contrast to earlier research, dyads surpassed the combined baseline level defined by the aggregated performance by members of the dyad working alone. In Study 3 we used the generalized model Sigma to illustrate how change in task environments (linear vs. nonlinear) can shape the knowledge representation that is used. We expected that people are not able to use cue abstraction when judging objects with a non-linear structure between the visual cues (features) of the objects and the criterion, and therefore they are forced to use exemplar-based processes. The results showed that participants in the non-linear conditions used an exemplar-based process. Both cue abstraction and exemplar memory processes were used in the linear condition suggesting that cue abstraction and exemplar memory can be used in a linear multiple cue judgment task with the same achievement. Taken together, the results of these
studies indicate that differences that characterize typical categorization and multiple-cue judgment tasks are conducive of qualitatively different cognitive processes, and that the task environment plays an important role for which cognitive processes are used in multiple cue judgments.

**Key words:** Multiple-cue judgment, categorization, cognitive processes, knowledge representations, exemplar-based models, cue abstraction models.
ACKNOWLEDGEMENT

Many people have made the completion of my thesis possible by help and support. I would like to thank the following persons.

To Peter Juslin, my almighty mentor and supervisor: Thanks for all your support and guidance throughout my time as a doctoral student. You have inspired and taught me what is important to be a good scientist; be motivated and staying focused even when research issues seem to be unsolvable. I am also highly indebted to Henrik Olsson for help in statistical, methodological and theoretical matters. The amazing analytic ability Peter and Henrik possess has been invaluable for me and has sharpened my eyes as a researcher. I would also like to thank my research project fellows Patrik Hansson, Linnea Karlsson and Håkan Nilsson for providing helpful comments on my work and for all “Friday research lunches”. Working in the project with you guys, has been valuable, useful and enjoyable in many ways and you have become good friends of mine. I am grateful to Jörgen Garvill for his comments and questions on an earlier version of this thesis, which have been very helpful in the refinement process.

I thank Daniel Broman for his supporting talks about work and all other stuff and for being a great colleague. Thanks to Pehr Andersson for being a superior personal trainer at the IKŠU gym, now I am fit and ready for fight and my biceps are even more impressive! I thank all fellow doctoral students and colleagues at the department for providing a friendly, humorous and inspiring working. Many important issues have been discussed during lunch and coffee breaks involving all sorts of things.

My very special thanks to my family, my loving and wonderful parents Roland and Margareta, my sister Marlene and her family, Thomas, Josefine and Johanna, my brother Conny, that I wished could share this with me, and his lively children Dennis and Malin. Thank you for always standing by my side. Fun times with close friends have let me not only focus on work. You made my personal life more important and rich. Thanks to the apple of my eye, my snoop-doggy-dog Gibson for enlightening my life. Finally I wish to thank my beloved Per-Erik for all his patience and love during this work and life and for all other reasons.

Umeå, April, 2004

Anna-Carin Olsson
LISTS OF PAPERS
This doctoral thesis is based on the following studies:


Figures and tables are reproduced with permission from the American Psychological Association.
INTRODUCTION

The cognitive system is complex and contains different cognitive processes. We collect information in the environment in order to solve problems and judge different objects or situations in the correct way and there appears to exist several alternative cognitive processes that may underlie these judgments (Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994, Nosofsky & Palmeri, 1997; Sloman, 1996; Tulving, 1983). A central position in many social sciences is occupied by the idea of maximization of expected utility according to Neumann and Morgenstern’s (1947) formulation of classical rationality. Economists explain and anticipate behavior in terms of rational optimization, the so called “Homo economicus”. These paradigms assume that humans are rational decision makers. Research in cognitive psychology has therefore focused on the question of human rationality – do we have the cognitive capacity demanded by this idea of optimization? The conclusion has often been that we rely on simplifying heuristics that conflict with the demands for optimization (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). These results are often interpreted as problematic for models of rational optimization.

The conclusions drawn about human rationality and the ability to make rational decisions are often stated in general and universal terms (Juslin & Olson, 2004). Human rationality, as discussed by the different approaches below, which lay the foundation for decision making and judgment can, however, also be related to the different knowledge processes and representations that are used. In these studies, I wanted to investigate what factors that affect which cognitive processes that underlie judgments, with the idea that, qualitatively different processes can underlie the judgments in different situations. In particular, I will consider two kinds of processes that have been emphasized in previous research, exemplar memory (Estes, 1994; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984; Nosofsky & Johansen, 2000; Nosofsky & Palmeri, 1997) and rule-based cue abstraction (Einhorn, Kleinmuntz, & Kleinmuntz, 1979).

For an illustration of exemplar memory, consider a physician who makes the diagnosis of a new patient by retrieving similar patients with known diagnoses. For an illustration of cue abstraction, imagine a physician that has abstracted specific symptom-diagnosis rules (e.g., fever co-varies with infection) from experience or previous instruction that are retrieved and integrated to make a diagnosis. How do these two different kinds of thinking affect the judgment process? As a preliminary to answering these questions this thesis aims to investigate what factors affect the cognitive processes in a multiple-cue judgment task. In the following, I briefly review previous research on multiple-cue judg-
ment and categorization, thereafter I summarize the results of three studies that investigate factors that may affect whether judgments are dominated by exemplar memory or cue abstraction.

BACKGROUND

Judgment Research

Multiple-Cue Judgment

Research on judgment and decision making began in the 1940s and the 1950s and two different research programs were introduced. One program was concerned with how people decide on a course of action and compared human behavior to normative models (Kahneman & Tversky, 2000). Another program was based on ideas from perception, and investigated how judgments rely on external information in the environment to make a judgment (Hammond & Stewart, 2001). In the latter program, inspired by Brunswikian psychology, it was stressed that the accuracy of a judgment depends on the nature of the task and how we identify and use information from the environment to make inferences. This latter program is often addressed by multiple-cue judgment. A multiple-cue judgment task requires participants to use a number of cues to infer a continuous or binary criterion. A typical example of multiple-cue judgment is stockbrokers that judge the probability that the stock will rise or fall on the basis of a few company features. Brunswikian-inspired paradigms emphasize the need to understand both the organism and the environment where the organism functions, as well as the relationship between them, an approach Brunswik called probabilistic functionalism (Brehmer, 1988). This was a new approach instead of focusing on the organism alone, as emphasized in research at the time.

In general, multiple-cue judgments are well captured by multiple linear regression models (Brehmer, 1994; Cooksey, 1996). The results show that the judgment is often a linear additive function of the cues, that few cues are used, that judgments are inconsistent, that judges have insufficient knowledge about the processes underlying the judgment, and often a process with a mix of analytic and intuitive thought termed quasi-rationality is observed (Brehmer, 1994). It is often stressed that these regression models are ratiomorphic (the predictability of lawful phenomena in the environment), statistical descriptions rather than process model (viz. Hoffman, 1960). Yet, in research on multiple-cue judgment (Brehmer, 1994; Cooksey, 1996) the explicit or implicit cognitive interpretation has often been that people abstract explicit knowledge of cue-criterion relations that is retrieved and mentally integrated into a judgment (Einhorn, Kleinmuntz, & Kleinmuntz, 1979). Moreover it is often claimed that the imperfections in the judgments arise from working memory capacity.
limitations (Baddeley, 1986). The picture emerging is generally one of controlled thought processes operating on explicit rule-based representations, although it is admitted that the processes and representations in multiple-cue judgments largely remain uncharted (Brehmer, 1994). It is important to note, however, that the fit of a linear multiple-regression model can not tell whether the judgments derive from exemplar memory or mental cue abstraction as further illustrated in Figure 3C and 3D below.

The question of whether judgments are based on intuition or analysis soon saw the light, and both positive and negative views of intuition and analysis were expressed among researchers (Hammond, 1996). Researchers with a positive view of intuition as an uncontrolled cognitive mode emphasized its role in great discoveries, while researchers negative to intuition claimed that these judgments are always poorer than those made by analysis and often biased. Emphasis on analytic cognition became the dominant view but those who were negative to analysis identified several problems with this process. A neutral position argued that intuition and analysis are the outcomes of two different information processing systems that work integratively, in the name of reliance on the one or the other in different situations. How much reliance that is put on one system is determined by different components, as the individual’s disposition, demands by the task, and experience (Epstein, Pacini, Denes-Raj, & Heier 1996; Pacini & Epstein, 1999). In the Brunswickian tradition, Cognitive Continuum Theory developed by Hammond (1996) suggested that the intuitive and the analytical modes are the outcomes of two different information-processing systems that work in an integrative way. The focus lies on the distinction between an intuitive and analytic way of thinking and these modes are located at each end of a continuum. Therefore, intuition and analysis occur at different locations on this continuum and in the middle both modes are active, termed quasi-rationality. In the Brunswik-paradigms the cognitive continuum theory focuses on the relationship between the cognitive system and the task system.

Heuristics and Biases

In the age of enlightenment it was commonly assumed that humans are rational. People update their beliefs in the light of new evidence and the model that is usually considered normative in the psychological literature is Bayes’ theorem, which multiplies the prior odds by the likelihood ratio to produce the posterior odds (Kahneman & Tversky, 1973). Already during the romanticism critics of rationality emphasized the irrationality of unconscious factors that control our behavior, for example in the form of Sigmund Freuds’ influential psychoanalysis. The starting point for the first psychological models was the
normative one, which was then modified to account for human behavior (Peterson & Beach, 1967). In the latter half of 20th century, however, even stronger critics of human rationality appeared suggesting that, even if we are not affected by strong unconscious impulses, we do not have the mental capacity to optimize and behave in accordance with normative models (Kahneman & Tversky, 1982; Gilovich et al., 2000). It was proposed that the cognitive processes that underlie judgment and decision making do not coincide with normative principles, instead they are based on simplifying heuristics that cause cognitive biases, systematic errors relative to normative models (Tversky & Kahneman, 1982). Relative to the earlier research one change was an increased emphasis on how people perform these tasks. Tversky and Kahneman argued that the psychological processes underlying judgment bore little or no resemblance to normative models. The idea of bounded rationality presented by Herbert Simon (1992) suggested that we are, at best, rational given the information and the limitations we have: problem solving is not optimized, but rather satisfied. Moreover, we tend to choose adaptively between different decision rules depending on cost-benefit considerations (Payne, Bettman, & Johnson, 1993).

Studies of rationality want to determine if behavior conform with the axioms of probability theory and decision theory, in other words, if the behavior is coherent. One example of such a norm is that preferences should be transitive (if A>B and B>C then A>C). If you do not obey this rule your preferences are intransitive and incoherent. Likewise, probability judgments should be additive (the probability for all exclusive events in the outcome space must sum to 1) and more generally fulfil the rules of probability theory (Tversky & Koehler, 1994).

Later research provided further interpretations of the idea of bounded rationality. Tversky and Kahneman’s (2000) Prospect theory predicts that our experiences of losses and gains differ. The theory assumes two different phases, an early editing phase and a second evaluation phase. The editing phase consists of a preliminary analysis of the offered prospects, which often yields a simpler representation of these prospects. In the second phase, the edited prospects are evaluated and the prospect with the highest value is chosen. The function of the editing phase is to organize and simplify the judgment. This simplification often leads to irrational choices. Another interpretation of bounded rationality is support theory (Tversky & Koehler, 1994), a psychological version of probability theory, in other words, a modified axiomatic model of probability.

The emphasis in the heuristic-and-biases program on studying the cognitive processes underlying judgment and decision making behavior represents important progress. Comparing the latter half of 20th century perspective and the research during the 90s there are differences. The latter stresses the importance of environment in determining what is normative and why people behave
as they do. Content and context matter in normative and descriptive ways. A task might have multiple reasonable normative responses and the focus shifts from whether or not responses are correct to what is the best explanation of the behavior (Einhorn & Hogarth, 1981; Gigerenzer, 1991a; McKenzie, 2003, Oaksford & Chater, 1994).

Ecological Rationality

Despite the success of the heuristics-and-bias paradigm, a significant amount of criticism was raised around 1980 (Gigerenzer & Murray, 1987). In recent years ecological rationality has provided a successful alternative interpretation of the idea of bounded rationality (Gigerenzer, 1993; Gigerenzer & Goldstein, 1996). This program criticizes the previous conclusions about human irrationality, the vagueness of the heuristics, and the lack of specificity in regard to when a given heuristic is used (Gigerenzer & Murray, 1987; Wallsten 1980). It is proposed that the negative view of human performance implied by this research is misleading (see Einhorn & Hogarth, 1981 for an example). Previous approaches claim either that human judgments coincide with normative models or deviate from normative models producing human irrationality (the heuristics and biases program). Both the programs of classical rationality and the heuristics and biases program thus accept probability theory, statistics and decision theory as normative, but disagree about whether humans can satisfy the norms. The traditional heuristics and biases paradigm, however, ignores the role that the environment plays in shaping human behavior. Ecological rationality takes a step back and addresses the question of how an algorithm can exploit the structure of real environments (Gigerenzer, Todd, & the ABC-group, 1999). Focusing on environmental structure to understand behavior is not new, of course (Brunswik, 1956; Gibson, 1979; Hammond, 1955; Marr, 1982; Simon, 1955, 1956, Toda, 1962, Tolman & Brunswik, 1935). Gigerenzer (2001) was inspired by Herbert Simons (Simon, 1956) idea that “Bounded rationality” consists of two central components. First, because of limitations in time, information, and computational ability, we cannot rely on complex optimization procedures, instead we have to use heuristics. Second, these heuristics are adjusted to specific structures in the environment which they exploit, and they cannot be understood independently of these structures. Gigerenzer and colleagues suggest that in the heuristics and biases program the ecological part of the theory has been ignored. Maybe one can view the heuristics and biases program as studying cognition independently of environmental considerations, as in a vacuum, and this can lead to misleading conclusions (Lamberts & Goldstone, 2003). The theory of ecological rationality is inspired by evolutionary theory as it commonly occurs in biology and ecology and, also, now more often in cogni-
tive psychology (Klein, Cosmides, Tooby, & Chance, 2002). Similar ideas have also been applied to explain the overconfidence bias in general knowledge (Gigerenzer, Hoffrage & Kleinbölting, 1991; Juslin, 1994; Juslin, et al., 2000), hindsight bias (Winman et al., 1998) and base-rate neglect (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995).

Categorization Research

Early categorization research either stressed category prototypes; humans are supposed to average their experience into a category prototype and classify the new item as a specific category memory if it is similar enough (Homa, Rhouds, & Chambliss, 1979; Mervis & Rosch, 1981; Posner & Keele, 1968, 1970; Rosch, 1976) or rules; humans mentally integrate cue-criterion rules referring to a number of specific features of the objects when presented with a new exemplar (Bruner, Goodwin, & Austin, 1966). Later evidence suggested that prototype models do not account for the cognitive processes completely and that prototypes are insufficient as a principle when organizing objects into categories. For example, if categorization was only based on prototypes, humans would be unable to learn many categories (Medin, Dewey, & Murphy, 1983; Medin & Schaffer, 1978; Medin & Smith, 1981).

Recent models often stress exemplar memory (Estes, 1994; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984; Nosofsky & Johansen, 2000; Nososky & Palmeri, 1997) or multiple representation levels (Ashby et al., 1998; Erickson & Kruschke, 1998). Overall, it appears that exemplar models have proven extremely useful in many research areas (Delosh, Busemeyer, & McDaniel, 1997; Dougherty, Gettys, & Ogden, 1999; Hintzman, 1986; Juslin & Persson, 2002; Smith & Zarate, 1992).

Recent research on categorization and memory has produced several models that assume that people have several qualitatively distinct and competing levels of knowledge representations (Ashby et al., 1998; Erickson & Kruschke, 1998; Sloman, 1996, Nosofsky, Palmeri, & McKinley, 1994, Nosofsky & Palmeri, 1997). Logan (1988) proposed a model that posits functionally independent rule-based and exemplar-based systems that race to completion (see also Palmeri, 1997). He proposes an instance-based model of automaticity in which people start with explicit algorithms for performing skilled actions. If the skilled actions are successfully performed, they are laid down in memory as instances and these instances are later retrieved and used to perform the tasks. A rule-plus-exception model (RULEX) of classification learning was proposed by Nosofsky, Palmeri, and McKinley (1994) in which people learn to classify objects by forming logical rules and remembering occasional exceptions to these rules. Palmeri (1997) found evidence for shifts from rules to exemplars in a
Erickson and Kruschke (1998) similarly presented evidence for both rule induction and exemplar memory, as well as a connectionist model that specifies the mechanism for combining rule and exemplar representations. Ashby et al. (1998) postulate a conscious rule-based level and a procedural level where responses are directly associated with perceptual inputs. Rule-based and exemplar-based modules may be functionally independent, but the outputs of these systems may compete based on strength of evidence rather than completion time (for another recent example, see Klein et al., 2002). However, only recently have theories about competing levels of knowledge representation been related to judgment and decision making (see Kahneman & Frederich, 2002).

The most successful cognitive models in categorization and multiple-cue judgment are profoundly different in terms of the computations they assume, for example, whether knowledge comes as explicitly abstracted cue-criterion relations (Einhorn et al., 1979) or as exemplar memory (Nosofsky & Johansen, 2000). However, the structure of the task in categorization and multiple-cue judgment is similar. Both tasks involve known aspects of an object or a situation (“features” or “cues”) and presume elaboration of these aspects in one way or another to infer an unknown variable (e.g., a “category” or a continuous “criterion variable”). In spite of the similarity between the two tasks, there have been few connections between these areas.

**Purpose of the Thesis**

The purpose of the thesis is to investigate what factors affect which cognitive processes that are used, assuming that people have several qualitatively distinct and competing levels of knowledge representations. Study 1 investigates why the theoretical conclusions are so different in categorization and multiple-cue judgment, considering that the task structure is similar in the two domains. Study 2 examines if social factors can shape which process mode that is used and in Study 3 the purpose is to investigate the effects of different task structures. In this thesis I focus on the exemplar-based model and the cue abstraction model because they are the most successful models in research on categorization and multiple-cue judgment.
TASK, COGNITIVE MODELS, AND PREDICTIONS

A Model Task

The task used in the studies was a simple model task developed to investigate knowledge representations in the lab and requires the participants to use four binary cues to infer a continuous criterion (Jones, Juslin, Olsson, & Winman, 2000; Juslin et al., 2003; Juslin, Olsson, & Olsson, 2003: Study 1). The cover story involves judgments of the toxicity of subspecies of the exotic (but fictitious) Death Bug. The subspecies vary in concentration of poison from 50 to 60 ppm (a continuous criterion), where a concentration below 55 ppm is harmless but a concentration above 55 ppm is lethal (a binary criterion, harmless vs. dangerous). Toxicity can be inferred from four cues of the subspecies (e.g., length of their legs, color of the back), see Figure 1 for examples of the stimuli.

The task structure is summarized in Table 1. The binary cues $C_1$, $C_2$, $C_3$, and $C_4$ take on values 1 or 0. If the binary cue has value 1, it suggests high toxicity level, and if the cue has value 0, it suggests low toxicity level. The toxicity $c$ of a subspecies is a linear, additive function of the cue values:

$$c = 50 + 4 \cdot C_1 + 3 \cdot C_2 + 2 \cdot C_3 + 1 \cdot C_4,$$

(1)

where $C_1$ is the most important cue with a coefficient of 4 (i.e., a relative weight .4), $C_2$ is the second to most important cue with a coefficient of 3, and so forth. The binary criterion $b$ is formed from the continuous criterion by assigning $c<55 \ b=0$ (harmless), $c>55 \ b=1$ (dangerous), and $c=55$ randomly as $b=1$ or $b=0$. A subspecies with feature vector $(0, 0, 0, 0)$ thus has 50 ppm and is harmless; a subspecies with feature vector $(1, 1, 1, 1)$ has 60 ppm and is dangerous. The criteria for all 16 subspecies (i.e., possible cue configurations) are summarized in Table 1.

In Study 3 a non-linear function of the cue values is also considered. In these experiments, a quadratic function of the criteria in the linear condition (Eq. 1) with the same range (50 to 60 ppm) was the function in the non-linear judgment task. The criterion $c$ is again computed by assigning the most important cue, $C_1$, the largest weight and the least important cue the smallest weight, $C_4$. For example a bug with features 1 0 1 1 has a toxicity of 58 in the linear condition and a toxicity of 56.4 in the non-linear condition. A normally and independently distributed random error was added in some conditions. The variance of the random error was chosen to produce a .9 correlation between the four cues and the observed criterion (see Juslin et al, 2003: Study 1).
In a *training phase*, the participants encounter 11 subspecies. Participants make either *binary judgments* about the toxicity of each subspecies (i.e., “harmless” or “dangerous”), or *continuous judgments* about the toxicity of each subspecies (e.g., “The amount of poison is 57 ppm”) as illustrated in Figure 1. In a *test phase*, the participants make judgments for all the 16 subspecies and without feedback (see Table 1). In all three studies reported below either exemplar memory or cue abstraction can be applied to solve the task.

*Figure 1. Panel A. An analogue subspecies example of the Death bug with the binary question. Panel B: A propositional subspecies example of the Death bug with the continuous question.*
Table 1. Structure of the Task Used in the Experiments. In the Constrained Training Sets, Exemplars Denoted with "Training" are Presented in the Training Phase, Where as Exemplars Denoted with "Exp." and "Interp." are Introduced in the Test Phase

<table>
<thead>
<tr>
<th>Exemplar</th>
<th>Cues</th>
<th>Criteria</th>
<th>Constrained training set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C₁</td>
<td>C₂</td>
<td>C₃</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>


Cognitive Models

Cue Abstraction Model and Exemplar Model

It is common to distinguish between two basic levels of knowledge representation, the levels of cue abstraction and exemplar-based representations (Erickson & Kruschke, 1998; Hahn & Chater, 1998; Nosofsky & Johansen, 2000; Nosofsky, Palmeri, & McKinley, 1994; Smith, Patalano, & Jonides, 1998). The cue abstraction representation suggests that human categorization relies on abstraction of sets of mental rules, referring to a number of specific features of the objects. On this level we expect intentional, conscious, and controlled cognitive processes and representations that are explicit cue-criterion rules that rep-
resent abstract knowledge, retrieved from semantic memory (Schacter & Tulving, 1994). In this thesis we represent rule-based judgments by the cue abstraction model. The cue abstraction model assumes that the participants abstract explicit cue-criterion relations at the time of training that become the objects of mental cue integration at the time of judgment. The rules for each cue are stored in memory and specify the sign of the relation and the importance of the cue with a cue weight. When presented with a new exemplar, the participants retrieve rules connecting cues to the criterion. The judgment is a linear additive combination of such cue criterion relations. The binary judgment involves classification of subspecies into two categories based on their continuous criterion \( c, c < 55 \) is harmless \((b=0)\) and \( c > 55 \) is dangerous \((b=1)\). In this case, the cue abstraction model suggests that the participants consider the cues, retrieve their weight or importance from memory, and perform a mental analogue of logistic regression (see Figure 2, and Study 1 for further details). This view of the cognitive processes is often implied in research on multiple-cue judgment (e.g., Einhorn et al., 1979).
Figure 2. Illustration of how the continuous judgments are computed according to the exemplar model (the upper part of the figure) and the cue abstraction model (the lower part of the figure). From “Exemplar Effects in Categorization and Multiple-Cue Judgment” by P. Juslin, H. Olsson, & A-C., Olsson, 2003. Journal of Experimental Psychology: General. Copyright from the American Psychological Association. Copyright by the American Psychological Association. Adapted with permission.
Exemplar models assume that the participants make judgments by retrieving similar exemplars from long-term memory. Exemplar memory is rapid, similarity-based, and relies on holistic memory traces primarily retrieved from episodic memory (Schacter & Tulving, 1994). According to the context model of perceptual classification (Medin & Schaffer, 1978) the probability of categorization equals the ratio between the summed similarity of the judgment probe to the exemplars of the target category and the summed similarity to all exemplars. As can be seen in upper part of Figure 2, the judgment is a weighted average of the criteria of the stored exemplars, where the weights are the probe-exemplar similarities. In categorization research, exemplar memory has been emphasized in recent years (Estes, 1994; Krusckhe, 1992; Medin & Schaffer, 1978; Nosofsky, 1984; Nosofsky & Johansen, 2000; Nosofsky & Palmeri, 1997). Basically, with cue abstraction, the rules refer to single cues as abstracted across exemplars and exemplar representations correspond to concrete, holistic representations of the individual objects in a category.

The computations implied by the exemplar and cue abstraction models are distinctly different. Figure 2 represents the input data as a matrix where the rows correspond to encountered exemplars of the bug and the columns to cues, for example, long green legs or short dark blue nose. The middle section of Figure 2, assumes the experience of four such bug exemplars, each described by presence of absence of the four binary cues (the “Experience” box). In this task the continuous criterion equals $50 + 4 \cdot \text{cue 1} + 3 \cdot \text{cue 2} + 2 \cdot \text{cue 3} + 1 \cdot \text{cue 4}$. The exemplar model implies representation of the concrete bug exemplars that have been encountered, as illustrated in the upper part in Figure 2 (episodic memory box). When a new bug (probe $p$) is presented its similarity to each of the stored bug exemplars is computed by the similarity rule of the original context model (Medin & Schaffer, 1978) and the judgment is a weighted average of the criteria (i.e., of the toxicity of the stored bug exemplar).

The cue abstraction model implies representation of abstracted cue weights of the bugs. It is assumed that the importance or weight of each of the four cue dimensions has been abstracted in training, as illustrated in the lower part of Figure 2 (semantic memory box). We assume that the intercept (50) and the optimal weights have been derived (i.e., 4, 3, 2, 1) and the sum of the cue values multiplied by these weights thus produce the criteria. When probe $p$ [$1, 1, 1, 1$] is presented the judgment is a weighted average of the cues of the probe plus the intercept, where the weights signify the importance of each cue (i.e., $50 + 4 \cdot 1 + 3 \cdot 1 + 2 \cdot 1 + 1 \cdot 1 = 60$). The judgment is thus made by retrieving abstracted cue weights that are integrated. Both models involve a linear additive combination, although in different ways: The exemplar model implies combination of criteria of concrete exemplars. Cue abstraction implies combi-
nation of the observed cues, and produces explicit representations of the structural relations between cues and criterion.

A growing insight is that judgments also can be based on a mix between analytic and intuitive thought—so called quasi-rationality, a key-point of cognitive continuum theory of multiple-cue judgment (Hammond, 1996). A complete account of categorization and multiple-cue judgment therefore involve the interplay between multiple qualitatively distinct representations including both explicit rule-based processes and similarity-driven memory processes (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Ell, 2001; Erickson & Kruschke, 1998; Jones, Justlin, H. Olsson, & Winman, 2000; Justlin, Jones, H. Olsson, & Winman, 2003; Logan, 1988; Nosofsky, Palmeri, & McKinley, 1994; Sloman, 1996). Kahneman (2003) suggests that the two systems have operating characteristics, proposing that intuitive judgments deal with both concepts and percepts and can be evoked by language. The intuitive operation generates impressions of the attributes of the object. Those impressions are neither voluntary nor verbally explicit. Intuition is applied to impression judgments that are not modified by analytic or reasoning thought. Judgments are always intentional and explicit in a sense when they are not verbally expressed (Kahneman, 2003). The paradigm of multiple representational levels is used in many areas. One example is cognitive neuropsychology where multiple memory system is investigated related to multiple categorization system (Ashby et al., 1998). Another example is Reber’s (1989) task of implicit learning in artificial grammar learning where participants cannot explain the rules they are using.

Cue-criterion relations are more often probabilistic in research on multiple-cue judgment. Probabilism is one of the theoretical keypoints of the Brunswikian tradition (Cooksey, 1996). In a probabilistic task the same cues and criteria do not reoccur which suggests that exemplar memory becomes less prevalent in such tasks. Moreover, Yamauchi and Markman (2000b) reported that the participants had greater difficulty learning a category when the instances varied from trial to trial. In contrast, a deterministic task allows perfect accuracy by retrieving exemplars that have been observed previously and the same exemplars with exactly the same cues and criteria reoccur time after time which increase the probability that those exemplars are stored and lead to more use of exemplar memory. Most category learning experiments involve deterministic tasks, with a small set of stimuli that is repeated. Smith and Minda (2000) noted that these conditions invite a mundane form of exemplar memory, implying that participants memorize each stimulus with its category name. The process then becomes a name retrieval for each stimulus without establishing a category in a more profound sense. However, even if determinism is well
captured with exemplar memory, it can facilitate for both exemplar memory and cue abstraction.

Quantitative Predictions

Figure 3 illustrates the predictions by the cue abstraction model and the exemplar model for either judgments of a binary criterion (Figures 3A & 3B) or a continuous criterion (Figures 3C & 3D) when applied to all 16 exemplars in Table 1. With the binary criterion you classify or categorize the objects as dangerous or harmless, while the continuous criterion involves judgment of a continuous variable, toxicity. Figure 3 illustrate that in certain conditions both models predict perfectly accurate judgments. The reasons for this accuracy are different in the two models. With the cue abstraction model, the accuracy derives from correct knowledge of the cue weights and error-free integration of this knowledge into a judgment. With the exemplar model, on the other hand, the accuracy derives from retrieval of stored exemplars, where only identical exemplars are allowed to have a strong effect on the judgment. Binary judgments are typically relevant to studies of categorization where exemplar models have been influential and continuous judgments applies to studies of multiple-cue judgment where cue abstraction models has been the most common implicit or explicit interpretation.

In studies where all exemplars are presented in the training phase it is virtually impossible to discriminate the different models from each other, because they produce the same predictions (see Figure 3). To discriminate the models from each other, exemplars in the training phase need to be withheld and presented for the first time in the test phase, as illustrated in Figure 4, where a constrained training set of exemplars has been used. The essence in the discrimination between the two models concerns the ability to extrapolate (Delosh, Busemeyer, & McDaniel, 1997). If you rely on cue abstraction, and you have identified the cues in the right way and weight them together appropriately, judgments for the most extreme exemplars is always more extreme than the judgments for the second to most extreme exemplar – you have the ability to extrapolate (see Figure 4A and 4C). For example, if you have identified all four cues that go with an increase in toxicity, you will make a more extreme judgment for an exemplar with all four cues present than for an exemplar with only three cues present, even if you have never seen the exemplar with four cues before. On the other hand, if you rely on exemplars you have difficulty with judging the new extreme exemplar in the right way, because the retrieval of exemplars with values in the interval 51-59 can never produce a value outside of this range – inability to extrapolate (see Figure 4B and 4D). Moreover, the exemplar model predicts that the judgments for the new exemplars in
the middle will be judged erroneously because those exemplars are more similar
to objects that are members of the other category. With an exemplar-based
model, the judgments are always more exact for old than for new exemplars,
because old exemplars can benefit from stored identical exemplars with the true
criterion value. Because the cue abstraction model can allow old-new differ-
ences at the cost of extremely poor fit to the other exemplars, these measures
need to be complemented by model fit to the entire data pattern for all exem-
plars, the reader is referred to Study 1 (Juslin, Olsson, & Olsson, 2003) for fur-
ther details on how these predictions are computed.
Figure 3. Predictions for judgments of a binary (A, B) and a continuous (C, D) criterion with the complete training set. Panel A and C: Cue abstraction model with no noise and noise for the complete training set. Panel B and D: exemplar model with similarity parameters only allowing retrieval of identical exemplars ($s = .0001$) or similar exemplars ($s = .1$). From “Exemplar Effects in Categorization and Multiple-Cue Judgment” by P. Juslin, H. Olsson, & A-C., Olsson, 2003. *Journal of Experimental Psychology: General*. Copyright by the American Psychological Association. Adapted with permission.
Figure 4. Predicted judgments for a binary (A, B) and a continuous (C, D) criterion with the constrained training set. Panel A and C: Cue abstraction model with noise for the constrained training set. Panel C and D: Exemplar model with similarity parameter $s = .1$. From “Exemplar Effects in Categorization and Multiple-Cue Judgment” by P. Juslin, H. Olsson, & A-C. Olsson, 2003. *Journal of Experimental Psychology: General*. Copyright from the American Psychological Association. Copyright by the American Psychological Association. Adapted with permission.
EMPIRICAL STUDIES

In the last decades, two research paradigms have emerged in cognitive science that addresses the issues of object judgment, categorization learning and multiple-cue judgment. Both tasks involve known aspects of an object or a situation, presume elaboration of these aspects in one way or another to infer an unknown variable (a category or a criterion variable). While categorization research, with its focus on cognitive modelling continues to flourish the more descriptively oriented multiple-cue judgment research largely disappeared in the 1980’s (Brehmer, 1994; Cooksey, 1996). On principled grounds it is often difficult to tell the tasks apart, and yet there is rarely cross-reference between the two literatures. As we have seen, in spite of the similarity of the tasks, different cognitive theories have dominated in these two paradigms. The three studies in this thesis, Study 1, 2, and 3 are based on these paradigms and the idea that there exist multiple representational levels (e.g., Ashby et al. 1998). The studies investigate why the theoretical conclusions are so different in categorization and multiple-cue judgment, considering that the task structure is similar in the two domains. In the three studies we also investigate if different factors such as stimuli format, social interaction, and the structure of the task environment shape which process that is used.

Study 1

The question addressed was why the theoretical conclusions from categorization and multiple-cue judgment research are so different, considering that the task structure is similar in the two domains. There are two alternative possibilities; (a) the differential emphasis arises from accidental differences in the research and modeling methods (b) the differences that characterize the tasks as they typically appear in the two domains are conducive of genuine differences in the cognitive processes. With Study 1, we wanted to investigate if the two judgment tasks indeed promote a shift from exemplar memory to abstraction of explicit cue-criterion relations when the task is changed from a typical categorization task to a typical multiple-cue judgment task. This presumes a cognitive system with multiple qualitatively distinct representations that compete to control behavior depending on the task requirements.

Study 1 contained 3 experiments and the multiple-cue judgment task presented earlier was used. The general method was the same for Experiments 1 and 2 except for the difference in the criterion variable, binary versus continuous. The participants in Experiment 1 judged the binary criterion and the participants in Experiment 2 judged the continuous criterion. The participants in
Study 1 were given written instructions that informed them that there were different subspecies of a Death bug that was to be classified into harmless or dangerous. The subspecies differed in toxicity between 50 and 60 ppm, while toxicity below 55 was harmless and toxicity above 55 was dangerous (Experiment 1). In Experiment 2 the toxicity of the subspecies was directly estimated as a number between 50 and 60. The training phase in Experiment 1 provided trial-by-trial outcome feedback about the binary criterion (“This bug is dangerous”) and in Experiment 2 about the toxicity (“This bug has toxicity 57 ppm”). The design of Experiment 1 and 2 was a 2x2 factorial design where presentation format at training (analogue vs. propositional: between subjects) was crossed with presentation format at test (analogue vs. propositional: within subjects). In both experiments, half of the participants were trained with analogue stimuli and the other half with propositional stimuli. All participants were tested with both presentation formats. Two different training sets were used and counterbalanced in the two experiments (see Table 1 above). The analogue format presented the cue values in terms of a picture of the Death bug and the propositional format provided the four cue values stated as four propositions (see Figure 1). Experiment 3 was a between group design and the same as Experiment 2 with regard to instruction and stimuli except that an independently distributed random error was added to the continuous criteria (probabilism). Only one training set was used in Experiment 3 and all participants also made both training and test judgments with the same presentation format, half of them consistently receiving propositional stimuli and the other half analogue stimuli (see Figure 1).

Experiment 1 involved binary criterion judgments, a condition that provides a baseline where the exemplar model is expected to dominate (see Nosofsky & Johansen, 2000). With a continuous criterion in Experiment 2, the feedback is more informative about the task structure and we expected this to increase the prevalence of cue abstraction. We proposed that the judgments of a continuous criterion are more conducive of mental cue abstraction because, while the feedback for binary judgments is often insufficient to induce the task structure, the feedback for continuous judgments generally provides constraints that are informative for inferring the relations between cues and criterion. Multiple-cue judgment in general involves the judgment of a continuous criterion (Brehmer, 1994; Cooksey, 1996; Hammond, 1996). In view of memory principles like encoding specificity (Tulving, 1983) and transfer appropriate processing (Morris, Bransford, & Franks, 1977) we also expected a direct use of memory in the form of retrieval of similar exemplars to be particularly important when the test and training conditions match. The training context would affect the cue abstraction model less because of its reliance on abstracted representations and explicit processes. Because Experiment 1 contains the less in-
formative feedback, however, it might conceal the beneficial effect of matching training and test stimuli by making exemplar memory dominating regardless of the training-test match. Experiment 2 is more amenable to both processes and should be more sensitive for testing this hypothesis. Moreover, we predicted that exemplar processes should be especially pervasive in the analogue (holistic) stimulus format. Experiment 3 involved judgments of a continuous criterion, like Experiment 2. While the task is deterministic in Experiment 2, it is probabilistic in Experiment 3; a normally and independently distributed random error is added to the criteria to produce a correlation of .9 between cues and criteria. We expected exemplar memory to be more prevalent in a deterministic task, because the repetition of identical exemplars particularly invites exemplar memorization (Smith & Minda, 2000).

Results

In Experiment 1, training with propositional stimuli produced more accurate judgment in the test phase. One possible explanation for the marginally higher accuracy and somewhat better fit of the exemplar model in the propositional condition is that the propositional format elicited more efficient training. Because automatic behavior often is aligned with exemplar memory, the behavior may have become more automatic in the propositional condition, producing higher accuracy and a clearer dominance for exemplar memory (Logan, 1988; Nosofsky & Palmeri, 1997). In all conditions the judgments of the binary criterion were dominated by exemplar memory, with large differences between old and new exemplars. This conclusion was confirmed by negative exemplar indices and quantitative model fits.
In Experiment 2, the model fits and the graph (see Figure 5B) suggest a mix of the two processes, in contrast to the dominance of exemplar processes in Experiment 1. The performance data suggest similar results as in Experiment 1. The propositional conditions produced better performance, especially when training and test conditions coincide. The function relating mean judgments to the criterion is fairly linear as predicted by the cue abstraction model but the local slopes and the exemplar index suggest the presence of exemplar processes. Individual participant data and analysis in terms of exemplar indices, suggest large individual differences in Experiment 2, with some participants clearly relying on cue abstraction while others rely on exemplar memory as illustrated in Figure 6. In contrast to our hypothesis, there was no significant increase in the use of exemplar memory when test and training format match or when the presentation format was analogue in Experiment 1 and 2.
The results of Experiment 3 suggest that cue abstraction increases in a task where the cue-criterion relations are probabilistic. Figure 7 suggests stronger exemplar effects in the deterministic condition (Experiment 2) than the probabilistic condition (Experiment 3) and a clear inability to extrapolate in the deterministic condition.

Figure 6. Panel A: A participant relying on exemplar memory along with his or her exemplar index. Panel B: A participant guided by cue abstraction along with his or her exemplar index. A negative exemplar index suggests exemplar processes, see Study 1 for details. From “Exemplar Effects in Categorization and Multiple-Cue Judgment” by P. Juslin, H. Olsson, & A-C., Olsson, 2003. Journal of Experimental Psychology: General. Copyright by the American Psychological Association. Adapted with permission.

The results of Experiment 3 suggest that cue abstraction increases in a task where the cue-criterion relations are probabilistic. Figure 7 suggests stronger exemplar effects in the deterministic condition (Experiment 2) than the probabilistic condition (Experiment 3) and a clear inability to extrapolate in the deterministic condition.

Figure 7. Panel A: Mean judgments for all exemplars in Experiment 2 (deterministic condition). Panel B: Mean judgments for all exemplars in Experiment 3 (probabilistic condition). From “Exemplar Effects in Categorization and Multiple-Cue Judgment” by P. Juslin, H. Olsson, & A-C., Olsson, 2003. Journal of Experimental Psychology: General. Copyright by the American Psychological Association. Adapted with permission.
The model predictions for Experiment 1, 2 and 3 were computed from the data from the latter half of the training phase (the last 110 training trials) for the 11 training exemplars; based on mean judgments in the continuous conditions and response proportions in the binary conditions. The models were then applied with the parameters fitted to the training data to all 16 subspecies in the test phase. This method minimizes the risk of overfit, and gives cross-validation for old items and genuine predictions for new items. The benefit of model fit is that it takes the overall structure of the data into consideration. The graphs of the data show the performance, in contrast, model fit suggests which knowledge process that is used. The cue abstraction model corresponds to a regression model and allows analytic derivation of the best fitting parameters, while parameters for the exemplar model with minimal squared sum of error as error function is obtained by the Quasi-Newton method. The model fit measures are RMSD, root mean square deviation and \( r^2 \), the coefficient of determination for the mean judgments (continuous condition) or the response proportions (binary condition). The model fits is dominated by exemplar memory in Experiment 1 with partial support for both kinds of processes in Experiment 2. Model fits in Experiment 3 clearly suggest that cue abstraction increases in a task where the cue-criterion relations are probabilistic (Figure 8).

Figure 8. Model fit for each of the experiments in Study 1 with data collapsed across analogue and propositional training. The left axis provides the Root Mean Square Deviation (RMSD) between predictions and data points; the right axis the coefficient of determination (\( r^2 \)) (i.e., the proportion of variance linearly accounted for by the models).
Discussion

The question addressed in Study 1 was why the theoretical conclusions from categorization and multiple-cue judgment research are so different, considering that the task structure is similar in the two domains. Research on categorization emphasizes the role of exemplar memory (Nosofsky & Johansen, 2000) and multiple-cue judgment research has stressed mental cue abstraction (Einhorn et al., 1979). The two kinds of processes are qualitatively different in terms of the computations, the cognitive processes, and the neural substrate they imply, despite the similar structure of the tasks. We hypothesized that the differences that characterize the tasks as they typically appear in the two domains are conducive of qualitatively different cognitive processes. The results in Experiment 1 and 2 suggest that the change from a binary to a continuous criterion indeed produces a shift towards mental cue abstraction, but exemplar memory continues to play a role as long as the task is deterministic and there are few exemplars. Thus, when the feedback is informative enough, the participants appear to exploit the possibility to infer explicit representations of cue-criterion relations.

The encoding specificity principle (Tulving, 1983) suggests that a direct use of exemplar memory should be particularly pervasive when test and training condition match, and processes that rely on controlled thought and abstract representations of cue-criterion relations were expected to be more detached from original context of learning. We expected that the use of exemplar memory should be pervasive when test and training match, but we found no consistent support for the hypothesis when looking at exemplar index and model fits across Experiment 1 and 2. Maybe different manipulations are needed to promote a shift in the cognitive processes, for example by changing the presentation format for an expert from that used in his or her daily expertise activity. In regard to the effects of stimulus format (analogue and propositional), the results showed no main effects on the balance between exemplar memory and cue abstraction, and the tendencies observed often pointed in the opposite direction to the one predicted – towards stronger exemplar effects with the propositional format. This may be a side effect of the more efficient learning with the propositional format (Logan, 1988). Participants paid more attention to the task and the stimuli resulting in a more efficient learning that leads to automatization of the processes (Logan, 1988, Nosofsky & Palmeri, 1997). The results also revealed considerable individual differences, with some participants predominantly relying on exemplar memory or mental cue abstraction and others on a mix between the two processes. This mix of processes within and between participants provides one explanation and interpretation of the notion of quasirationality that occurs in multiple-cue judgment research (Brehmer, 1994),
where intuition and analysis interplays. These results also concur with the increased attention paid to multiple and qualitatively distinct representation levels in categorization research.

The interpretation of Experiment 3 is that the deterministic task, where participants are exposed to a small set of exactly the same exemplars, promotes memorization of these exemplars that can later be used in judgment. In a probabilistic task, the same exemplars do not reappear except by chance and this is less conducive of exemplar memory, creating a shift towards cue abstraction (see Smith & Minda, 2000, for similar arguments).

The results reported in this study suggest that people have an inclination to abstract explicit representations whenever possible (a rule bias, cf. Ashby et al., 1998), where exemplar memory acts as a back-up system in tasks where explicit representations of cue-criterion relations cannot be abstracted or when behavior has become automatic (Logan, 1988).

**Study 2**

In Study 2, we wanted to investigate if social factors can also shape which process mode that is used. Most theories in decision making have focused solely on one kind of process and thus assumed that all processes work in similar ways. Moreover, virtually all research on category learning and multiple-cue judgment is based on single individuals. If the knowledge representation can differ between individuals, can it also be affected by co-operative situations? Dyads often outperform individuals in different tasks, but in general they do not reach the base-line predicted by the statistical aggregation of the performance by members of the dyad working alone (Andersson, 1996). For example, person A may remember ten out of twenty words and person B may also remember ten out of twenty words. When they are working in a dyad they remember twelve out of twenty words, but when we combine the individual recall we would expect fifteen words out of twenty. It has been suggested that this is caused by factors such as social loafing and lack of co-operation between the members of the group (Brown, 1988; Harkins & Szymanski, 1989; North, Linley & Hargreaves, 2001; Sheffard, 1993). These negative effects can be reduced by psychological independence (Budescu & Rantilla, 2000; Ariely, Bender, Dietz, Gu, Wallsten, & Zauberman, 2001). Factors such as group size (Mortensen, 1972) and friends versus non friends (Andersson, 2001) can affect the quality of the communication.

In the context of the cue abstraction and exemplar models of concern here, consideration of the task suggested a number of alternative ways in which people can adapt to the demand for learning to make judgments in dyads. A first possibility is what we refer to as an exemplar pooling effect. This effect is plausi-
ble in a task where the individual participants rely strongly on exemplar memory. When co-operating in dyads, they can store more exemplars in memory together, and this exploitation of memory should lead to both improved performance and superior fit for the exemplar models in the dyad conditions (where the exemplar storage refers to a dyad of individuals, not a single individual). The exemplar pooling effect comes both in a weaker statistical version and a stronger synergetic version. The weaker statistical version refers to the mere aggregation effect when the exemplar memories of two individuals are combined. The synergetic version of the exemplar pooling effect implies beneficial effects of working in dyads over and above the improvement expected from mere aggregation, for example, because of better encoding or storage of exemplars during training or more efficient retrieval at test. The hypothesis of a social abstraction effect predicts a shift towards more cue abstraction in the condition where the participants train to make judgments in dyads. If we abstract explicit representations of cue-criterion relations, this provides knowledge that is explicit and easily communicated by verbal means. It will be easy to verbally explain what specific cues go with high or low toxicity, while it is difficult to communicate the memory of exemplars. Social interaction with the verbal interchange should therefore promote a shift from exemplar-based processes to processes of cue abstraction. Evidence for the social abstraction effect is shown if the data for individual participants is best accounted for with an exemplar model and the data for dyads is best accounted for with the cue abstraction model.

The multiple-cue judgment task was the same as in Study 1. The design in Experiment 1 was a between-group design, where all participants made both training and test judgments with the same presentation format (analogue), with half of them trained and tested individually and the other half trained and tested in dyads. Experiment 2 was a 2x2 factorial design where judgment situation at training and test was factorially manipulated (between subjects). Both factors can take the values “single” or “dyad”, implying 4 cells with two cells where training and test do not coincide (training alone-tested in dyads and training in dyads-tested alone). The hypothesis of a social abstraction effect received a more sensitive test in Experiment 2, where a continuous criterion affords a better chance for cue abstraction, and also allows us to compare statistical dyads based on means of individual judges to real dyads. In both experiments dyads were constructed by participants that already knew each other to lessen the social loafing effect and to facilitate for maximization of verbal communication.

We thus investigated if other processes and knowledge representations are developed if people make judgments individually or in dyads and how performance differs between participants working alone and in dyads. Other ques-
tions were if the exemplar pooling effect is just a statistical aggregation effect at test, and if verbal interaction promotes more cue abstraction.

Results

In Experiment 1 the results indicated that participants in the dyad condition performed better than the individuals (see Figure 9), but dyads are not significantly more consistent than individuals. The exemplar indices provided no support for a social abstraction effect and the data from both conditions were best explained by the exemplar model. The models were fitted to the data by identifying the free parameters in each model that minimized the sum square of error between the proportion of dangerous decisions and the predictions by the models. The model fits in Experiment 1 verify the conclusion from the analysis of exemplar indices by suggesting that the exemplar model is a superior account of the data in both conditions. This together with the improvement in performance suggests an exemplar pooling effect. Because there is no non-arbitrary way to aggregate the binary judgments by two individuals – especially when they disagree – the binary task is not well suited for testing the synergetic version against the statistical version of the exemplar pooling effect. On the other hand, Experiment 2 allows us to compare statistical dyads with real dyads to investigate if mere statistical aggregation explains the better performance in dyads.

Figure 9. Panel A: Mean response proportions as a function of the continuous criterion for the individual participants. Panel B: Mean response proportions as a function of the continuous criterion for the dyads. Compare with predictions in Figure 4A and 4B.
The cue abstraction and exemplar models were fitted to the data from Experiment 2 by identifying the free parameters in each model that minimized the sum square of error between the mean judged toxicity and the predictions by the models. The magnitude of the interaction in the model fits is modest and suggests no signs of increased cue abstraction in the dyads. If anything, the results suggest an increased use of exemplar memory (see Figure 10).

**Figure 10.** Model fit in Experiment 2 of Study 2. The left axis provides the Root Mean Square Deviation (RMSD) between predictions and data points; the right axis the coefficient of determination ($r^2$) (i.e., the proportion of variance linearly accounted for by the models).

The results in Experiment 2 also indicated that participants trained and tested in dyads made judgments more accurately than participants trained and tested individually. This beneficial effect arises from statistically significant additive effects both of training in dyads – supporting the synergetic version of the exemplar pooling effect – as well as of testing in dyads – supporting the aggregation effect implied by the statistical version. Exemplar indices show exemplar effects with an inability to extrapolate in all four cells.

The results showed a beneficial effect of working in dyads and, in contrast to previous research, the performance of dyads therefore surpass the average judgment by two individuals (see Figure 11).
Discussion

We suggested that the effect of social interaction is of importance for understanding our judgments and we wanted to investigate if this factor has an effect on which specific knowledge system that is used in a multiple-cue judgment task. The results in Experiment 1, indicated that the participants in the dyad condition performed better than the individuals as a result of working together. Experiment 1 that involved judgment of a binary criterion, was dominated by exemplar memory. To further examine the hypothesis of a social abstraction effect participant in Experiment 2 made judgments with a continuous criterion and two more conditions were added, individual training-dyad testing and dyad training-individual testing that allowed us to investigate if the beneficial effects of working in dyads, arise in training, at test, or both. The results in Experiment 2 showed clear differences in performance between the individual and dyad conditions and indicated that the participants that were trained and tested in dyads learned to make judgments more accurately than the participants that were trained and tested individually. We found no support for the social abstraction effect, perhaps because of lack of interactive communication between the individuals working in dyads. Further research should concentrate on promoting the communication aspect of the task more carefully to make sure that the participants interact. We may find more evidence for the social abstraction effect in situations where there is even more verbal interaction. Figure 11
showed lower (better) RMSE (Root Mean Square Error between judgments and criteria) in the real dyads training – dyads test condition than for the statistical dyads. This and the significant main effect of training alone or in dyad speaks against that the difference in performance is due to mere statistical aggregation and supports the synergetic exemplar pooling hypothesis that implies additional beneficial effects of working in dyads. In contrast to earlier research, dyads reached the combined base-line level obtained by statistical aggregation of two individuals. One interpretation is that in contrast to previous abstract memory tasks (e.g., remembering word lists), this task draws on remembering in a more meaningful context of problem solving.

The performance is different in dyads and individuals, but we were unable to detect any differences in the representation of knowledge between participants working alone or participants working in dyads. The results suggest that co-operating in dyads can store more exemplars in memory leading to a more efficient exploitation of memory with exemplar-processes dominating the judgments. Another possibility is that when working in dyads the communication between the members of the dyad makes them work more with every exemplar, regarding time and quality of learning, resulting in better storage of exemplars. In sum, we can say that working in dyads increases the ability to correctly learn and classify objects and, in the tasks that we have used, the cognitive processes underlying dyad judgments seem to be primarily exemplar-based.

**Study 3**

Environments contain different structures that people respond to by using either analytic or the intuitive processes, and some environment structures are more difficult to judge than others. The purpose of Study 3 was to examine which cognitive processes people use in linear vs. non-linear task environments. A new cognitive process model, Sigma (Juslin, Karlsson, & Olsson, 2004), was used to derive predictions about how the deep-lying structure of the task affects what representational mode dominates the judgments. Sigma integrates a wide range of models, for example, multiple linear regression (Cooksey, 1996; Hammond & Stewart, 2001), the belief-adjustment model (Hogarth & Einhorn, 1992), and the exemplar models (Nosofsky & Johansen, 2000). The process in Sigma is a sequential adjustment of the judgment, compatible both with a linear, additive cue-integration rule and with linear additive integration of the criteria of retrieved exemplars. The Sigma model is a process model with both cue abstraction and exemplar memory as special cases. The model specifies the interplay between exemplar model and cue abstraction model to make a judgment. Sigma suggests that the judgment can depend on factors in the environment. In short, in the Sigma model the previous estimate from a cue or an
exemplar is adjusted every time a new piece of evidence is considered. Therefore, cue abstraction involves sequential adjustment of an estimate based on cues, where the cue weight is relative to the cues presented so far. In contrast, the weight assigned to an exemplar is given by the similarity of a specific exemplar relative to the similarity to the exemplars retrieved so far. Because of the integration of stored criterion values instead of cue-criterion relations, Sigma is not confined to linear additive task structure and the model can represent any task structure as long as similar exemplars have similar criteria, allowing judgments in non-linear and multiplicative environments (Juslin, Karlsson, & Olsson, 2004). Moreover, in training, because of the limitation to comparing and adjusting two estimates before and after new evidence has been considered, humans are limited to induce the difference between two estimates which corresponds to a linear slope. Therefore, humans can only add the linear effect of single cues. Those limitations allow high performance with cue abstraction in linear environments, but fails in non-linear environments. The key-point is that by considering the characteristic structural properties of a task environment, we are able to predict representational shifts. In additive environments, cue abstraction should dominate, while exemplar model should dominate in multiplicative environments (Juslin et al., 2004).

The questions addressed in this study therefore were: What structure of a categorization task triggers analytic thinking and what structure triggers intuition? Is it possible to predict and control the knowledge processes by manipulating structures in cue-criterion relations? Are humans limited to only obtain explicit knowledge in linear, additive tasks? We tested the following hypotheses: In a probabilistic non-linear task participants are not able to use cue abstraction, because of the limitations to inducing the difference between two estimates and adding the effect of single cues. Instead, they are forced to use exemplar memory. Because probabilism is likely to be a critical factor for performance in a non-linear judgment task, there should be an increase in the performance and an improvement in the fit of an exemplar model when the probabilism is removed creating a deterministic condition. However, Sigma implies that even if the non-linear task is simplified, the participants should not be able to use cue abstraction in a non-linear multiple-cue judgment task. The linear slopes of the estimates imply that non-linear relationships can not be apprehended and the participants are forced to use intuition in the form of exemplar memory to handle the task. The multiple-cue judgment task was the same as in the previous studies, while in some conditions the task has been changed into a non-linear structure. In Experiment 1 we wanted to see if people are able to use rule-based processes when judging objects in a multiple-cue judgment task with a complex non-linear structure. Experiment 1 involves judgments of a continuous criterion in two different conditions. First, a probabilistic linear judgment
task (Juslin, Olsson, & Olsson, 2003: Study 1), and second a non-linear judgment task with the additive linear task as a model. Experiment 2 involved a task intended to facilitate cue abstraction by manipulating the frequency of exemplars in training. Low level exemplars were viewed more often and the function was also changed from a probabilistic to a deterministic task. We expected the deterministic condition to expose Sigma to a stronger test, because it should facilitate cue abstraction, although the condition is still non-linear. The design of Experiment 1 and Experiment 2 was a between-group design, where all participants made the judgments in training and test with the same presentation format. In both experiments, there were two conditions. In Experiment 1, the first condition was a probabilistic non-linear judgment task, while the second condition was a probabilistic linear judgment task. The function in the non-linear judgment task was a quadratic function of the criteria in the linear condition, with the same range (50 to 60 ppm) and maximum value (60 ppm) as in the linear function.

\[ c_{NL} = -2 \cdot c_L^{3/5} + 44 \cdot c_L - 1150, \]

The function \( C_{NL} \) is the level of poison in the non-linear task and \( C_L \) is the linear criterion in Eq. 1 above. For example a bug with features [1101] has a toxicity level of 58 in the linear condition and a toxicity level of 56.4 in the non-linear condition, see Table 1. A normally and independently random error was added, the variance of which was selected to produce a .9 correlation between cues and criterion (see Juslin et al, 2003: Study 1). In Experiment 2, the frequency of the exemplars in training phase was changed by manipulating the frequency in both conditions, where low level exemplars were viewed more often so as to facilitate abstraction of the non-linear relation in the non-linear probabilistic condition. The first condition was a probabilistic non-linear task and the second condition was a deterministic non-linear task.

Results

Experiment 1 involved judgments of a continuous criterion in two different conditions, a probabilistic linear judgment task and a non-linear judgment task modelled on the additive linear task. The performance (Figure 12) was better in the linear condition, suggesting the use of cue abstraction model on group level. Performance for old exemplars was better than for new exemplars in the interpolation and extrapolation range supporting the exemplar model. These results suggest a mixed result within and between participants.
The model fits were analyzed as in Study 1 and reveal a good fit for both exemplar model and cue abstraction model in the linear condition and poorer fit for both models in the non-linear condition (Figure 15). Therefore, model fit was analyzed on the individual level in an attempt to verify the conclusions from the model fit on group level. In the model fit based on the individual level, the data from training for each participant were used to predict the judgments from the test. The results of the individual model fits was very poor because of large deviations from one judgment to the other, and no conclusions of which model that dominated could be drawn. No model is in significant domination within the linear condition. One tentative explanation of these results in the linear condition is the use of the mix of exemplar memory and mental cue abstraction which provides one interpretation of quasi-rationality (Hammond, 1996). The non-linear condition, on the other hand, reveals poor fit by the cue abstraction model. If anything, almost all measures support exemplar memory. One possible explanation of these results, is that the judgment curve has a very flack inclination, and the task is thus to difficult to learn. The result from Experiment 1 supports the hypotheses that the non-linear judgment task is too difficult for judgments based on mental integration of cues and suggests that even exemplar-based processes are hard to adopt.

In Experiment 2 almost all statistical analyses show that the deterministic condition has better results than the probabilistic condition. The deterministic non-linear condition also has better overall performance as compared to the probabilistic non-linear condition in Experiment 1. In contrast, the probabilistic condition has a deviation in the extrapolation range (Figure 14).
Figure 14. Study 3, Experiment 2: Means judgments of the exemplar model and the cue abstraction model in non-linear environment.

Figure 15. Model fit for each experiment in Study 3. The left axis provides the Root Mean Square Deviation (RMSD) between predictions and data points; the right axis the coefficient of determination ($r^2$) (i.e., the proportion of variance linearly accounted for by the models).
The model fits provide support for the exemplar model in both conditions. Model fits on individual level was made but follows the same results as in Experiment 1, with very bad fits caused by large deviations from one judgment to the other, suggesting lack of learning in the training phase. When comparing the two conditions in Experiment 2 with the non-linear condition in Experiment 1 we see that the deterministic condition suggests the best fit for both the exemplar model and the cue abstraction model, apparently refuting the hypotheses that participants in the deterministic condition should produce better fit specifically for exemplar memory (Figure 15). The results in Experiment 2 support the hypotheses that the deterministic condition promotes better performance than the probabilistic conditions in Experiment 1 and 2, with improved fit for the exemplar model. Because the probabilistic condition in Experiment 2 shows poorer results than the non-linear condition in Experiment 1, the frequency manipulation of the exemplars in the training phase does not appear to make the task less difficult. Apparently, the important factor here is whether the task is probabilistic or deterministic (Juslin, Olsson, & Olsson, 2003, Study 1; Smith & Minda, 2000).

Discussion

This study emphasizes the role of the environment in shaping what cognitive process that is used (Brehmer, 1994; Cooksey, 1996; Hammond & Stewart, 2001). People appear to give priority to rule-based processes in many judgment situations (Ashby et al., 1998). The results of Study 3 suggest that cue abstraction is limited to linear and additive structures where a sequential step by step correction of the estimate on the basis of retrieved cue criterion relations allows accurate performance, while exemplar memory works as a back-up system when the judgment task is too complex for cue abstraction (Juslin, Olsson & Olsson, 2003; Study 1). The results in Experiment 1 suggested that the probabilistic non-linear task lead to more use of exemplar memory, supporting the hypothesis and it seems that the use of different cognitive modes is affected by the cue-criterion relations in the environment. Non-linear single cue judgment extrapolation is possible to a degree that supports the use of both cue abstraction and exemplar memory (Delosh et al, 1997). In contrast, a multiple-cue non-linear judgment task using multiplicative cue-criterion relationship forces participants to use an exemplar-based strategy (Karlsson, 2002). Some kind of cue abstraction, perhaps in the form of quasi-rationality, seems to have been used in the linear condition, while if anything exemplar memory was used in the non-linear condition, with no traces of cue abstraction.

In Experiment 2, there were two manipulations; a change of the frequencies of exemplars in the training phase to increase the possibility to abstract the
cue-criterion relations and a deterministic condition removing the error from the criterion in the training phase. This should lead to repetition of identical exemplars and therefore promote exemplar memory (Juslin, Olsson, & Olsson, 2003; Study 1: Smith & Minda, 2000) but also to facilitate for cue abstraction. The results from Experiment 2 support the hypothesis that a deterministic non-linear task will produce better performance results and better model fit for an exemplar-based model than a probabilistic non-linear judgment task. The probabilistic non-linear condition in Experiment 2 has poorer results than the non-linear condition in Experiment 1. Even when the non-linear judgment task is simplified, by frequency manipulation and determinism, it is still too complex for cue abstraction and exemplar memory is the working strategy. Thus, both experiments suggest that cue abstraction is extremely difficult in non-linear judgment tasks where the multiple cues have a quadratic function behind the cue-criterion relationship.

CONCLUSIONS

This thesis indicates that the differences that characterize typical categorization and multiple-cue judgment tasks are conducive of qualitatively different cognitive processes. The change from a binary to a continuous criterion appears to produce a shift towards processes of explicit cue abstraction. Exemplar memory, however, continues to play an important role when the task is deterministic and there are few exemplars. The tendency to infer explicit representations of cue criterion relations appears when the feedback is informative enough. This suggests that exemplar memory acts as a back-up system in tasks where cue criterion relations cannot be abstracted or behavior has become automatic.

We also proposed that the notion of quasi-rationality can be explicated as the mix of rule-based and exemplar-based processes within and between participants, which also concurs with the recent attention paid to multiple and qualitatively distinct representational, levels (Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994; Sloman, 1996).

Working in dyads increased the ability to correctly learn and classify objects. There was a beneficial effect of working in dyads that surpasses the average judgment by two individuals. It seems like criterion change does not produce an equally strong shift towards mental cue abstraction when participants are working in dyads, even though the use of language should promote cue abstraction. Rather dyads appear to exploit memory more efficiently leading to improved performance and superior fit for the exemplar model. These results differ from earlier research in the sense that dyads reached the combined baseline level obtained by statistical aggregation of two individuals (Ariely et al., 2000; Johnson, Budescu, & Wallsten, 2001). In contrast to previous abstract
memory tasks, this task draws on remembering in a more meaningful context of problem solving, which could be one possible interpretation for these results.

The structure of the task decides if we rely on intuitive or analytic thought. It seems that non-linear structures trigger exemplar memory and linear structures favor analytic thinking, in the form of cue abstraction (Juslin et al., 2004). Apparently, this contrast with the predictions by the cognitive continuum theory, which suggests that non-linear condition should induce analytic thinking (Hammond, 1997). As predicted by Sigma cue-criterion relations in the form of quadratic function, makes abstraction of rules behind the cue-criterion relations very difficult. Even if the non-linear task is simplified by determinism, we are limited to use the exemplar memory because of our inability to consider more than two estimates. This makes it difficult to apprehend non-linear relations.

Rationality in a biological sense can never be understood independently of the environmental structures. These three studies suggest that humans rationally adapt to the environment, as suggested by the notion of ecological rationality (Gigerenzer et al., 1999). When the switch between cognitive processes depends on the nature of the task, the rationality seems to be different in different environments. We are not completely void of the mental capacity making us irrational, but neither we can be regarded as completely rational, considering that there are boundaries in our computational capacity, for example in working memory.

Potential criticism may point to the difficulty of distinguishing between cue abstraction and exemplar-based processes and question the support for the assumption that exemplar-based processes are rapid, unconscious and automatized, in contrast to the rule-based processes that are expected to be intentional, conscious and controlled. The most straightforward evidence for that distinction is the striking dissociation between the neural regions activated in the rule and memory conditions, where the known functionality of the activated regions show that only rule application involves selective attention and working memory, while visual-perceptual and visual-memory processes may be common to episodic memory (Smith, Patalano, & Jonides, 1998). The evidence is consistent with the view that rule-application and exemplar-similarity are qualitatively different.

Another criticism may question if we can conclude the use of exemplar-based processes from the inability to extrapolate. One may point out that participants may be reluctant to give a response value outside the training range (51-59). Therefore, the extrapolation effects are not evidence of exemplar-based processes, but logical reasoning by the participants, asking “why give a response value outside the training range when I never did it before?” If this criticism is correct, the responses for the extreme exemplars in test phase should not be
higher than 51 or lower than 59, instead the response values should be 51 or 59. This criticism does not explain why the new extreme exemplars are less extreme than the old extreme exemplars and why there exist old-new differences in the middle of the judgment range.

In previous research participants have made their judgments based on knowledge about only a few cues. In real environments there exist a great number of cues that we have to take into account when making a judgment about objects. We also have to identify the most important cues, because it is impossible to take all of the environmental cues into account. This is a shortcoming of previous studies that has to be further investigated. It is easier to figure out the relations among cues when they are a few and they are all remembered. Therefore, judgments made with a few cues could be different from judgments that are made on a great number of cues (Gigerenzer et al., 1999). Further research should investigate if the number of cues can affect which cognitive process that is used in judgments.

Maybe learning by presentation of a number of object exemplars with feedback in 220 trials in a lab is a defective way to investigate cognitive knowledge processes. In the real world this does not happen. In real world settings we may be presented with many exemplars but we do not receive an immediate feedback and the presentation of exemplars occurs in a much longer range of time. In further research the time between training and test should be longer, maybe a few days or a week to investigate if this affects the cognitive processes that is used. Different manipulations of feedback can be useful to complete the knowledge of cognitive process in some important ways.

The insights about multiple memory systems (see Schacter & Tulving, 1994 for an example) that support different representations have not yet been sufficiently projected into research on the use of this stored knowledge in judgment. It seems evident that categorization and multiple-cue judgment is likely to be mediated by different processes depending on the task and circumstances (Ashby & Ell, 2001). Therefore, we conclude that research should concentrate on understanding the factors determining how different knowledge systems are applied to make adaptive judgments as a function of individual and environmental characteristics.

Manipulations inducing shifts in the cognitive modes, for example, by introducing tasks that disrupts the functioning of working memory could be one possible way to further investigate the effects of task and circumstances on the cognitive processes. Humans have a bias for rules (suggested by Ashby et al., 1998): whenever the environment allows for it, there will be an abstraction of explicit cue-criterion relations and exemplar-based process will serve as back-up system. This predicts that a linear additive task should induce an analytic cue abstraction process and a non-linear, non-additive task should induce an exem-
pler-based process (Juslin Karlsson, & Olsson 2004). It has indeed been found that the analytic part of a judgment process is affected most in a dual task (Cooksey, 1996) and working memory is more involved in a rule-based process than in an exemplar-based process. Therefore, when introducing a distracter of working memory it is expected that performance in the additive task should deteriorate more than in the multiplicative task (Karlsson, 2002).

Moreover, in categorization research the participants have in general been forced to decide between two alternatives or categories. The research in this thesis has also focused on either exemplar-based or rule-based models and largely rejected other possible cognitive processes such as prototype-based models. In future research, we will perform experiments that involve choice between more than two categories, to investigate which knowledge representation is used in such complex tasks. This may show the presence of prototype-based rather than exemplar-based knowledge, and allow us to compare the similarity between prototype model and cue abstraction model for binary judgments.

In Study 2 we concluded that the more meaningful remembering task could be one explanation of why dyads reached the combined base-line level. I suggest that even more meaningful real-world based tasks could be used, for example, that participants pretend to be physicians that make diagnoses of patients with different symptoms. There is also a lack of control of communication between the dyads in Study 2 and this can explain the absence of the social abstraction effect. To increase this effect, maybe a more real-world problem solving would increase the communication between the two participants.

The participants in the studies have all been students and criticism can be addressed to this. It could, however, be argued that processes at basic levels of cognition like this, is the same for all human beings. What should be further investigated, though, is if there exist differences between different work domain groups, social groups and cultures in making judgments, like theoretical versus practical working domains. This would also be a step towards studying the differences between novices and experts. Earlier, it has been assumed that experts rely more on exemplar-based processing than novices, because of their well-informed knowledge and a great deal of experience of different objects within their domain. Is it possible that experts rely on different cognitive processes depending on their specific domain?

This thesis is a preliminary answer to the question “What factors shaping which cognitive processes and representations that are used?” Still, there are questions that must be asked and answered to complete this important issue.

40
REFERENCES


