Training Complex Judgment:
The Effects of Critical Thinking and Contextual Interference

The research that is reported on in this dissertation was carried out at:

TNO Defence, Security and Safety, in cooperation with Open Universiteit Nederland

ISBN: 978-90-9023669-8

Cover Design: Ariaen Zimmerman

Printed bij Print Partner Ipskamp bv te Enschede
on FSC certified paper

Training Complex Judgment: The Effects of Critical Thinking and Contextual Interference

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Open Universiteit Nederland
op gezag van de rector magnificus
prof. dr. ir. F. Mulder
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 7 november 2008 te Heerlen
om 13.30 uur precies

door

Annechien Sarah Helsdingen

Geboren op 9 februari 1972 te Amersfoort

Promotor: Prof. dr. J. J. G. van Merriënboer, Open Universiteit Nederland

Co-promotor: Dr. T. van Gog, Open Universiteit Nederland

Overige leden beoordelingscommissie:
Prof. dr. J. M. Pieters, Universiteit Twente
Prof. dr. J. M. C. Schraagen, Universiteit Twente & TNO Defensie en Veiligheid
Prof. dr. F. R. H. Zijlstra, Universiteit Maastricht
Prof. dr. J. von Grumbkow, Open Universiteit Nederland
Dr. J. H. Kerstholt, TNO Defensie en Veiligheid

Voorwoord

Ik heb het schrijven van dit proefschrift ervaren als een zeer bijzonder en prettig. Het heeft me in staat gesteld om de resultaten van een aantal van mijn onderzoeken bij TNO voor specifieke opdrachtgevers, in meer detail te beschouwen en te bespreken in het licht van psychologische theorieën over beslissen en leren. Dankzij de steun en het vertrouwen van Ariaen heb ik me een jaar lang exclusief kunnen richten op mijn proefschrift en het lopen van twee marathons. Dat je me de gelegenheid hebt gegeven om me te storten op zulke zeer individuele activiteiten vind ik heel bijzonder en tegelijkertijd tekenend voor onze relatie waarin we elkaar optimaal steunen in persoonlijke ontwikkeling en ontplooiing. Aan deze ontwikkeling heb je niet alleen steun maar ook directe medewerking verleend door het ontwerpen van de kaft en het corrigeren en drukklaar maken van het proefschrift. Ook belangrijk was de hulp die ik ontving bij de opvang van Joris en Marnix, van familie en vrienden, in het bijzonder van mijn ouders, Marijke en Rob Helsdingen, en van mijn schoonmoeder, Willy Oosterhof-Blom. Mijn paranimfen Saskia Helsdingen en Janneke Lijkendijk dank ik voor hun vriendschap en het feit dat ze me willen bijstaan tijdens de verdediging, en daarvoor beloofd hebben het proefschrift helemaal te lezen!

Aan TNO ben ik veel dank verschuldigd, in de eerste plaats omdat ik onderzoekopdrachten zodanig heb kunnen uitvoeren dat ze niet alleen bruikbare resultaten opleverden voor de opdrachtgevers, maar ook geschikt waren voor een wetenschappelijke publicatie. In de tweede plaats vanwege haar prachtige onderzoeksfaciliteiten. Maar bovenal ben ik mijn TNO collega's dankbaar voor hun steun en collegialiteit. Karel van den Bosch voor de begeleiding in het bedenken, opzetten en uitvoeren van zowel de lab- als de veldexperimenten, en voor z'n zeer grondige corrigeerwerk van mijn teksten. En natuurlijk voor zijn vriendschap en voor de lol die we hadden bij de operationele defensie cursussen, tijdens de veldexperimenten en alle jaren bij TNO. Voor het ontwikkelen van de laboratoriumtaak bedank ik Antoon Wennemers en Koen Sanne. Antoon was zeer geduldig bij het steeds maar weer aanpassen van de taak en uit Koen z'n brein ontsproten in eerste instantie de zeer gewelddadige beschrijvingen van misdrijven zoals die zijn gebruikt in de labexperimenten. Daarnaast heb ik prettig samengewerkt in de uitvoering en analyse van de experimenten met Marlous de Beer, Martin van Schaik, Martijn van Emmerik, en Ralf Sluimer. Met Marlous heb ik culinair Den Helder verkend en alle sport en ontspanningsfaciliteiten van de Nieuwe Haven geprobeerd. Martin en ik hebben voor de experimenten veel avonden samengewerkt, waarbij de grootste uitdaging er soms in lag ons door het gebouw te kunnen verplaatsen zonder ergens opgesloten te raken. Ralf en Martijn ben ik erkentelijk om hun methodologische en statistische ondersteuning, en Martijn daarbuiten ook nog om z'n

sportieve steun in onze gezamenlijke voorbereiding en ons succesvolle uitlopen van de Marathon van Amsterdam! Verder wil ik Leny van der Boon en Dymphie van der Heyden bedanken bij het zoeken naar relevante literatuur. En natuurlijk Hans Korteling voor zijn vriendschap, zijn relativerende en zeer kritische vragen, altijd goedbedoelde maar niet altijd goed ontvangen commentaar, en professionele steun in het realiseren van mijn *sabbatical* jaar.

Dit proefschrift beschrijft zowel labexperimenten als veldexperimenten. Zulke experimenten uitvoeren is een leuke en spannende bezigheid, waarbij je echt afhankelijk bent van de mensen in het veld, in mijn geval de begeleiders bij de Krijgsmacht, de instructeurs en leerlingen van de Nederlands/Belgische Operationele School van de Koninklijke Marine, en de medewerkers van de Groep Gebonden Wapens De Peel van de Koninklijke Luchtmacht. In het bijzonder wil ik Rob Dekker, Peter van Mourik, Dick van Ingen, en Hans Bergsma bedanken voor hun enthousiaste medewerking bij de totstandkoming van de veldexperimenten.

Tijdens het feestje ter gelegenheid van mijn afstuderen vertelde mijn toenmalige afdelingshoofd en scriptiebegeleider Herke Schuffel aan mijn ouders dat ik waarschijnlijk wel binnen enkele jaren zou kunnen promoveren, wellicht zelfs vóór mijn dertigste. In de eerste jaren bij TNO heeft Herke geregeld hierop aangedrongen en me in de gelegenheid gesteld gericht te werken aan een wetenschappelijke promotie. Ik had echter eerst andere dingen te doen, ik wilde werken aan toegepaste projecten, op wereldreis, bij de Koninklijke Marine aan de slag, kinderen krijgen… Pas jaren later voelde ik de behoefte om me meer te verdiepen in de theorie en me te concentreren op het schrijven van wetenschappelijke publicaties op basis van de onderzoeken die ik tot dan toe had uitgevoerd. Het is echter niet eenvoudig om een rode lijn te ontdekken in je eigen werk, en dit te plaatsen in een relevant wetenschappelijk kader. De begeleiding van Jeroen van Merriënboer en Tamara van Gog van de Open Universiteit is hierbij van wezenlijk belang geweest. Dat jullie, ondanks mijn haperende planning en gebrekkige kennis van het alfabet, in staat waren zo gestructureerd en snel mijn ideeën te becommentariëren en te corrigeren gaf me het gevoel dat jullie full-time beschikbaar waren om mij te begeleiden. Zonder jullie opbouwende en geïnteresseerde medewerking was dit proefschrift nog niet geschreven.

Aan allen en nog vele anderen: Dank jullie wel!

Table of Contents

Chapter 1
Introduction

Professional environments in which complex judgment and decision making skills play a crucial role, such as the military and international business, are characterized by increasing interdependency and complexity. Therefore, judgment nowadays has to be exercised on matters with more important consequences than was ever the case in the past (Klein, Phillips, Rall, & Peluso, 2007; Van den Bosch & de Beer, 2007). In such professional environments, the decisions of only one or a few persons can affect the livelihoods of very large groups of people. Consider, for example, a central executive officer (CEO) of a multinational company deciding whether to move a production facility from The Netherlands to India in order to maximize profit, stock value, and long-term prospects for the company. In making this judgment, s/he may take the costs of labour, the infrastructure, the political stability, and the availability of adequate personnel in both countries into account, as well as the consequences of distributed management processes, of creating jobs in developing countries, and of closing a production facility in The Netherlands. Clearly, this CEO is faced with an extremely difficult and complex decision-making process, having to consider all these interrelated, often uncertain or unknown, and highly dynamic factors. Moreover, there is the pressing awareness that the final decision affects the lives of many people both in The Netherlands and in India.

Too often, in hindsight complex judgments and decisions are suboptimal. Especially novices with relatively little experience consider aspects of the situation literally rather than conceptually. They regard cues, factors, and features separately and treat them as being independent of each other (Zsambok & Klein, 1997). A central executive officer with relatively little experience, for instance, may decide to move the production facility to India because this country scores higher than The Netherlands on most of the identified factors. Yet, this might prove to be a bad decision later on because serious communication problems develop between the Indian production facility and other production facilities in Europe, due to differences in language and culture . Whereas the novice CEO carefully compared the advantages and disadvantages of setting up the facility in both countries, s/he forgot about the wider context of communication and cooperation between different establishments of the same multinational company. In short, novices have great problems with complex judgment and decision making because they have difficulties applying their knowledge and skills in quickly changing and highly complex situations (Hoffman, 2007).

Because of the complexity and required flexibility of complex judgment and decision making skills, it is of utmost importance that professional decision makers are well prepared for their tasks and acquire a level of understanding that enables them to operate adequately on different tasks and in different contexts, that is, to transfer their knowledge and skills from the training task or context to new tasks or contexts. Research on judgment and decision making has identified several defining features of experienced decision makers. Their experience enables them to recognize a large number of situations as familiar, and has allowed them to develop typical solutions or strategies available for those situations (Klein & Calderwood, 1991). Furthermore, when confronted with new, unfamiliar problem situations, experienced decision makers critically evaluate the available evidence, try to explain events in a comprehensive story that incorporates all available observations, attempt to identify any gaps in such a story, seek additional information and make assumptions about those gaps, and continuously test these assumptions in a critical way whenever new evidence is found (Alison & Barrett, 2004; Cohen, Freeman, & Thompson, 1998; Innes, 2003; Pennington & Hastie, 1992; Robinson & Hawpe, 1986). Thus, an expert CEO would probably rely on his or her knowledge of successful and unsuccessful moves of production facilities from one country to another country, and use this knowledge base for an in-depth analysis of advantages and disadvantages of a possible move in the given situation.

In order to help novices become expert decision makers, *critical thinking instruction* for complex judgment and decision making skills may be developed based on the known characteristics of experienced decision makers (Cohen et al., 1998). Simultaneously, critical thinking instruction aims to present learners with many prototypical problem situations to help them build a large and integrated body of knowledge to facilitate the process of recognition-primed decision making (Klein, 1989). Such a knowledge base includes the cognitive strategies that may be applied in novel and ambiguous situations, facilitating transfer of knowledge and skills from one type of decision problems to other types of problems, conditions, and decision making tasks (Klein, 1998; Klein et al., 2007).

As stated before, this transfer of knowledge and skills from learning tasks to real world tasks or from one task domain to another domain, is paramount for professional decision makers working in continuously changing environments. The degree of transfer depends both on the degree to which task activities call upon the same procedural skills and on the learner's level of understanding of a task domain (Detterman & Sternberg, 1993; Holyoak & Koh, 1987; Mayer & Wittrock, 1993). Van Merriënboer (1997) describes these two processes as, in order, rule-based transfer and schema-based transfer and their combined use as 'reflective expertise'. A process of abstraction is needed to gain deep understanding of a domain, by proceeding from knowledge of the task at hand to a higher level of generalized knowledge, which is in turn required for transfer to different tasks or domains. *Contextual interference* may enhance such abstraction processes (De Croock, Van Merriënboer, & Paas, 1998). Contextual interference refers to the interference a task performer experiences between consecutive tasks; high contextual interference occurs for example when several variations of the task are practiced in a random order. This high interference has proven to benefit transfer of knowledge and skills (for an overview, see Brady, 1998; Schmidt & Bjork, 1993).

Both critical thinking instruction and contextual interference have been proven to be effective for transfer of knowledge and skills. However, the effects of critical thinking instruction have mostly been studied in real-world educational programs, that is, with small numbers of participants and with little experimental control over the sequence of learning events and the timing of instructor interventions (e.g., Cohen et al.1998; Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000; Freeman & Cohen, 1997). The effects of contextual interference on learning, retention and transfer, on the other hand, have been extensively studied in controlled experiments. It has been studied in many domains, however, almost no attention has been paid to the effects of contextual interference on learning complex judgment and decision making skills (Brehmer, 1973, 1977, 1979). Moreover, the combined effects of critical thinking instruction and contextual interference have not been investigated before. Yet, it might be the case that both approaches can strengthen each other or, on the contrary, interfere with each other.

Concluding, this dissertation presents a series of field studies and laboratory experiments that systematically investigate the effects of critical thinking instruction and contextual interference, separately and in combination, on learning, retention, and transfer of complex judgment and decision making skills.

Overview of the Dissertation

The studies presented in this dissertation seek to answer the following research questions:

1. Does critical thinking instruction improve learning of a complex judgment and decision making task?
2. Does the contextual interference effect manifest itself in learning a complex judgment task?
3. What are the combined effects of contextual interference and critical thinking prompts on acquisition and transfer test performance of a complex judgment task?

Chapter 2 provides a review of the characteristics of complex judgment tasks and of the theoretical underpinnings of the expected effects of critical thinking instruction and contextual interference on learning, retention, and transfer. Chapter 3 addresses the first research question. Two field studies are presented that assessed the effects of critical thinking instruction on learning complex judgment tasks. Chapter 4 focuses on the second research question. Two experiments are described that investigated the potential benefits of a random practice schedule (i.e., high contextual interference) compared to a blocked practice schedule (i.e., low contextual interference) for retention and transfer test performance of a complex judgment task. Chapter 5 addresses the third research question. An experiment is presented that investigated the effects of critical thinking prompts in blocked and random practice schedules on acquisition, retention, and transfer test performance of a complex judgment task. Chapter 6 presents a general, overall discussion of the findings of the studies presented in Chapters 3, 4, and 5. This chapter also discusses the implications of our findings for further theory development and educational practice.

References

Alison, L., & Barrett, E. (2004). The interpretation and utilisation of offender profiles: A critical review of 'traditional' approaches to profiling. In J. R. Adler (Ed.), *Forensic psychology: Concepts, debates and practice* (pp. 58-72). Devon: Willan Publishing.

Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. *Quest, 50*, 266-293.

Brehmer, B. (1973). Note on the relation between single-cue probability learning and multiple-cue probability learning. *Organizational Behavior and Human Decision Processes, 9*, 246-252.

Brehmer, B. (1977). General and specific effects of practice in using functional rules in inference tasks. *Umea Psychological Reports, 114*, 1-13.

Brehmer, B. (1979). Effect of practice on utilization of nonlinear rules in inference tasks. *Scandinavian Journal of Psychology, 20*, 141-149.

Cohen, M. S.,Freeman, J. T., & Thompson, B. B. (1998). Critical thinking in tactical decision making: A model and a training strategy. In J. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for training and simulation* (pp. 155-191). Washington, DC: American Psychological Association.

Cohen, M. S., Thompson, B. B., Adelman, L., Bresnick, T. A., Shastri, L., & Riedel, S. L. (2000). *Training critical thinking for the battlefield. Volume II: Training system and evaluation*. Arlington, VA: Cognitive Technologies, Inc. Retrieved January 2007 from: http://www.cog-tech.com/papers/mentalmodels/Vol_IITraining.pdf

De Croock, M. B. M., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior, 14*, 249-267.

Detterman, D. K., & Sternberg, R. J. (Eds.) (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex Publishing.

Freeman, J. T., & Cohen, M. S. (1996). Training for complex decision-making: A test of instruction based on the recognition / metacognition model. *Proceedings of the 3rd International Command and Control Research and Technology Symposium 1996*. Retrieved May 12, 2005 from www.cog-tech.com/Publications/TrainingPubs.htm

Hoffman, R. R. (Ed.) (2007). *Expertise out of context: Proceedings of the sixth International Conference on Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*, 332-340.

Innes, M. (2003). *Investigating murder: Detective work and the police response to criminal homicide*. Oxford: Oxford University Press.

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine system research* (Vol. 5, pp. 47-92). Greenwich, CT: JAI Press, Inc.

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. A., & Calderwood, R. (1991). Decision models: Some lessons from the field. *IEEE Transactions on Systems, Man and Cybernetics, 21*, 1018-1026.

Klein, G. A., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data/frame theory of sensemaking. In R. R. Hoffman (Ed.), Expertise out of context: Proceedings of the sixth International Conference on Naturalistic Decision Making (pp. 113-158). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*, 189-206.

Robinson, J. A., & Hawpe, L. (1986). Narrative thinking as a heuristic process. In T. R. Sarbin (Ed.), *Narrative psychology: The storied nature of human conduct* (pp. 111-125). New York: Praeger Publications.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Van den Bosch, K., & De Beer, M. M. (2007). Play a winning game: An implementation of critical thinking training. In R. R. Hoffman (Ed.), Expertise out of context: Proceedings of the sixth International Conference on Naturalistic Decision Making (pp. 177-1980). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Zsambok, C. E. & Klein, G. A. (1997). Naturalistic Decision Making: Where are we now?, In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 3-16). Mahwah, NJ: Lawrence Erlbaum Associates.

Chapter 2
Research on Learning and Transfer of Complex Judgment Skills

Abstract

A review of the literature on Social Judgment Theory (SJT) and Naturalistic Decision Making (NDM) is provided here, which reveals that learning complex judgment and decision making involves: (1) learning what the relevant cues upon which to base a decision are and how they are related to each other, and (2) learning cognitive strategies that guide the learning and decision making processes. Within both the SJT and NDM research approach, the effects of several interventions aimed at learning either cues and their interrelations or cognitive strategies have been investigated. These studies focussed mostly on performance during learning and/or on retention, that is, later performance on the same tasks. However, empirical evaluations of the effects on transfer, that is, later performance on novel tasks were scarce. Therefore, this chapter continues with a discussion of measures that may be effective for enhancing transfer of complex judgment skills, such as critical thinking instruction and contextual interference.

People make numerous choices, decisions, and judgments throughout their lives. Every day they decide what to eat for breakfast, which shoes to wear, how to commute to the office and by what route, and whether it is safe to cross the street. Somewhat less frequently they make decisions with higher impact, such as accepting jobs, buying houses, and marrying life partners. These are just a few examples that illustrate how many choices, decisions, and judgments people have to make during everyday life.

The goal of research on judgment and decision making is to investigate how people make those judgments and decisions, how accurate they are in doing so, and how the process can be improved. The findings from this research can be used to design or improve training programs for decision makers. Training these skills is of utmost importance because professional decision makers make judgments on a wide variety of highly complex issues, often with far-reaching consequences for large groups of people. Therefore, a major aim of training programs is to establish transfer of decision making skills from the learning tasks to professional tasks in real-life contexts.

The main goal of this chapter is to describe theoretical perspectives on learning and teaching complex judgment and decision making. The first section describes two prominent approaches in decision making research: Social Judgment Theory (SJT) and Naturalistic Decision Making (NDM). Thereafter, instructional methods that enhance transfer are described, incorporating a description of specific methods such as critical thinking instruction and increasing contextual interference between learning tasks. The chapter ends with concluding remarks on how these methods might work in training complex judgment.

Studying Decision Making: SJT and NDM Approaches

Although the processes of making decisions and judgments have been extensively studied, there is not one single scientific approach or method. Different research paradigms or guiding theories can be distinguished, for example, social judgment theory (Raiffa, 1997; Sherif & Hovland, 1961) and naturalistic decision making (Klein, Orasanu, Calderwood, & Zsambok, 1993), but also decision theory (Lehman, 1950; Raiffa, 1997), behavioral decision theory (Einhorn & Hogarth, 1981), attribution theory (Heider, 1958), utility theory (Fishburn, 1970), prospect theory (Kahneman & Tversky, 1979), et cetera. These approaches differ on several aspects. First, they differ in their focus on *what* is studied, which can be, for example, the elements of decision problems, the process of generating alternatives, the process of choosing between alternatives, predictive judgment, evaluative judgment, dynamic decisions, static choices, the process of acquiring and aggregating information, or how and when people deviate from rationality. Second, there are differences in *how* the research is conducted, for example, in the laboratory, in the field, by formal analysis, empirical analysis, or experimental manipulations. Third, researchers studying judgment and decision making come from different disciplines, such as economics, psychology, operations research, statistics, and so forth. And lastly, there are differences between the goals of research, which can be either to describe and explain human judgment and decision making, to improve training programs, or to develop formal models prescribing how decisions should best be made.

In all different paradigms, the question of representativeness of the research conditions for real-life decision making and the resulting generalizability of the conclusions is a common concern (Hammond, McLelland, & Mumpower, 1980). On one side of the continuum, there are the simple choice problems in a laboratory setting to study the effects of several parameters on the choices people make. The problem with these choice problems is that they are often not representative of the many decision problems people encounter in their daily or professional life. However, these laboratory experiments do provide the opportunity to exercise strict experimental control over the research conditions, to investigate large groups of participants, and to manipulate many parameters; aspects that increase the generalizability of the conclusions. On the other end of the continuum are pre-experimental and quasi-experimental studies conducted in field settings. In these settings, professional judgment and decision making processes can be studied under natural conditions, although there is often a lack of experimental control, a small number of participants, and a limited opportunity for experimental manipulations.

Two approaches, using elements from both ends of the continuum, are applied in this dissertation and will be discussed in the next sections. They are chosen because they both focus on how people acquire judgment and decision making skills and emphasize the representativeness of the study methods. These approaches are Social Judgment Theory (SJT; Brehmer & Joyce, 1988; Sherif & Hovland, 1961) and Naturalistic Decision Making (NDM; Klein et al., 1993). Within SJT, a major research focus is on so-called ‘multiple functional learning’ experiments in the laboratory, whereas in the NDM approach real world decision making is studied in the field.

Social Judgment Theory: Multiple Functional Learning

In Chapter 1, a hypothetical decision problem of a central executive officer (CEO) of a large company was sketched, namely, the CEO must decide on the best strategy to optimize his or her company’s profits, survivability, and stock value: Moving production to India or keeping it in The Netherlands? When the CEO thinks about such a complex judgment problem points of reference may be considered that are expected to be related to the target, such as labor costs, infrastructure, inflation rates, and so forth. These points of reference are part of the CEO’s mental representation of the network of relationships between elements (objects, events) in the environment and the event to be predicted. Accuracy of the judgment depends on the extent to which the CEO’s mental representation matches the real network of relationships (Hogarth, 1980). It is this match, or lack thereof, that is the object of study within SJT (Brehmer & Joyce, 1988; Hammond, McClelland, & Mumpower, 1980).

SJT is modeled after Brunswik’s theory of perception (Brunswik, 1943, 1955). According to that theory, a person does not have access to any direct information about the objects in the environment. Instead, perception is seen as an indirect process, mediated by a set of proximal cues (i.e., points of reference). The perceptual system uses these cues to make inferences about distal objects. In accordance with this view, SJT defines judgment as a process that involves the integration of information from a set of cues into a judgment about some distal state of affairs (Hogart, 1980). SJT emphasizes that the probabilistic and interdependent relations among variables in the natural environment create cognitive difficulties for the person attempting to understand it. For example, in his studies of clinical judgment, Hammond (1955, 1996) noted that the information a clinician receives from a patient is dependent upon the actions performed on or questions posed to that patient. Thus, the clinician must be aware of this interaction to be able to make an accurate diagnosis. Furthermore, s/he needs to learn that different symptoms may lead to the same judgment since one patient may report one set of symptoms, whereas another patient having the same disease may report another set of symptoms. Brunswik (1955) called this capacity to make use of alternatives *vicarious functioning*. For human perception – and cognition – as well as for human action, vicarious functioning seems a necessary capacity to cope with inconsistent, unexpected, incomplete, and imperfect events (Gigerenzer & Kurz, 1999).

Within the SJT research paradigm, an experimental method was devised to study how people learn such difficult judgment tasks: The (Multiple) Cue Probability Learning experiment (MCPL; Björkman, 1965; Brehmer, 1972; Brehmer & Brehmer, 1988; Brunswik & Herma, 1951; Hammond, Hursch, & Todd, 1964; Hursch, Hammond, & Hursch, 1964; Smedslund, 1955), also referred to as Multidimensional Functional Learning (MFL; P. J. Hoffman, Earle, & Slovic, 1981). During a typical MFL experiment a person makes judgments based on a number of probabilistic cues over a series of trials. Feedback may be given on each trial, or feedback may be given after subsets of trials. The aim is to correctly predict the quantitative or categorical criterion value on each trial. Cues differ in terms of their relevance (ecological validity) to the criterion. For example, being nauseated is rarely a symptom of having a brain tumor, and vice versa, when a patient has a brain tumor, nausea is an infrequent symptom (Chandana, Movva, Arora, & Singh, 2008), thus, nausea is a low validity cue for a brain tumor.

MFL studies have focused on how people learn to discover cues and judge the importance of these cues (Klayman, 1988a). In particular, it has been studied how learning and transfer performance are affected by parameters of the task (e.g., linearity of relationships between variables, predictability of cues, meaningful labels, time pressure; Edland, 1993; Koh, 1993), the nature and timing of given feedback (e.g., delayed feedback, cognitive feedback, outcome feedback, feedforward; Balzer, Sulsky, Hammer, & Sumner,1992), and characteristics of the task performer (e.g., age, goal setting, prior knowledge or experience; Alm & Brehmer, 1982; Chasseigne, Mullet, & Stewart, 1997; P. J. Hoffman et al., 1981). From these studies, it became clear that performance is higher when (a) linear relationships exist between cues and the criterion (Alm, 1982a; Brehmer, 1979, 1987; Hammond & Summers, 1965), (b) these relationships are positive rather than negative (Björkman, 1965; Brehmer, 1977; Sheets & Miller, 1974), (c) cues have meaningful rather than abstract labels (Koele, 1980; Muchinsky & Dudycha, 1975; Ruble & Cosier,1990), (d) there is no time pressure (Rothstein, 1986), (e) positive feedback is given rather than negative feedback (Klayman, 1988b), and (f) participants set goals for themselves (DeShon & Alexander, 1996).

Also, interactions have been found between task difficulty and feedback strategy on complex judgment performance (Balzer, Doherty, & O'Connor, 1989; Lindell, 1976; Tsao, 1994; Wigton, Patil, & Hoellerich, 1986). For complex tasks, cognitive feedback often renders best performance, but in simple tasks outcome feedback is better. Whereas outcome feedback only provides information on the correctness of a decision, cognitive feedback concerns characteristics of the task as well as characteristics of the person's cognitive processes to reach a decision (Hammond et al., 1964). First, cognitive feedback may include information on the relationships between the criterion and the cues, or the ecological validity of the cues (Doherty & Balzer, 1988). Second, information may be given on the correlation between the participants' *estimated* relationships between cues and the criterion and the *actual* relationships between those cues and the criterion. Third, cognitive feedback may also provide the learner with information on his or her own learning process (e.g., weights a participant ascribes to cues or the stability of the applied judgment strategy).

Cognitive feedback may include either one or more of the aspects mentioned above. For example, in a study by Gattie and Bisantz (2006), which investigated the effects of different types of cognitive feedback on task performance in a dental diagnosis task, participants had to judge whether a dental condition was benign or malicious on the basis of a patient's age, gender, tumor growth location, growth size, growth color, and cancer risk. In the first cognitive feedback condition, participants received information on the weight they ascribed to each of those indicators, calculated continuously based on regression analysis of their decisions. In the second condition, participants received information on the task: The relationships between cues and criterion, the validity of cues, and so forth. In the third condition, participants received information on the correlation between their own decisions and the actual task properties. Both the information on decision strategies (first condition) and on the task (second condition) improved participants' performance. Feedback on the decision strategies was especially helpful for novice participants, which suggests that participants who are unfamiliar with the experimental task or the specific domain may need to understand their own decision policies in order to be able to learn. Furthermore, the task information provided may have made individuals more aware of the appropriate cues that should be examined. The study failed to find any benefit of the third cognitive feedback condition, compared to the outcome feedback condition.

Regrettably, in most MFL studies including the Gattie and Bisantz (2006) study discussed above, participants' performance was only measured during the learning phase. This is problematic, because observed performance during the learning phase can be a notoriously poor guide to predicting learning outcomes, that is, post-training performance (Bjork, 1994; see also the next section on 'Enhancing Learning and Transfer'). Only a few MFL studies measured test performance after a retention interval (e.g., Alm, 1982b; Bauer, 1971; Brehmer & Lindberg, 1973) or transfer of learning to new, unfamiliar tasks (e.g., Andersson & Brehmer, 1977; Brehmer, 1977, 1979; Brehmer & Almqvist, 1977; Lindberg & Brehmer, 1976). However, transfer of knowledge and skills across domains is a prerequisite for vicarious functioning, that is, the ability to perform adequately in different contexts and use different cues. And vicarious functioning is a necessary capacity to cope with inconsistent, unexpected, incomplete, and imperfect events (Gigerenzer & Kurz, 1999).

Concluding, MFL experiments use representative experimental tasks that adequately capture the major characteristics of complex judgment tasks. But although many task-related, feedback-related, and learner-related aspects were investigated for their effects on performance during learning and retention, few studies have focused on transfer effects. Furthermore, the effects of cognitive feedback that have been found suggest that measures that improve understanding may enhance learning of complex judgment tasks. Such measures are, for example, specific instructional methods to increase understanding through elaborative processing, such as providing critical-thinking prompts or increasing task interference.

Naturalistic Decision Making

The previous section described SJT and its MFL experiment as a powerful approach to study complex judgment tasks. However, the design of typical MFL experimental tasks, with a fixed set of cues and a trial-by-trial presentation of cue values and feedback, departs from two of Brunswik's central ideas on human performance described earlier: Vicarious functioning and the idea that a person's actions affect the information s/he collects. Therefore, this dissertation not only presents the results of MFL studies, but also empirical evaluations of learning naturalistic decision making tasks in the field of military command and control. Because of the interactive nature of command and control decision making, the commander's decisions influence the situation and the information s/he collects. Furthermore, a military commander has to prepare for a wide variety of decision problems and task environments (R. R. Hoffman, 2007). Thus, any training program should aim to present many different scenarios because the learners must learn to deal with many different alternatives (i.e., vicarious functioning).

Naturalistic Decision Making (NDM) is a relatively recent approach in decision making research (Klein et al., 1993; Zsambok & Klein, 1997). It focuses on the question of how people make use of their expertise in real world judgment and decision making tasks. Whereas many laboratory experiment approaches to judgment and decision making tend to focus on the process of weighing options and making choices, NDM investigates judgment and decision making as the complete process from collecting information to generating actions. Moreover, this process is studied in professional environments in which experience is not viewed as a confounding variable, as it often is in traditional judgment and decision making studies, but as an object of study in its own right (Klein, 1997; Klein & Calderwood, 1991).

The NDM paradigm is explanatory in nature: Research aims to identify the cognitive processes and representations that underlie people's judgments and decisions. Heuristics, or rules of thumb, are considered to be adaptive strategies that enable people to behave adequately in quickly changing and complex situations (Todd & Gigerenzer, 2001). These heuristics can be false or may prove to be inadequate in certain situations, but nevertheless they are valuable because they allow for flexibility in new situations and increase the chance that an acceptable decision or judgment is made. Thus, NDM suggests that normative decision models should be based on cognitive and behavioral criteria, not on a formal analysis of the decision problem.

Orasanu and Connolly (1993) identified eight aspects of tasks that make them relevant experimental tasks within NDM research: (1) ill structured problems, (2) uncertain, dynamic environments, (3) changing, ill defined, or conflicting goals, (4) multiple feedback loops, (5) time pressure, (6) major interests, (7) several parties involved, and (8) individual goals have to be weighed against the organizational goals. Klein (1998) states that NDM research using this kind of tasks has confirmed that decision making is a process of situation assessment rather than weighing options and choosing. Experienced decision makers do not seem to be troubled by weighing options: Their real challenge lies in acquiring an adequate and realistic overview of the situation. Moreover, the so-called *zone of indifference* could render studies with an exclusive focus on weighing options and making choices trivial: If two options are very alike the choice between the two becomes harder, whereas the consequences of choosing one option above the other become negligible.

On the basis of NDM studies on real-world decision making, Klein and colleagues (Klein, 1989; Klein, Calderwood, & Clinton-Cirocco, 1986) have formulated the Recognition Primed Decision (RPD) model. This model describes how people use their experience to assess situations and make decisions. They match their observations of the current situation with what they remember from earlier experiences (i.e., their mental model), and then generate a prototypical option or a sequence of options they can evaluate one by one (i.e., by running a mental simulation of their mental model). If the option is satisfactory, it is selected. Thus, selection of options is not based on a parallel weighing of options, but on a sequential consideration of the adequacy of one option at the time (Drillings & Serfaty, 1997). The RPD model further encompasses a diagnostic function (Klein, 1998), which people use to explain the observed events in a situation and to indentify causal relationships between these events.

This diagnosis function serves as an evaluative process in uncertain situations and intervenes when explanations, or possible matches between observations and memory, diverge.

Recognition and situation assessment are important functions in NDM models. Klein, Moon, and R. R. Hoffman (2006) distinguished two levels in their data/frame theory of sensemaking: The level of *mental model formation* which is backward looking and explanatory, and the *mental simulation* which is forward looking and anticipatory. They posit that decision makers always apply a frame, that is, some mental model based on experiences, when they observe and interpret the world around them. This frame serves the two levels of sensemaking: It guides both explanation of the situation and prediction of future events. Expertise is characterized by large amounts of relevant representations of prototypical experiences in memory as well as efficient structuring and chunking of this information to facilitate instant retrieval (Chase & Simon, 1973). It allows experts to represent problems in terms of deep theoretical principles rather than surface features as novices commonly do (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980; H. G. Schmidt, Norman, & Boshuizen, 1990). In other words, experts have better mental models allowing them to identify and select relevant cues and patterns in a situation (Stout, Cannon-Bowers, & Salas, 1997; Vincente, 1988; Vincente & Wang, 1998) and to perform more effective searches for further information (Lipshitz & Ben Shaul, 1998; Serfaty, Macmillan, Entin, & Entin, 1998).

This role of mental models and expertise is also emphasized in the recognition/metacognition model of Cohen, Freeman and Wolf (1996). In many situations, people's decisions are based on recognition of aspects of a situation, matching recognized aspects with earlier experiences, and forming a mental model of the current situation that implies a prototypical or sufficient decision option. However, this mental model will often be incomplete or inconsistent because of missing information, conflicting evidence, and unreliable assumptions. Critically testing and evaluating one's mental model are therefore considered paramount in the decision process, especially when high stakes are involved, when problems are dynamic and complex, or both. These critical thinking skills typically are not part of a training program for professional decision making; they are mainly acquired as a result of experience in the field (Anderson, 1993; Klein, 1998; Klein et al., 1993). However, on the basis of their recognition/metacognition model, Freeman and Cohen (1996) developed an instructional strategy to teach these critical thinking skills.

Critical thinking is conceptualized as higher order thinking that is purposeful, reasoned, and goal directed. It is involved in solving problems, formulating inferences, calculating likelihoods, and making decisions (Frijters, Ten Dam, & Rijlaarsdam, 2008; Halpern, 2003). Within the critical thinking research community, there is an ongoing debate on the issue of generalizability versus specificity of critical thinking skills (Siegel, 1991). Some studies have shown that special critical thinking programs usually do not result in long-lasting effects (Tsui, 1999). But at the same time, several authors point out that some general principles of critical thinking transcend specific subjects (e.g., Klaczynski, 2001; Stanovich & West, 2000; Tsui, 1999). As mentioned before, transfer of skills across different tasks and knowledge domains is important because today's jobs require professionals to work in continuously changing contexts. Still, the development of expertise in judgment and decision making skills is thought to be largely domain-specific (Ten Dam & Volman, 2002; Tsui, 1999). In their own domain, experienced decision makers quickly identify meaningful factors, realize what information they may be missing, recognize typical problems, and recall appropriate actions; in unfamiliar domains, the same decision makers may be at loss as to what the relevant factors are, thus making it impossible to realize missing information or come up with appropriate actions (Klein & Calderwood, 1991). Therefore, the challenge is to teach critical thinking in the context of specific meaningful subject matter, but yet in such a way that transfer to other tasks and domains becomes possible (Brown, 1997; Frijters et al., 2008).

Enhancing Learning and Transfer

As mentioned before, many SJT and NDM studies measure performance during the learning phase to draw conclusions about how different instructional methods stimulate or hamper learning. However, learning is best understood as the observable change or modification in behaviour as a result of experience (Skinner, 1950; Thorndike, 1910). This definition reveals the inherent difficulty in establishing whether a person has learned something: Performance is observed, but learning is inferred. And performance during a learning or acquisition phase can be a notoriously poor guide to infer actual learning outcomes. Many studies have shown that although no apparent changes in performance were observed between conditions A and B during learning tasks, or inferior performance was found in, say, condition A compared to condition B, performance after a (short) interval on subsequent test tasks can yet be substantially *higher* in condition A – indicating that more or better learning must have taken place during the learning phase in this condition (Bjork & Bjork, 2006). The converse has been found true as well: When performance during the learning phase improves quickly and substantially, this improvement may no longer appear on the test. This shows that substantial changes in performance during a learning phase (i.e., apparent learning) are not always good indicators of actual learning (Bjork, 1994; Bjork & Bjork, 2006).

The same "paradox" applies to transfer of learning, which –as mentioned before- is usually the ultimate goal of training programs: Instructional strategies that appear to slow the learner's progress during instruction or training often lead to better transfer test performance (Van Merriënboer, De Croock, & Jelsma, 1997). A distinction is often made between *near* transfer and *far* transfer. Near transfer tasks share structural features but differ on superficial or surface features from the learning tasks, whereas far transfer tasks differ from the learning tasks on both surface and structural features (Quilici & Mayer, 1996, provide a good description). Thus, the degree of transfer depends on the degree to which task activities call upon the same type of skill, but also on the learner's level of understanding of a task (Detterman & Sternberg, 1993; Holyoak & Koh, 1987; Mayer & Wittrock, 1993). Near transfer can be obtained if the learner knows *what to do when* (procedural knowledge), whereas far transfer is usually possible only if the learner also knows *why* he needs to do that, that is, has the knowledge to interpret whether particular actions may be effective or not (conceptual knowledge; Van Merriënboer, 1997). To gain such deep understanding required for far transfer, a process of generalization and/or abstraction is needed, from knowledge of the task at hand to a higher level of knowledge, for instance, of the general principles of a domain or the general procedures for performing a class or even several classes of tasks (Van Merriënboer & Paas, 1990).

Instructional measures such as reflection (Boud, Keogh, & Walker, 1985), prompted self-explanations during learning (Chi, 1996; Renkl, 1997; Stark, Mandl, Gruber, & Renkl, 2002), contextual interference between learning tasks (Battig, 1979; Magill & Hall, 1990), practice variability (Burke & Hutchins, 2007), the use of degraded stimuli (Kunen, Green, & Waterman, 1979), and infrequent feedback (Wulf & Shea, 2002) may enhance this process of generalization and abstraction. Critical thinking instruction as developed by Freeman and Cohen (1996) encompasses elements of reflection and self-explanation and thus may also be expected to enhance transfer via processes of generalization and abstraction. In this dissertation, studies are reported that focus on the effects of critical thinking and contextual interference in learning complex judgment and decision making tasks. These two instructional measures will be discussed in more detail in the next sections.

Critical Thinking

Critical thinking instruction as it was developed by Freeman and Cohen (1997) on the basis of their recognition/metacognition theory aims to develop two types of skill: Recognition skills and metacognitive skills. Recognition is the mechanism that enables experienced decision makers to select all relevant cues from the situation to activate or form an accurate mental model of this situation. To develop the experience necessary to recognize a vast amount of situations, an individual needs to be confronted with many different situations and discover the relevant cues, rather than being told what aspects or cues are important in what situations (Stout et al., 1997). Training should therefore be focused on presenting as many relevant problem situations as possible, and each situation should incorporate one or more relevant cues. Stout et al. developed specific cueing training, in which they support novices by ‘augmenting’ relevant cues. This augmentation can be visual, for example by increasing contrast values on a screen, or auditory, for example by presenting a warning signal. But it may also be accomplished by asking specific questions or providing feedback during training in order to focus a learner’s attention on specific cues.

The (meta)cognitive skills that are part of the recognition/metacognition model involve skills to indentify evidence-conclusion relationships, criticize the mental model, adapt the mental model, and perform so called quick tests. Quick tests investigate whether there is sufficient time and opportunity to continue with elaborative processing, or whether the current mental model should serve as the basis for an immediate decision (Freeman & Cohen, 1996). Therefore, the training program for military officers that Freeman and Cohen developed not only involved the presentation of many different problem situations, but also contained four specific steps to instruct and prompt the elaborative processes:

1. Develop a story (i.e., form a mental model) of the situation. Incorporate history, intentions and capacities of all parties involved in your story to explain all your observations and predict future events.

2. Test your story for conflicting and/or missing information. Try to explain all observations within one comprehensive story, even if these observations do not seem to be related to your story. Identify gaps in your story and make explicit assumptions to cover these gaps.

3. Evaluate your story. There is the devil’s advocate that tells you—part of—your story is false. Try to come up with an alternative story that can also explain your observations. Which story is more plausible?

4. Develop plans and contingencies for the weakest assumptions in your story.

Analyses of historic events may serve as training materials for story building. In such analyses, the decision processes of professionals involved are often documented, and as a consequence, all elements of a comprehensive story can be easily identified. During story building, testing, and evaluating, it is necessary to constantly monitor whether critical thinking is still useful and wise in the current situation, or whether immediate action is required. The three preconditions that call for continued critical thinking are:

1. The risk of delay is acceptable.

2. The costs of a possible mistake are high.
3. The situation is not routine, but new and/or complex.

The critical thinking instruction starts with an initial instruction in the method, that is, an introduction to the theoretical background and the practical relevance of the approach, an explanation of the four critical thinking steps, and a demonstration of critical thinking being applied in a situation assessment and decision problem. It continues with prompts being provided during practice to initiate the critical thinking processes in learners. The instruction method has been empirically tested in several studies (e.g., Cohen, Freeman, & Thompson, 1998; Freeman & Cohen, 1996, but these studies were less than optimal. They were conducted in simplified training environments and compared the performance of critical thinking participants with the performance of participants who did not receive any training at all. Since only performance during learning was assessed, it was not established whether critical thinking instruction had differential effects on retention and transfer. Cohen and colleagues (Cohen et al., 1998; Freeman & Cohen, 1996) attributed the benefits of critical thinking instruction to acquiring an appropriate decision making strategy. When students are taught critical thinking skills, they are less likely to make the typical mistakes in making judgments, such as giving in to confirmation bias (i.e., interpret new information so that it complies with the things you already know) or neglect of probability (i.e., disregard probability when making a decision under uncertainty; Halpern, 1997). Thus, they will show better overall performance on judgment and decision making tasks. In addition, it is expected that participants also gain a deeper, more abstract level of understanding of the task content that specifically enables them to solve decision problems different from the learning problems on both a superficial and structural level. Critical thinking instruction may facilitate processes such as self explanation and reflection, that have been proven to enhance learning and transfer (Chi, De Leeuw, Chiu, & LaVancher, 1994; Ferguson-Hessler & de Jong, 1990; Lee & Hutchinson, 1998; Nathan, Mertz, & Ryan, 1994).

Contextual Interference

As stated before, contextual interference may enhance generalization and abstraction processes (De Croock, van Merriënboer, & Paas, 1998) and as a result, improve transfer across tasks or knowledge domains. Contextual interference may be manipulated by the scheduling of learning tasks. Blocked task sequences, that is, sequences of learning tasks organised in blocks, with only one variation of a task being practised in each block (e.g., AAA-BBB-CCC), have low contextual interference. These have often been found to lead to higher performance during the learning phase than random practice schedules (e.g., A-B-C-B-C-A-A-C-B), which have high contextual interference (Schneider, Healy, & Bourne, 1998, 2002). However, random practice schedules often result in better retention and transfer of skills to related tasks and situations (Greeno, 1964; Magill & Hall, 1990). These effects of high contextual interference have been observed and studied extensively in the learning of motor tasks (Cross, Schmitt, & Grafton, 2007; Lee & Magill, 1983; Shea & Morgan, 1979; Simon, 2007), but are not unique to the motor learning domain. Studies on learning procedural tasks (Carlson, 1989; Carlson & Schneider, 1989; Carlson, Sullivan, & Schneider, 1989; Carlson & Yaure, 1990), cognitive operational tasks, such as interacting with automatic teller machines (Jamieson & Rogers, 2000), foreign vocabulary learning (Jacoby, 1978; Schneider, Healy, & Bourne, 1998, 2002), logical rules (Schneider, Healy, Ericsson, & Bourne, 1995), learning problem solving from worked examples (Paas & Van Merriënboer, 1994), or learning high-level cognitive tasks, such as troubleshooting a complex simulation of a chemical plant (De Croock et al., 1998), demonstrate that contextual interference is a general phenomenon that applies to a variety of learning tasks and contexts.

In their explanation of the effect of contextual interference, Shea and Morgan (1979) formulated the elaboration hypothesis: Under high contextual interference conditions, different procedures for different task variations need to be kept in mind simultaneously, thus providing the opportunity to identify similarities and differences between task variations and those procedures. This more elaborate and distinctive processing of the tasks would facilitate remembering of the appropriate procedure for each task because more details of the original context may serve as retrieval cues.

A different explanation is provided by Lee and Magill (1983). In their reconstruction hypothesis, they theorize that under conditions of high contextual interference, the appropriate procedure for each task variation has to be reconstructed from (long term) memory since not all procedures can be kept active simultaneously. In the repetitive presentation of the same task variation in a blocked schedule, these repeated reconstructions are not necessary because the solution to the problem is still readily available. Thus, repeated reconstructions are considered to be responsible for the learning advantages of random practice schedules.

Yet another account for the differences in acquisition and transfer performance following random or blocked practice schedules is given by R. A. Schmidt and Bjork (1992). They claim that the additional information processing demands in a random practice schedule are the retrieval processes: A random practice schedule prevents learners from generating a stable set of responses for a particular task, and forces the learner to retrieve necessary information for task performance from memory. This practice in the process of retrieving information from memory is critical for test performance, therefore learners will perform better on transfer tests after a random practice schedule.

For all three hypotheses there are empirical data that seem to support them. However, it is still unclear which one is most viable (Li & Wright, 2000; Wulf & Shea, 2002). What all these '*extra processing*' hypotheses have in common, though, is that they mainly explain the benefits of random practice over blocked practice. Although this implies some consideration of the constraints of blocked practice, focussing primarily on the latter could lead to a fourth explanation for the contextual interference effect. Deteriorated learning in a blocked practice schedule compared to a random practice schedule might also be due to an *illusion of competence* learners experience as a result of the repetitive presentation of the same type of task (Koriat & Bjork, 2005). Illusions of competence may inhibit search and application of alternative task strategies and so block gaining a deeper level of understanding (Kornell & Bjork, 2007). As a consequence, transfer test performance, particularly far transfer, will be hampered in a blocked sequence compared to a random sequence. In a random sequence illusions of competence are far less likely to occur, for learners are confronted each time with a new task.

Conclusion

Learning complex judgment and decision making requires two types of skill: (1) development of recognition skills based on subject matter expertise representing the relevant cues, the criterion value that needs to be predicted, and the interrelationship between the different cues and the criterion value (i.e., a mental model), and (2) higher order critical thinking skills (called metacognitive skills in the recognition/metacognition model of Cohen, Freeman, & Wolf, 1996) that serve to increase understanding. Designing training for complex judgment and decision making skills using critical thinking instruction and contextual interference is expected to improve learning and transfer through elaboration of the task content (e.g., by generalisation, discrimination, or abstracting away from it).

Although both critical thinking and contextual interference are believed to increase transfer of judgment skills, the former has not yet been studied in well-controlled experiments and the latter has not yet been studied with complex judgment tasks. Moreover, the combined effects of critical thinking and contextual interference have not yet been investigated. The next chapters present the results of several field and laboratory studies that investigated the effects of critical thinking instruction (Chapter 3), contextual interference (Chapter 4), and their combination (Chapter 5). Chapter 6 presents a general discussion of all conducted studies.

References

Alm, H. (1982a). Learning of probabilistic inference tasks: Effects of uncertainty and function form. *Umea Psychological Reports, 9*, 1-36.

Alm, H. (1982b). Effects of pretraining on the construction of complex rules in probabilistic inference tasks. *Umea Psychological Reports,162*, 1-13.

Alm, H., & Brehmer, B. (1982). Hypotheses about cue-criterion relations in linear and random inference tasks. *Umea Psychological Reports, 164*, 1-20.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Andersson, H., & Brehmer, B. (1977). Transfer effects in the SJT interpersonal learning paradigm. *Umea Psychological Reports, 131*, 1-42.

Balzer, W. K., Doherty, M. E., & O'Connor, R. Jr. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106*, 410-433.

Balzer, W. K., Sulsky, L. M., Hammer, L. B., & Sumner, K. E. (1992). Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? *Organizational Behavior and Human Decision Processes, 53*, 35-54.

Battig, W. F. (1979). The flexibility of human memory. In L. S. V. Cermak & F. I. M. Craick (Eds.), *Levels of processing in human memory* (pp. 23-44). Hillsdale, NJ: Lawrence Erlbaum.

Bauer, M. (1971). Prediction and probability judgment in a cue-probability learning task during two non-feedback phases. *Umea Psychological Reports, 54*, 1-36.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L-G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological Perspectives* (pp. 109-133). New York: Psychology Press.

Björkman, M. (1965). Studies in predictive behavior: Explorations into predictive judgments based on functional learning and defined by estimation, categorization, and choice. *Scandinavian Journal of Psychology, 6*, 129-156.

Boud, D., Keogh, R., & Walker, D. (Eds.) (1985). *Reflection: Turning experience into learning*. London: Kogan Page.

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75-114). Amsterdam, The Netherlands: North-Holland Elsevier.

Brehmer, B. (1972). Cue utilization and cue consistency in multiple-cue probability learning. *Organizational Behavior and Human Decision Processes, 8*, 286-296.

Brehmer, B. (1977). General and specific effects of practice in using functional rules in inference tasks. *Umea Psychological Reports, 114*, 1-20.

Brehmer, B. (1979). Effect of practice on utilization of nonlinear rules in inference tasks. *Scandinavian Journal of Psychology, 20*, 141-149.

Brehmer, B. (1987). Note on subjects' hypotheses in multiple-cue probability learning. *Organizational Behavior and Human Decision Processes, 40*, 323-329.

Brehmer, B., & Almqvist, P. (1977). Learning two multiple-cue probability learning tasks in succession. *Umea Psychological Reports, 130,* 4-32.

Brehmer, B., & Joyce, C. R. B. (Eds.) (1988). *Human judgment: The SJT view*. Amsterdam, The Netherlands: North-Holland Elsevier.

Brehmer, B., & Lindberg, L. A. (1973). Retention of single-cue probability learning tasks as a function of cue validity, retention interval, and degree of learning. *Journal of Experimental Psychology, 101*, 404-406.

Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist, 52*, 399-413.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review, 50*, 255-272.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193-217.

Brunswik, E., & Herma, H. (1951). Probability learning of perceptual cues in the establishment of a weight illusion. *Journal of Experimental Psychology*, *41*, 281-290.

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review, 6*, 263-296.

Carlson, R. A. (1989). Processing nonlinguistic negation. *American Journal of Psychology*, *102*, 211–224.

Carlson, R. A., & Schneider, W. (1989). Acquisition context and the use of causal rules. *Memory and Cognition*, *17*, 240–248.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 517–526.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 484-496.

Chandana, S. R., Movva, S., Arora, M., & Singh, T. (2008). Primary brain tumors in adults. *American Family Physician,* 77, 1423-1430.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Chasseigne, G., Mullet, E., & Stewart, T. R. (1997). Aging and multiple cue probability learning: The case of inverse relationships. *Acta Psychologica, 97*, 235-252.

Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10,* 35-49.

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Cohen, M. S., Freeman, J. T., & Thompson, B. B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. In J.A. Cannon-Bowers & E. Salas (Eds.). *Making decisions under stress: Implications for individual and team training* (pp. 155-190). Washington, DC: American Psychological Association.

Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Meta-recognition in time stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*, 206-219.

Cross, E. S., Schmitt, P. J., & Grafton, S. T. (2007). Neural substrates of contextual interference during motor learning support a model of active preparation. *Journal of Cognitive Neuroscience, 19*, 1854 - 1871.

De Croock, M. B. M., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior, 14*, 249-267.

DeShon, R. P., & Alexander, R. A. (1996). Goal setting effects on implicit and explicit learning of complex tasks. *Organizational Behavior and Human Decision Processes, 65*, 18-36.

Detterman, D. K., & Sternberg, R. J. (Eds.) (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex Publishing.

Doherty, M., & Balzer, W. (1988). Cognitive feedback. In B. Brehmer & C. Joyce (Eds.), *Human Judgment: the SJT view* (pp. 163-197). Amsterdam, The Netherlands: Elsevier North Holland.

Drillings, M., & Serfaty, D. (1997). Naturalistic decision making in command and control. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 71-80). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Edland, A. (1993). The effects of time pressure on choices and judgments of candidates to a university program. N. J. Castellan Jr. (Ed), *Individual and group decision making: Current issues* (pp. 43-64). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, *32*, 53-88.

Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying physics text: Differences in study processes between good and poor performers. *Cognition and Instruction*, *7*, 41-54.

Fishburn, P. C. (1970). *Utility theory for decision making*. Huntington, NY: Robert E. Krieger Publishing Co.

Freeman, J. T., & Cohen, M. S. (1996). Training for complex decision-making: A test of instruction based on the recognition / metacognition model. *Proceedings of the 3rd International Command and Control Research and Technology Symposium 1996*. Retrieved May 12, 2005 from www.cog-tech.com/Publications/TrainingPubs.htm.

Frijters, S., Ten Dam, G., & Rijlaarsdam, G. (2008). Effects of dialogic learning on value-loaded critical thinking. *Learning and Instruction, 18,* 66-82.

Gattie, G., & Bisantz, A. M. (2006). The effects of integrated cognitive feedback and task conditions on training in a dental diagnosis task. *International Journal of Industrial Ergonomics, 36*, 485-498.

Gigerenzer, G., & Kurz, E. (1999). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 342-348). New York: Oxford University Press.

Greeno, J. G. (1964). Paired-associate learning with massed and distributed repetitions of items. *Journal of Experimental Psychology, 67, 286-295.*

Halpern, D. F. (1997). Students need 'reality checks'. *Education Digest, 63,* 24-26.

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking* (4th ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62*, 255-262.

Hammond, K. R. (1996). How convergence of research paradigms can improve research on diagnostic judgment. *Medical Decision Making, 16*, 281-287.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71*, 438-456.

Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). *Human judgment and decision making: Theories, decisions, and procedures*. New York: Praeger Publications.

Hammond, K. R., & Summers, D. A. (1965). Cognitive dependence on linear and non-linear cues. *Psychological Review, 72*, 215-224.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons Ltd.

Hoffman, P. J., Earle, T. C., & Slovic, P. (1981). Multidimensional functional learning (MFL) and some new conceptions of feedback. *Organizational Behavior and Human Performance, 27*, 75-102.

Hoffman, R. R. (Ed.) (2007). *Expertise* out of context: Proceedings of the sixth International Conference on Naturalistic Decision Making. Mahwah: Lawrence Erlbaum Associates, Inc.

Hogarth, R. M. (1980). *Judgment and choice*. New York: John Wiley and Sons Ltd.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*, 332-340.

Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review, 71*, 42-60.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

Jamieson, B. A., & Rogers, W. A. (2000). Age-related effects of blocked and random practice schedules on learning a new technology. *The Journals of Gerontology series B: Psychological Sciences and Social Sciences, 55*, 343-353.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-292.

Klaczynski, P. (2001). Framing effects on adolescent task representations, analytic and heuristic processing, and decision making: Implications for the normative/descriptive gap. *Applied Developmental Psychology, 22*, 289–309.

Klayman, J. (1988a). Cue discovery in probabilistic environments: Uncertainty and experimentations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14,* 317-330.

Klayman, J. (1988b). On the how and why (not) of learning from outcomes. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The Social Judgment Theory view* (pp. 115-162). Amsterdam, The Netherlands: North-Holland

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine system research* (Vol. 5, pp. 47-92). Greenwich, CT: JAI Press, Inc.

Klein, G. A. (1997). The current status of the naturalistic decision making framework. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: Emerging themes and applications* (pp. 11-28). Aldershot, UK: Ashgate.

Klein, G. A. (1998) *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. A., & Calderwood, R. (1991). Decision models: Some lessons from the field. *IEEE Transactions on Systems, Man and Cybernetics, 21*, 1018-1026.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the 30th Annual Human Factors Society* (Vol. 1, pp. 576-580). Dayton, OH: Human Factors Society.

Klein, G. A., Moon, B. A., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*, 70-73.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (Eds.) (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishing Co.

Koele, P. (1980). The influence of labeled stimuli on nonlinear multiple-cue probability learning. *Organizational Behavior and Human Decision Processes, 26*, 22-31.

Koh, K. (1993). Induction of combination rules in two-dimensional function learning. *Memory and Cognition, 21*, 573-590.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 187-194.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Kunen, S., Green, D., & Waterman, D. (1979). Spread of encoding effects within the nonverbal visual domain. *Journal of Experimental Psychology: Human Learning and Development, 5*, 574-584.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Lee, A. Y., & Hutchison, L. (1998). Improving learning from examples trough reflection. *Journal of Experimental Psychology: Applied, 4*, 187-210.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 730-746.

Lehmann, E. L. (1950). Some principles of the theory of testing hypotheses. *Annals of Mathematical Statistics, 21*, 1-26.

Li., Y., & Wright, D. L. (2000). An assessment of the attention demands during random and blocked training schedules. *The Quarterly Journal of Experimental Psychology, 53*, 591-606.

Lindberg, L. A., & Brehmer, B. (1976). Transfer in single-cue probability learning. *Organizational Behavior and Human Decision Processes, 16*, 177-192.

Lindell, M. K. (1976). Cognitive and outcome feedback in multiple-cue probability learning tasks. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 739-745.

Lipshitz, R., & Ben Shaul, O. (1997). Schemata and mental models in recognition-primed decision making. In C. E. Zsambok & G. A. Klein (Eds.), *Naturalistic decision making,* (pp. 293-303). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Magill, R. A., & Hall, K. G. (1990). A review of contextual interference effect in motor skill acquisition. *Human Movement Science, 9*, 241-289.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.

Muchinsky, P. M., & Dudycha, A. (1975). Human inference behavior in abstract and meaningful environments. *Organizational Behavior and Human Performance, 13* , 377-391.

Nathan, M. J., Mertz, K., & Ryan, B. (1994). *Learning through self-explanation of mathematical examples: Effects of cognitive load.* Paper presented at the 1994 Annual Meeting of the American Educational Research Association.

Orasanu, J., & Connolly, T. (1993), The reinvention of decision making. In G. A. Klein, J.Orasanu, R.Calderwood, & C. E. Zsambok (Eds), *Decision making in action: Models and methods* (pp. 3-21). NJ: Ablex Publishing Corporation.

Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86, 122-1*33.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*, 144-161.

Raiffa, H. (1997). *Decision analysis: Introductory readings on choices under uncertainty*. New York: McGraw Hill.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1-29.

Rothstein, H. G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes, 37*, 83-92.

Ruble, T. L., & Cosier, R. A. (1990). Effects of cognitive styles and decision setting on performance. *Organizational Behavior and Human Decision Processes, 46*, 283-295.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*, 611-621.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schneider, V. I., Healy, A. F., & Bourne, L. E. Jr. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 78-90). Mahwah, NJ: Erlbaum.

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention and transfer. *Journal of Memory and Language, 46*, 419-440.

Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E., Jr. (1995). The effects of contextual interference on the acquisition and retention of logical rules. In A. F.

Healy & L. E. Bourne Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 95-131). Thousand Oaks, CA: Sage Publications.

Serfaty, D., MacMillan, J., Entin, E. E., & Entin, E. B. (1997). The decision-making expertise of battle commanders. In C. E. Zsambok & G. A. Klein (Eds.), *Naturalistic decision making* (pp. 233-246). New York: Lawrence Erlbaum.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179-187.

Sheets, C. A., & Miller, M. J. (1974). The effect of cue-criterion function form on multiple-cue probability learning. *American Journal of Psychology, 87,* 629-641.

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven, MA: Yale University Press.

Siegel, H. (1991). The generalizability of critical thinking. *Educational Philosophy and Theory*, *23*, 18-30.

Simon, D. A. (2007). Contextual interference effects with two tasks. *Perceptual and Motor Skills, 105, 177-183.*

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review, 57*, 193-216.

Smedslund, J. (1955). *Multiple-probability learning: An inquiry into the origins of perception*. Oslo, Norway: Oslo University Press.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, *23*, 701-726.

Stark, R., Mandl, H., Gruber H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction, 12*, 39-60.

Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1997). A team perspective of situational awareness: Cueing training. In *Proceedings of the 19th Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL* (pp. 174-182) [CD ROM].

Ten Dam, G., & Volman, M. (2002). Critical thinking as a citizenship competence: Teaching strategies. *Learning & Instruction, 14*, 359-370.

Thorndike, E. L. (1911). The contribution of psychology to education. *Journal of Educational Psychology*, *1*, 5-12. Retrieved august 2008 from http://psychclassics.yorku.ca/Thorndike/education.htm

Todd, P., & Gigerenzer, G. (2001). Putting naturalistic decision making into the adaptive toolbox. *Journal of Behavioral Decision Making*, *14*, 353-384.

Tsao, C. J. (1994). Factors affecting multiple-cue-probability learning: Evaluative feedback and time pressure. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 55*, 34-58.

Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education, 40*, 185-200

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., De Croock, M. B. M., & Jelsma, O. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual and Motor Skills, 84*, 784-786.

Van Merriënboer, J. J. G., & Paas, F. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior, 6*, 273-289.

Vicente, K. J. (1988). Adapting the memory recall paradigm to evaluate interfaces. *Acta Psychologica*, *69*, 249-278.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*, 33-57.

Wigton, R. S., Patil, K. D., & Hoellerich, V. L. (1986). The effect of feedback in learning clinical diagnosis. *Journal of Medical Education, 61*, 816–822.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychological Bulletin & Review, 9*, 185-211.

Zsambok, C. E., & Klein, G. A. (Eds.) (1997). *Naturalistic decision making*. Mahwah, N.J.: Lawrence Erlbaum Associates, Inc.

Chapter 3
The Effects of Critical Thinking Instruction on Learning Complex Judgment and Decision Making

This chapter is based on Helsdingen, A. S., Van den Bosch, K., Van Gog, T., & Van Merriënboer, J. J. G. (2008). The effects of critical thinking instruction on learning complex judgment and decision making. Submitted to *Applied Cognitive Psychology*.

Abstract

Two field studies on the effects of critical thinking instruction were conducted. Study 1 investigated the effects on learning and near transfer and Study 2 investigated the effects on far transfer. Participants received a training with scenario-based exercises in both simplified (Study 1) and high fidelity (Study 2) learning environments. In both studies, half of the participants received instruction in critical thinking. The other half received the same exercises, but without critical thinking instruction. After the training, test scenarios were administered to both groups. Results provide support for the hypotheses that critical thinking instruction has a positive effect on transfer test performance and that this benefit is greater for far transfer performance than for near transfer performance.

Studies on decision strategies of experienced decision makers in complex environments (e.g., military command and control, crisis management) have shown that experienced decision makers not only have a large body of knowledge, but also apply deliberate problem-solving strategies that differ significantly from novices' strategies (Endsley, Hoffman, Kaber, & Roth, 2007). When faced with a complex and unfamiliar problem, experienced decision makers collect and critically evaluate the available evidence, seek for consistency, and test assumptions underlying their assessment of the problem (Klein, Moon, & Hoffman, 2006). Novices, on the other hand, tend to focus and act upon single observations without trying to build and test a comprehensive model of the situation. Usually, expert problem solving strategies are not part of a training program for professional decision making; they are only acquired as a result of experience in the field (Anderson, 1993; Klein, 1998; Lipshitz, 1993).

Cohen and colleagues (Cohen, Freeman, & Thompson 1997; Cohen, Freeman & Wolf, 1996; Freeman & Cohen, 1996) developed an instructional concept that combined instruction aimed at acquisition and application of domain knowledge with instruction on explicit problem-solving strategies: Critical thinking instruction. Several evaluation studies showed promising results: After critical thinking instruction, trainees considered more observations, identified more cause and effect relations, provided better arguments, and made better decisions according to subject-matter experts (e.g., Cohen, Freeman, & Thompson, 1998; Freeman & Cohen, 1996). However, these studies had some limitations. First of all, they were conducted in simplified learning environments. Moreover, performance of participants who received critical thinking instruction was compared to that of participants who did not receive any instruction or practice at all. And finally, it was not established whether critical thinking instruction had differential effects on near and far transfer performance. Therefore, we conducted two field studies on the effects of critical thinking instruction on learning and transfer of complex judgment and decision making in tactical command.

Complex Judgment and Decision Making in Tactical Command

Military commanders are responsible for the preparation, execution, and management of military operations in unstable and complex conditions. They have to decide on deployment of sensor systems, weapons, and personnel. Therefore, they have to assess situations quickly, identify threats, and make assumptions on intentions of enemies and other parties to predict what will happen and what consequences their own actions will have. In the present studies, the effects of critical thinking instruction are tested in two training programs for tactical command and control: (1) ground-to-air defence operations for air force officers, and (2) naval air and surface warfare for naval officers.

In ground to air operations, the task is to plan and manage the deployment of several weapon systems (e.g., Patriot, Hawk, Stinger) to provide air defence for a geographical area or specific object of interest. The decision maker has to prioritize elements in the area s/he has to defend, taking into account enemy intent, weapons capabilities, current capacity, and so forth. During an engagement, the decision maker may have to revise the defence plan, command the damage repair according to priorities, command casualty transportation, decide which targets to engage, and so forth.

In naval air and surface warfare the decision maker has to plan manoeuvres of his or her frigate in such a way that the ship provides maximum defence against air and surface threats to an object of priority (e.g., another ship carrying humanitarian goods) while still continuing the mission. The decision maker has to monitor the situation, anticipate threats, decide on deployment of passive or active sensor systems, command forward patrol and reconnaissance aircrafts (e.g., helicopters), decide on engagements of targets, and so forth.

In summary, these tactical command and control tasks are complex and dynamic, set in quickly changing, uncertain and interactive environments. Education and training programs in military command and control used to focus mainly on decision making through analysis and deductive reasoning (Czerwinsky, 1996). Commanders were taught to collect lots of information, generate all possible options, and weigh probabilities and expected effects to select an optimal course of action. Such an approach, however, is based on the false assumption that the environment is largely stable, predictable, and knowable. And the operational environment of military command and control simply does not provide enough time to decision makers to generate and weigh all options. Experienced decision makers apply other reasoning strategies that enable them to quickly chose an appropriate option. This is not necessarily the optimal course of action, but one that usually suffices (Johnson & Raab, 2003). Therefore, naturalistic decision making research has provided different approaches for learning command and control decision making (Cohen, et al., 1997; Klein, McKloskey, Pliske, & Schmidt, 1997). These approaches focus on teaching reasoning strategies of experienced decision makers and on making the implicit information processing explicit.

Critical Thinking Instruction

Experienced decision makers differ from novices on two aspects. First, experienced decision makers have more and better organized knowledge of the domain they are working in. Their knowledge is represented and organized in cognitive schemas, that is, networks of abstract mental concepts (Anderson, 1977; Bartlett, 1932; Iran-Nejad, 2000; Minsky, 1975; Neisser, 1967; Rumelhart, 1975). These schemas facilitate recognition and categorization of problem situations, thus guiding the identification of an appropriate response (Graesser, Gordon, & Sawyer, 1979). The second difference between novices and experienced decision makers concerns the (meta)cognitive processes that guide and critically review the decision-making process (Cohen, et al., 1996; Ertmer & Newby, 1996). In their assessment of the situation, experienced decision makers integrate the available information into its context, which may include elements such as the history of events leading to the current situation, and the intentions, capacities, and opportunities of others. Novices, in contrast, often consider aspects of the situation separately and independently (Zsambok & Klein, 1997). Furthermore, experienced decision makers critically test assumptions, actively search for inconsistencies or conflicts, and reflect upon their own decision strategies (Olsen & Rasmussen, 1989).

Critical thinking instruction aims at teaching both recognition-primed decision making and (meta)cognitive techniques (Cohen, et al. 1996; Freeman & Cohen, 1996)This requires exposure to many prototypical problems during learning along with specific instruction in critical thinking. Typical critical thinking instruction involves an explanation of the decision making process before the learning phase. In addition, an instructor will typically prompt the learner to reflect on his or her cognitive strategy and initiate one of the critical thinking steps during the learning phase. These steps include (Freeman & Cohen, 1996):

Creating a story: A story is a comprehensive assessment of the situation, in which all existing evidence is incorporated and explained and assumptions are made about uncertain aspects of the situation. Past, present and future are addressed in the story. The purpose of story building is to keep participants from assessing situations solely on isolated events. Instead, participants are taught how they can integrate the available information into its context. For example, consider a naval commander aboard a frigate intercepting a brief emission of an enemy helicopter radar. The frigate is in silent mode, that is, no sensor or weapon systems are active, only passive sonar is deployed since presence of an enemy submarine is expected, and the frigate is sailing in a shipping lane thus mimicking a freighter. Novice commanders may focus solely on the interception, interpret that as a sign that a helicopter is close by and has detected their frigate, and has switched off its radar to approach the frigate to make a visual identification. The result of that interpretation would probably be that novice commanders immediately go into an active state, deploying all sensors and weapons, planning to destroy the helicopter before it can communicate with other enemy forces. However, experienced commanders may also consider information from sonar, the fact that it is the first interception, and the intentions and capacities of the enemy. They may reason that although the helicopter has detected their presence, it is probably still at a large distance and is probably classifying all ships in the shipping lane. These considerations combined with the absence of any enemy submarines on the sonar images may lead the experienced commander to decide not to go into active state, but remain silent and unsuspicious.

Testing a story: Participants are instructed how to identify inconsistency and uncertainty, and how to adjust or refine their story by deliberate testing. They have to correct these problems by collecting more data, retrieving knowledge from memory, making assumptions about the missing piece of the story, or by resolving conflicts in the argumentation. Consider, again, the frigate commander having intercepted the helicopter radar emission. The experienced commander remains silent because s/he expects the helicopter to be at a large distance, not within weapons range. S/he has to identify gaps in the story, one of which is information on the helicopter's distance. By closely monitoring the bearing of the interception, seeing how it changes, combining that information with the exact moment the emission is lost and with information on the position of other ships, s/he can estimate its likely location. The commander may also identify conflicting evidence: The presence of the helicopter is usually an indication that an enemy submarine is in the area, however, the sonar images show nothing. To resolve this conflict, s/he may manoeuvre the ship in such a way that another sonar image is got that might reveal an enemy submarine. Or s/he may assume that the helicopter is on a regular reconnaissance mission instead of being a forward sensor for a submarine.

Evaluating a story: After a story is constructed and tested for gaps, it should be evaluated for its plausibility. The decision maker has to take a step back, identify critical or hidden assumptions and play the devil's advocate by falsifying these assumptions, that is, explaining how an assumption can be false and build an alternative story. In our previous example, the commander had estimated the helicopter's likely position. S/he will also time the period between two intermissions to estimate at what distance from any ship the helicopter switches off its radar. The idea behind this approach is that the helicopter is building a tactical picture of the area, and switches off its radar when it approaches a ship for a visual identification and classification. At a certain point, the commander may assume that the frigate may be next to be identified and classified. Based on these assumptions on the helicopter's speed, it can be calculated that the helicopter may be within weapons range 2 minutes after it switches off its radar. Thus, s/he plans to go active state two minutes after the next time the helicopter switches off its radar. This will give the opportunity to engage the helicopter before the helicopter can communicate the frigate's position and direction to other enemy forces. S/he realizes that the critical assumption is the current distance of the helicopter. By monitoring its behaviour, the commander has assumed that the helicopter is now close, but this may be wrong. S/he looks at the tactical picture of the area and tries to find out whether the helicopter's manoeuvring history can also be explained in another way that would have resulted in the helicopter now being at a greater distance. S/he also realizes that evasive manoeuvres and an engagement with an enemy submarine must be planned if the helicopter for some reason cannot be destroyed.

Quick test: Critical thinking is not always appropriate. Decision makers have to evaluate the time available and the consequences of their actions. In stressful situations such as those often encountered by professional decision makers, there is usually little time to spare. The decision makers should act immediately *unless* the risk of delay is acceptable, the costs of errors are high (i.e., when lives are at stake), and/or the situation is non-routine or problematic (i.e., complex or novel problems). For easy, routine problems that are not important, critical thinking is not expected to increase decision quality (Cohen et al., 1998). In our example throughout the steps above, the experienced commander initially took the extra time to continue to assess the situation and took action when s/he thought there was an urgency to keep the initiative. The novice commander, however, went to active state immediately, probably only to find that the helicopter was too far to engage. But the active sensor and weapon systems may have warned the enemy helicopter and possible submarines in the area of the frigate's presence, making it a so called 'sitting duck'. In this case, s/he could have delayed a decision, and the cost of error was high.

Cohen and colleagues (Cohen, et al., 1996; Cohen et al., 1998; Freeman & Cohen, 1996) attributed the benefits of critical thinking instruction to participants having acquired an appropriate decision making strategy. In addition to that, we assume that critical thinking instruction may help participants acquire more elaborate knowledge representations, or schemas, and a deeper, more abstract level of understanding of the learning content because it forces learners to compare different pieces of information, identify cause and effect relationships between events, and differentiate between observations and assumptions (cf.

Biggs, 1985; Berthold, Nückles & Renkl, 2007; Russel, 2002; Schön, 1983). In short, learners are required to focus not just on the superficial aspects of situations or on mere observations, but on the underlying structures and causes. This enhanced understanding, in turn, should lead to better transfer.

Transfer of Training

Transfer-of-training refers to the transfer of skills learned under specific conditions or from practicing a specific task to other conditions or tasks (Mayer & Wittrock, 1996; Roscoe & Williges, 1980). Transfer of knowledge and skills from the learning tasks to professional tasks is usually the ultimate goal of education and training programs. A distinction is often made between *near* and *far* transfer. Near transfer tasks share structural features but differ on superficial or surface features from the learning tasks, whereas far transfer tasks differ from the learning tasks on both surface and structural features (Holyoak & Koh, 1987; Quilici & Mayer, 1996). Thus, the degree of transfer depends on the degree to which task activities call upon the same skill, but also on the learner's level of understanding of a task (Detterman & Sternberg, 1993; Mayer & Wittrock, 1993). Near transfer can be obtained if the learner knows *what* to do, whereas far transfer is usually possible only if the learner also knows *when* and *why* s/he needs to do that (Van Merriënboer, 1997). To gain such deep understanding required for far transfer, a process of abstraction is needed, from knowledge of the task at hand to a higher level of knowledge, for instance, of the general principles of a domain or the general procedures for performing a certain type of task (Van Merriënboer & Paas, 1990).

Critical thinking instruction is expected to deepen understanding of learning materials, and consequently improve far transfer performance because, as stated before, learners are required to focus not on the superficial aspects of situations or on mere observations, but on the underlying structures and causes. Also, with critical thinking instruction learners are expected to explain their reasoning processes and reflect on them. And several studies have shown that processes such as self-explanation may enhance reflection and consequently transfer (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; Ferguson-Hessler & de Jong, 1990; Lee & Hutchinson, 1998; Renkl, 1997).

The Present Studies

We present two field studies that assess the effects of critical thinking instruction on learning complex judgment tasks. The first study focuses on skill acquisition and transfer performance of individual air force officers learning to command and control ground-to-air defence operations. The second study investigates naval (petty) officers learning to command

and control air and surface warfare. It is hypothesized that critical thinking instruction has a positive effect on transfer performance and that this benefit is greater for far transfer performance than for near transfer performance.

Study 1

The first study was conducted in the Tactical Command Station of a ground-to-air defence battalion of the Royal Netherlands Air Force. The effects of critical thinking instruction on learning processes and transfer test performance were assessed.

Method

Participants and Design

Sixteen officers of the Royal Netherlands Air Force volunteered to participate in the experiment (15 male, 1 female; age M = 32.31, SD = 3.81). Prior experience in ground-to-air defence tasks ranged from 1 to 5 years (M = 3.69, SD = 2.14). The supervising project officer matched participants according to their tactical education and experience, and assigned participants from each pair randomly to the critical-thinking condition (n = 8) or the control condition (n = 8).

Materials

Practice scenarios. Six practice scenarios were developed. The paper-and-pencil scenarios encompassed a starting point and a description of events specified in time. They required the participants to make assessments of the geographical area, identify priorities in the area to be defended, assess threats and perform certain actions such as planning engagements of targets, deploy sensor and weapon systems, make a damage repair plan, plan for transportation of casualties, reallocate resources, and make new priorities. The scenario leader introduced the scenario events and provided outcome-feedback at moments that were pre-specified in the scenario. Events were specified in minutes from the start (e.g., X = start scenario, event Y starts at X + 5 min., event Z at X + 10 min., etc.) and feedback moments were specified relative to participants' reaction to events (e.g., after participants' assessment of the situation, provide feedback on the accuracy of their assessment). Feedback referred to the accuracy of the participants' assessment and actions, and if necessary information on the correct assessment or action to continue the scenario. Participants played the role of battle captain and the scenario leader played all other roles. Each scenario took approximately 45 minutes to complete.

Critical thinking instruction. In the critical thinking condition, participants received a paper-based instruction explaining the four steps in the process of critical thinking: (1) creating a story of the situation, (2) testing that story, (3) evaluating its plausibility and finding alternative stories (i.e., contingency plans), and (4) quick consideration of the need to decide immediately or spend more time on the critical thinking process. Participants were instructed to think aloud during the whole process (see Ericsson & Simon, 1993). Besides the textual explanation, a hand-out with a graphic representation of the four steps (see Figure 1) and two demonstrations of critical thinking in real-world situations were provided. In each demonstration, two scenario leaders showed how critical thinking should be used in the scenarios; one of them played the role of participant (i.e., the battle captain) and one of them played the other roles.

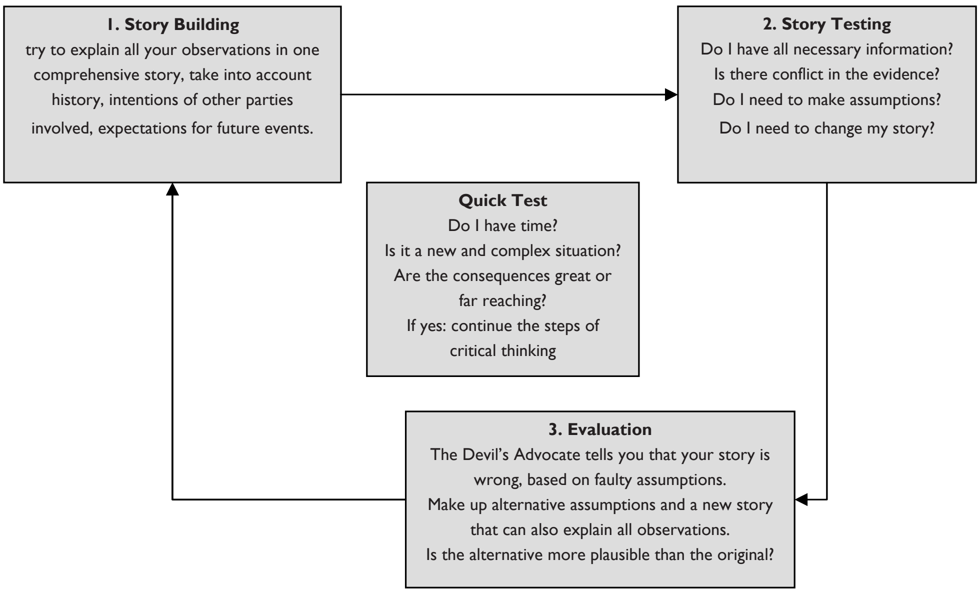

Figure 3.1: Critical thinking steps (cf. Freeman & Cohen, 1996)

Feedback during the execution of the practice scenarios consisted of process feedback, focused on the participant's decision strategy and critical thinking processes. Instructors closely monitored participants' verbalizations of their decision making processes and matched that with the required assessments as specified in the scenario. If they identified gaps or missing or wrong assessments, they would ask open ended questions to encourage the participant to critically reflect on his own decision process, acknowledge the gap, or the missing or wrong assessment, and take recuperating actions. These open ended questions focused on one of the four steps of the critical thinking process (see Table 3.1). When participants were adequately engaged in the critical thinking process, that is, following the steps of critical thinking and arriving at good assessments, instructors only made encouraging remarks (e.g., 'very well' or 'I understand'). Instructors did not provide any other feedback or information

Table 3.1

Example Questions for Instructors to Encourage Participants to Engage in the Steps in Critical Thinking

Critical thinking step	Questions
1. Story building	Can you describe what the situation is? What caused all this? What are the intentions of the enemy and why? What do you expect to happen?
2. Story testing	What additional information do you need to make a reasonable prediction? Did you identify any conflicting evidence? Did you incorporate all aspects of the situation, all observations, in your story?
3. Story evaluation	Could there be an alternative explanation for all these observations? How would the situation unfold if this alternative explanation were true?
4. Quick test	Why do you need to make a decision immediately? Is the situation complex and are consequences great or far reaching? Can you spend more time thinking?

Control group instruction. The control group participants were given a presentation of the organization and projects of the research institute where two of the authors worked to engage them during the critical thinking instruction of the other group. They also received a short introduction regarding the purpose and context of the tactical practice scenarios they were about to receive. Furthermore, two example scenarios were demonstrated by two scenario leaders. These demonstration scenarios were being played as normal command post exercises, without critical thinking. During execution of the practice scenarios, they received outcome feedback only (e.g., “that was a good assessment”, “you made the right decision”, “you should not have done that”, “the rules of engagement do not allow you to take these actions”, “you failed to destroy the target”).

Test scenarios. The test scenarios were two paper-and-pencil scenarios in which the participants again played the role of battle captain and the scenario leader covered all other roles. Neither outcome feedback nor process feedback was provided during the execution of the test scenarios. Participants were required to think aloud during execution of the test scenarios.

Procedure

The experiment had a duration of approximately 10 hours, and was run in five sessions of approximately two hours each. The sessions were held at five consecutive days with one or two participants at the same time. In the first session, participants of the critical thinking group received the critical thinking instruction followed by the two demonstration scenarios including the critical thinking steps. The control group was given the general presentation and witnessed the same two demonstration scenarios but without critical thinking. In the three subsequent learning sessions (sessions 2-4), two practice scenarios were played each time. The participants played the role of battle captain and a scenario leader played the other roles. There was a 15-minute break between the scenarios. The order of the 6 practice scenarios was randomly selected from all possible scenario sequences for each participant. Consecutive practice scenarios were facilitated by different scenario leaders, that is, participants practiced three scenarios with one scenario leader and the other three with another scenario leader. The participants were instructed to think aloud during the scenarios to provide the scenario leader insight into the decision strategies and critical thinking process, enabling the evaluation of participants' performance and provision of feedback when a required assessment was made or missed. About two to three times during a practice scenario, the scenario leader provided outcome or critical thinking feedback, during which the scenario was 'frozen' for a few minutes. Scoring of the performance indicators was done continuously during the scenario, while the behaviour was exhibited by participants. The performance measures were graded at the end of a scenario.

In the fifth session two test scenarios were played. Order and assignment of the test scenarios to scenario leaders was balanced. Again, all participants were asked to think aloud, however, no feedback was given. Participants' verbalizations of their decision and critical thinking strategies were used to grade performance.

Data Analysis

Two types of performance measures were gathered during the practice and test scenarios: (1) outcome measures to assess the quantity and quality of the end result, and (2) process measures, to evaluate the strategies, steps and procedures used to accomplish the task. Conform the 'Command and Control Performance Measurement Tool' (Van Berlo & Schraagen, 2000), performance indicators were specified for both the outcomes and process measures, enabling the evaluators to observe and interpret the decision making processes and outcomes. These performance indicators were formulated concisely, and were easily scored in terms of whether the behaviour was observed or not; this was indicated in the respective column (Yes or No) and an overall grading (ranging from 1 = very poor, to 10 = excellent) for that performance measure was determined. Prior to the experiment proper, scenario leaders had used the results of a pilot-study (using the same scenarios but with different participants) to reach a common understanding of assigning grades.

Quality of actions and contingency plans were used as outcome measures. Performance indicators for quality of actions were: Concise, unambiguous, stating the right priorities, conform expert solution, and logical result of the assumptions and argumentation. Performance indicators for quality of contingency plans involved: The availability of contingency plans, concise plan, unambiguous plan, and a logical result of (expressed) conflicting evidence or weak assumptions underlying the plan. Information processing and argumentation served as process measures. For information processing, the performance indicators were: Being able to distinguish between relevant and irrelevant information, recognition of conflicting evidence, and identification of missing information. For argumentation, the performance indicators were: Recognition of cause and effect relationships, incorporation of history of the situation, explication of expectations for the future, plausibility of these expectations, recognition of uncertainties, and differentiation between observations and assumptions.

These performance measures, that is, information processing, argumentation, quality of actions, and quality of contingency plans, were graded on a scale between 1 to 10. The grades were used for analysis of the effects of critical thinking instruction on trainee performance during learning scenarios and on the test.

Results

The grades for information processing, argumentation, quality of actions, and quality of contingency plans were calculated for each participant for the six learning scenarios and the two test scenarios, and were analyzed using the Wilcoxon signed ranks test for two dependent samples. The members of the matched pairs differed in instruction (critical thinking vs. no critical thinking). In all analyses reported below, a significance level of .05 is set.

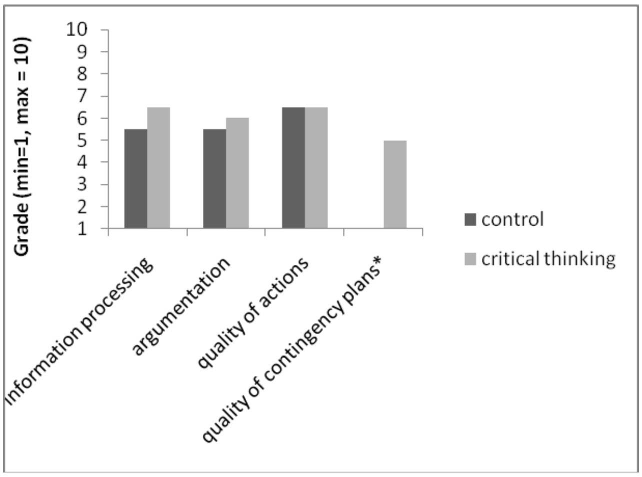

Figure 3.2: Median performance grades of participants of the critical thinking group and the control group in the learning phase.

Learning phase

The median scores for information processing, argumentation, action plan quality, and contingency plan quality during the learning phase are presented in Figure 3.2. Wilcoxon signed ranks tests showed that critical thinking had an effect on quality of contingency plans (Z = -2.20, p = .03), with participants in the critical thinking group reaching higher grades (*Mdn* = 5, *range* = 2) than participants in the control group (*Mdn* = 1, *range* = 9). However, no significant effects were found of critical thinking in the learning phase on information processing (Z = -1.17, p=.09), argumentation (Z = -1.14, p=.26), or quality of actions (Z = -1.00, p=0.32).

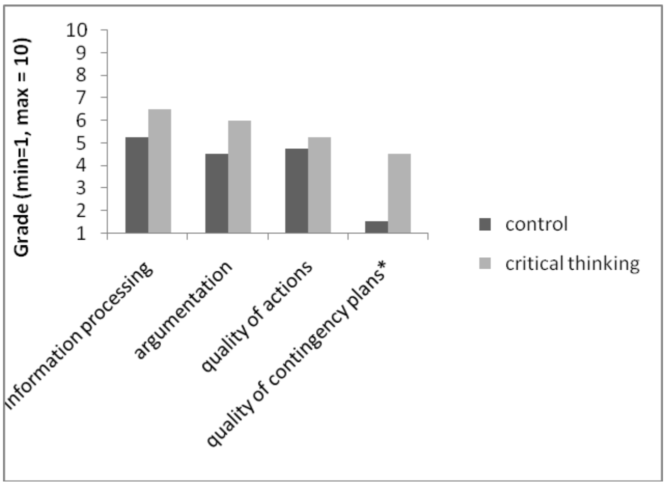

Figure 3.3: Median performance grades of participants of the critical thinking group and the control group in the test phase.

Test phase

The median scores for information processing, argumentation, action plan quality, and contingency plan quality on the test are presented in Figure 3.3. As with the results of the learning phase, Wilcoxon signed ranks tests showed an effect of critical thinking instruction on contingency plan quality (Z = 2.21, p = .03), with participants in the critical thinking group performing better (*Mdn*=4.5, *range* = 3) than participants in the control group (*Mdn*=1.5, *range*=7). Again, no effects were found of critical thinking on information processing (Z = -1.16, p=.25), argumentation (Z = -1.61, p=.11), and quality of actions (Z = -1.13, p=.235).

Discussion

Participants in the critical thinking condition already developed better contingency plans than the control group during the learning phase, and this carried over into the test phase. This is easy to understand, since the critical thinking approach aims to make participants aware of critical assumptions, conflict, and missing information in their situation assessment, and to prepare for contingency. Although the other performance measures failed to show statistically significant effects of critical thinking, a trend can be observed from Figures 3.2 and 3.3 suggesting that participants from the critical thinking group obtained slightly higher grades than the control group. The results suggest that the critical thinking approach might be a useful tool for improving the quality of learning tactical judgment, as earlier explorative studies indicated (e.g. Cohen, Freeman, & Wolf, 1996; Freeman & Cohen, 1996; Klein et al., 1997). However, this study had some limitations. First, scenario leaders' scoring may have been biased by their knowledge of the instructional intervention and the design of the study. Second, the effect of the instruction has been studied in a simplified task environment. Eventually, critical thinking skills need to be applied in the real world. For reasons of transfer it is necessary to investigate whether critical thinking skills can be successfully trained in high-fidelity task environments. Finally, in this study we did not investigate whether there were differential effects of critical thinking instruction on near and far transfer test performance. When critical thinking leads to better and more general understanding of the learning material, we would expect that benefits of this type of instruction and support would be stronger on a far transfer test than on a near transfer test. Hence, a second study was conducted to address these issues, using independent assessments and a more realistic task environment.

Study 2

This study was conducted in the domain of *anti air warfare* and *anti-surface warfare* at the Operational School of the Royal Netherlands Navy. Teams of one participant officer and one petty officer played single ship / single threat scenarios in a high fidelity tactical simulator.

Method

Participants

Sixteen officers of the Royal Netherlands Air Force volunteered to participate in the experiment (15 male, 1 female; age M = 36.08, SD = 6.07), with prior sailing experience ranging from 2 to 17 years (M = 10.23, SD = 5.57). The supervising project officer matched participants according to their tactical education and experience, and assigned participants from each pair randomly to the critical-thinking condition (n = 8) and control condition (n = 8).

Materials

Practice scenarios. Four practice scenarios of approximately 120 minutes each were developed. Two of these scenarios were paper-and-pencil scenarios, two were scenarios to be run in the high fidelity tactical simulator. All four scenarios had a similar structure: They encompassed a starting point, contained events specified in time, and required assessments or actions by the participants. The scenario leader introduced the scenario events and provided feedback at moments that were marked in the scenario. Events were specified in minutes from the start (cf. Study 1) and feedback moments were specified relative to participant's reaction to events (cf. Study 1). Feedback referred to the accuracy of the participant's assessment and actions, and if necessary, provided information on the correct assessment or action to continue the scenario. Participants played the role of principal warfare officer and assistant and the scenario leader played all other roles. The assistant was responsible for tactical picture compilation, interacted with the operators or the system to gather all information and presented that graphically to the principal warfare officer, who interpreted the picture and decided on which actions (manoeuvres, engagements) to take. In working together, assistant and warfare officer usually came to a common understanding of the situation and necessary actions to take, however, sometimes they disagreed. In such cases, the principal warfare officer was responsible and decided on the course of action.

Critical thinking instruction. Similar to Study 1.

Control group instruction. Similar to Study 1.

Transfer tests. Two transfer test scenarios were developed: A near transfer test and a far transfer test. The near transfer test scenario was a variation of the learning scenarios, involving the same background information, the same environment, but with a different timing and direction of threatening events.

The far transfer test scenario was different from the learning scenarios on both superficial and structural level. The scenario's historical, political, and geographical background differed from the learning scenarios, and furthermore, the nature and timing of events was different. Both near and far transfer test scenarios were played in the tactical simulator. No feedback was provided during execution of the test scenarios.

Procedure

Prior to the experiment proper, instructors assigned to critical thinking teams were extensively briefed on the critical thinking instruction, as well as on how to support the teams in the application of the critical thinking processes. Instructors assigned to control teams were not informed about the concept of critical thinking. They were told to support the teams as they would normally do in practice scenarios. The experimenter also briefed the two independent subject matter experts who were evaluators for the transfer test performance evaluation on the scoring procedure and how to use the scale. Evaluators were not informed about the concept of critical thinking instruction nor of the purpose and design of the study. Again, scenarios of a pilot study were used to arrive at a common interpretation of performance indicators.

The experiment had a duration of approximately 20 hours and was run in five daily sessions of approximately four hours. In the first session, participants received the instruction to play the scenarios as a normal command and control exercise. Participants of the critical thinking group received a critical thinking tutorial, followed by a demonstration in which the instructor showed how critical thinking should be used in the scenarios. The control group witnessed a presentation of the research institute where two of the authors work. After that, a demonstration was given of the same example scenario as the critical thinking group received, now being played as a normal command and control exercise without critical thinking. In the subsequent learning sessions (2 and 3), two practice scenarios were played each day, with a 1-hour break in between the scenarios. The order of practice scenarios was predefined and scenario leaders were assigned to scenarios since specific expertise was required for specific scenarios. Participants were asked to think aloud during the scenarios to provide the scenario leader insight into the decision strategies and critical thinking process, enabling the evaluation of participants' performance and provision of feedback when a required assessment was made or missed. About four to five times during a practice scenario, the scenario leader provided outcome or critical thinking feedback, and 'froze' the scenario for a few minutes. Scoring of the performance indicators was done during the work on the scenario. The performance measures were graded four times during the scenario; markers in the scenario description prompted the instructor to do this.

On the fourth session the test scenarios were played in the simulator. Order of the test scenarios was balanced per session: Thus, half the participants played a far transfer test prior to the near transfer test, and the other participants first played the near transfer test and then the far transfer test. Evaluators were assigned to specific scenarios on the basis of their expertise. All participants were asked to think aloud but no support or feedback was given to them. An independent subject matter expert evaluated participants' individual performance on these test scenarios on the basis of the verbalizations of participants. The expert received the scenarios on paper. He scored the performance indicators during execution of the scenario when a participant exhibited the behaviour, and specific markers in the scenario description prompted the evaluator to grade all performance measures at that particular moment, four times per scenario.

Data analysis

The same outcome and process measures as in Study 1 were used. In addition, performance with respect to time management (performance indicators were being aware of deadlines, taking into account the importance and impact of mistakes, knowing when to think more, anticipate events) and team skills were graded. The performance indicators for team skills were adequate information exchange, concise and unambiguous communication, and teammates showing supportive behaviour and initiative/leadership (cf. Smith-Jentsch, Zeisig, Acton, & McPherson, 1998). Because instructors and evaluators were used to using the official NATO 4-point scale, it was decided to use this 4-point scale which ranges from 1 = Unsatisfactory, 2 = Marginal, 3 = Satisfactory, to 4 = Excellent (see Van Berlo & Schraagen, 2000).

Participants were graded individually, thus, within the same team of principal warfare officer and assistant, both could get a different grade on one of the performance measures in the same scenario. Again, the performance indicators served merely as support for the evaluators, only the grades on the six performance measures (measured 4 times per scenario) were taken for data analysis.

Results

The median grades on information processing, argumentation, quality of action plans, quality of contingency plans, time management, and team processes were calculated for each participant for the four learning scenarios, the near transfer test, and the far transfer test. They were analyzed using Wilcoxon signed ranks test for two dependent samples. The members of the matched pairs differed in instruction (critical thinking, no critical thinking). In all analyses reported below, a significance level of .05 is set.

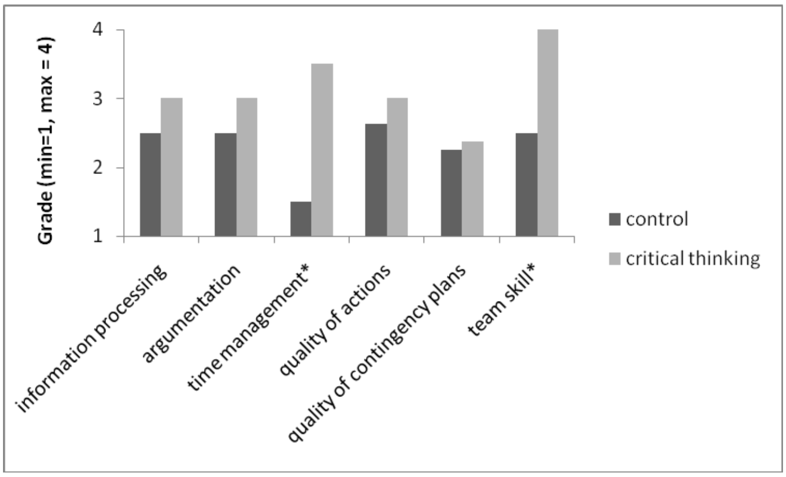

Figure 3.4: Median performance grades of participants of the critical thinking group and the control group in the learning phase.

Learning Phase

The median grades on information processing, argumentation, time management, quality of actions and contingency plans, and team skills for both conditions during the learning phase are presented in Figure 3.4. Analysis of the median grades on information processing (Z = -1.71, p = .09), argumentation(Z = -1.67, p = .09), quality of actions(Z = -1.80, p = .07), and quality of contingency plans (Z = -0.63, p = .53) showed no significant effects of critical thinking in the learning phase. Critical thinking did improve time management and team skills (Z = -2.37, p = .02, and Z = -2.21, p = .03, respectively). Participants in the critical thinking group received higher grades for time management (*Mdn* =3.5 , *range*=2) and for team skills (*Mdn*= 4, *range*=1) than participants from the control group (*Mdn* = 2, *range* =2 and *Mdn* =2.5 , range = 2, for time management and team skills respectively).

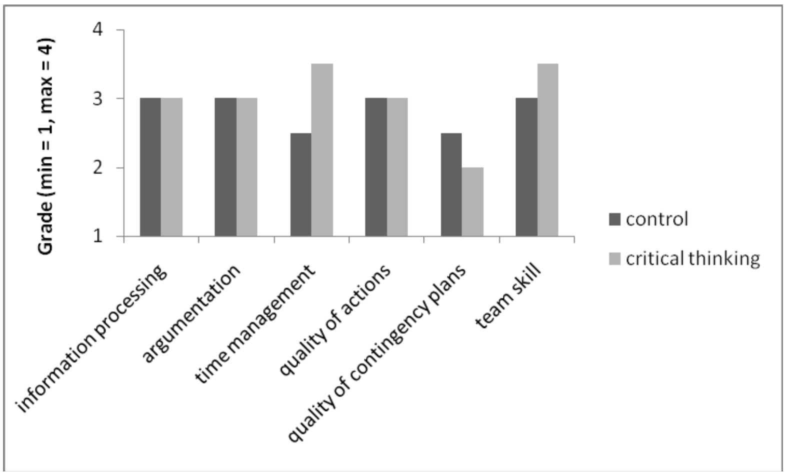

Figure 3.5: Median performance grades of participants of the critical thinking group and the control group on the near transfer test

Near Transfer Test

Figure 3.5 presents the median grades for information processing, argumentation, time management, quality of actions and contingency plans, and team skills for both conditions in the near transfer test. None of the performance measures were affected by critical thinking instruction, as was shown by the results from the Wilcoxon signed ranks test: information processing (Z = -0.33, p=0.739), argumentation (Z = -0.82 , p=.41), time management(Z =-.68 , p=.49), quality of actions (Z = -.65, p= .52), quality of contingency plans (Z = -.81, p=.416), and team skills (Z = -1.55, p = .12).

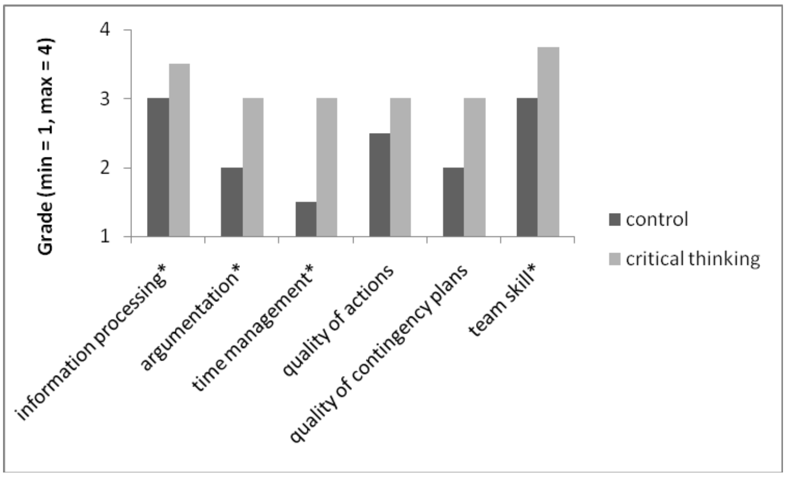

Figure 3.6: Median performance grades of participants of the critical thinking group and the control group on the far transfer test.

Far Transfer Test

The median grades on information processing, argumentation, time management, quality of actions, quality of contingency plans, and team skills for both conditions on the far transfer test are presented in Figure 3.6. Critical thinking instruction had an effect on information processing, Participants from the critical thinking group received significantly higher grades on information processing (Z = -2.33, p = .02), argumentation (Z = -2.06, p = .04), time management (Z = -2.27, p =.02) and team skills (Z = -2.16, p = .03), but there was no difference between the groups for quality of actions (Z = -1.84, p = .07) and quality of contingency plans (Z = -1.61, p = .11). The grades for information processing were *Mdn* =3.5 *range* =1 and *Mdn* =3, *range* =2; for argumentation *Mdn* =3, *range* =0, and *Mdn* =2, *range* =3; for time management *Mdn* =3, *range* =1, and *Mdn*=1.5, *range*=3; and for team skills *Mdn* =3.5 *range* =1 and *Mdn* =3, *range* =2, for critical thinking group and control group respectively.

Discussion

In line with our hypothesis, this study showed clear beneficial effects of critical thinking instruction on *far* transfer: Participants from the critical thinking group received higher grades on the majority of the performance measures. During the learning phase critical thinking mainly improved time management and team skills. This might be a result from asking participants to think aloud, which provided their teammate with information on their reasoning. This in turn may have facilitated the support given to each other or the efficiency of the communication.

This study investigated whether critical thinking instruction could benefit learning in dynamic, complex task environments such as a high fidelity tactical simulator. Although the dynamic and interactive nature of the simulator environment sometimes provided participants with little opportunity to reflect critically on their own task behaviour, it was possible to apply this approach by means of in-team discussion and by the introduction of pauses in the simulator scenarios. Still, we assume that it is important that participants are provided with ample opportunity to get acquainted with the critical thinking approach prior to the simulator sessions, which we achieved by starting with paper-and-pencil scenarios.

General Discussion

The two field studies reported in this paper examined the effects of critical thinking instruction on learning complex judgment and decision making in tactical command and control. With respect to participants' performance, it was hypothesized that: (1) critical thinking instruction yields better learning outcomes than training without critical thinking, that is, higher test performance, and (2) this performance benefit is greater for far transfer than for near transfer test performance.

Regarding the first hypothesis, there is a trend in both studies for participants who received critical thinking instruction to perform better on the test than participants who received no critical thinking instruction before and during the exercises, but this did not always reach significance. With regard to the second hypothesis, the results from Study 2 show that participants from the critical thinking group performed better than the control group on the majority of the performance measures in the *far* transfer test, but not on the near transfer test.

Our findings are consistent with those of studies on the effects of critical thinking instruction on command and control decision making (Freeman & Cohen, 1996), nursing (Forneris & Peden-McAlpine, 2006), medical diagnosis (Klein, 1998), and management (Yeo, 2007). We have shown that critical thinking instruction can also be successfully applied in learning complex judgment skills in interactive and dynamic high fidelity simulators and renders better results than merely practicing complex judgment in these simulator environments.

The performance gains can probably be partly ascribed to participants being able to apply a better decision strategy as a result of the critical thinking instruction. As Halpern (1997) has shown, when students are taught critical thinking skills, they are less likely to make the typical judgment mistakes caused, for example, by confirmation bias or failures to reason adequately about probabilities. Thus, they will show better overall performance on judgment and decision making. However, the fact that we found critical thinking instruction to have a greater effect on far transfer test performance than on near transfer test performance may be an indication that apart from having acquired a better decision making strategy, participants have also gained more general and abstract schemas and thus a deeper understanding of the task content, which enables them to solve new decision problems that differed from the learning problems on both a superficial and a structural level. An interesting question for future research remains what aspects of critical thinking instruction affect the deeper level of understanding that was observed, or whether it is the entire instruction. In the Introduction, we already mentioned possible candidates, namely self-explanation and reflection, but on the basis of our results we cannot pinpoint exactly which process is responsible for the improved level of comprehension after critical thinking instruction. It might be interesting to provide participants with only reflection or self-explanation prompts during a learning task, to disentangle possible separate effects on learning and transfer.

In combination with previous studies, our field studies seem to warrant the implementation of critical thinking instruction in education and training programs. However, implementation studies with larger numbers of participants are needed to provide answers to important questions, concerning how critical thinking instruction can be integrated into an existing curriculum, what instruction and practice observers/trainees need for successful application, and what the long term effects are of critical thinking instruction on judgment and decision making performance.

References

Anderson, J. R. (1977). Memory for information about individuals. *Memory and Cognition, 5*, 430-442.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Bartlett, F. C. (1932). *Remembering.* Cambridge, UK: Cambridge University Press.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*, 564-577.

Biggs, J., (1985). The role of metalearning in study processes. *British Journal of Educational Psychology,* 55. 185-212.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, *13,* 145-182.

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Cohen, M. S., & Freeman, J. T. (1998). Improving critical thinking. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: Emerging themes and applications* (pp. 161-169). Brookfield, VT: Ashgate Publishing Co.

Cohen M. S., Freeman, J. T., & Thompson, B. T. (1997). Training the naturalistic decision maker. In C. Zsambok & G.A.Klein (Eds.), *Naturalistic Decision Making* (pp. 257-268). Mahwah, NJ: Erlbaum.

Cohen, M. S., Freeman, J. T., & Thompson, B. T. (1998). Critical thinking skills in tactical decision making: a model and a training strategy. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 155-190). Washington, DC: APA.

Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Meta-recognition in time stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*, 206-219.

Czerwinsky, T. J. (1996). Command and control at the crossroads. *Parameters*, *26*, 121-132.

Detterman, D. K., & Sternberg, R. J. (Eds.) (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex Publishing.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.

Endsley, M. R., Hoffman, R., Kaber, D. B., & Roth, E. (2007). Cognitive engineering and decision making: An overview and future course. *Journal of Cognitive Engineering and Decision Making, 1,* 1-21.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated and reflective. *Instructional Science, 24,* 1-24.

Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying physics texts: Differences in study processes between good and poor performers. *Cognition and Instruction, 7*, 41-54.

Forneris, S. G., & Peden-McAlpine, C. J. (2006). Contextual learning: A reflective learning intervention for nursing education. *International Journal of Nursing Education Scholarship, 3, 1-18.*

Freeman, J. T., & Cohen, M. S. (1996). Training for complex decision-making: A test of instruction based on the recognition / metacognition model. *Proceedings of the 3rd*

International Command and Control Research and Technology Symposium 1996. Retrieved May 12, 2005 from www.cog-tech.com/Publications/TrainingPubs.htm.

Graesser, A. C., Gordon, S. E., & Sawyer, J. D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *Journal of Verbal Learning and Verbal Behavior, 18*, 319–332.

Halpern, D. F. (1997). Students need 'reality checks'. *Education Digest, 63,* 24-26.

Hamm, R. M. (1998). Characterizing individual strategies illuminates non optimal behavior. *Psycoloquy, 9*(49). Retrieved February 9, 2008 from http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?9.49

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*, 332-340.

Iran-Nejad, A. (2000). *Brain, knowledge, and self-regulation*. New York: Institute Mind and Behavior.

Johnson, J., & Raab, M. (2003). Take the first: Option generation and resulting choices. *Organizational Behavior and Human Decision Processes, 91,* 215-229.

Klein, G. A. (1998). *Sources of power: How people make decisions.* Cambridge, MA: MIT Press.

Klein, G. A., McClosky, M., Pliske, R., & Schmitt, J. (1997). Decision skills in training. In *Proceedings of the Human Factors Society, 41th Annual Meeting, Albaquerque, AL* (pp., 182-185). Santa Monica, CA: HFES.

Klein, G. A., Moon, B. A., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*, 70-73.

Lee, A. Y., & Hutchison, L. (1998). Improving learning from examples trough reflection. *Journal of Experimental Psychology: Applied, 4*, 187-210.

Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 103-137). Norwood, NJ: Ablex.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211-280). New York: McGraw Hill.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Olsen, S. E., & Rasmussen, J. (1989). The reflective expert and the pre-novice: Notes on skill-, rule-, and knowledge-based performance in the setting of instruction and training. In L. Bainbridge & S. A. Ruiz-Quintanilla (Eds.), *Developing skills with information technology* (pp. 9-33). New York: John Wiley & Sons Ltd.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*, 144-161.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1-29.

Roscoe, S. N., & Williges, B. H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), *Aviation Psychology* (pp. 182-193). Ames, IA: Iowa State University Press.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 185-210). New York: Academic Press.

Russell, A. (2002). The role of epistemic agency and knowledge building discourse to foster interprofessional practice in a Canadian hospital. In B. Hall (Ed.), *The value and validity of knowledge building pedagogy and technology to foster professional development: Symposium conducted at the American Educational Research Association Annual Conference, New Orleans*. Washington, DC: American Educational Research Assaciation.

Schön, D. (1983). *The reflective practitioner*. New York: Basic Books.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-298). Washington DC: APA.

Van Berlo, M. P. W., & Schraagen, J. M. C. (2000). A generic assessment tool for evaluating C2 exercises. In *Proceedings Interservice / Industry Training, Simulation and Education Conference (I/ITSEC) 2000*, Orlando FL (pp. 652-660) [CD ROM].

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., & Paas, F. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior, 6*, 273-289.

Yeo, R. K. (2007). (Re)viewing problem-based learning: An exploratory study on the perceptions of its applicability to the workplace. *Journal of Managerial Psychology, 22, 369-391*.

Zsambok, C. E. & Klein, G. A. (1997). Naturalistic Decision Making: Where are we now?, In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 3-16). Mahwah, NJ: Lawrence Erlbaum Associates.

Chapter 4
The Effects of Practice Schedule on Learning a Complex Judgment Task

This chapter is based on Helsdingen, A.S., Van Gog, T., & Van Merriënboer, J. J. G. (2008). The effects of practice schedule on learning a complex judgment task. Manuscript submitted to *Learning and Instruction*.

Abstract

The effects of contextual interference on learning a complex judgment task are investigated. In Experiment 1, participants' judgment accuracy on a retention test was higher after a random practice schedule (high contextual interference) than after a blocked or operational practice schedule. Experiment 2 demonstrated that judgment on a transfer test was also better after a random practice schedule than after a blocked schedule. Both experiments failed to show any effects of contextual interference on performance *during* learning. These findings show that the benefits of contextual interference for retention and transfer, which have been reported in the literature for perceptual-motor tasks, also apply to learning complex judgment skill, and moreover, may be achieved without detrimental effects on performance during practice.

Most educational programs aim to achieve two major goals: Adequate *post* training performance (retention) and transfer to related tasks and situations. However, very often, those goals are confused with enhancing performance and speed of skill acquisition *during* training. Research has shown that the opposite is often true: Interventions that enhance performance during training may have detrimental effects on retention and transfer performance, and conversely, instructional manipulations that degrade performance during skill acquisition may support the long term goals of training (for an overview, see Schmidt & Bjork, 1992).

An example of the latter is increasing contextual interference, that is, interference between learning tasks, through a practice schedule where different variations of the learning tasks are sequenced randomly as opposed to sequenced in separate blocks (Shea & Morgan, 1979). Increasing contextual interference may degrade performance during the learning phase but lead to better post training performance and transfer (Lee & Simon, 2003). This technique has mainly been studied in motor tasks (Brady, 1998; Cross, Schmitt, & Grafton, 2007; Guadagnoli & Lee, 2004; Lee & Magill, 1983; Magill & Hall, 1990; Shea & Morgan, 1979; Simon, 2007). However similar findings have been obtained with for example procedural tasks (Carlson, 1989; Carlson & Schneider, 1989; Carlson, Sullivan, & Schneider, 1989; Carlson & Yaure, 1990), cognitive operational tasks, such as interacting with automatic teller machines (Jamieson & Rogers, 2000), language learning (Jacoby, 1978), foreign vocabulary learning (Schneider, Healy, & Bourne, 1998, 2002), learning logical rules (Schneider, Healy, Ericsson, & Bourne, 1995), learning problem solving from worked examples (Paas & Van Merriënboer, 1994), and troubleshooting tasks (De Croock, Van Merriënboer, & Paas, 1998).

Only few studies have been conducted in the past to investigate the effects of task interference on learning complex judgment and decision making, in which the goal is to learn the complex relationships between several phenomena and predict the value of a distal variable (e.g., clinical diagnosis, weather forecast, threat assessment; Brehmer, 1973, 1977, 1979). Moreover, these studies measured performance during training, not on retention or transfer tests. Since training principles that are effective with relatively simple tasks are not necessarily effective for complex tasks as well (Wulf & Shea, 2002), research is required to establish whether contextual interference is effective for retention and transfer in learning this type of complex judgment task. It is important to establish the most optimal training sequence for such tasks, because of the far reaching consequences that for example wrong clinical diagnoses or military judgments may have (Hogarth, 1980). Therefore, the present study explores the effects of contextual interference on learning and transfer of complex judgment tasks.

Complex Judgment Tasks

In the numerous choices or judgments people make every day, two distinct classes can be identified: Value judgments, which express their preferences, and predictions, which reflect what they expect to happen (Hogarth, 1980). Value judgments encompass for example a choice for one house over another, or one pair of shoes over another. Predictions concern future outcomes, such as for example expectations regarding how someone might react to what you say or do, or who will win the next presidential elections. In this study we focus on predictive judgment tasks.

Making predictive judgments usually involves combining different information sources and considering many factors. For example, a judgment concerning the value of a stock portfolio next year is the result of a consideration of factors such as the current value of the stock, the results of the companies that have been invested in, industry trends, interest rates, political climate, et cetera. The accuracy of a prediction depends on the extent to which all relevant factors, or cues, and their relative importance for the judgment, have been correctly identified. Learning to make accurate predictive judgments is difficult, because often we cannot know all factors involved, causing the environment in which these judgments have to be made to appear as probabilistic (Brunswik, 1943).

A well known research method for studying how people learn these complex judgment tasks is the multiple cue probability learning experiment (MCPL; Brehmer, 1980; Hammond, Stewart, Brehmer, & Steinman, 1975; Smedslund, 1955), also known as multidimensional functional learning (MFL; Hoffman, Earle, & Slovic, 1981). The basic elements of MFL learning tasks are: A criterion (something a learner must learn to predict), cues (information a learner is given from which to make a prediction), and feedback. Learning the relationships between one or more cues and the criterion requires that the learner compares different situations, or different tasks, with respect to the cue values and the criterion.

Contextual Interference

Contextual interference may be manipulated by the scheduling of learning tasks. Blocked task sequences, that is, sequences of learning tasks organised in blocks, with only one variation of a task being practised in each block (e.g., AAA-BBB-CCC), have low contextual interference. These have often been found to lead to better performance during training than random practice schedules (i.e., high contextual interference; e.g., A-B-C-B-C-A-A-C-B; see e.g., Schneider, Healy, & Bourne, 2002). However, random practice schedules often result in better retention and transfer of skills to related tasks and situations (Greeno, 1964; Healy et al., 2002). As mentioned above, these beneficial effects of high contextual interference have been observed in a variety of domains and tasks.

To explain the benefits of contextual interference for retention and transfer, several hypotheses have been formulated, such as the elaboration hypothesis (Shea & Morgan, 1979), the reconstruction hypothesis (Lee & Magill, 1983), and the retrieval hypothesis (Schmidt & Bjork, 1992). The major line of argument of these hypotheses is that in a random practice schedule, learners are challenged to compare the different procedures associated with the different tasks, whereas in a blocked schedule, only one task procedure has to be kept in mind during a block of tasks, forsaking the need for extra processing activities such as elaboration, reconstruction, or retrieval of procedures from long-term memory. These extra processing activities lead to abstraction, that is, richer mental representations and more general knowledge about principles and procedures. In MFL, high contextual interference conditions may encourage learners to abstract cue-criterion relations from the learning tasks, whereas in low contextual interference conditions learners may rely on memory of specific cue-criterion observations from prior learning tasks, without attempting to abstract the underlying relationships between cues and criterion. Therefore, it is expected that although low contextual interference conditions may enhance performance during the learning phase of an MFL task, high contextual interference conditions will eventually lead to better retention and transfer.

Operational Practice Schedules

Studying the effects of different practice schedules on learning and transfer of complex judgment and decision making skills may also provide insight into the effectiveness of the *train as you fight* paradigm that is being widely applied in decision making training programs (e.g., for military command and control, crisis management, and general leadership and management). In this training approach, the real world sequence and frequency of events serves as a basis for the scheduling of practice events (i.e., operational practice schedule). Within the research community, it is often considered an ineffective training methodology for several reasons (Farmer, Van Rooij, Riemersma, Jorna, & Moraal,1999). One of the objections against *train as you fight* is that it may not provide the opportunity to practice rare or unusual tasks, which may yet be critical to effectively deal with emergency situations. However, if the real world sequence of events increases contextual interference because it is in random order, it might lead to adequate post training performance. Little research has addressed the effects of the *train as you fight* or operational practice schedules, especially in comparison with random or blocked schedules (e.g., Beaubien, Palev, Shadrick, Ennis & Jacklin, 2006; Lussier, Shadrick & Prevou, 2003).

The experiments presented here investigate the effects of different practice schedules on learning, retention (Experiment 1 and 2), and transfer (Experiment 2) of complex judgment skills. It is hypothesized that a blocked practice schedule yields better performance *during* the learning phase, whereas regarding learning *outcomes*, measured in terms of retention and transfer test performance, random (Experiment 1 and 2) and operational (Experiment 1) practice schedules yield better performance than a blocked practice schedule.

Experiment 1

This experiment investigates the effects of random, blocked, and operational practice schedules on performance during the learning phase and on retention tasks.

Method

Participants and Design

Fifty-four individuals volunteered to participate in this experiment (21 male, 33 female; age *M* = 21.0 years, *SD* = 2.9), with no prior knowledge or experience concerning the experimental tasks. Participants received 32 Euros for their participation in the experiment and could gain an additional bonus of between 0 and 12 Euros, based on their level of performance during the experiment. Participants were randomly assigned to the blocked practice schedule (*n* = 18), random practice schedule (*n* = 18), and operational practice schedule (*n* = 18).

Materials

Learning tasks. Three sets of 6 cases were developed. One dealt with injury cases, one with damage cases, and one with traffic cases. Participants were presented with one set of cases, and had to prioritize each case on the urgency for the police to deal with it. Case type was balanced over conditions, that is, in each condition 6 participants worked on injury cases, 6 participants worked on damage cases, and 6 participants worked on traffic cases. Priorities depended on the values of two different cues. One cue had three possible values, the other cue had two different values, yielding six combinations of cue values. For injury cases these cues were: (1) condition of victim (light injury, heavy injury, dead) and (2) weapon (firearm, no firearm). For damage cases the cues were: (1) level of damage (light, heavy, irretrievable) and (2) nature of the crime (burglary, arson). For traffic cases the cues were: (1) nature of the offence (speeding, driving without insurance, drunk driving) and (2) history (first time, recidivist). Tables 4.1, 4. 2, and 4.3 present the priority scores for each combination of cue values.

Table 4.1

Cue Values, Priority Scores, and Presentation Frequency in the Operational Practice Schedule of the Injury Cases

Condition of victim	Weapon	Priority score	Presentation frequency in operational practice schedule
Light injury	No Firearm	18	45%
	Firearm	32	5%
Heavy injury	No firearm	48	25%
	Firearm	62	10%
Dead	No firearm	78	3%
	Firearm	100	12%

The cases were presented one by one on a computer screen (see Figure 4.1). Each participant received 96 learning tasks (i.e. each of the 6 cases was presented 16 times) . The first 6 tasks were introduced by a short description of the crime, the subsequent 90 tasks were presented as a set of cue values only. On the lower half of the screen, beneath the presentation of the case, a slide bar was presented covering the whole range of priority scores (1-100). Using the computer mouse, participants could manipulate an indicator on the slide bar to mark the priority of the case.

Table 4.2

Cue Values, Priority Scores, and Presentation Frequency in the Operational Practice Schedule of the Damage Cases

Nature of crime	Level of damage	Priority score	Presentation frequency in operational practice schedule
Burglary	Light	19	18%
	Heavy	25	31%
	Irretrievable	31	21%
Arson	Light	49	3%
	Heavy	55	12%
	Irretrievable	61	15%

Table 4.3

Cue Values, Priority Scores, and Presentation Frequency in the Operational Practice Schedule of the Traffic Cases

Nature of offense	History	Priority score	Presentation frequency in operational practice schedule
Speeding	First time	4	30%
	Recidivist	20	45%
Uninsured	First time	24	7%
	Recidivist	40	3%
Drunk driving	First time	44	9%
	Recidivist	60	6%

Feedback during the training session consisted of a second slide bar with the indicator at the position of the true priority score. This feedback slide bar was presented above the first one, immediately after participants had indicated their priority score. The bonus that participants had earned ($bonus = \sum_{i=1}^{96} \left(1 - \left(\frac{|true\ priority_i - estimate_i|}{true\ priority_i} \right) \right) \times .125$) was continuously visible on the screen.

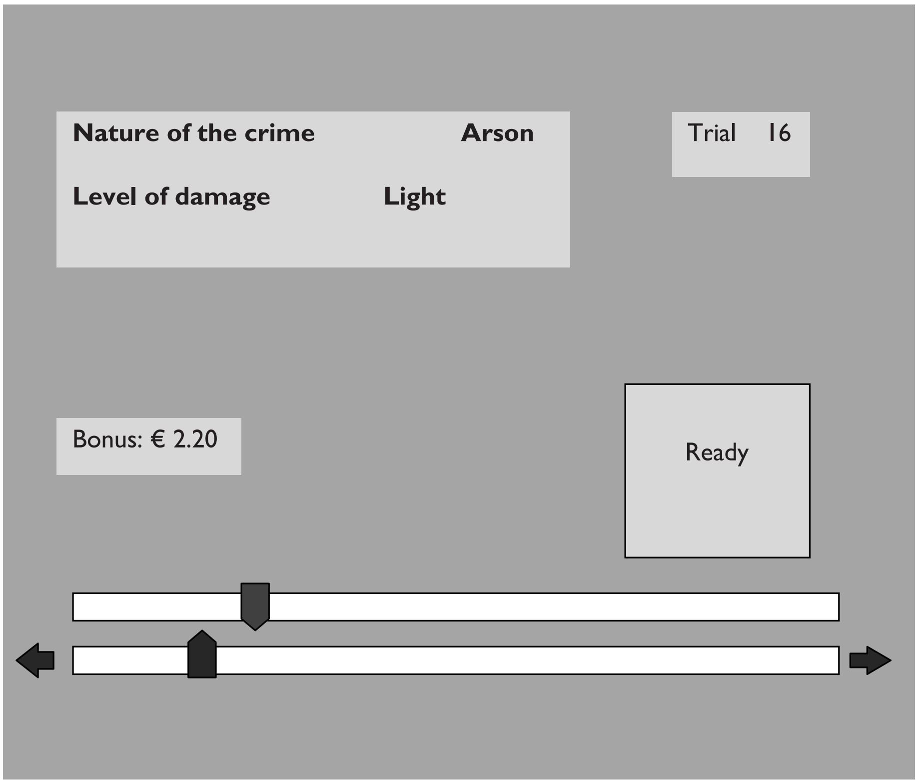

Figure 4.1: Presentation of a case, including feedback on its true priority, on the participant's computer screen in Experiment 1 (translation, original in Dutch).

Practice schedules. In the random practice condition, the cases were presented in a fully randomized order (without replacement) and each combination of cue values was presented an equal number of times. Thus, during the 96 learning tasks, each of the six combinations of cue values (i.e. each case) was presented 16 times. In the blocked practice condition, each combination of cues also appeared 16 times in total. The learning tasks in the first block of the learning phase (12 tasks) were sequenced in such a way that only one cue (the cue with 3 values) changed value from one task to the next, and the value of the second cue was kept constant. In the second block (12 tasks), the second cue had changed value, and again only the first cue changed value from one task to the next. In the third block (8 tasks) the first cue was kept constant and the second cue (with two values) changed value from one task to the next. In the fourth block (8 tasks), the value of the first block had changed, and again a sequence of tasks was presented where only the second cue changed value from one task to the next. In the fifth block (8 tasks), the value of the first cue changed again, and the second cue changed value from one task to the next. In the sixth block (24 tasks) and seventh block (24 tasks) two cues simultaneously changed their values from one task to the next. The operational task sequence was realized by a random order of tasks in which some tasks were more likely to be presented to participants than others (i.e., the higher the occurrence in real-life situations, the more likely the case was to be presented). See Tables 4.1, 4.2 and 4.3, for the probability of selecting each case in the operational sequence.

Test tasks. The retention test tasks consisted of a random selection of 24 tasks from the learning tasks, presented in random order that was the same for each participant. Each combination of cue values was presented approximately four times. No feedback was given during the test.

Procedure

The experiment had a duration of approximately 1.5 hours and was run in sessions with at most six participants. Before the experiment started, participants read a short instruction explaining how to rate the priority of the cases by moving the indicator on the slide bar. Then, the experimenter assigned each participant to a computer and started the learning phase. After participants had practiced the 96 learning tasks, they had a short break of approximately 10 minutes after which they completed the 24 tasks of the retention test. For both the learning phase and the test phase, participants could take as much time as they needed.

For each task in the learning phase and the test phase, participants' deviation score, defined as the absolute difference between the estimated priority and the true priority of the task, was automatically stored in a logfile. For each participant, time-on-task for the total session (learning phase and test together) was logged.

Results and Discussion

Table 4.4 presents the judgment performance (i.e., deviation) scores in the learning phase and the test phase per condition. Time-on-task did not differ between conditions, $F(2, 51) < 1$, $p > .20$, and type of cases (injury, damage or traffic) did not affect the deviation scores, $F(2, 51) < 1$, $p > .20$. In the analyses reported here, a significance level of .05 is set, and partial eta-squared is reported as a measure of effect size.

An ANOVA on the mean deviation scores during the learning phase showed that practice schedule did not have a significant effect on performance during the learning phase, $F(2, 51) = 0.94$, *ns*. That is, contrary to our expectations, participants in the blocked condition did not perform better than participants in the random or operational conditions during the learning phase.

Table 4.4

Means and Standard Deviations of the Deviation Scores for All Conditions in Experiment 1

	Blocked practice schedule		Random practice schedule		Operational practice schedule	
	M	*SD*	*M*	*SD*	*M*	*SD*
Learning phase	7.25	3.88	6.15	2.17	6.01	2.61
Retention test	5.48	2.81	3.12	1.84	4.58	2.41

Regarding the mean deviation scores during the test phase, an ANOVA did show a significant effect of practice schedule, $F(2, 51) = 4.46$, $MSE = 5.71$, $p = 0.02$, $\eta_p^2 = 0.15$. Post hoc Tukey tests indicated that in line with our expectations, participants who followed a random practice schedule performed better (i.e., lower deviation scores) on the retention test than participants who followed a blocked practice schedule ($|M_{diff}| = 2.36$, $p = 0.01$). The other differences between conditions were not statistically significant.

Concluding, the expectation that a blocked practice schedule generates better performance during learning than a random or operational practice schedule was not confirmed by our results. This may be due to the presentation of a large amount of learning tasks, such that participants in both the blocked and the random condition reached similar performance levels at some point during the learning phase, masking possible initial positive effects of a blocked practice schedule. There are some other studies that also failed to find effects of practice schedule on performance during learning: Whereas some failed to find effects on test performance as well (e.g., French, Rink, & Werner, 1990; Jones & French, 2007), others did show effects on test performance, in line with our results (Immink & Wright, 1998; Ollis, Button, & Fairweather, 2004; Wrisberg & Liu, 1991). That is, we did find the expected benefits for participants in the random condition over the blocked condition on the retention test. However, the operational condition did not outperform the blocked practice group. When practice tasks were presented in an operational order, that is, randomly but with some cases being presented more frequently than others, the benefits of a random practice schedule for retention performance disappeared. Because some cases appear more often than others in an operational schedule, not all cases that were presented in the test had been trained to the same extent, which might explain why the operational schedule did not outperform the blocked schedule as expected. So even though operational schedules are advocated by some practitioners as a means of providing more realistic training scenarios while increasing contextual interference, and as a consequence post training performance, this experiment shows that a 'normal' random schedule is to be preferred in terms of reaching good retention performance.

In this experiment, only effects on retention were measured. However, the aim of many education or training programs is to attain transfer of learning, that is, the adequate application of skills or knowledge acquired under specific conditions or with specific tasks in different tasks or conditions (Mayer & Wittrock, 1996; Roscoe & Williges, 1980). It has been argued that transfer is higher following a random practice schedule than following a blocked schedule, because in a random schedule participants may compare different tasks continuously, which may not only lead to a deeper understanding of the relationships between the different cues and the criterion, but also to abstract knowledge of how to approach this type of task (Wulf & Shea, 2002). Therefore, Experiment 2 investigates the effects of a random and blocked practice schedule on transfer. Because an operational practice schedule did not generate any performance difference during learning and on the test in the first experiment, this condition is excluded from Experiment 2.

Experiment 2

This experiment studies the effects of random and blocked practice schedules on performance during learning and on transfer of complex judgment skill.

Method

Participants and Design

Sixty-four participants (32 male, 32 female; age *M* = 22.0 years, *SD* = 4.4) were randomly assigned to two groups that received either random practice or blocked practice. Their prior level of schooling ranged from senior vocational education to university masters level. All participants received 32 Euros for their participation in the experiment and could gain an additional bonus of between 0 and 12 Euros, based on their level of performance during the learning phase.

Materials

Learning tasks. Learning tasks consisted of 32 descriptions of crimes. Each crime had to be prioritized on the urgency for the police to deal with it. Priorities could be determined on the basis of the dichotomous values on four different cues that occurred in each crime description: (1) the condition of the victim (injured, dead); (2) the use of a weapon (no firearm, firearm); (3) the nature of the crime (burglary, violence/holdup), and (4) available information concerning the perpetrator (description, known to the police). Table 4.5 presents the priority scores for each combination of cue values. Two tasks were developed for each of the 16 combinations of cue values, resulting in 32 tasks.

Table 4.5

The Effect of Cue Values on the Priority Score in Experiment 2

Condition of victim	Use of weapon	Nature of crime	Information on perpetrator	Priority score
Injury (10)	No Firearm (0)	Burglary (0)	Description (0)	10
			Known to police (8)	18
		Violence/holdup (12)	Description (0)	22
			Known to police (8)	30
	Firearm (4)	Burglary (0)	Description (0)	14
			Known to police (8)	22
		Violence/holdup (12)	Description (0)	26
			Known to police (8)	34
Death (43)	No Firearm (0)	Burglary (0)	Description (0)	43
			Known to police (19)	62
		Violence/holdup (29)	Description (0)	72
			Known to police (19)	91
	Firearm (9)	Burglary (0)	Description (0)	52
			Known to police (19)	71
		Violence/holdup (29)	Description (0)	81
			Known to police (19)	100

The crime descriptions were presented to the participants one by one on a computer screen (see Figure 4.2). The task was presented on the upper half of the screen. Below that, the four cues and possible cue values with tick boxes as well as a blank space for participants' estimate of the priority of the crime (to be given as a numerical value between 1 and 100) were presented. The correct cue values had to be ticked before entering priority scores. Feedback in terms of the true priority of a crime was presented on the screen after completion of each crime. The bonus that participants had earned ($bonus = \sum_{i=1}^{36}\left(1-\left(\frac{|true\ priority_i - estimate_i|}{true\ priority_i}\right)\right)\times .375$) was continuously visible on the screen.

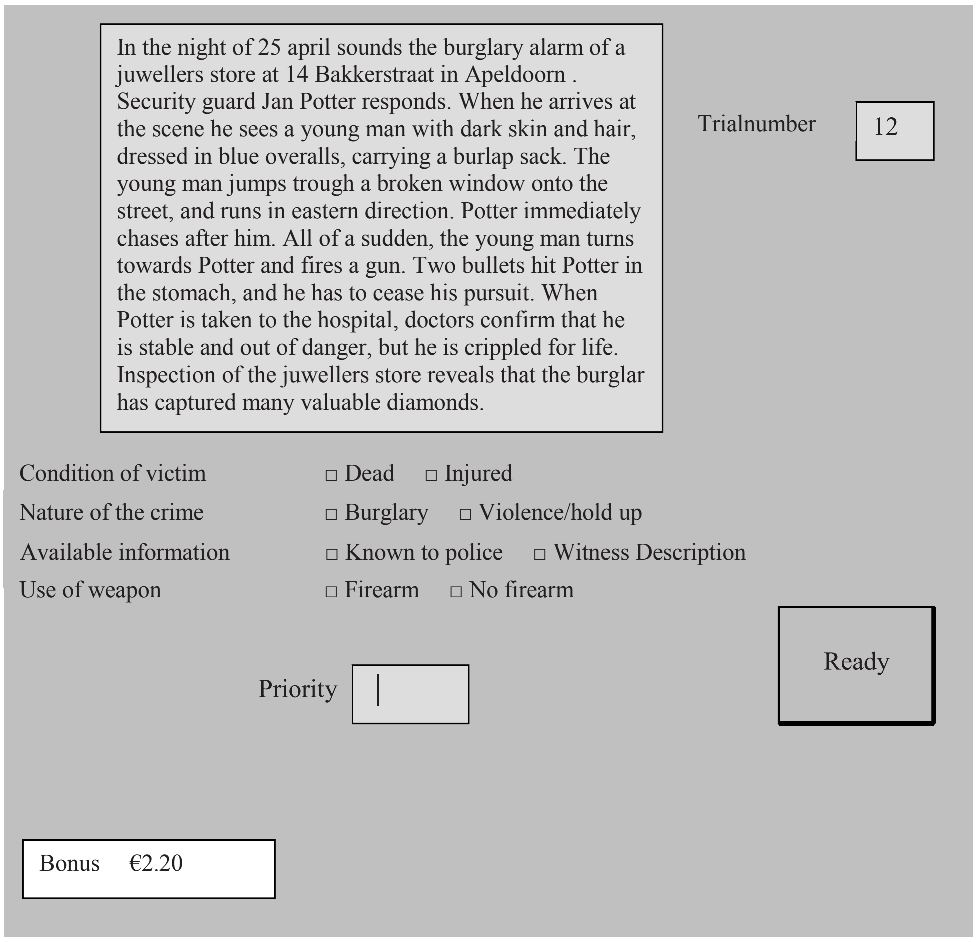

Figure 4.2:Presentation of a case on the participant's computer screen in Experiment 2 (translation, original in Dutch).

Practice schedules. In the random practice group, the task order was determined by random selection without replacement from the 32 available tasks. In the blocked practice group, the tasks in the first block of 8 tasks focused the participants' attention on the most influential cue, namely, the condition of the victim. The tasks in the second block (tasks 9-16) simultaneously varied the values of the condition of the victim and one other cue. The tasks in the third block (tasks 17-24) simultaneously varied the values of the condition of the victim and of the two other cues not yet used in the second block. The tasks in the fourth and final block (tasks 25-32) simultaneously varied the values of the condition of the victim and the three other cues.

Transfer test. The test tasks consisted of eight transfer tasks, that is, tasks that were similar to the training tasks in structural features (same combination of cues) but different with regard to surface features (cover stories). The test tasks were presented in a random order that was the same for all participants. No feedback was given during the test.

Procedure

The experiment lasted about one hour, with at most four participants per session. Before the experiment started, participants read an instruction explaining how to rate the priority of the tasks, and two exemplary task descriptions were discussed. Then, the experimenter assigned each participant to a computer and started the learning phase. The test was conducted a few minutes after the last task in the learning phase was finished. For each task, participants' deviation scores, defined as the absolute difference between the estimated priority and the true priority of the task, were automatically stored in a log file. Time-on-task for the total session (learning phase and test) was logged for each participant.

Results and Discussion

Table 4.6 presents the mean deviation scores per condition in the learning phase and the test phase. Time-on-task did not differ between conditions, $t(62) = -.26$, $p > .20$. In the analyses reported below, a significance level of .05 is set, and Cohen's *d* is reported as a measure of effect size.

Table 4.6

Means and Standard Deviations of the Deviation Scores for the Conditions in Experiment 2

	Blocked practice schedule		Random practice schedule	
	M	*SD*	*M*	*SD*
Learning phase	15.30	5.70	16.89	6.39
Transfer test	15.09	6.92	10.78	6.82

A *t*-test for independent samples showed that as in Experiment 1, practice schedule did not have a significant effect on performance during the learning phase, $t(62) = -1.05$, *ns*. In line with our expectations, participants in the random practice schedule outperformed (i.e., lower deviation scores) participants in the blocked practice schedule on the transfer test, $t(62) = 2.51$, $p = 0.01$, $d = 0.63$.

In sum, this experiment examined the effects of random and blocked practice schedules on learning and transfer of a multidimensional functional learning task. We expected that a blocked practice schedule would generate equal or better performance than a random practice schedule during the learning phase. However, as in Experiment 1, performance during the learning phase was equal, but not better in the blocked practice schedule compared to the random practice schedule. This unexpected finding is further discussed in the general discussion. In line with our expectations, this second experiment showed that a random practice schedule not only results in better retention as was shown in Experiment 1, but also in higher transfer test performance than a blocked schedule.

General Discussion

The purpose of this study was to determine the effects of contextual interference on learning, retention and transfer of complex judgment tasks and to investigate the effectiveness of an operational practice schedule stemming from the *train as you fight paradigm* (i.e., a random schedule but with some tasks being more likely to be presented than others, based on their frequency of occurrence in reality). For this, a multidimensional functional learning experiment (Experiment 1) was conducted in which participants had to learn how to judge the priority of crimes on the basis of a set of cues. We had expected that the blocked group would outperform both random groups during training and both random groups would perform better during the retention test. Indeed, we found that the random group performed better than the blocked group on the retention test. However, performance of the operational group did not differ from performance of the other groups and no differences were found on performance during the learning phase. A second experiment comparing random and blocked schedules on learning and transfer, also failed to find an effect during the learning phase, but did show beneficial effects of a random schedule for transfer.

Regarding the finding that blocked practice did not result in better performance during the learning phase, we initially thought this might be due to the large number of tasks in Experiment 1. However, we also failed to find differences during the learning phase in Experiment 2, in which the learning phase consisted of much less tasks, so this is not the most likely explanation. As mentioned before, other studies have noted lack of performance differences during the learning phase with blocked and random practice schedules, but sometimes these studies also failed to find effects on a test (e.g., French et al., 1990; Jones & French, 2007). However, some other studies also failed to find an effect of contextual interference on performance during learning, but did find positive effects on transfer, as in this study. Such findings were reported in a study conducted by Wrisberg and Liu (1991) on comparison of contextual interference levels on learning of badminton skills. They demonstrated better retention and transfer under random practice but no difference was found between groups during acquisition. And using knot-tying skills in professional fire-fighters training, Ollis et al. (2004) found that the detriment to acquisition performance as a result of high contextual interference was not as great as previous laboratory findings would suggest. Similarly, Immink and Wright (1998) failed to find any performance detriment during acquisition of a movement task as a result of a random practice sequence. They hypothesized, in accordance with the reconstruction hypothesis by Lee and Magill (1985), that a blocked schedule benefits performance during acquisition because it obviates the need to replan movements between tasks, whereas participants in a random schedule need to engage in these time consuming replanning activities. They subsequently showed that when participants in a random schedule were given sufficient time in between learning tasks to plan the upcoming response, the acquisition benefit often apparent with blocked practice disappeared, while keeping the benefits for random practice for transfer performance. As participants could work self-paced, a similar mechanism might underlie the lack of performance difference between random and blocked schedules we found in the acquisition of the highly cognitive task of complex judgment skill. However, whether this assumption is correct remains an open question, but an interesting one for future research.

As mentioned above, the retention test results in our first study showed an improvement as result of a random practice schedule compared to a blocked schedule, and no improvement after having followed an operational practice schedule. This operational schedule, representative of the “train as you fight” paradigm, failed to show higher performance compared to a blocked schedule. Therefore, as intuitively attractive as operational schedules may seem, a random practice schedule seems preferable over an operational schedule.

In considering the practical implications of our results, we have to take into account our operationalization of a blocked practice schedule, which is somewhat different from blocked schedules as they are traditionally designed in contextual interference studies. In a blocked schedule, the sequence of learning tasks is such that only one task variation per block is practiced (e.g. block 1: AAAAAA, block 2: BBBBB, where A and B are variations of the learning task). As we stated in the Introduction, in complex judgment, the task for the judge is to predict a future outcome on the basis of a few cues. In learning to predict such future outcomes, the judge has to learn which cues are relevant (cue selection), how each cue relates to the criterion to be predicted, and whether cues are intercorrelated (i.e., cue-criterion functions). Thus, the creation of blocks for these complex judgment tasks is less straightforward than it would be for less complex tasks. The blocked schedules in our experiments were such that in the first blocks one cue changed value from one task to the next. In terms of predicting the future outcomes, this would not seem to be a blocked sequence as the resulting sequential tasks are different task variations. However, in terms of the learning task, that is, to learn cue-criterion functions, participants are provided the opportunity to learn the effects of one cue-criterion function before moving on to learning the next. And in the last blocks, two (or more) cues changed value from one task to the next, providing the opportunity to learn the interaction between the cues. As such, this is a blocked practice schedule, and moreover, one that resembles real world training approaches for complex judgment and decision making. For example, sonar image operators, in a typical training program for identifying and judging sonar contacts, first learn how a sonar image depends on ocean bottom patterns, then how water temperatures influence sonar image, and only after that, how ocean bottom pattern and water temperatures interact and how that influences the sonar image (see e.g., www.mosaichydro.com). The data presented here, however, show that such training programs better present the trainee with a random sequence of tasks, leaving it to the trainee to identify and categorize cues and cue-criterion relationships.

To conclude, this study showed that high contextual interference benefits both retention and transfer of complex judgment skill, and that these effects may occur without detrimental effects on performance during practice. This study focused on retention and near transfer test performance, i.e. the transfer test tasks were different from the learning tasks on a superficial level, but not on a deep structural level. Far transfer of complex judgment skills, when the transfer test tasks also differ from the learning tasks on a deep, structural level, was not investigated, but remains an interesting subject for future research.

References

Beaubien, J. M., Shadrick, S. D., Paley, M. J., Badugu, S., Ennis, C. W., & Jacklin, S. (2006). *Using deliberate practice to train military-civilian interagency coordination. Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), 2785,* 1-11.

Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. *Quest, 50*, 266-293.

Brehmer, B. (1973). Note on the relation between single-cue probability learning and multiple-cue probability learning. *Organizational Behavior and Human Decision Processes, 9*, 246-252.

Brehmer, B. (1977). General and specific effects of practice in using functional rules in inference tasks. *Umea Psychological Reports, 114*

Brehmer, B. (1979). Effect of practice on utilization of nonlinear rules in inference tasks. *Scandinavian Journal of Psychology, 20*, 141-149.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, *45*, 223-241.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review, 50*, 255-272.

Carlson, R. A. (1989). Processing nonlinguistic negation. *American Journal of Psychology*, *102*, 211-224.

Carlson, R. A., & Schneider, W. (1989). Acquisition context and the use of causal rules. *Memory and Cognition*, *17*, 240-248.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 517-526.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 484-496.

Cross, E. S., Schmitt, P. J., & Grafton, S. T. (2007). Neural substrates of contextual interference during motor learning support a model of active preparation. *Journal of Cognitive Neuroscience, 19*, 1854-1871.

De Croock, M. B. M., Van Merriënboer, J. J. G., & Paas, F. (1998). High vs low contextual interference in simulation based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior, 14*, 249-267.

French, K. E., Rink, J. E., & Werner, P. H. (1990) Effects of contextual interference on retention of three volleyball skills. *Perceptual and Motor Skills, 71*, 179-186.

Greeno,, J. G. (1964). Paired-associate learning with massed and distributed repetitions of items. *Journal of Experimental Psychology, 67,* 286-295..

Guadagnoli, M. A., & Lee, T. D. (2004) Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*, 212-224.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinman, D. O. (1975). Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes: Formal and mathematical approaches* (pp. 271-312). New York: Academic Press.

Healy, A. F., Buck-Gengler, C. J., Barshi, I., Parker, J. T., Schneider, V. I., Raymond, W. D., LaVoie, N. N., Bowles, A. R., Pauli, P., Fisher, J. A., & Bourne, L. E. Jr. (2002). Optimizing the durability and generalizability of knowledge and skills. In S. P. Shohov (Ed.), *Trends in cognitive psychology* (pp. 123-192). Hauppauge, NY: Nova Science Publishers.

Hoffman, P. J., Earle, T. C., & Slovic, P. (1981). Multidimensional functional learning (MFL) and some new conceptions of feedback. *Organizational Behavior and Human Performance, 27*, 75-102.

Hogarth, R. M. (1980). *Judgment and choice: The psychology of decision*. New York: John Wiley & Sons, Ltd.

Immink, M. A., & Wright, D. L. (1998). Contextual interference: A response planning account. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 51A,* 735-754.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649-667.

Jamieson, B. A., & Rogers, W. A. (2000). Age-related effects of blocked and random practice schedules on learning a new technology. *The Journals of Gerontology series B: Psychological Sciences and Social Sciences, 55*, 343-353.

Jones, L. I., & French, K. E. (2007). Effects of contextual interference on acquisition and retention of three volleyball skills. *Perceptual and Motor skills, 105*, 883-890.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 730-746.

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In D. Goodman, R. B. Wilberg, & I. M. Franks (Eds.), *Differing perspectives in motor learning, memory, and control* (pp. 3-22). Amsterdam, The Netherlands: Elsevier.

Lee, T. D., & Simon, D. A. (2004). Contextual interference. In A. M. Williams & N. J. Hodges (Eds.), *Skill acquisition in sport: research, theory and practice* (pp. 29-44). London, UK: Routledge.

Lussier, J., Shadrick, S., & Prevou, M. (2003). *Think Like A Commander prototype: Instructor's guide to adaptive thinking*. Research product 2003-02. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Magill, R. A., & Hall, K. G. (1990). A review of contextual interference effect in motor skill acquisition. *Human Movement Science, 9*, 241-289.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Simon & Schuster Macmillan.

Ollis, S., Button, C., & Fairweather, M. (2005). The influence of professional expertise and task complexity upon the potency of the contextual interference effect. *Acta Psychologica, 118*, 229-244.

Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86,* 122-133.

Roscoe, S. N., & Williges, B. H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), *Aviation Psychology* (pp. 182-193). Ames, IA: Iowa State University Press.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schneider, V. I., Healy, A. F., & Bourne, L. E. Jr. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 78-90). Mahwah, NJ: Erlbaum.

Schneider, V. I., Healy, A. F., & Bourne, L. E. Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46,* 419-440.

Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E. Jr. (1995). The effects of contextual interference on the acquisition and retention of logical rules. In A. F. Healy & L. E. Bourne Jr. (Eds), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 95-131). Thousand Oaks, CA: Sage Publications, Inc.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179-187.

Simon, D. A. (2007). Contextual interference effects with two tasks. *Perceptual and Motor Skills, 105,* 177-183.

Smedslund, J. (1955). *Multiple-probability learning: An inquiry into the origins of perception.* Oxford, UK: Akademisk.

Wrisberg, G. A., & Liu, Z. (1991). The effect of contextual variety on the practice, retention, and transfer of an applied motor skill. *Research Quarterly for Exercise and Sport, 62,* 406-412.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychological Bulletin & Review, 9*, 185-211.

Chapter 5
The Effects of Practice Schedule and Critical Thinking Instruction on Learning and Transfer of a Complex Judgment Task

This chapter is based on Helsdingen, A. S., Van Gog, T., & Van Merriënboer, J. J. G. (2008). The effects of practice schedule and critical thinking instruction on learning and transfer of a complex judgment task. Manuscript submitted to *Journal of Educational Psychology*.

Abstract

Many studies have shown beneficial effects of random over blocked practice on transfer of learning, even though blocked practice often leads to better performance during the learning phase. In a 2 x 3 factorial experiment (N = 120) with the factors practice schedule (random, blocked) and critical thinking prompts (proactive, retrospective, none), this study investigates whether this also applies to complex judgment tasks, and whether critical thinking prompts can enhance the effectiveness of those practice schedules. It is hypothesized that retrospective prompts yield best transfer in a random practice schedule, whereas proactive prompts yield best transfer in a blocked schedule. In line with our hypothesis, a blocked schedule led to better performance than random practice during learning, but not on the transfer test, where a random schedule was beneficial. The hypothesized interaction effect was also found: whereas retrospective critical thinking prompts have surplus value in a random schedule, transfer test performance following a blocked schedule is enhanced through proactive critical thinking prompts.

Within current practices of training and education, much effort is invested in measures that enhance training efficiency and performance. Unfortunately, many of these measures focus on performance gains *during* the training. This can be problematic, because those measures do not always contribute to, or may even deteriorate, the transfer of knowledge and skills from the learning tasks to the operational tasks. A technique that has been shown to lead to better transfer is contextual interference. However, this technique has mainly been studied in relatively simple problems and not in relation with other measures that could establish interference, such as prompting critical thinking. The present study investigates the effects of contextual interference and critical thinking prompts on learning and transfer of a complex judgment task.

Many studies have shown that conditions that increase contextual interference, such as a practice schedule in which several task variations are randomly presented, deteriorate performance during the learning phase but lead to better transfer when compared to conditions of low interference (Battig, 1972, 1979; Lee & Magill, 1983; Magill & Hall, 1990; Shapiro & Schmidt, 1982; Shea & Morgan, 1979). The beneficial effects of high contextual interference have been observed in the learning of motor tasks, troubleshooting tasks, procedures, and language. It has hardly been studied for learning complex judgment tasks, in which the goal is to learn the complex relationships between several phenomena (i.e., multidimensional functional learning; e.g., clinical diagnosis, weather forecast, threat assessment). However, it is important to do so, because training principles developed on the basis of learning simple tasks do not always generalize to learning complex tasks, as Wulf and Shea (2002) concluded after reviewing the literature on learning motor tasks. They found, for example, that for retention and transfer of some highly complex motor tasks, blocked practice schedules (low contextual interference) were more beneficial than random practice schedules (high contextual interference). They hypothesized that the additional processing demands imposed by conditions of high interference would compete with the essential processing required by the execution of a highly complex task. It is unclear whether this is also true for non-motor tasks, because in learning a highly complex judgment task, for example, increasing interference by prompting learners to engage in critical thinking is known to yield better transfer of judgment skills (Cohen, Freeman, & Thompson, 1998; Cohen, Freeman, & Wolf, 1996; Freeman & Cohen, 1996; Schwalbe, 2004; Helsdingen, Van den Bosch, Van Gog, & Van Merriënboer, 2008 – Chapter 3). This seems to suggest that increasing interference through random practice might also be beneficial for such tasks, and a recent experiment by Helsdingen, Van Gog, and Van Merriënboer, 2008 – Chapter 4) shows that this is indeed the case, at least with regard to retention and *near* transfer.

Effects of contextual interference on *far* transfer have not yet been studied for complex judgment tasks, and effects of critical thinking instruction have mostly been studied in real world training programs, that is, with a small number of participants and with low experimental control over the sequence of learning events and timing of instructor interventions (Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000; Helsdingen, Van den Bosch, et al., 2008 – Chapter 3). Moreover, the combined effects of critical thinking and contextual interference have not yet been investigated. Therefore, as we will argue in more detail below, it would be interesting to study whether using the techniques together is more effective than using them separately, and whether there is a differential effect of critical thinking in conditions of low or high contextual interference.

Complex Judgment Tasks

A well known paradigm for laboratory studies into learning complex judgment tasks in dynamic and uncertain conditions is multiple cue probability learning (MCPL; Brehmer, 1980; Hammond, Stewart, Brehmer, & Steinman, 1975), also known as multidimensional functional learning (MFL; Hoffman, Earle, & Slovic, 1981). The basic elements of MFL learning tasks are: A criterion (something a learner must learn to predict), cues (information a learner is given from which to make a prediction), and feedback. The learner's goal is to learn how the criterion values relate to the cue values, or, in other words, to predict a future outcome on the basis of a few indicators. Consider threat assessment as an example of such a task. When an air defence officer needs to judge the threat of an object, such as a hostile airplane, s/he has to do so based on a few cues or attributes, such as its speed, position, altitude, direction, and identity. To do this adequately, the defence officer needs to assess the value and importance of all cues separately, relative to the other cues, and in combination. For example: How important is speed/identity/position? Is speed more important than identity when an object is near one's own position? What constitutes a higher threat, an object that is fast and far away but moving towards one's own position, or an object that is slow and very close but moving away? Et cetera.

Learning the relationships between one or more cues and the criterion requires that the learner compares different situations, or different tasks, with respect to the cue values and the criterion. This may be realized by instructional interventions that call the learner's attention to different cue and criterion values of different tasks, such as a random practice schedule or critical thinking prompts.

Contextual Interference

Transfer-of-training refers to the observation that knowledge and skills learned under specific conditions or from practicing a specific task, transfer to other conditions or tasks (Mayer & Wittrock, 1996; Roscoe & Williges, 1980). Transfer of knowledge and skills from the learning tasks to the operational tasks is usually the ultimate goal of training programs. A distinction is often made between *near* and *far* transfer. Near transfer tasks share structural features but differ on superficial or surface features from the learning tasks, whereas far transfer tasks differ from the learning tasks on both surface and structural features (Quilici & Mayer, 1996). Thus, the degree of transfer depends on the degree to which task activities call upon the same skill, but also on the learner's level of understanding of a task (Mayer & Wittrock, 1996): Near transfer can be obtained if the learner knows *what* to do (same use of the same—specific—knowledge), whereas far transfer is usually possible only if the learner also knows *when* and *why* he needs to do that (other use of the same—general—knowledge; Van Merriënboer, 1997). To gain such deep understanding required for far transfer, a process of abstraction is needed, from knowledge of the task at hand to a higher level of knowledge, for instance, of the general principles of a domain or the general procedures for performing a class or even several classes of tasks (Van Merriënboer & Paas, 1990).

Contextual interference may enhance such abstraction processes (De Croock, Van Merriënboer, & Paas, 1998). To explain the benefits of contextual interference, several hypotheses have been formulated, such as the elaboration hypothesis (Shea & Morgan, 1979), the reconstruction hypothesis (Lee & Magill, 1983), and the retrieval hypothesis (Schmidt & Bjork, 1992). The major line of argument of these hypotheses is that in a random practice schedule, learners are challenged to compare the different procedures associated with the different tasks, whereas in a blocked schedule, only one task procedure has to be kept in mind, forsaking the need for extra processing activities such as elaboration, reconstruction, or retrieval of procedures from long-term memory. These extra processing activities lead to abstraction, that is, richer mental representations and more general knowledge about principles and procedures.

Empirical data lend support to all three hypothesis above, however, it is still unclear which one is most viable (Wulf & Shea, 2002). What all these '*extra processing*' hypotheses have in common, though, is that they mainly explain the benefits of random practice over blocked practice. Although this implies some consideration of the constraints of blocked practice, focussing primarily on the latter could lead to a fourth explanation for the contextual interference effect. Deteriorated learning in a blocked practice schedule compared to a random schedule might also be due to an *illusion of competence* the learner experiences as a result of the repetitive presentation of the same type of task (Koriat & Bjork, 2005). Illusions of competence may inhibit search and application of alternative task strategies and so hinder gaining a deeper level of understanding (Kornell & Bjork, 2007). As a consequence, transfer test performance, particularly on far transfer, will be hampered compared to a random sequence where illusions of competence do not occur.

As mentioned before, other instructional interventions, such as critical thinking prompts, may lead to similar effects as high contextual interference, thereby preventing illusions of competence and enhancing abstraction.

Critical Thinking Instruction

Critical thinking aims at teaching recognition based decision making and (meta)cognitive techniques that guide and critically review this decision-making process (Cohen, Freeman, & Wolf, 1996). Judgment based on recognition of a situation requires exposure to many different problems during training. Furthermore, training should foster the (meta)cognitive skills that allow decision makers to verify and improve the results of recognition-primed decision-making processes, by identifying evidence-conclusion relationships (i.e., arguments), critiquing the arguments used to support a conclusion (i.e., identifying incompleteness, conflict, or unreliability), and correcting the errors in argumentation (Cohen, Freeman, & Wolf, 1996).

Typical critical thinking instruction involves an explanation of the decision making process before the learning phase and an instructor who prompts the learner to reflect on his or her cognitive strategy and initiate one of the critical thinking steps during the learning phase (see for a more detailed discussion Helsdingen, Van den Bosch, et al., 2008 – Chapter 3). Instructors usually deliver prompts at random during the training. In the literature, positive effects are mentioned of, for example, planning prompts (Davis, 2003), driving questions

(Dochy, 1992; Douglas, Hosokawa, & Lawler, 1988; Morgan & Saxton, 1991; Naidu & Bernard, 1992; Orlich, Harder, Callahan, Kauchak, & Gibson, 1994; Rowntree, 1992), and deep level reasoning questions (Craig, Sullins, Witherspoon, & Gholson, 2006), all given *before or during* execution of a learning task. However, positive effects of reflection prompts provided *after* task execution have also been reported (e.g., Belland, Glazewski, & Richardson, 2007; Davis & Linn, 2000; Van den Boom, Paas, & Van Merriënboer, 2007; Van den Boom, Paas, van Merriënboer & van Gog, 2004).

In several studies, positive effects of critical thinking instruction on transfer of complex judgment skills have been found (Cohen et al., 1998; Freeman & Cohen, 1996; Helsdingen, Van den Bosch, et al., 2008 – Chapter 3). One of these studies (Helsdingen, Van den Bosch, et al., 2008 – Chapter 3) focused on a group of naval warfare officers learning to identify, categorize, and react to naval surface threats in a dynamic and probabilistic simulated environment. On the far transfer test, which included real-world events presented in a simulated environment, officers who received critical thinking instruction were better at selecting relevant information, identifying missing or conflicting evidence, provided better arguments for their assessments, were better at planning of their actions in time, , and displayed better team skills han officers who received the same amount of judgment training but without critical thinking instruction.

Cohen and colleagues (Cohen et al., 1998; Freeman & Cohen, 1996) attributed the benefits of critical thinking instruction to participants having acquired an appropriate decision making strategy. When students are taught critical thinking skills, they are less likely to make the typical judgment mistakes that are caused by for example confirmation bias or failures to reason adequately about probabilities (Halpern, 1997). Thus, they will show better overall performance on judgment and decision making. However, in the study by Helsdingen, Van den Bosch, et al. (2008 – Chapter 3) described above, critical thinking instruction was found to have effects primarily on *far* transfer test performance. This may be an indication that apart from having acquired a better decision making strategy, participants also gained a deeper, more abstract level of understanding of the task content that enabled them to solve decision problems that differed from the learning problems on a superficial and a structural level. Critical thinking instruction may have this effect because it might facilitate processes such as self explanation and reflection, that have been proven to enhance learning and transfer (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; Ferguson-Hessler & De Jong, 1990; Lee & Hutchinson, 1998; Renkl, 1997). Chi et al. (1989) attribute the enhanced learning and transfer as a result of self explanation during learning tasks, to learners being able to judge their learning progress adequately, not having developed an illusion of competence. Learners may develop such an illusion as a consequence of several instructional measures that aim to enhance their performance during training. An example might be the presentation of learning tasks in a blocked schedule, which provides a condition of low contextual interference, where only one task variation needs to be practiced before moving on to the next (Simon & Bjork, 2001).

However, in order to be effective in combination with different practice schedules, timing of the critical thinking prompts seems crucial. In a blocked practice schedule, critical thinking processes would need to be prompted *before* execution of the task (i.e., proactive), so that critical thinking can occur concurrently with task execution. When prompted after task execution in a blocked practice schedule, learners may not be inclined to critically reflect, since they are under the impression that they already understand the task. In a random practice schedule, the key factor in successful planning of critical thinking prompts is the learners' limited processing capacity: When learners already experience high processing demands, they will not be able to critically reflect upon their task strategies concurrently with task execution. Because in a random practice schedule learners would not be hampered by the illusion of competence, prompting critical thinking after execution of the learning task (i.e., retrospective) could have beneficial effects on learning without leading to processing capacity problems.

The Present Study

The present study examines the effects of practice schedule (random or blocked) and critical thinking prompts (none, proactive, or retrospective) on judgment performance during learning and on a far transfer test, using an MFL-task in the domain of criminal investigation. Participants' read a case description and judge its priority for the police on the basis of several cues present in each case description. Both participants' performance, that is, judgment accuracy, and the degree of certainty about the accuracy of their judgments is measured, the latter is expected to be an indicator of illusion of competence.

It is hypothesized that a blocked practice schedule generates better performance and less uncertainty *during* the learning phase than a random practice schedule. When a blocked practice schedule is combined with proactive critical thinking prompts, however, uncertainty is expected to be at the same level as in a random schedule. Regarding learning *outcomes*, measured in terms of transfer test performance, it is hypothesized that learners in a blocked practice schedule profit more from proactive prompts than from retrospective or no prompts, while in a random practice schedule, learners benefit more from retrospective prompts than from proactive or no prompts.

Method

Participants and Design

Participants were 120 students from Dutch universities. Their age varied from 20 to 25 years (M = 21.58, SD = 1.59). Participation was voluntary and was rewarded with 45 Euros. Participants were randomly assigned to one of the six conditions in a 2 x 3 factorial design with the factors Practice Schedule (Blocked or Random) and Critical Thinking Prompts (Proactive, Retrospective, or None): (1) blocked practice schedule with proactive critical thinking prompts (n = 20), (2) blocked practice schedule with retrospective critical thinking prompts (n = 20), (3) blocked practice schedule without prompts (n = 19), (4) random practice schedule with proactive critical thinking prompts (n = 20), (5) random practice schedule with retrospective critical thinking prompts (n = 21), and (6) random practice schedule without prompts (n = 20).

Materials

Learning tasks. A set of 32 case descriptions of crimes was developed for this experiment. Each case had to be prioritized on the urgency for the police to deal with it. Priorities could be determined on the basis of the dichotomous values on three different cues that occurred in each case description: (1) the condition of the victim (injured, dead), (2) the nature of the crime (burglary, violence/holdup), and (3) the use of a weapon (no firearm, firearm). Table 1 presents the priority scores for each combination of cue values. Four crime descriptions were developed for each of the eight combinations of cue values, resulting in 32 cases (see Figure 5.1 for an example of a case).

Table 5.1

Schematic Overview of the Cues, Cue Values and Subsequent Priorities

Learning and near transfer cases (crimes)			
Condition of victim	Weapon	Nature of crime	Priority
Dead	Firearm	Burglary	62
		Violence	91
	No firearm	Burglary	58
		Violence	72
Injured	Firearm	Burglary	29
		Violence	43
	No firearm	Burglary	10
		Violence	39
Far Transfer cases (traffic offenses)			
Nature of the offense	Condition of Offender	History	Priority
Speeding	Drunk	First offence	84
		Recidivist	94
	Not drunk	First offence	64
		Recidivist	74
Driving without insurance	Drunk	First offence	32
		Recidivist	40
	Not drunk	First offence	12
		Recidivist	20

The cases were presented one by one on a computer screen. On the lower half of the screen, beneath the presentation of the case, a slide bar with an indicator of priority was presented. The slide bar covered the whole range of priority scores: 1-100. There were no numbers or other reference points on or near the slide bar. An indicator on the slide bar could be manipulated with the computer mouse to mark the priority of a case. In addition, the margin of the area on the right and the left from the indicator could be adjusted (minimum = 2, maximum = 50) to indicate certainty about the accuracy of the response.

Feedback during the learning session consisted of a second slide bar with the indicator at the position of the true priority score. This feedback slide bar was presented above the first one, directly after participants indicated their priority and certainty score. If the true priority score was within the certainty margin of the participant's response, a green rectangle was presented around the feedback slide bar. If the true priority score lay outside this certainty margin, a red rectangle was presented around the feedback slide bar.

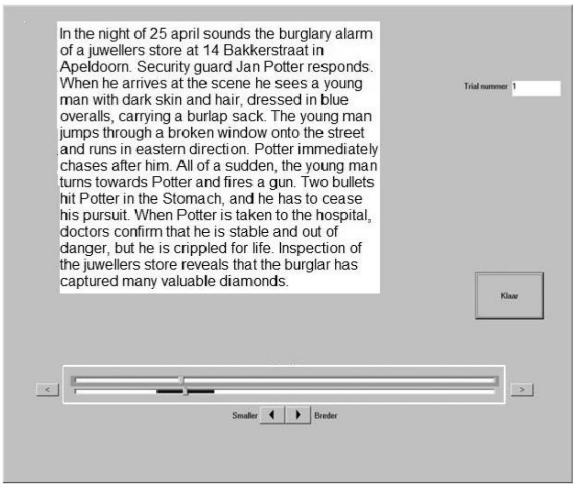

Figure 5.1: Example of the task for participants, as it was presented on the computer screen (translated, original in Dutch).

Practice schedules. In the blocked practice schedule, the order of the 32 learning tasks was the same for all participants. The 32 cases were sequenced in such a way that in the first block of 12 cases only one cue changed its value from one case to the next, and the cue that changed its value was alternated from case to case. Thus, if the first case involved a dead victim (cue 1) after a burglary (cue 2) and the use of a firearm (cue 3), the second case could be a dead victim (cue 1: unchanged) after a burglary(cue 2: unchanged) *without* the use of a firearm (cue 3: changed), and the third case could subsequently be an *injured* victim (cue 1: changed) after a burglary (cue 2: unchanged) without the use of a firearm (cue 3: unchanged). In the second block of 12 cases two cues simultaneously changed value from one case to the next. In the last block of 8 cases all three cues changed value from one case to the next. In the random practice schedule the learning tasks were presented in a randomized order that was the same for all participants in this schedule.

Critical thinking instruction and prompts. Participants in all critical thinking conditions were introduced to critical thinking with a paper-based instruction explaining four steps: (1) creating a model of the relationships between the cues and the priorities, (2) testing a model, (3) evaluating a model, and (4) finding alternative explanations (i.e., contingencies). Apart from this textual explanation, a hand-out with a graphic representation of the four steps (Figure 2) and two short stories of critical thinking in real-world situations were provided.

Both the 12 proactive and the 12 retrospective prompts were formulated as open-ended questions provided on the computer screen, which intended to focus the participant's attention to particular cues and how their priority changed (for examples, see Table 5.2). The proactive prompts preceded every two or three cases, similar retrospective prompts followed those same cases, thus, only the timing relative to the learning tasks was different for both critical thinking conditions. Participants were told that the open ended questions could help them learn the task and that they should therefore answer the questions for themselves.

Table 5.2

Critical Thinking Prompts

Proactive prompts	Reflective prompts
Are there any similarities between the following two cases? Which are they? And what is different between these cases?	Were there any similarities between the last two cases? Which were they? And what were the differences?
The following two cases will differ from the former two cases on one specific aspect. Can you tell me what that is? And does it have an effect on the priority of a case?	The last two cases were different from the two preceding cases on one specific aspect. Can you tell me what that was? And did it have an effect on the priority of a case?
Do you consider some aspects to be more important than others when estimating a case's priority?	Do you consider some aspects to be more important than others when estimating a case's priority?

Near and far transfer tests. The near transfer test consisted of 8 cases that were similar to the learning tasks in structural features (same combination of cues and same domain, i.e., crimes) but different to the learning tasks in surface features (i.e., cover stories). The near transfer cases were presented in the same random order to all participants and the sequence contained no critical thinking prompts and no feedback.

The far transfer test consisted of 32 cases describing traffic offenses that differed in both structural and surface features from the learning tasks. Priorities depended on: (1) nature of the offence (speeding, driving without insurance), (2) condition of the offender (drunk, not drunk), and (3) history (first offence, recidivist). These far transfer cases were scheduled in random order that was the same for all participants and no critical thinking prompts were given. However, feedback was provided because this new domain contained new cues to base the priority on. See Table 5.1 for an overview of the cues and priorities of the learning, near transfer, and far transfer cases.

Procedure

The experiment had a duration of approximately two hours and was run in sessions with at most six participants. Before the experiment started, participants read a short instruction explaining how to rate the priority of the cases by means of the slide bar and how to indicate the margin of certainty. All participants received paper and pencil to make notes. In addition, the critical thinking conditions received a written instruction about the four steps of the critical thinking method, the hand-out with the graphical representation of the four steps of critical thinking (cf. Figure 5.2), and two example stories of critical thinking in real world situations. Then, all participants were assigned to a computer and started to work on the learning tasks of their respective condition. The near transfer test was administered right after completion of the learning tasks, and the far transfer test immediately followed the near transfer test.

Participants could choose their own pace; the time they took to complete the whole experiment was logged. For both learning tasks and transfer test cases, participants' *deviation score*, defined as the absolute difference between the estimated priority and the true priority of the case, and the *uncertainty margin* they indicated were automatically stored in a logfile. After completion of the far transfer test open exit interviews were conducted with the participants.

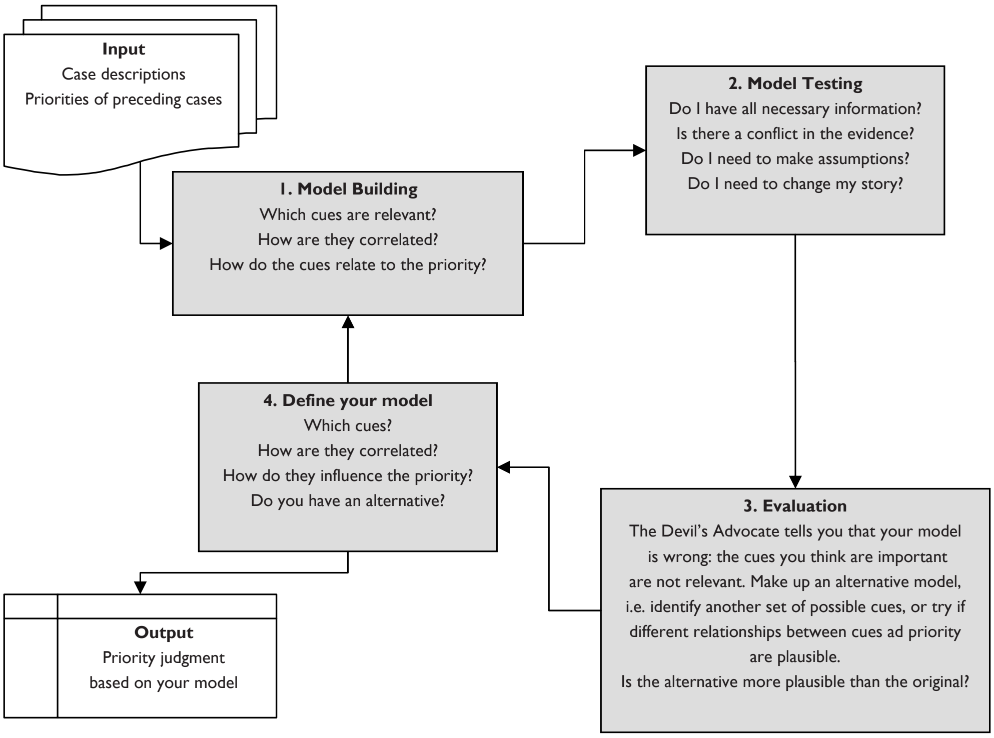

Figure 5.2: Schematic overview of the critical thinking steps (translated, original in Dutch).

Results

The mean deviation scores and mean uncertainty margins were calculated for each participant. See Tables 5.3, 5.4, and 5.5 for those data from the learning phase, near transfer test, and far transfer test, respectively. Time-on-task did not differ between conditions, $F(1, 114) = .01$, $p > .20$. In all analyses reported below, a significance level of .05 is set, and partial eta-squared is reported as a measure of effect size.

Learning Phase

Table 3 presents the means and standard deviations of the deviation scores and the uncertainty margins in the learning phase for all conditions. A 2 x 3 ANOVA on the mean deviation scores showed that practice schedule had a significant effect in the learning phase, $F(1,114) = 14.43$, $MSE = 191.21$, $p < .01$, $\eta_p^2 = .11$ In line with our hypothesis, participants who worked in a blocked practice schedule performed better (lower deviation scores) during the learning phase than participants in the random practice schedule. There was no significant effect of critical thinking prompts, $F(2, 114) = 2.23$, *ns*, but there was a trend towards an interaction effect between critical thinking prompts and practice schedule, $F(2, 114) = 2.64$, $MSE = 35.02$, $p = .08$, $\eta_p^2=.04$. Separate analyses within practice schedules showed no significant effect of critical thinking prompts on the deviation scores in the blocked practice schedule, $F(2, 56) = .05$, *ns*, but a significant effect of critical thinking prompts in the random practice schedule, $F(2, 58) = 4.34$, $MSE = 65.30$, $p = .02$, $\eta_p^2 = .13$, with post hoc Tukey tests indicating that retrospective critical thinking prompts yielded lower deviation scores than both proactive critical thinking prompts ($|M_{diff}| = 2.94$, $p = .05$) and no critical thinking prompts ($|M_{diff}| = 3.20$, $p = .03$).

Table 5.3

Means and Standard Deviations of the Deviation Scores and Uncertainty Margins for All Conditions in the Learning Phase

	Blocked practice schedule						Random practice schedule					
	Retrospective prompts		Proactive prompts		No Prompts		Retrospective prompts		Proactive prompts		No prompts	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Deviation	18.43	2.81	18.48	3.59	18.16	3.67	18.83	5.41	21.78	2.79	22.04	2.70
Uncertainty	13.32	4.83	8.79	1.48	13.28	4.96	12.71	3.00	13.15	2.96	10.48	2.76

Analyses of the uncertainty margins showed that practice schedule had no effect on the uncertainty margins, $F(1, 114) = 0.24$, *ns*. Critical thinking prompts significantly affected participants' uncertainty margins, $F(2, 114) = 3.43$, $MSE = 42.75$, $p = .04$, $\eta_p^2 = .06$, and there was a significant interaction between critical thinking prompts and practice schedule, $F(2, 114) = 10.75$, $MSE = 133.84$, $p < .01$, $\eta_p^2=.16$. Therefore, the effects of critical thinking prompts were analyzed separately for both practice schedules. In a blocked practice schedule, critical thinking prompts had a significant effect on the uncertainty margins, $F(2, 56) = 8.14$, $MSE = 134.83$, $p = .01$, $\eta_p^2 = .23$, with proactive critical thinking prompts leading to smaller uncertainty margins than both retrospective critical thinking prompts ($|M_{diff}| = 4.54$, $p < 0.01$) and no critical thinking prompts ($|M_{diff}| = 4.50$, $p < .01$). In a random practice schedule, critical thinking prompts also significantly affected the uncertainty margins, $F(2, 58) = 4.89$, $MSE = 41.43$, $p = 0.01$, $\eta_p^2 = .14$: Participants who received no critical thinking prompts had smaller uncertainty margins than participants who received either retrospective critical thinking prompts ($|M_{diff}| = 2.24$, $p = .04$) or proactive critical thinking prompts ($|M_{diff}| = 2.68$, $p = .01$).

Near Transfer Test

Table 5.4 presents the means and standard deviations of the deviation scores and uncertainty margins on the near transfer test. A 2 x 3 ANOVA on the mean deviation scores showed that critical thinking prompts had no main effect on performance on the near transfer test, $F(2, 114) = 2.59$, *ns*. There was a main effect of practice schedule, $F(1, 114) = 4.23$, $MSE = 120.39$, $p = .04$, $\eta_p^2 = .04$: Participants in the random practice schedule reached lower deviation scores on the near transfer test (i.e., better performance) than participants in the blocked practice schedule ($|M_{diff}| = 2.01$, $p = .04$). However, there was also a significant interaction between practice schedule and critical thinking prompts, $F(2, 114) = 3.26$, $MSE = 92.78$, $p = .04$, $\eta_p^2 = .05$. Analysis of the practice schedules separately showed a significant effect of critical thinking prompts in the random schedule, $F(2, 58) = 5.98$, $MSE = 153.70$, $p < .01$, $\eta_p^2 = .17$, with retrospective prompts yielding significantly better results than both proactive prompts ($|M_{diff}| = 5.23$, $p < .01$) and no prompts ($|M_{diff}| = 3.964$, $p = .05$). In the blocked practice schedule, critical thinking prompts had no effect on the deviation scores of the near transfer test, $F(2, 56) = .48$, *ns*.

Analysis of the uncertainty margins showed a main effect of practice schedule, $F(1, 114) = 4.58$, $MSE = 63.83$, $p = .03$, $\eta_p^2 = .04$, indicating that participants in the random practice schedule had smaller uncertainty margins on the near transfer test than participants in the blocked practice schedule ($|M_{diff}| = 1.53$, $p = .03$). There was also a main effect of critical thinking prompts on uncertainty margins, $F(2, 114) = 4.19$, $MSE = 69.75$, $p = .02$, $\eta_p^2 = .07$. Tukey post hoc tests showed significantly higher uncertainty margins in the condition without critical thinking prompts than in the condition with proactive critical thinking prompts ($|M_{diff}| = 2.414$, $p = .02$). However, the significant interaction between practice schedule and critical thinking prompts, $F(2, 114) = 3.94$, $MSE = 59.89$, $p = .02$, $\eta_p^2 = .07$, indicated that critical thinking prompts had a differential effect in the random and blocked practice schedule. In the random practice schedule, critical thinking prompts had no effect on the uncertainty margins, $F(2, 58) = .81$, *ns*. In the blocked practice schedule, critical thinking prompts did have a significant effect, $F(2, 56) = 6.15$, $MSE = 113.17$, $p < .01$, $\eta_p^2 = .18$, with post hoc Tukey tests indicating that proactive prompts yielded smaller uncertainty margins than both retrospective prompts ($|M_{diff}| = 4.20$, $p = .01$) and no prompts ($|M_{diff}| = 4.07$, $p = .01$).

Table 5.4

Means and Standard Deviations of the Deviation Scores and Uncertainty Margins for All Conditions on the Near Transfer Test

	Blocked practice schedule						Random practice schedule					
	Retrospective prompts		Proactive prompts		No Prompts		Retrospective prompts		Proactive prompts		No prompts	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Deviation	16.00	5.06	15.18	5.79	16.94	5.91	10.97	5.81	16.21	4.45	14.94	4.81
Uncertainty	13.26	4.70	9.06	2.80	13.13	5.07	9.64	3.71	10.21	3.36	11.03	3.55

Far Transfer Test

Table 5.5 presents the means and standard deviations of the deviation scores and uncertainty margins on the far transfer test. A 2 x 3 ANOVA on the mean deviation scores showed a main effect of practice schedule on the far transfer test, $F(1, 114) = 38.21$, $MSE = 977.51$, $p < .01$, $\eta_p^2 = .25$. Participants in the random practice schedule performed better (i.e., lower deviation scores) than participants in the blocked practice schedule. Critical thinking prompts had no main effect, $F(2, 114) = 1.84$, *ns*, however, there was a significant interaction between practice schedule and critical thinking prompts, $F(2, 114) = 8.97$, $MSE = 229.57$, $p < .01$, $\eta_p^2 = .14$. Further investigation of this interaction by analyzing random and blocked practice schedules separately, showed that critical thinking prompts significantly affected participants' far transfer score, both after a random practice schedule, $F(2, 58) = 4.13$, $MSE = 116.73$, $p = .02$, $\eta_p^2 = .12$, and a blocked practice schedule, $F(2, 56) = 6.94$, $MSE = 158.18$, $p < 0.01$, $\eta_p^2 = .20$. In line with our hypothesis, in a random practice schedule, retrospective critical thinking prompts yielded better scores on the far transfer test than both proactive prompts ($|M_{diff}| = 4.14$, $p = .04$) and no prompts ($|M_{diff}| = 4.19$, $p = .04$), whereas in a blocked practice schedule, proactive prompts lead to better scores than both retrospective prompts ($|M_{diff}| = 5.36$, $p < .01$) and no prompts ($|M_{diff}| = 4.18$, $p = .02$).

Table 5.5

Means and Standard Deviations of the Deviation Scores and Uncertainty Margins for All Conditions on the Far Transfer Test

	Blocked practice schedule						Random practice schedule					
	Retrospective prompts		Proactive prompts		No prompts		Retrospective prompts		Proactive prompts		No prompts	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Deviation	23.16	5.48	17.80	3.03	21.98	5.44	12.49	4.95	16.63	5.66	16.68	5.29
Uncertainty	11.33	4.29	7.94	1.32	12.06	4.40	7.44	2.72	7.49	2.80	8.00	2.65

Analyses of the uncertainty margins showed a main effect of practice schedule, $F(1, 114) = 22.97$, $MSE = 234.47$, $p < .01$, $\eta_p^2 = .17$, indicating that participants in the random practice schedule had lower uncertainty margins than participants in the blocked practice schedule ($|M_{diff}| = 2.80$, $p < .01$). There was also a main effect of critical thinking prompts on the uncertainty margins, $F(2, 114) = 5.63$, $MSE = 57.48$, $p < .01$, $\eta_p^2 = .90$. Tukey post hoc tests showed that proactive prompts yielded significantly lower uncertainty margins than both retrospective prompts ($|M_{diff}| = 2.27$, $p < .01$) and no prompts ($|M_{diff}| = 0.17$, $p = .05$). The interaction between practice schedule and critical thinking prompts, $F(2, 114) = 4.12$, $MSE = 42.07$, $p = .02$, $\eta_p^2 = .07$, was significant, therefore, the effects of critical thinking prompts on the uncertainty margins were analyzed separately for the random and blocked practice schedules. In a random practice schedule, critical thinking prompts did not affect the uncertainty margins on the far transfer test, $F(2, 58) = .261$, *ns*. However, in a blocked practice schedule, critical thinking prompts significantly affected the uncertainty margins on the far transfer test, $F(2, 56) = 7.31$, $MSE = 95.48$, $p < .01$, $\eta_p^2 = .21$, with post hoc Tukey tests indicating that proactive prompts resulted in smaller uncertainty margins than both retrospective prompts ($|M_{diff}| = 3.39$, $p = .01$) and no prompts ($|M_{diff}| = 4.13$, $p < 0.01$).

Discussion

This study examined the effects of proactive and retrospective critical thinking prompts provided in combination with random and blocked practice schedules on learning a multidimensional functional learning task. With respect to participants' performance and uncertainty about the accuracy of their judgments, it was hypothesized that: (1) A blocked practice schedule yields better performance and less uncertainty during the learning phase than a random practice schedule, (2) a blocked practice schedule combined with proactive critical thinking prompts shows the same level of uncertainty as a random schedule without prompts, and (3) a blocked practice schedule yields higher transfer when participants receive proactive critical thinking prompts rather than retrospective or no prompts, but a random practice schedule yields higher transfer when participants receive retrospective prompts rather than proactive or no prompts.

In line with our first hypothesis, results showed that in the learning phase, participants who followed a blocked practice schedule performed better than participants who practiced according to a random schedule. This is consistent with research findings on the effects of contextual interference on the acquisition of motor skills (Brady, 1998; Magill & Hall, 1990), troubleshooting skills (De Croock et al., 1998; Jelsma, Van Merriënboer, & Bijlstra, 1990), and language skills (Schneider, Healy, & Bourne, 2002). However, contrary to our expectation, practice schedule did not affect the reported uncertainty margins, and even led to opposite results in the blocked practice schedule with proactive prompts: Uncertainty margins were significantly *lower* than in the retrospective or no prompts conditions, while we hypothesized that the proactive prompts would prevent participants in the blocked group to develop an illusion of competence, which would *increase* uncertainty about their judgments. There are several explanations for this unexpected finding. First, it might be the case that illusions of competence did not occur and that the explanations for the contextual interference effect should indeed be sought in the beneficial effects of random practice rather than the detrimental effects of blocked practice. Second, uncertainty margins might not be an adequate measure of illusions of competence.

Looking at near transfer test performance, we see that participants who followed a random practice schedule benefited most from retrospective prompts as we hypothesized. However, the results of participants who followed a blocked practice schedule are not in line with our hypothesis: Their near transfer test performance was not affected by critical thinking prompts. The explanation for this finding probably lies in the nature of the near transfer tasks. Because these tasks did not differ from the learning tasks on structural features, participants in all blocked conditions could rely on their specific memories of similar cases from the learning phase. That is, the near transfer tasks do not measure the effects of practice schedule on deep understanding. If this explanation is correct, the beneficial effects of proactive prompts with blocked practice should show up on far transfer test performance.

And indeed, in line with our hypothesis, the results on the far transfer test showed superior performance when a random schedule was combined with retrospective critical thinking prompts, and when a blocked schedule was combined with proactive prompts. In the latter condition, transfer test performance was even comparable to that of participants who followed a random practice schedule without critical thinking prompts. This finding has two interesting implications: (a) Combining both techniques to increase interference has surplus value, provided that the prompts are appropriately timed, and (b) critical thinking is an instructional technique that probably also increases interference and that can be just as effective as providing a random practice schedule, again, provided that the prompts are appropriately timed. However, when prompts are not appropriately timed participants' performance need not be negatively affected, as the results of the random condition showed us: Adding proactive prompts did not deteriorate transfer test performance when compared to the no prompts condition.

The second implication, suggesting that critical thinking may replace a random practice schedule, is especially interesting if one considers that there seem to be situations in which tasks are so complex that a random practice schedule starts to hamper learning and blocked practice is the preferred option (cf. Wulf & Shea, 2002). Future research should establish whether under such very high complexity conditions, critical thinking prompts could still be successfully implemented to enhance transfer of learning.

The findings of this study are highly relevant for training settings, and are relatively easy to implement. However, in considering the practical implications of our results, we have to take into account our operationalization of a blocked practice schedule, which is due to the complexity of our task somewhat different from blocked schedules as they are traditionally designed in contextual interference studies. In a traditional blocked schedule, the sequence of learning tasks is such that only one task variation per block is practiced (e.g. block 1: AAAAAA, block 2: BBBBB, where A and B are variations of the learning task). As we stated in the Introduction, in complex judgment, the task for the judge is to predict a future outcome on the basis of a few cues. In learning to predict such future outcomes, the judge has to learn which cues are relevant (cue selection), how each cue relates to the criterion to be predicted, and whether cues are intercorrelated (i.e., cue-criterion functions). Thus, the creation of blocks for these complex judgment tasks is less straightforward than it would be for less complex tasks. The blocked schedules in our experiments were such that in the first blocks one cue changed value from one task to the next. In terms of predicting the future outcomes, this would not seem to be a blocked sequence as the resulting sequential tasks are different task variations. However, in terms of the learning task, that is, to learn cue-criterion functions, participants are provided the opportunity to learn the effects of one cue-criterion function before moving on to learning the next. And in the last blocks, two (or more) cues changed value from one case to the next, providing the opportunity to learn the interaction between the cues. As such, this is a blocked practice schedule, and moreover, one that resembles real world training approaches for complex judgment and decision making. For example, sonar image operators, in their training to identify and judge sonar contacts, first learn how a sonar image depends on ocean bottom patterns, then how water temperatures influence sonar image, and only after that, how ocean bottom pattern and water temperatures interact and how that influences the sonar image (see e.g., www.mosaichydro.com). The data presented here, however, clearly show that such training programs better present the trainee with a random sequence of cases, best combined with retrospective prompts, and leaving it to the trainee to identify and categorize cues and cue-criterion relationships.

Unfortunately, despite the large body of research showing beneficial effects of random practice, instruction is more often than not still structured in such a way that task variety A has to be mastered before moving on to task variety B. However, studies such as this one show that both students and their teachers may think that a task is sufficiently mastered in such a blocked schedule when this is not actually true, because performance during the learning phase is a poor indicator of actual learning outcomes and, especially, of far transfer. The present study shows, nonetheless, that the intuitively appealing blocked practice format does not have to be discarded completely: By combining it with proactive prompts effects on learning are enhanced. This is also good news for instructors using training technologies such as commercially available PC games, of which the event sequences cannot always be adapted. Utilizing techniques such as proactive critical thinking in blocked practice schedules, may improve the educational value of such games.

References

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Vos (Eds.), *Topics in learning and performance* (pp. 131-159). New York: Academic Press.

Battig, W. F. (1979). The flexibility of human memory. In L. S. V. Cermak & F. I. M. Craick (Eds.), *Levels of processing in human memory* (pp. 23-44). Hillsdale, NJ: Erlbaum.

Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology, Research and Development, 56,* 401-422.

Brady, F. (1998). A theoretical and empirical review of the contextual intereference effect and the learning of motor skills. *Quest, 50*, 266-293.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, *45*, 223-241.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13,* 145-182.

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477. .

Cohen M. S., Freeman, J. T., & Thompson, B. T. (1997). Training the naturalistic decision maker. In C. Zsambok & G.A.Klein (Eds.), *Naturalistic Decision Making* (pp. 257-268). Mahwah, NJ: Erlbaum.

Cohen, M. S., Freeman, J. T., & Thompson, B. T. (1998). Critical thinking skills in tactical decision making: a model and a training strategy. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 155-190). Washington, DC: APA.

Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Meta-recognition in time stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*, 206-219.

Cohen, M. S., Thompson, B. B., Adelman, L., Bresnick, T. A., Shastri, L., & Riedel, S. L. (2000). *Training critical thinking for the battlefield. Volume II: Training system and evaluation*. Arlington, VA: Cognitive Technologies, Inc. Retrieved January 2007 from: http://www.cog-tech.com/papers/mentalmodels/Vol_IITraining.pdf

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning questions effect: The role of dialogue and deep-level reasoning questions during vicarious learning. *Cognition and Instruction, 24*, 565-591.

Davis, E. A. (2003). Untangling dimensions of students' beliefs about scientific knowledge and science learning. *International Journal of Science Education, 25*, 439-468.

Davis, E. A., & Linn, M. C. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education, 22*, 819-837.

De Croock, M. B. M., Van Merriënboer, J. J. G., & Paas, F. (1998). High vs low contextual interference in simulation based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior, 14*, 249-267.

Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning*. Utrecht, The Netherlands: Lemma BV.

Douglas, M., Hosokawa, M., & Lawler, F. (1988). *A practical guide to clinical teaching in medicine.* New York: Springer.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.

Farmer, E. W., Van Rooij, J. C. G. M., Riemersma, J. B. J., Jorna, P. G. A. M., & Moraal, J. (1999). *Handbook of simulator-based training*. London, UK: Ashgate.

Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying physics texts: Differences in study processes between good and poor performers. *Cognition and Instruction, 7*, 41-54.

Freeman, J. T., & Cohen, M. S. (1996). Training for complex decision-making: A test of instruction based on the recognition / metacognition model. *Proceedings of the 3rd International Command and Control Research and Technology Symposium 1996*. Retrieved May 12, 2005 from www.cog-tech.com/Publications/TrainingPubs.htm.

Halpern, D. F. (1997). Students need 'reality checks'. *Education Digest, 63,* 24-26.

Hamm, R. M. (1998). Characterizing individual strategies illuminates non optimal behavior. *Psycoloquy, 9*(49). Retrieved February 9, 2008 from http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?9.49

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinman, D. O. (1975). Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes: Formal and mathematical approaches* (pp. 271-312). New York: Academic Press.

Helsdingen, A., Van Gog, T., & Van Merriënboer, J. J. G. (2008). *The effects of practice schedule on learning a complex judgment task*. Manuscript submitted for publication.

Helsdingen, A., Van den Bosch, K., Van Gog, T., & Van Merriënboer, J. J. G. (2008). *The effects of critical thinking instruction on learning complex judgment and decision making*. Manuscript submitted for publication.

Hoffman, P. J., Earle, T. C., & Slovic, P. (1981). Multidimensional functional learning (MFL) and some new conceptions of feedback. *Organizational Behavior and Human Performance, 27*, 75-102.

Jelsma, O., Van Merriënboer, J. J. G., & Bijlstra, J. P. (1990). The ADAPT design model: Towards instructional control of transfer. *Instructional Science, 19*, 89-120.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187-194.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Lee, A. Y., & Hutchison, L. (1998). Improving learning from examples trough reflection. *Journal of Experimental Psychology: Applied, 4*, 187-210.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 730-746.

Magill, R. A., & Hall, K. G. (1990). A review of contextual interference effect in motor skill acquisition. *Human Movement Science, 9*, 241-289.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.

Morgan, N., & Saxton, J. (1991). *Teaching, questioning and learning*. London, UK: Routledge.

Naidu, S., & Bernard, R. M. (1992). Enhancing academic performance in distance education with concept mapping and inserted questions. *Distance Education, 13*, 218-233.

Orlich, D. C., Harder, R. J., Callahan, R. C., Kauchak, D. P., & Gibson, H. W. (1994). *Teaching strategies: A guide to better instruction* (4th Ed.). Toronto, Canada: Heath and Company.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*, 144-161.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1-29.

Roscoe, S. N., & Williges, B. H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), *Aviation Psychology* (pp. 182-193). Ames, IA: Iowa State University Press.

Rowntree, D. (1992). *Exploring open and distance learning*. London, UK: Kogan Page.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention and transfer. *Journal of Memory and Language, 46*, 419-440.

Schwalbe, C. (2004). Re-visioning risk assessment for human service decision making. *Children and Youth Services Review, 26*, 561-576.

Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 113-150). Hoboken, NJ: John Wiley and Sons, Ltd.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 907-912.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179-187.

Van den Boom, G., Paas, F., & van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction, 17*, 532-548.

Van den Boom, G., Paas, F., van Merriënboer, J. J. G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior, 20*, 551-567.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., & Paas, F. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior, 6*, 273-289.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychological Bulletin & Review, 9*, 185-211.

Chapter 6
General Discussion

Abstract

This chapter discusses the findings of the empirical studies reported in Chapters 3, 4, and 5 in terms of theoretical and practical implications, limitations, and suggestions for further research. It can be concluded that the experiments reported in Chapters 3 and 5 prove that critical thinking instruction significantly benefits learning complex judgment tasks, although it has not yet been established whether the entire critical thinking instruction or specific processes evoked at certain moments, such as reflection and self explanation, were responsible for the learning and transfer benefit. Similarly, in Chapters 4 and 5 a benefit of contextual interference for learning complex judgment was reported. In Chapter 4, this benefit arose without detrimental effects on performance during learning, whereas in Chapter 5 detrimental effects during the learning phase were observed of a random (high contextual interference) compared to a blocked schedule. The collected data cannot fully explain this incongruence. Despite these limitations, it is suggested that critical thinking can be fruitfully implemented in current and future training programs for complex judgment and decision making. With a correct timing of critical thinking prompts, the effectiveness not only of random but also of blocked practice schedules can be greatly enhanced.

Professional decision makers have to operate more and more in complex and highly interactive, quickly changing environments. Their decisions may affect the lives of many, and therefore, it is paramount that they are adequately prepared for their tasks. Experience is important, but given the far-reaching consequences of suboptimal decisions, relying solely on experience is dangerous. Prior or additional (on-the-job) training is necessary. Therefore, the main question this dissertation tried to answer is how acquisition of complex judgment and decision making skills can best be fostered.

Complex judgment and decision making is a highly demanding task. It involves collecting and integrating information from different sources (cues) to predict a future outcome. Studies on decision making have provided knowledge of how experienced decision makers approach such complex decision problems (Anderson, 1993; Klein, 1998). It seems to involve at least two types of skill: (1) recognition skills based on subject matter expertise (i.e., acquired cognitive schemas), that is, knowledge of relevant cues, their mutual interrelationships and the relationships with the criterion value that needs to be predicted, and (2) higher order critical thinking skills that serve to increase understanding by means of generalization and abstraction. Targeting training at these skills may improve learning and transfer of subject matter expertise through elaboration of the content (e.g., by generalisation, discrimination, or abstracting away from it). The main aim of this dissertation was to study whether instructional methods for implementing critical thinking (Cohen, Freeman, & Wolf, 1996) and contextual interference (Battig, 1972, 1979; Lee & Magill, 1983; Magill & Hall, 1990;

Shapiro & Schmidt, 1982; Shea & Morgan, 1979), separately and in combination, can increase transfer of judgment skills to new tasks and contexts. Recapitulating, the research questions were:

1. Does critical thinking instruction improve learning of a complex judgment and decision making task?
2. Does the contextual interference effect manifest itself in learning a complex judgment task?
3. What are the combined effects of contextual interference and critical thinking prompts on acquisition and transfer test performance of a complex judgment task?

In the following, first a discussion of the main findings is presented, encompassing a discussion on critical thinking instruction, the effects of contextual interference, the combined effects of critical thinking and contextual interference and a summary. Then, the limitations, theoretical implications and questions for future research are discussed. Finally, the chapter concludes with summarizing the practical implications.

Discussion of Main Findings

Critical Thinking Instruction

An evaluation of the literature showed that positive effects of critical thinking instruction on learning, retention and transfer of complex judgment and decision making have not been established in controlled experimental studies. This may be a consequence of the fact that critical thinking instruction largely stems from naturalistic decision making (NDM) research. As described in Chapter 2, this research approach to decision making focuses on studying behavior of experienced decision makers in natural environments (e.g., at the workplace). The NDM approach provided several descriptive models of decision making, such as the Recognition-Metacognition model (Cohen, Freeman, & Thompson, 1998; Freeman & Cohen, 1996), which placed recognition at the heart of decision making. Therefore, a prerequisite for critical thinking instruction resulting from this model of decision making is that learners must already have, or be able to simultaneously acquire, basic knowledge of the domain in which they are taught the critical thinking skills.

With respect to implementation of critical thinking instruction, it is required that the skills are taught as part of learning a specific task or a specific piece of knowledge. Indeed, despite the ongoing debate on the generalizability of certain types of cognitive skills, the evidence that (critical) thinking skills can be taught separately from any application in a specific domain is meager (Gagné, 1984). Although experienced decision makers seem to have specific critical thinking skills that enable them to operate in a multitude of situations, these skills were not trained or acquired independent from a domain (see, e.g., Ten Dam & Volman, 2002; Tsui, 1999). What seems more plausible is that task specific knowledge and strategies are acquired and then tried across different tasks and domains to arrive, after many years of experience, at a level of general and abstract critical thinking skills (Ten Dam & Volman, 2002; Toplak & Stanovich, 2002). In this dissertation, the question was studied whether it is feasible to deliberately teach these expert skills, embedded in domain specific tasks, to relative novices. Some researchers questioned the viability of such an approach, claiming that novices in a particular domain may experience too high a workload from understanding the rules and principles of the domain and therefore do not have the cognitive resources available for (learning) higher order thinking skills (e.g., Kanfer & Ackerman, 1989; Winne, 1995). This high workload may be especially problematic, as research inspired by cognitive load theory (Sweller, Van Merriënboer, & Paas, 1998; Van Merriënboer & Sweller, 2005) has shown that adding additional instructions under high workload conditions (i.e., in which learners have little if any cognitive resources available) may not just be ineffective but actually hamper learning in the domain (for a review see, e.g., Kalyuga, Ayres, Chandler, & Sweller, 2003).

Whereas one may question the value of critical thinking instruction for novice learners on the basis of cognitive load considerations, both the field studies (Chapter 3) and the laboratory experiment (Chapter 5) presented in this dissertation showed that critical thinking instruction enhances retention and, in particular, transfer test performance. In the field studies, critical thinking instruction was embedded in professional training in command and control judgment and decision making for relative novices in the domain with only a few years of experience. In the laboratory experiment the participants were genuine novices (i.e., they had no prior experience whatsoever with the learning task), but they were still able to profit from the critical thinking instruction and reflection prompts while learning a complex judgment task, provided the prompts were appropriately timed in their particular practice schedule: Pro-active in a blocked schedule and retrospective in a random schedule (see Chapter 5). This is an interesting outcome in the light of Winne's (1995) discussion of experiencing a high workload and not having the resources available for higher order thinking skills. When critical thinking prompts were *not* appropriately timed, participants' performance indeed did not benefit from critical thinking instruction, possibly indicating they were not able to use the prompts due to high workload. But their performance was not negatively affected, which suggests that in high workload conditions learners may be able to ignore given prompts all together and focus on learning the specific rules and principles of the task without being distracted by them.

Contextual Interference

In Chapter 2, based on the literature several explanations were described for the benefit of random practice schedules (i.e., high contextual interference) compared to blocked practice schedules (i.e., low contextual interference). This benefit has been found for retention and transfer across tasks relying on many types of skill and knowledge (e.g., Brady, 1998; Wulf & Shea, 2002). Different explanations for the positive effects of random practice (e.g., Lee & Magill, 1983; Shea & Morgan, 1979; Schmidt & Bjork, 1992) have in common that they assume that a random practice schedule calls for more elaborative processing between or within tasks, and that these elaborative processes lead to deeper understanding and, consequently, better retention and transfer test performance.

The results of our laboratory studies on contextual interference (Chapter 4) demonstrated the hypothesized benefits of random practice schedules on retention and transfer of complex judgment tasks, without showing the performance detriment during the learning phase often reported as a result of random practice. This result was fully in line with findings of Immink and Wright (1998). With regard to performance detriments during learning, they found that the time between the randomly scheduled learning tasks is crucial: The decline does not occur when enough time is provided. They assumed this was the case because learners need to replan their actions between randomly sequenced tasks and need time to do so. In their study, replanning involved reconstruction of an action plan for the next movement in pressing a sequence of keys on a keyboard. In the studies reported in Chapter 4, the start of each learning task was self paced, possibly giving participants sufficient time to replan their response in a random schedule. Whereas Immink and Wright provide a likely explanation, it should be noted though that their study involved a very different task. It still needs to be ascertained whether a similar mechanism applies to the acquisition of the complex cognitive task of judgment and decision-making used in our study, especially because a detrimental effect of random compared to blocked practice during the learning phase *was* found in the experiment reported in Chapter 5 (see below).

In Chapter 5 another explanation for the differential effects of random and blocked practice schedules on retention and transfer was given, focusing less on the benefits of random practice and more on the potential drawbacks of blocked practice. This potential explanation concerns an *illusion of competence* learners may experience as a result of the repetitive presentation of the same type of task in a blocked practice schedule (Koriat & Bjork, 2005). Illusions of competence may inhibit search and application of alternative task strategies and thus deteriorate gaining a deeper level of understanding (Kornell & Bjork, 2007). As a consequence, transfer and, in particular, far transfer test performance might be hampered in a blocked sequence compared to a random sequence, where illusions of competence are less likely to occur because learners are confronted with constantly varying learning tasks.

The experiment reported in Chapter 5 showed similar benefits of random practice for near and far transfer performance. As mentioned previously, with regard to performance during the learning phase, the results from the last experiment (Chapter 5) were not consistent with those of the previous experiments (Chapter 4). In the final experiment, participants' performance during the learning phase was better in a blocked practice schedule than in a random practice schedule, even though they were again allowed to control their own pace, as in the earlier experiments. The incongruence may be due to a difference in the judgment tasks between the experiments of Chapters 4 and 5: In the experiment in Chapter 5, participants were required to identify the relevant cues in each case description by themselves, whereas in the former experiments those relevant cues were given to them. Thus, in the former experiments the task was to integrate information from the cues and to predict a future outcome, while in the final experiment this task was expanded with the identification of relevant cues. Cue identification may have been far more difficult in a random schedule than in a blocked schedule, since in a blocked schedule only one cue changed value from one case to the next, making it relatively easy to spot that cue. In a random schedule, cues and values changed randomly, making it much harder to identify the relevant cues. However, whether this difference in the complex judgment task has indeed caused the different results for the learning phase performance between the experiments of Chapter 4 and 5 remains uncertain. This is an interesting question for future research, as it might imply that Immink and Wright's (1998) finding that availability of time between learning tasks is the crucial factor, may actually depend on the type of task.

Combining critical thinking instruction and contextual interference

In the experiment of Chapter 5, critical thinking instruction was combined with random and blocked practice schedules to investigate it's combined effects on learning complex judgment. To deal with learners' limited processing capacity and illusions of competence, it was hypothesized that the timing of critical thinking prompts is crucial in order to be effective. In a blocked schedule, critical thinking processes would need to be prompted *before* execution of the task (i.e., proactive), so that critical thinking may occur concurrently with performing the learning task. When prompted after task execution in a blocked practice schedule, learners are not expected to critically reflect, since they are under the impression that they already understand the task (illusion of competence). In a random practice schedule, the key factor for successfully planning critical thinking prompts would be the learners' limited processing capacity: When learners already experience high processing demands, they will not be able to critically reflect on their task strategies concurrently with performing the learning task. Therefore, it was expected that prompting critical thinking *after* execution of the learning task (i.e., retrospective) would have beneficial effects on learning without leading to processing capacity problems.

In line with our hypotheses, it was found that proactive prompting in a blocked schedule benefited learning, as did retrospective prompting in a random schedule. A blocked practice schedule with proactive critical thinking prompts even brought far transfer test performance at a similar level as a random practice schedule without critical thinking instruction. These results suggest that under conditions where blocked practice is the only viable option (e.g., because the learning tasks are too complex, so that providing learners with a random schedule would increase workload to an unacceptable level), presenting proactive critical thinking prompts might provide enough interference to enhance transfer. However, future research should establish if under very high task complexity conditions, critical thinking prompts might still be successfully implemented to enhance transfer test performance.

It is interesting to note that the results presented in Chapter 5 suggest that critical thinking instruction has similar effects as random practice, that is, it can raise learners' performance after a blocked schedule to the same level reached in a random schedule without prompts. Furthermore, critical thinking instruction can amplify the effects of a random practice schedule, that is, random practice combined with retrospective prompts is more effective than random practice without prompts. On the one hand, critical thinking instruction may lead learners to become engaged in more elaborative processing of particular tasks and series of tasks, comparable to the explanation for benefits of random practice schedules. On the other hand, critical thinking instruction may also prevent learners from developing an illusion of competence, because it forces them to focus on testing hypotheses, critiquing arguments, and developing contingency plans. Thus, preventing illusions of competence may possibly by reached by random practice as well critical thinking instruction.

To monitor whether participants developed an illusion of competence during learning, in the experiment of Chapter 5, participants' uncertainty margins were measured as an indication of their confidence in the accuracy of their own judgments. It was expected that during the learning phase, uncertainty would be lower in a blocked practice schedule, as a result of an illusion of competence that learners would develop as a consequence of the repetitive presentation of similar learning tasks. However, contrary to the expectation, practice schedule did not affect the reported uncertainty margins: Uncertainty was at a similar level in random and blocked practice schedules. At the same time, results showed that proactive critical thinking prompts decreased uncertainty, indicating that participants were *more* certain of the accuracy of their judgments. This was also opposite to what was hypothesized, since proactive critical thinking prompts were expected to provide participants with a more realistic view of their own performance during blocked practice and thereby increase uncertainty. Yet, in accordance with our expectation, test performance of the blocked group with proactive prompts was better than that of the blocked group that received no prompts or retrospective prompts. So, it might still be true that an illusion of competence was prevented by proactive prompts. This did, however, not show in the uncertainty scores in the way it was expected to, indicating that these scores might not represent a good measure of illusions of competence. To find out what learners consider when they rate priorities, future research should develop and use alternative measures of illusions of competence. In addition, concurrent or retrospective reports (Ericsson & Simon, 1993) or –under high workload conditions –cued retrospective reports (Van Gog, Paas, Van Merriënboer, & Witte, 2005) may help to recognize illusions of competence.

Limitations and Future Research

The studies presented in this dissertation established that an approach such as critical thinking instruction enhanced learning and transfer in complex judgment – both in the field and in the laboratory. However, a limitation of these studies in terms of theoretical explanations is that both in the field and in the laboratory the comprehensive critical thinking instruction was given. This made it impossible to establish whether it is the instruction in its entirety or specific aspects of critical thinking, such as reflection or self-explanation, which are responsible for the benefits on learning and transfer. Future research should compare the entire critical thinking instruction with for example reflection prompts or self-explanation prompts only, to study the effects of these aspects separately.

A second limitation concerns the explanation for the benefits of contextual interference on learning complex judgment. Explanations have been suggested that focus on the elaboration and extra processing of information, which are necessary under conditions of high contextual interference but not under conditions of low contextual interference. However, an alternative suggestion was made, namely, that elaboration and extra processing are not inherent to conditions of high contextual interference, but may be hampered in low contextual interference practice schedules as a result of illusions of competence. The two types of explanations are not mutually exclusive, that is, it could be that high contextual interference leads to extra elaboration and, at the same time, low contextual interference leads to illusions of competence. The results do not support one explanation over the other, since (a) it could not be determined whether a random practice schedule required more processing and elaboration than a blocked practice schedule, and (b) it could not be established whether participants actually developed any illusions of competence in a blocked practice schedule. Both aspects are interesting for future research. For that, it is first important to determine and integrate an adequate and non-intrusive measure to determine participants' judgments of learning in order to establish whether they develop an illusion of competence (Koriat & Bjork, 2005). Second, time on task and/or measures of cognitive effort (cf. Paas, 1992) should be measured for each separate learning task and test task, as an indication of additional and elaborative processing.

A third limitation follows from the differences between the laboratory experiments as was discussed earlier: In the three laboratory experiments, the tasks trained in the first two experiments (Chapter 4) were slightly different from the task trained in the third experiment (Chapter 5). Apart from aggregating information from cues and estimating the priority of a case on the basis of that information, in the third experiment participants first had to identify relevant cues. This difference in tasks to be learned might have been the cause for differential effects on performance during the learning phases. However, future research is necessary to establish whether contextual interference has indeed differential effects on performance during the learning phase for different types of tasks. A positive consequence of the difference between the tasks in Chapter 4 and 5, is that all studies indicated that contextual interference is effective for transfer, so this effect held even under more complex conditions and in combination with critical thinking (which may further increase processing demands).

Practical Implications

Despite the limitations of the presented studies and the open questions that remain, the results provide important practical implications. First, instruction based on expert judgment and decision making strategies, such as critical thinking, increases learners' understanding of the learning tasks and improves their post-training performance. It is therefore advisable to make critical thinking instruction an integral part of a training program for professional decision makers, embedded as a general approach but with learning tasks that feature domain-specific judgment and decision problems. Such an approach could encompass the four critical thinking steps described in Chapters 2 and 3.

Second, practice in complex judgment and decision making, for example in the domain of military command and control, should be randomly sequenced to realize optimal post training performance. Unfortunately, despite the large body of research showing beneficial effects of random practice, instruction is more often than not still structured in such a way that task A has to be mastered before moving on to task B. Studies such as those presented in this dissertation show that both teachers and students working in a blocked schedule may easily think that they master a task whereas this is not actually true. Performance during the learning phase in a blocked schedule may be high, but this is often a poor indicator of actual learning outcomes and, especially, far transfer.

Thirdly, when a task is highly complex and a random practice schedule would provide novice learners with too high a workload, a blocked practice schedule might be combined with proactive critical thinking prompts to enhance learning and post training performance. In addition, the finding that the effects of blocked practice can be enhanced by proactive prompts is also good news for instructors using training software such as commercially available training packages or games, of which the event sequences cannot always be adapted. Utilizing techniques such as critical thinking may improve the educational value of such packages or games.

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Vos (Eds.), *Topics in learning and performance* (pp. 131-159). New York: Academic Press.

Battig, W. F. (1979). The flexibility of human memory. In L. S. V. Cermak & F. I. M. Craick (Eds.), *Levels of processing in human memory* (pp. 23-44). Hillsdale, NJ: Lawrence Erlbaum.

Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. *Quest, 50*, 266-293.

Cohen, M. S., Freeman, J. T., & Thompson, B. T. (1998). Critical thinking skills in tactical decision making: a model and a training strategy. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 155-190).

Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Meta-recognition in time stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*, 206-219.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Freeman, J. T., & Cohen, M. S. (1996). Training for complex decision-making: A test of instruction based on the recognition / metacognition model. *Proceedings of the 3rd International Command and Control Research and Technology Symposium 1996.* Retrieved May 12, 2005 from www.cog-tech.com/Publications/TrainingPubs.htm.

Gagné, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist, 39,* 377-385.

Immink, M. A., & Wright, D. L. (1998). Contextual interference: A response planning account. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 51A,* 735-754.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). Expertise reversal effect. *Educational Psychologist*, *38*, 23-31.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude treatment approach to skill acquisition. *Journal of Applied Psychology, 74,* 657-690.

Klein, G. (1998). Sources of power: *How people make decisions*. Cambridge, MA: MIT Press.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187-194.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 730-746.

Magill, R. A., & Hall, K. G. (1990). A review of contextual interference effect in motor skill acquisition. *Human Movement Science, 9*, 241-289.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429-434.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and co-ordination*. New York: Wiley.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on acquisition, retention and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179-187.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10, 251-296.*

Ten Dam, G., & Volman, M. (2002). Critical thinking as a citizenship competence: Teaching strategies. *Learning and Instruction, 14*, 359-370.

Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 94*, 197-209.

Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education, 40*, 185-200.

Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*, 237-244.

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147-177.

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, *30*, 173-187.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychological Bulletin* & *Review, 9*, 185-211.

Summary

Many of today's work environments are dynamic and complex sociotechnical systems in which professional decision makers need to exercise judgment on issues that may have far reaching consequences. These decision makers should be well prepared for their jobs, and adequate education and training in complex decision making and judgment is therefore paramount. The aim of such education and training programs often is to establish transfer of knowledge and skills from the learning tasks to the professional tasks. However, many training programs do not take the results of empirical studies into account regarding instructional measures and strategies that may enhance transfer. The studies reported in this dissertation were undertaken to determine to what extent novice decision makers can benefit from specific instruction in judgment strategies as employed by experienced decision makers, and to study how contextual interference in practice schedules affects learning and transfer in complex judgment and decision making.

Chapter 1 introduces the research questions and provides an overview of the dissertation. In Chapter 2 the theoretical background of the studies presented in this dissertation is elaborated. First, an overview of two different approaches to decision making research is provided: Social judgment theory (SJT) and naturalistic decision making (NDM). Both SJT and NDM research emphasize that learning complex judgment and decision making involves, on the one hand, learning what the relevant cues upon which to base a decision are and how they are related to each other, and, on the other hand, learning (meta)cognitive strategies that guide the learning and decision making processes. Within both research approaches the effects of several interventions aimed at learning either cues and their interrelations or (meta)cognitive strategies have been investigated. These studies focussed mostly on performance during learning and/or on retention, that is, later performance on similar tasks. However, empirical evaluations of the effects on transfer, that is, later performance on *novel* tasks were scarce. The chapter continues with a discussion of transfer of training and measures that may be effective for enhancing transfer of complex judgment, such as critical thinking instruction and contextual interference.

Critical thinking instruction resulted from studying strategies of experienced decision makers and has been proven to enhance complex judgment and decision making skills in several field studies, at least with regard to performance during learning and retention. Critical thinking presumably leads to elaborative processes, such as reflection and/or self explanation of similarities and differences between tasks and task contexts, thereby leading to more abstract or generalized knowledge of the learning tasks, facilitating retention and transfer. However, critical thinking instruction has not been studied yet in well-controlled experiments, and transfer test performance was not specifically assessed in earlier evaluative field studies.

The second measure that might lead to enhanced retention and transfer test performance in complex judgment and decision making is contextual interference. Contextual interference is the interference between learning tasks as a result of a practice schedule where different variations of the learning tasks are sequenced randomly (high interference) as opposed to sequenced in separate blocks (low interference). The effects of contextual interference have been studied for performance on many types of tasks, and effects have been measured during learning, on retention, on near transfer (i.e., transfer to tasks that differ from the learning tasks on superficial aspects but are similar in structural aspects), and on far transfer (i.e., transfer to tasks that are different from the learning tasks on both superficial and structural aspects). High contextual interference (i.e., random practice) has proven to lead to enhanced transfer, especially far transfer, in several domains. Chapter 2 describes an overview of studies on contextual interference and some mechanisms that may underlie the benefits of random practice, such as elaboration processes that lead to more abstract and generalized knowledge of the learning tasks and learners' higher involvement to become engaged in elaboration. This chapter concludes with establishing that several interesting research questions result from the above: (a) What are the effects of critical thinking on transfer? (b) What are the effects of contextual interference on learning complex judgment tasks? and (c) What are the combined effects of critical thinking and contextual interference on learning, on near transfer, and on far transfer of a complex judgment task? In order to address those questions, the subsequent chapters present five experiments in all – two field studies on the effects of critical thinking in Chapter 3; two experimental studies on the effects of contextual interference in Chapter 4, and a final experiment on the combination of effects of critical thinking instruction and contextual interference in Chapter 5.

In Chapter 3 two field studies are reported that investigated the effects of critical thinking instruction on complex judgment and decision making. The first study focused on acquisition and transfer performance of air force officers learning to command and control ground-to-air defense operations; the second study on naval (petty) officers learning to command and control air and surface warfare. Participants received a training with scenario-based exercises in both simplified (Study 1) and high fidelity (Study 2) learning environments. In both studies, half of the participants received instruction in critical thinking. The other half received the same exercises, but without critical thinking instruction. After training, test scenarios were administered to both groups. The results provided support for the hypotheses that critical thinking instruction has a positive effect on transfer test performance and that this benefit is greater for far transfer than for near transfer.

Chapter 4 presents two experiments that investigated the potential benefits of a random practice schedule relative to a blocked schedule for retention and transfer test performance in a complex judgment task. Participants' task was to read a description of a fictitious crime and to judge its priority on the basis of several cues present in each case description. In Experiment 1, participants' judgment accuracy on a retention test was higher after a random practice schedule when compared to a blocked or operational practice schedule, that is, a practice schedule that is modelled according to the real world sequence and frequency of events. Experiment 2 demonstrated that judgment performance on a transfer test was also better after a random practice schedule than after a blocked practice schedule. Both experiments failed to show any detrimental effects of random practice on performance during the learning phase, which has often been reported in other studies on contextual interference. Our findings show that contextual interference also improves retention and transfer of complex judgment skill, and that these benefits may be achieved without detrimental effects on performance during practice.

Chapter 5 contains a study that investigated the combined effects of critical thinking instruction in blocked and random practice schedules on learning and transfer of complex judgment skills. Similar to Chapter 4, participants' task was to read a description of a fictitious crime and to judge its priority on the basis of several cues present in each case description, however, this time, participants had to discover the relevant cues themselves. Critical thinking prompts (proactive, retrospective, none) were manipulated: Proactive prompts were given before the learning task they referred to, thereby facilitating critical thinking concurrently with task execution; retrospective prompts were given after execution of the tasks they referred to, facilitating reflection upon those tasks. Furthermore, practice schedule (random, blocked) was manipulated again to study whether the different critical thinking prompts differentially affected random and blocked practice schedules. Results showed that during the learning phase a blocked practice schedule led to better performance than random practice, which may possibly be due to the difference in tasks compared to Chapter 4: The task was now more complex because learners also had to identify the relevant cues. During the transfer test phase a random schedule was superior. As expected, an interaction effect was also found: Whereas in a random schedule retrospective critical thinking prompts have surplus value over proactive prompts, transfer performance following a blocked schedule can better be enhanced through proactive critical thinking prompts.

Chapter 6 presents a General Discussion of these findings in terms of the theoretical and practical implications, limitations of the studies, and suggestions for further research that result from the work presented in this dissertation. It concludes that the experiments reported in Chapters 3 and 5 prove that critical thinking instruction significantly benefits learning complex judgment tasks, although it has not yet been established whether the entire critical thinking instruction or specific processes evoked at certain moments, such as reflection and self explanation, were responsible for the learning and transfer benefit.

Similarly, in Chapters 4 and 5 a benefit of contextual interference for learning complex judgment was reported. In Chapter 4, this benefit arose without detrimental effects on performance during learning, whereas in Chapter 5 detrimental effects during the learning phase were observed of a random (high contextual interference) compared to a blocked schedule. The collected data cannot fully explain this incongruence. Despite these limitations, it is suggested that critical thinking can be fruitfully implemented in current and future training programs for complex judgment and decision making. With a correct timing of critical thinking prompts, not only random but also blocked practice schedules can greatly enhance learning and transfer performance.

Samenvatting

Veel hedendaagse werkomgevingen zijn zeer dynamische en complexe, sociotechnische systemen waarin professionals besluiten moeten nemen met soms verstrekkende gevolgen. Deze besluitvormers moeten goed voorbereid zijn op hun taken en daarom is adequate opleiding en training in beoordelen en beslissen van groot belang. Het doel van zulke opleidings- en trainingsprogramma's is meestal het bewerkstelligen van transfer van kennis en vaardigheden van de leertaken naar de operationele taken. Daarbij worden echter regelmatig de bevindingen van empirische studies naar het verbeteren van transfer genegeerd. De studies die dit proefschrift rapporteert zijn uitgevoerd om te bepalen of onervaren besluitvormers baat kunnen hebben bij specifieke instructie in de beoordelingsstrategieën van ervaren besluitvormers ('kritisch-denken instructie'). Daarnaast wordt onderzocht of contextuele interferentie in trainingschema's het leren en de transfer van complexe besluitvormingsvaardigheden positief beïnvloedt.

Hoofdstuk 1 introduceert de onderzoeksvragen en biedt een overzicht van het proefschrift. De theoretische achtergrond van de studies die in dit proefschrift worden gepresenteerd staat in Hoofdstuk 2. Allereerst wordt een overzicht gegeven van twee verschillende benaderingen van beslisonderzoek: *Social Judgment Theory* (SJT) en *Naturalistic Decision Making* (NDM). Beide methoden benadrukken dat het voor complexe besluitvorming van belang is te leren wat de relevante factoren in een situatie zijn en hoe deze met elkaar samenhangen. Tevens wordt het belang benadrukt van het leren van (meta)cognitieve vaardigheden die het leer- en besluitvormingsproces sturen. In beide onderzoeksmethoden zijn de effecten van diverse interventies gericht op zowel het leren van factoren en hun onderlinge relaties als op het leren van (meta)cognitieve strategieën onderzocht. Deze studies richtten zich meestal op de prestatie tijdens het leerproces en op retentie, dat wil zeggen, prestatie op dezelfde taken na afloop van de training. Empirische evaluaties van de effecten van diverse interventies op transfer, dat wil zeggen, prestatie na de training op *nieuwe* taken, zijn schaars. Een discussie over transfer van training volgt, waarin ook maatregelen besproken worden die effectief zouden kunnen zijn voor het bevorderen van deze transfer, zoals *critical thinking* instructie en contextuele interferentie.

Critical thinking instructie is het resultaat van diverse studies naar de strategie van ervaren besluitvormers in hun natuurlijke taakomgeving. Het is een aanpak waarvan middels veldexperimenten is aangetoond dat het een meerwaarde genereert voor de prestatie gedurende het leerproces en de retentie van complexe besluitvormingsvaardigheden. De idee is dat *critical thinking* instructie leidt tot verdieping van de leerstof door middel van processen zoals bijvoorbeeld reflectie en het verklaren van de overeenkomsten en verschillen tussen taken en taakcontexten, waardoor een meer abstract of algemeen begrip van de leertaak bewerkstelligd wordt, wat de retentie en transfer bevordert. De effecten van *critical thinking* instructie zijn echter nooit getoetst in een gecontroleerde experimentele omgeving, en transfer van het geleerde naar andere taken of taakomgevingen werd in de eerdere veldstudies niet gemeten.

De tweede maatregel die tot betere retentie en transfer van complexe besluitvormingsvaardigheden zou kunnen leiden is contextuele interferentie. Contextuele interferentie verwijst naar interferentie als gevolg van een toevallige opeenvolging van leertaken (hoge interferentie; bijv. BAACBC), in tegenstelling tot een geblokte opeenvolging van leertaken (lage interferentie; bijv. AA-BB-CC). De effecten van contextuele interferentie op prestatie tijdens leren, retentie, *near transfer* (transfer naar taken die oppervlakkig verschillen van de leertaken, maar op structureel niveau gelijk zijn) en *far transfer* (transfer naar taken die oppervlakkig en structureel verschillen van de leertaken) zijn onderzocht voor vele verschillende typen taken. Uit deze studies komt naar voren dat hoge contextuele interferentie leidt tot betere transfer, met name *far transfer*, op verschillende complexe taken. In Hoofdstuk 2 wordt een overzicht gegeven van studies naar de effecten van contextuele interferentie en worden enkele mechanismen besproken die ten grondslag zouden kunnen liggen aan het waargenomen voordeel van een toevalsvolgorde van leertaken. Dat zouden bijvoorbeeld processen kunnen zijn, die leiden tot een meer abstract en algemeen begrip van de leertaak. Ook zouden leerlingen minder gemotiveerd kunnen zijn om zich te verdiepen in de leerstof als zij werken aan leertaken in een geblokte volgorde. Het hoofdstuk besluit met enkele onderzoeksvragen die voortvloeien uit de bovengenoemde discussie van relevant onderzoek (a) Wat zijn de effecten van *critical thinking* instructie op transfer van complexe besluitvormingsvaardigheden? (b) Wat zijn de invloeden van contextuele interferentie op leren en transfer van complexe besluitvormingsvaardigheden? En (c) wat zijn de gecombineerde effecten van *critical thinking* instructie en contextuele interferentie op het leren, de *near transfer* en de *far transfer* van complexe besluitvormingsvaardigheden? Om deze vragen te beantwoorden zijn in totaal 5 studies uitgevoerd. Hoofdstuk 3 doet verslag van twee veldstudies naar de effecten van *critical thinking* instructie. Hoofdstuk 4 behandelt twee experimentele laboratoriumexperimenten naar de effecten van contextuele interferentie op het leren van complexe besluitvorming, en Hoofdstuk 5 gaat over een experiment naar de gecombineerde effecten van *critical thinking* instructie en contextuele interferentie.

Hoofdstuk 3 rapporteert twee veldstudies die de effecten van *critical thinking* instructie op complexe beoordeling en besluitvorming onderzoeken. De eerste studie met luchtmachtofficieren was gericht op leren en de transfer van *command and control* van grondgebonden luchtverdedigingsoperaties. De tweede studie was gericht op marine-officieren en onderofficieren die *command en control* taken van luchtverdedigings- en oppervlakte-oorlogsvoeringsoperatie leerden. Deelnemers kregen scenariogebaseerde oefeningen in zeer eenvoudige (Studie 1) en dynamische, interactieve (Studie 2) leeromgevingen. In beide studies kreeg de helft van de deelnemers instructie in kritisch denken, de andere helft van de deelnemers deed dezelfde oefeningen maar dan zonder instructie in kritisch denken. Na de training werd bij beide groepen dezelfde test afgenomen. De resultaten onderschrijven de hypothese dat *critical thinking* instructie een positief effect heeft op de transfer van complexe besluitvormingsvaardigheden, en dat dit effect groter is op *far transfer* dan op *near transfer*.

Hoofdstuk 4 presenteert twee laboratoriumexperimenten naar de effecten van de ordening van oefenmateriaal op het leren. Een toevallige opeenvolging van leertaken (hoge contextuele interferentie) en een geblokte opeenvolging (lage contextuele interferentie) werden met elkaar vergeleken. De deelnemers lazen een beschrijving van een fictieve misdaad. Zij moesten op basis van expliciet gegeven aanwijzingen ('cues') in de tekst de prioriteit van deze misdaad voor een fictief opsporingsapparaat beoordelen . Uit de resultaten van Experiment 1 bleek dat de beoordelingsprestatie op de retentietest hoger was na de toevallige volgorde van leertaken dan na een geblokte of operationele volgorde, dat wil zeggen, een opeenvolging die is vastgesteld op basis van de frequentie en volgorde van taken in de werkelijkheid. Ook Experiment 2 toonde aan dat beoordelingsprestaties op een transfertest beter waren na een toevallige volgorde van leertaken dan na een geblokte volgorde. Tegen de verwachting in lieten beide experimenten niet zien dat een toevallige volgorde van leertaken leidde tot een verminderde prestatie gedurende het leren, zoals dat wel vaak werd gevonden in andere studies naar de effecten van contextuele interferentie. Tezamen tonen deze resultaten aan dat contextuele interferentie de retentie en transfer van complexe besluitvormingsvaardigheden bevordert en dat dit bereikt kan worden zonder een prestatievermindering gedurende het leren.

Hoofdstuk 5 beschrijft een studie naar de effecten van *critical thinking* instructie in combinatie met een toevallige of geblokte volgorde van leertaken op het leren en de transfer van complexe besluitvormingsvaardigheden. Evenals in de experimenten beschreven in Hoofdstuk 4, lazen deelnemers een beschrijving van een fictieve misdaad en de prioriteit van deze misdaad voor een fictief opsporingsapparaat beoordelen op basis van enkele aanwijzingen die in elke beschrijving voorkwamen. Echter, anders dan in de experimenten van Hoofdstuk 4, werden de aanwijzingen niet expliciet gegeven, maar moesten de deelnemers de aanwijzingen zelf ontdekken in de beschrijving. Het aanbieden van verschillende *critical thinking* prompts (pro-actief, retrospectief, geen) werd gemanipuleerd tussen condities: Proactieve prompts werden gegeven voorafgaand aan de taken waarop ze betrekking hadden, om zo het kritisch denken vooraf en tijdens de taakuitvoering te faciliteren; retrospectieve prompts werden gepresenteerd na de uitvoering van taken waarop ze betrekking hadden, om zo de reflectie op deze taken te bevorderen. Daarnaast werd, zoals in Hoofdstuk 4, de volgorde van leertaken gemanipuleerd (toevallig, geblokt) om te kunnen bestuderen of de effecten van *critical thinking* prompts verschilden in beide taakvolgorden. De resultaten toonden aan dat –in tegenstelling tot de resultaten bij Experiment 4- een geblokte volgorde van leertaken leidde tot een betere prestatie gedurende de leerfase. Wellicht kan dit toegeschreven worden aan de complexiteit van de taak: De taak was nu complexer dan in de vorige experimenten (Hoofdstuk 4) doordat de deelnemers ook de relevante aanwijzingen in de beschrijving moesten identificeren. De prestatie op de transfertest liet een duidelijk voordeel zien van een toevallige volgorde van leertaken. Zoals voorspeld werd ook een interactie-effect gevonden: In een toevallige volgorde van leertaken hadden retrospectieve prompts een meerwaarde op de retentie en transfer. In een geblokte volgorde van leertaken kunnen beter proactieve *critical thinking* prompts gegeven worden.

Hoofdstuk 6 presenteert de Algemene Discussie van de bevindingen van de verschillende studies en hun implicaties voor theorie en praktijk. Ook worden de beperkingen van de studies besproken en worden suggesties gegeven voor verder onderzoek. Geconcludeerd wordt dat *critical thinking* instructie een significant voordeel oplevert voor het leren van complexe besluitvormingsvaardigheden. Op basis van de studies in deze dissertatie kan niet worden vastgesteld of deze effecten veroorzaakt worden door de instructie als geheel, of door specifieke aanwijzingen in de instructie zoals bijvoorbeeld “reflecteer” of “verklaar”. Ook de voordelen van contextuele interferentie voor het leren van complexe besluitvormingsvaardigheden werden empirisch vastgesteld. Echter, dit voordeel voor retentie en transfer van een toevallige volgorde werd in de eerste experimenten gevonden zonder een prestatievermindering gedurende het leren ten opzichte van een geblokte volgorde, terwijl in een later experiment wel degelijk een prestatievermindering werd geconstateerd gedurende het leren als gevolg van een toevallige volgorde van leertaken. Dit verschil is lastig te verklaren, maar heeft –zoals eerder genoemd- mogelijk te maken met een verschil in taakcomplexiteit tussen de experimenten zoals gerapporteerd in de hoofdstukken 4 en 5. Ondanks deze beperkingen wordt geconcludeerd dat de studies gepresenteerd in dit proefschrift aantonen dat *critical thinking* instructie succesvol kan worden ingezet voor het leren van complexe besluitvorming. Met een correcte timing van *critical thinking* prompts kan bovendien worden gegarandeerd dat niet enkel een toevallige volgorde, maar ook een geblokte volgorde van leertaken leidt tot adequate retentie en transfer van complexe besluitvormingsvaardigheden.