Enhanced Memory for Faces of Cheaters?

A Critical Test of the *Social Contract Theory*

Inaugural-Dissertation
zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Bettina Mehl
aus Offenbach

Düsseldorf 2005

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. A. Buchner

Koreferent: HD Dr. M. Niedeggen

Tag der mündlichen Prüfung: 04.11.2005

Meinen Eltern

Table of Contents

Abstract

The present thesis provides an analysis and comprehensive empirical test of the so-called *Social Contract Theory* (Cosmides, 1989; Cosmides & Tooby, 1989). Within their theory, Cosmides and Tooby postulate the existence of neural circuits inherent in the human brain that are functionally specialized for the detection of potential cheaters. These mechanisms are presumed to have evolved in the course of our species' evolutionary history by means of natural selection. In a series of four experiments, the hypothesis of an enhanced memory for faces of cheaters due to evolved psychological mechanisms operating on the encoding and storage of information relating to social exchange was tested. Under the guise of a test-retest reliability study, students were asked to rate the attractiveness (Experiments 1 to 3) or sympathy (Experiment 4) of male facial photos. Each photo was presented with a fictional descriptive sentence giving information on the depicted person's threat potential (history of cheating, history of trustworthiness, irrelevant information) and social status (low, high; Experiments 1 to 3) or type of behavior (exceptional, ordinary; Experiment 4). In a subsequent recognition test, subjects were presented an equal number of old and new facial photos—this time, without descriptions—and were asked to indicate which of the photos they remembered. Given a face was classified as remembered, subjects were to specify the behavioral information previously associated with this face in a source-memory test. Multinomial source-memory models were applied to the data analysis (Experiments 3 and 4). The results of all four experiments were not consistent with the cheater-detection hypothesis. The predicted bias that subjects would preferentially recognize faces initially presented as those of cheaters was not confirmed. Moreover, neither social status nor type of behavior notably affected subjects' recognition performance. Solely, Experiment 4 yielded evidence consistent with a somewhat modified cheater-detection hypothesis positing an enhanced memory for source information on cheating. Independent of prior face recognition subjects showed an improved memory for source information on cheating compared to information not associated with potential threat. However, this result does not necessarily support the hypothesis of evolved psychological mechanisms guiding reasoning involved in detecting cheaters.

1 Introduction

From their earliest beginnings, human beings have been living in social groups (Cosmides & Tooby, 1989, 1997; Geary, 1998). While it is hardly conceivable from our modern point of view, our ancestors spent most of our species' evolutionary history living in hunter-gatherer societies (Cosmides & Tooby, 1997)[1]. Living in small, nomadic communities of kin, our Pleistocene hunter-gatherer ancestors were faced with a variety of day-to-day problems, which to solve was a matter of survival (Cosmides & Tooby, 1997)—consider, for example, hunting animals, gathering plant foods or defending themselves against offenders. Certainly, to cooperate on these tasks comprised considerable benefit for those involved in the interaction, enhancing the probability of their surviving. Actually, there is evidence that our Pleistocene hunter-gatherer ancestors did cooperate in a multitude of ways. For example, it is known that they engaged in extensive food sharing and cooperative hunting and that they shared tools as well as information on tool making (Cosmides & Tooby, 1989). Yet as every coin has its flip side, any situation in which individuals cooperate for mutual benefit holds the possibility of cheating. Consequently, to prevent oneself from deception may be regarded as a crucial factor for successful engagement in social exchanges. Based on this assumption, Leda Cosmides and John Tooby (Cosmides, 1989; Cosmides & Tooby, 1989, 1992) provided a broad theoretical framework for the analysis of social exchange. Within their so-called computational theory of social exchange, or *Social Contract Theory*, the authors claim the existence of neural circuits inherent in the human brain that are functionally specialized for the detection of potential cheaters. These mechanisms are presumed to have evolved in the course of our species' evolutionary history by means of natural selection.

However, their postulate, which has been derived from a Darwinist approach to psychology, largely rests upon its prima facie plausibility. Virtually all of the empirical evidence provided by Cosmides to support her hypotheses (Cosmides, 1989) is based upon one single paradigm, namely Wason's four-card selection task (Wason, 1966, 1968). The appropriateness of this task with respect to Cosmides' work is highly disputed. Note that it had originally been developed for an entirely different purpose (namely, as a test for inductive reasoning abilities). Furthermore, there are a variety of alternative explanations that may ac-

[1] Yet this may become more traceable considering that human sociality has at least a 25-to-35-million-year evolutionary history (Geary, 1998)—starting from the estimated age of the ancestor common to all extant primates—and taking into account that, for example, agriculture emerged not until 10.000 years ago (Cosmides & Tooby, 1997).

count for Cosmides' findings (e.g., Cheng & Holyoak, 1989; Fodor, 2000; Lawson, 2002; Liberman & Klar, 1996). The present work is concerned with a closer examination of Cosmides' theory of social exchange (Cosmides, 1989; Cosmides & Tooby, 1989, 1992). It is argued that if a cheater-detection mechanism as proposed by Cosmides and Tooby exists, then it should survive empirical tests that make use of diverse methodological approaches which are more closely related to situations of social exchange.

The first part of the present work provides a review of the theoretical framework of Cosmides' evolutionary theory of social exchange that constitutes the basis for the research that is later reported. First of all, some fundamental principles proposed within an evolutionary approach to psychology will be exemplified, since they build the scaffolding for the theoretical postulate of Cosmides and Tooby on social exchange (Cosmides, 1989; Cosmides & Tooby, 1989, 1992). Subsequently, a description of Cosmides' theory of social exchange (Cosmides & Tooby, 1989) will be provided, followed by an overview on the empirical evidence assumed to substantiate the propositions of the theory (Cosmides, 1989). Based on criticisms referring to the methodological approach of Cosmides—that is, Wason's selection task (Wason, 1966, 1968)—an alternative method for the analysis of hypotheses derived from Cosmides' theory of social exchange will be outlined. This approach, which is based on a face-recognition paradigm, has previously been applied in the context of only one single experiment (Mealey, Daood, & Krage, 1996). Finally, own research will be illustrated in detail in the second part of the present work.

2 An Evolutionary Approach to Psychology

When Charles Darwin published his theses on the origin of species in 1859 (Darwin, 1859/1998), he was already confident that his work would be of vital importance for the advancement of neighboring research disciplines. Although initially chary about the application of his ideas to humans, towards the end of the origin he made a bold prediction for the future of psychology (Darwin, 1859/1998, p. 647):

> In the future I see open fields for fare more important researches. Psychology will be securely based on the foundation […] of the necessary acquirement of each mental power and capacity by gradation. Much light will be thrown on the origin of man and his history.

In fact, more than 100 years later Darwin's fundamental assumptions on how natural selection shaped the process of evolution have been seized by a small group of psychologists and embedded into a new approach to psychological work: *evolutionary psychology*. Evolutionary psychologists share a notion of their research discipline as a branch of evolutionary biology (Cosmides & Tooby, 1997; Tooby & Cosmides, 1989), synthesizing knowledge from a variety of neighboring disciplines such as anthropology, paleontology, behavioral ecology, cognitive science, and neuroscience (Cosmides & Tooby, 1994, 2002). Basically, evolutionary psychology is concerned with the application of biological principles to the investigation of psychological phenomena, supposedly providing a new foundation for psychological research—at least from the point of view of its proponents (Cartwright, 2000; Tooby & Cosmides, 1989). According to its adherents, the primary research objective in evolutionary psychology must be regarded as the investigation of the development, structure, and functioning of the human mind (Cosmides & Tooby, 1997). Criticizing traditional approaches to the field of cognitive psychology, some evolutionary psychologists boldly raise the claim that a consistent application of evolutionary psychology can "repair many of the deficiencies that have hampered progress in the social sciences," and thus enhance the emergence of a "true social science" (Cosmides & Tooby, 1987, p. 278). However, in a sense their rationale appears somewhat overzealous, and one might criticize that the plausibility of some of the arguments they bring forward largely stems from their triviality.

The fundamental proposition held by evolutionary psychologists is that since humans were created in the process of evolution—just as all the other more or less sophisticated organisms—all inherent human characteristics must be regarded as products of the evolutionary process (Tooby, 1987). More precisely, it is assumed that "human minds, human behavior, human artifacts, and human culture are all biological phenomena—aspects of the phenotypes of humans and their relationships with one another" (Tooby & Cosmides, 1992, p. 20f). Based on this probably oversimplified assumption that we are all Darwinians (Symons, 1987), Cosmides and Tooby provided a broad theoretical framework for the study of mind and brain (Cosmides, 1989; Cosmides & Tooby, 1987, 1989, 1994, 1997, 2002; Tooby & Cosmides, 1989). The key principles underlying their "powerful manifesto on evolutionary psychology"—as Cartwright (2000, p. 193) terms it—are trivial in some respects and in few particulars even fail to meet accordance with the facts. Nevertheless, since Cosmides and Tooby attach great importance to these principles, they will be specified in the following.

2.1 Five Constitutive Principles of Evolutionary Psychology

Principle 1: The Human Mind Is What the Brain Does

Evolutionary psychologists view the human mind as an information-processing device operating analogous to a computer in that it is receiving inputs and generating outputs (Cartwright, 2000; Cosmides & Tooby, 1997). Before turning towards a closer examination of these inputs and outputs, the understanding of *mind* and *brain* as it prevails in the evolutionary-psychology literature needs to be clarified. As Cosmides and Tooby (1997) point out, for cognitive scientists the terms *brain* and *mind* refer to the same system. However, this system may be accounted for from two complementary viewpoints. While *brain* refers to the system's physical properties, the term *mind* implies the information-processing operations conducted by the system. In other words, as Cartwright (2000, p. 193) puts it, "[t]he human mind is what the brain does." Thus, the brain is conceptualized as a physical system that is composed of a multitude of organic devices, namely neurons, just as a computer is made up of silicon chips. Via electrochemical processes these neurons are provoked to fire, causing the transmission of information. Furthermore, it is assumed that neurons are connected to one another and organized in neural circuits that are similar to the circuits inherent in a computer. This is, of course, a simplification of the standard of knowledge. Disregarding the actual complexity of the human brain's structure and functionality, it is assumed that, just as a computer gains input from its user (e.g., via keyboard or mouse) our neural circuits are provided with information on the outer world and other parts of the body by a variety of receptors. Consequently, a plentitude of information is processed by our neural circuits and transmitted to the brain, from which they may be further processed to motor neurons that are connected to our muscles and cause them to move. This movement is the output of our computational system, namely behavior (Cosmides & Tooby, 1987, 1997)[2].

Thus, the basic argument within the evolutionary framework is that our neural circuits were designed to generate behavior responding to information, or demands, from the environment. This is per se a plausible assumption. Beyond it, however, it is argued that the challenge is not just to produce any kind of behavior, but behavior that is *appropriate*. Yet what are the characteristics of an appropriate behavioral response? As Cosmides and Tooby (1997) point out, appropriateness must be understood with respect to an organism's specific needs and environmental demands. To substantiate their postulation the authors illustrate their view by comparing behavior of humans and dung flies in the presence of feces in the environment. Unsurprisingly, appropriate behavior for the dung fly considerably differs from what may be appropriate for humans. To behave appropriately therefore requires a computational machine that analyzes an organism's needs and demands and enters accordant parameters into the decision-making (Cartwright, 2000). In the view of evolutionary psychologists, the information-processing machinery that is responsible for the mapping of informational input onto behavioral output is understood as a psychological mechanism. Accordingly, the psychology of an organism is defined as the entirety of proximate mechanisms that cause behavior (Cosmides & Tooby, 1987). Note that it is not the environment per se that determines what may or may not be appropriate behavior. Instead, Cosmides and Tooby (1997) argue that, as a matter of principle, any given environmental stimulus might be linked to any kind of behavior via suitably designed circuits. As will be outlined in the following section, the crucial aspect in defining a certain behavior as appropriate in terms of the evolutionary approach is the degree to which it is apt for solving so-called adaptive problems.

Principle 2: Human Problem-Solving Is Shaped by Natural Selection

As mentioned above, it is assumed that the neural circuits inherent in the human brain were designed for the purpose of producing, in a sense, appropriate behavior. Yet given

---

[2] As already mentioned, this is an oversimplified description of the human brain's functionality. However, it should be sufficient to give an impression of a fundamental viewpoint within evolutionary psychology, namely, considering the human mind and brain in terms of function. Note that, according to this view, both thought and behavior are cognitive processes (Cartwright, 2000).

this proposition, asking who or what actually accomplished this design seems justified, and, even more important, how it was done and why. Evolutionary psychology claims to provide answers to all of these questions. According to the Darwinian approach to psychology, there is only one force that is capable of designing machines as complex as the human brain, namely natural selection (Buss, 1999; Cosmides & Tooby, 1997). It is assumed that the neural circuits that constitute the human brain have evolved in the course of our species' evolutionary history to solve problems that our ancestors faced in their social and ecological environment. At first view, it seems plausible that "we have one set of circuits rather than another" because "the circuits that we have were better at solving problems that our ancestors faced during our species' evolutionary history than alternative circuits were" (Cosmides & Tooby, 1997); yet one may wonder about this manifest focus on our species' early past. It is further argued that, just as not any kind of behavior may be accepted as appropriate, not any kind of problem may count as relevant to the evolution of our information-processing devices. Evolutionary psychologists insist on the proposition that the only problems natural selection could design machines for solving are those that have an impact on reproductive rates (Cartwright, 2000; Cosmides & Tooby, 1997; Duchaine, Cosmides, & Tooby, 2001). The defining feature of such problems is that they are *adaptive*. Basically, adaptivity implies two things: first, adaptive problems are characterized in that they repeatedly occur in the course of a species' evolutionary history. Second, the effective solution of an adaptive problem is crucial with respect to an organism's successful reproduction (Cartwright, 2000; Cosmides & Tooby, 1997). As Cosmides and Tooby (1987) point out, adaptive problems may vary immensely with respect to their complexity. While some problems are easily solved and therefore require only simple cognitive mechanisms[3], most of the adaptive problems an organism may face are highly complex, and an even more complex machinery of cognitive mechanisms is necessary to solve these types of problems. To measure up to this complexity, Cosmides and Tooby rigorously demand the development of so-called *computational theories*, that is, task analyses that specify the nature of information-processing problems (Cosmides & Tooby, 1987, 1994; Tooby & Cosmides, 1989). Inspired by the pioneer work of Marr (Marr & Nishihara, 1978), the authors postulate that only by defining adaptive problems within the frame of computational theories, the cognitive programs to their solution may be adequately investigated. Their argument is

[3] As an example for an adaptive problem that is low in complexity and requires only simple information-processing mechanisms to be solved, Cosmides and Tooby refer to the problem of a newborn to gain sustenance from its mother. In this case, the authors point out, the cognitive programs "directly connect the perception of an environmental cue with an adaptively appropriate behavioral response" (Cosmides & Tooby, 1987, p. 286).

that since a computational theory specifies a given problem, it provides an explanation for why there is a device to solve it. Thus, a computational theory specifies the *function* of an information-processing mechanism (Cosmides & Tooby, 1994). Reflections on function are indispensable in the view of evolutionary psychology and are too often neglected, as Tooby and Cosmides (1989) note. Later in this chapter, the issue of computational theories will again be raised in the context of Cosmides' theory of social exchange (Cosmides & Tooby, 1989, 1992).

Principle 3: Consciousness Is Just the Tip of the Iceberg

With respect to the afore-mentioned complexity of adaptive problems, Cosmides and Tooby (1997) call attention to a dilemma. Alluding that "[c]onsciousness is just the tip of the iceberg", they argue that most of the processes that form the basis of our behavior must be conceived as unconscious or hidden from our awareness. Therefore, we may tend to misinterpret a variety of complex problems as easy to solve, since we are not aware of all the cognitive activities necessary to come up with these problems. As Cosmides and Tooby (1997) point out, what we actually become aware of are only "a few high level conclusions" that are, however, "passed by thousands and thousands of specialized mechanisms: some that are gathering sensory information from the world, others that are analyzing and evaluating that information, checking for inconsistencies, filling the blanks, figuring out what it all means." The authors argue that by investigating the human mind one may surely derive some hypotheses on its functionality from own conscious experience. Still, these intuitions may be misleading in that they may impart an impression of simplicity that is not warrantable.

Principle 4: The Modular Mind

The notion of the human brain as a complex formation of thousands of functionally specialized mechanisms providing for higher-level solutions (Buss, 1999) is probably the most disputed proposal of evolutionary psychology. The fundamental argument is equated with an engineering principle in that one specialized machine is rarely capable of solving two different problems equally well (Cosmides & Tooby, 1994, 1997). Following this, it is argued that "when the computational requirements of two tasks differ, one expects selection to have created a *different* computational system for accomplishing each—that is, two different functionally specialized adaptations" (Duchaine et al., 2001, p. 225). This argument draws some plausibility from the analogy with systemic organs (Buss, 1999; Cosmides & Tooby, 1994). Cosmides and Tooby argue that just as no one would assume the same organs to be responsible, for example, for breathing and digestion, "[t]he cognitive programs that govern how you choose a mate should differ from those that govern how you choose your dinner" (Cosmides & Tooby, 1994, p. 59). Yet this is, in a sense, prevarication; consider, for example, chemoreceptors that are evidentially involved in scent and taste as well as in mating. However, reverting to their above-outlined computer analogy, Cosmides and Tooby (1997) view the proposed specialized cognitive programs as "mini-computers," or modules, whose output must be functionally integrated to produce adaptive behavior. Based on their above-outlined assumptions, Cosmides and Tooby (1994) argue that such modules must be domain-specific in that the specialized design features that make them capable of successfully dealing with problems that arise in one domain (e.g., mating) make them inefficient in solving problems that come up in another (e.g., foraging). Moreover, the proposed modules are supposed to be content-dependent. Thus, within a specific domain (e.g., language acquisition) they are supposedly activated by different types of content (e.g., speech versus screams). The authors use the analogy of a Swiss army knife combining a large number of components designed for solving a variety of problems.

The supposition of domain-specific, content-dependent problem-solving circuits is at variance with another, more traditional notion of the human mind: that, rather than functionally specialized modules, there is a general-purpose learning process in the brain which is only for solving the problem of learning (Cosmides & Tooby, 1994). Later in this chapter, this position of content-general systems, as interpreted by adherents of evolutionary psychology, will be outlined in more detail.

Clearly, from the perspective of evolutionary psychology, the way in which the human brain solves the multitude of adaptive problems can by no means be through a handful of general problem-solving devices (Cosmides & Tooby, 1994; Duchaine et al., 2001; Geary, 1998). To substantiate their proposition of modularity, evolutionary psychologists frequently refer to the work of well-known cognitive scientists such as Fodor or Chomsky, who argued in terms of domain-specificity as well, yet not necessarily reverting to Darwinism. For example, Chomsky (1980), in the context of his studies on language acquisition, emphatically holds a view of the human mind as modular in character, composed of "a diversity of cognitive structures, each with its specific properties and principles" (Chomsky, 1980, p. 1). Referring to these structures as "mental organs, analogous to the heart or the visual system or the system of motor coordination and planning," Chomsky (1980, p. 3) points out:

> In short, there seems little reason to insist that the brain is unique in the biological world, in that it is unstructured and undifferentiated, developing on the basis of uniform principles of growth or learning—say those of some learning theory, or of some yet-to-be conceived general-purpose learning strategy—that are common to all domains.

Similarly, Fodor (1986, p. 1) argues in terms of *faculty psychology*, the view that "many fundamentally different kinds of psychological mechanisms must be postulated in order to explain the facts of mental life." Extending Chomsky's (1980) line of argumentation, he posits that some so-called *vertical faculties* may be regarded as modules that are "domain specific," "genetically determined," "associated with distinct neural structures," and "computationally independent" (Fodor, 1986, p. 21). He conceptualizes modules in terms of input systems, comprising highly specialized mechanisms[4]. For example, in the case of vision, he considers "mechanisms for color perception, for the analysis of shape, and for the analysis of three-dimensional spatial relations" as well as "task-specific 'higher-level' systems concerned with the visual guidance of bodily motions or with the recognition of faces of conspecifics" (Fodor, 1986, p. 47). Yet, as Symons (1987) concedes, Fodor's analysis of the mind is not informed by Darwinism. However, he argues, it may be hard to fault it on that ground, and therefore it may still be taken as substantiating the evolutionary approach.

Figuring the human mind as a puzzle made up of unique pieces, questioning the number of pieces seems justified. Obviously, one may easily assume multitudinous specialized modules cooperating for the purpose of adaptive behavior. However, as Cartwright (2000, p. 195) points out, "[a]t what level of discrimination and finesse have we reached an indivisible module?" In other words, a central problem along with the modular view of mind certainly is to know when to stop defining more and more subordinate modules. Geary (1998) provides an ostensive, albeit conjectured, hierarchical organization of domain-specific modules. He basically differentiates between social and ecological modules, which, as he posits, are likely to capture the core features of the evolved mind. As evident from Figure 2-1, he suggests that each of these superordinate modules may be subdivided into a number of submodules, which in turn are composed of even more specialized submodules[5].

Ecological modules, to begin with, are concerned with the processing of information derived from the biological and physical environment. The corresponding submodules are

[4] Fodor (1986) posits assumptions on a variety of features that these input systems share. For example he presumes that their operation is mandatory, and that there is only limited central access to mental representations which input systems compute. For a detailed illustration, see Fodor (1986, Part III).

[5] Note that Geary (1998) does not imply a one-to-one correspondence of modules and clearly defined brain regions. Instead, he suggests that specified cognitive competencies might be achieved by means of synchronized activities of localized ensembles of cells specialized for domain-specific information processing.

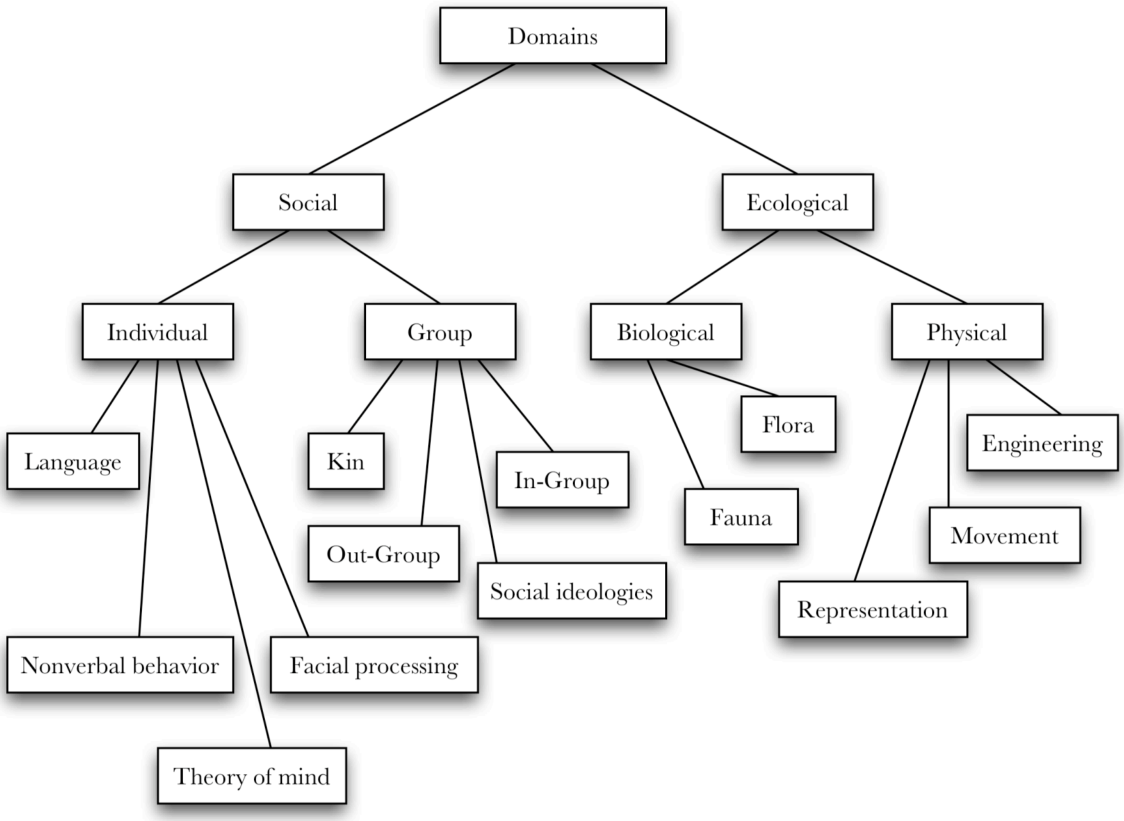

Figure 2-1: Some exemplary evolved domains of mind, adapted from Geary (1998, p. 181). Social modules are proposed to enable individuals to participate in social interactions in mutually beneficial ways. Ecological modules guide the human organism in exploring and exploiting the biological and physical environment.

thought to enhance an individual's survival and reproduction in that they enable the exploitation of biological and physical resources available in local habitats. According to Geary (1998, p. 189), biological modules are, at a general level, "designed to categorize and represent the flora and fauna in local ecologies." Physical modules, in addition, allow an individual to "behaviorally engage in the environment" (Geary, 1998, p. 189). This requires, for example, the development of mental representations of the environment's structure, the formation of memory for the relative location of specific objects, or the acquisition of knowledge on tool making and tool use.

The significance of the social modules is obvious if one considers that at all times humans have been living in social groups. As outlined above, the formation and maintenance of supportive interpersonal relationships may be regarded as being of crucial importance within our ancestors' evolutionary history. According to Geary (1998), individual-level social modules evolved for the purpose of regulating one-on-one social interactions with kin and friends. Plausibly, kin-based relationships are of considerable importance with respect to *inclusive fitness*. Inclusive fitness denotes an organism's overall genetic contribution to following generations—overall in that it refers to the number of replications of one's genes appearing in one's own offspring or in the offspring of kin (Cartwright, 2000; Geary, 1998). Natural selection is assumed to particularly favor those individuals who lend support to their kin and therefore aid their survival and reproduction (Geary, 1998). Postponing the controversy on whether or not the human mind is modularly organized, it seems plausible that instances concerned with verbal and nonverbal behavior—or, in other words, with communication—contribute to the successful engagement in social interactions. Likewise, Geary (1998) argues that a theory of mind is indispensable in that it provides a mental model of other individuals' intentions, beliefs, and emotional state from which predictions on their behavior may be drawn. This point of view may be qualified, however, taking into account that there are numerous other species who are known to socially cooperate yet lack the mental capacities to develop a theory of mind as defined above (consider, for example, insects such as bees and ants, but also mammals such as wolves or other carnivores engaging in cooperative hunting, let alone the diversity of symbioses appearing between various species). Yet confining himself to the human mind and its functionality, Geary (1998) further explicates that facial processing, in turn, essentially contributes to the formation of a theory of mind. For example, facial expressions may provide information due to which the mood of other individuals may be anticipated. Furthermore, face recognition is assumed to be crucial with respect to decisions on future interactions with others, as well as in terms of mate choice. Obviously, the proposed modules and submodules may considerably interact in that their informational output may function as input that activates other modules.

While individual-level modules are conceptualized with respect to one-on-one relationships, Geary (1998, p. 184) proposes almost complementary systems "designed to divide and organize the social universe." These group-level modules are thought to enable individuals to reciprocally engage in social interactions. Considering instances of kin-relationships and taking into account that the involved individuals are likely to differ with respect to their age, their knowledge, and their general skills and abilities, it is obvious that those relationships are not necessarily defined by reciprocity. However, long-term relationships with friends are reciprocal by definition (Geary, 1998; Trivers, 1971), just as are other interpersonal relationships established for the purpose of mutual support with respect to enhancing one's inclusive fitness. Thus, cognitive systems that regulate the understanding and accepting of social ideologies are just as important pre-conditions for reciprocal exchanges as modules concerned with in-group and out-group identification. Clearly there is some linkage to individual-level modules such as facial processing and theory of mind which contribute to the development of the afore-mentioned group-level modules. The issue of reciprocal exchanges will be further discussed later in this chapter, and special attention will be turned to violations of rules that govern social exchange.

Principle 5: Stone Age Minds In Modern Skulls

The evolution of the human brain and its inherent neural circuits has proceeded over an inconceivably long period of time. For almost all of this time our ancestors lived in hunter-gatherer societies (Cosmides & Tooby, 1987, 1994, 1997). Thus, their environment was completely different from our modern circumstances, and so were the problems our ancestors faced. Evolutionary psychologists boldly argue that, as a consequence, any cognitive mechanisms that evolved in the course of our species' evolutionary history were designed to solve problems resulting from Pleistocene environmental conditions. In other words, information-processing machines that nowadays account for our behavior are assumed to exist because they accomplished solving adaptive problems in the past. However, strictly speaking, this notion implies that since the Pleistocene age, no more evolutionary processes have taken place. Yet this is clearly contrary to Darwin's evolutionary theory. Still, Cosmides and Tooby (1987) argue that human psychological mechanisms should be adapted to those Pleistocene environments and not necessarily to the twentieth-century industrialized world. In a nutshell, what they postulate is that "our modern skulls house a stone age mind" (Cosmides & Tooby, 1997). At this point, it seems reasonable to question how we are able to deal with 21st century adaptive problems if our mental equipment is optimized for Pleistocene conditions. Could we actually have created our present-day world if Pleistocene mental modules had controlled us? Cosmides and Tooby (1997) argue that our ability to solve a variety of problems which considerably differ from those our ancestors faced—such as driving cars or solving complex mathematical equations—, must be regarded as a *by-product* of our evolved problem-solving circuits[6]. In the course of the present work, the vagueness of this argument will be further discussed.

[6] The authors comprehensibly illustrate their assumption with the following example (Cosmides & Tooby, 1997): "[…] when our ancestors became bipedal—when they started walking on two legs instead of four—they had to develop a very good sense of balance. And we have very intricate mechanisms in our inner ear that allow us to achieve our excellent sense of balance. Now the fact that we can balance well on two legs while moving means that we can do other things besides walk—it means we can skateboard or ride the waves on a surfboard." For a more specific discussion of adaptations and their concomitants, or by-products, see Tooby and Cosmides (1992, p. 55ff).

Résumé

In short, the basic principles of evolutionary psychology may be summarized as follows. First, the neural circuits inherent in our minds evolved for the purpose of generating adaptive behavior. Second, computational theories are needed to specify the nature of adaptive problems and draw some conclusions on the function of the problem-solving circuits engaged. Third, intuition oversimplifies processes at work and may be misleading. Fourth, the way in which the human brain accomplishes the issue of adaptive problem solving is through a set of domain-specific, content-dependent functionally specialized modules. Fifth, cognitive mechanisms optimized for Pleistocene conditions may not necessarily appear adaptive today.

Proponents of the evolutionary view of psychology bemoan that there has been only marginal use of these "Darwinian insights," preventing psychologists from the implementation of a fully Darwinian research program (Tooby & Cosmides, 1989). Thus, adherents of evolutionary psychology fault *traditional* psychology—a terminus that is used representative for cognitive psychology and behaviorism as well as clinical psychology and neighboring social sciences—for studying psychological phenomena in isolation (Solso, 2001). They argue that several factors may have contributed to a persistent reluctance against evolutionary psychology, most of which are explicable with respect to the history of psychology as a research discipline. For example, it is frequently argued on the part of evolutionary psychologists that in times affected by behaviorist dogmas calling for empirical observations rather than speculative theories, approaches such as the Darwinian were dismissed as meaningless and banned from laboratory science (for detailed historical reviews, see Cartwright, 2000; Fodor, 1986; Rozin & Schull, 1988). However, this detracts from the fact that behaviorists very well questioned their general-principle approach that "the same general principles of learning will be discovered regardless of what species of subject, what response, and what stimuli one chooses to study" (Mazur, 1998, p. 103). In fact, during the 1960s researchers began to report findings questioning the validity of the general-principle approach, indicating biological constraints on classical conditioning and other types of learning (Mazur, 1998). Moreover, proponents of evolutionary psychology find fault with the vagueness in which, supposedly, theories have been formulated and phenomena have been described within the field of psychology (Tooby, 1987). However, if at all, evolutionary psychology must be reproached with this same vagueness, since in its early beginnings it was characterized by an imprecise nature (Cartwright, 2000; Tooby, 1987). Even Tooby (1987) concedes that, initially, the crucial point of evolutionary theory (i.e., the nature of adaptations due to selection pressures) was barely elaborated. Finally, evolutionary psychologists such as Cosmides and Tooby take up the position that since the emergence of cognitive psychology, affected by the development of modern computer science and consequently capable of constructing more specified models of human information processing, the ground for the development of an integrated evolutionary psychology could be prepared. They criticize, however, that this progress of evolutionary psychology was hindered due to a "failure of the social sciences to explore or accept their logical connections to the rest of the body of science—that is, to causally locate their objects of study inside the larger network of scientific knowledge" (Tooby & Cosmides, 1992, p. 21). Obviously, the authors confuse the issue, lumping together various research disciplines as "the social sciences," disregarding, for example, that many psychologists draw on their discipline's self-conception as a natural science. The nineteenth-century scheme notwithstanding (Solso, 2001), Tooby and Cosmides insist on attacking what Cartwright (2000) calls the enemy of their manifesto, the *Standard Social Science Model*. In their opinion, the *Standard Social Science Model* has "served for a century as the intellectual framework for the organization of psychology and the social sciences and the intellectual justification for their claims of autonomy from the rest of science" (Tooby & Cosmides, 1992, p. 23). However, the authors allege a philosophy of science that is incommensurate with reality. To give an impression of what Tooby and Cosmides regard as the fundamental arguments and principles on which traditional psychological research has been supposedly based, their equivocal definition of the *Standard Social Science Model* will be outlined in the following.

2.2 The *Standard Social Science Model*

According to Cosmides and Tooby's (1997) notion of the *Standard Social Science Model*, the human mind is regarded as a *tabula rasa*, or as the authors depict it, a "blank slate, virtually free of content until written by the hand of experience", viewing all of the specific content inherent in the human mind as externally acquired (i.e., derived from an individual's social environment). This assumption reflects the authors' somewhat biased view of psychological research as a whole and of cognitive psychology in particular. However, it fits their argumentation in favor of evolutionary psychology.

The dubiety of Tooby and Cosmides' rationale becomes apparent considering their argument that "the only feature of the *Standard Social Science Model* that is correct as it stands" (Tooby & Cosmides, 1992, p. 23) implies that infants are uniformly equipped with the same developmental potential—a point of view that has been over the hill for a while. Of course, there have been controversies among psychologists with respect to this so-called nature-nurture debate, and some psychologists in fact held that infants were "blank tablets" upon which experiences of the outer world must be recorded. Yet the dominant view is that cognitive development is a matter of the interaction between built-in structures and the encouragement and demands of the physical and social environment (Solso, 2001). Furthermore, disregarding today's substantiated state of knowledge, Tooby and Cosmides (1992) argue that genetic variation cannot provide an explanation for any observed differences between racial groups[7].

Referring to Geertz (1973), who considered man, in physical terms, as an incomplete and unfinished animal, yet provided with the ability to learn, Tooby and Cosmides (1992, p. 26) conclude that a clear distinction is made within the *Standard Social Science Model* between two aspects of human mind and behavior:

> (1) the 'innate'(or inborn or genetically determined, etc.), which is supplied 'biologically' and is what you see in the infant, and (2) the social (or cultural or learned of acquired or environmental), which contains everything complexly organized and which is supplied by the social environment.

In this regard, the authors allege that it is further assumed by adherents of the *Standard Social Science Model* that, while an individual's mental organization is fully created by its social environment, the individual itself can by no means influence its social world (Tooby & Cosmides, 1992). Supposedly, within the *Standard Social Science Model* "the individual is the acted upon (the effect or the outcome) and the sociocultural world is the actor (the cause or the prior state that determines the subsequent state)" (Tooby & Cosmides, 1992, p. 26). However, Geertz, whom Tooby and Cosmides (1992, p. 25) provokingly term a "fully conventional modern adherent" of the *Standard Social Science Model*, provides a differentiated definition of culture and its role for human development (Geertz, 1973, p. 44):

> [...] culture is best seen not as complexes of concrete behavior patterns—customs, usages, traditions, habit clusters—as has, by and large, been the case up to now, but as a set of control mechanisms—plans, recipes, rules, instructions (what computer engineers call 'programs')—for the governing of behavior. [...] man is precisely the animal most desperately dependent upon such extragenetic, outside-the-skin control mechanisms, such cultural programs, for ordering his behavior.

[7] Note that they palliate their assertion, pointing out that this "depends on the existence of complex evolved psychological and physiological adaptations" (Tooby & Cosmides, 1992, p. 25).

Yet Cosmides and Tooby criticize the concept of culture remaining vague. Consequent from the above-outlined assumptions, which have supposedly been fundamental to socio-scientific research as Tooby and Cosmides (1992) insist, the role of psychology within the *Standard Social Science Model* is clearly defined, and the central concept in psychology must be regarded as learning. The authors strongly criticize the proposition that we may be endowed with "extremely general response capacities" (Geertz, 1973, p. 46) that, as Tooby and Cosmides (1992, p. 29) dispute, must be "constructed in such a way that they can absorb any kind of cultural message or environmental input equally well." Supporting the above-outlined principle of evolutionary psychology of domain-specific, functionally specialized mental modules, they challenge the conceptualization of the human mind as consisting of a small number of general purpose-mechanisms, working content-independent and domain-general and replacing genetically determined systems of behavior (Tooby & Cosmides, 1992). They strongly criticize the assumption that these general-purpose mechanisms by definition have neither a specific content, nor exhibit specialized features for the processing of particular content aspects. They conclude that, since no content information is processed by these mental mechanisms, our social environment implements any aspect of meaning, and any content of the human mind must therefore be regarded as social constructions (Cosmides & Tooby, 1992, 1997, 2002; Tooby & Cosmides, 1992).

Obviously, the above-described concept of the human mind in terms of few general problem-solving devices is not at all compatible with the principles proposed by evolutionary psychology. Cosmides and Tooby's drastic portrait of traditional cognitive psychology is notably contested. For example, Shapiro and Epstein (1998) reveal that indeed some, but by no means all, cognitive psychologists share the view of domain-general mental processes. The authors emphasize that even prominent cognitive scientists such as Chomsky (1980) or Fodor (1986) have been rigid advocates of domain-specific approaches to psychology. Thus, given that adherents to evolutionary psychology like to refer to their work as support for the evolutionary approach, it seems justified to argue in line with Shapiro and Epstein (1998, p. 174) that

> [i]t is for this reason ironic that evolutionary psychologists bemoan the 'traditional' cognitive psychologist's domain-general view of mind while at the same time applauding the Chomskian programme. If Chomsky is not central to the cognitive tradition, no one is.

Moreover, Shapiro and Epstein (1998) criticize Cosmides and Tooby's notion of cognitive psychology, turning towards a domain-specific direction (Cosmides & Tooby, 1987, 1989, 1992; Tooby & Cosmides, 1992). They point out that there is by no means consensus among cognitive psychologists with respect to the specificity of processes that form the basis of particular psychological abilities[8]. Yet as becomes clear from the debate between Shapiro/Epstein and Cosmides/Tooby (Shapiro & Epstein, 1998; Tooby & Cosmides, 1998), as well as from other disquisitions on the issue of evolutionary psychology, the most controversially discussed proposal of evolutionary psychology is the subject of domain-specificity. In fact, the notion of the human mind characterized by a plenteousness of functionally specialized mechanisms for any possible kind of problem seems far from economic. At least problems that are highly similar in content and structure should reasonably be handled by the same mechanism. Furthermore, simply to demonstrate that specific problems can be solved is not a sufficient proof of domain-specific mechanisms. Instead, general-purpose devices might be applied to a wide variety of problems. Before getting into a precipitate discussion of the manifold pros and cons of the above-outlined theoretical positions, the actual issue of the present work, Cosmides' theory of social exchange, needs to be introduced.

[8] The authors argue that this is true even for the area of language acquisition, which has "long been a stronghold for advocates of domain-specificity" (Shapiro & Epstein, 1998, p. 174).

3 An Evolutionary Theory of Social Exchange

Recapitulating the previous paragraphs, which conclusions can be drawn with respect to the validity of the evolutionary approach? On the one hand, the propositions made by Cosmides, Tooby, and other evolutionary psychologists are perspicuous. This may be due to the manner in which the authors tell their story—coherently, descriptively, and, above all, passionately. On the other hand, however, there are serious arguments militating not against the Darwinian view in principle, but against the proposition of domain-specific cognitive modules in particular. Keeping in mind the immense significance that is ascribed to evolutionary psychology by Cosmides and Tooby, especially with respect to revolutionizing cognitive psychology, there is clearly one thing missing to corroborate their point of view: empirical evidence. Yet this has been conceded by the authors themselves (Cosmides & Tooby, 1992, p. 163):

> Nonetheless, if such a view has a merit, it not only must be argued for on theoretical grounds—however compelling—but also must be substantiated by experimental evidence, as well as by converging lines of empirical support drawn from related fields such as neuroscience, linguistics, and anthropology.

For this reason, they conducted an experimental research program to explore evolutionary psychology's central hypothesis that, by means of natural selection, the human mind is comprised of domain-specific, content-dependent, and functionally specialized problem-solving devices. They concentrated on reasoning about social exchange for the following reasons (Cosmides & Tooby, 1992). First of all, the domain of social exchange is assumed to be of major significance with respect to the human species' evolutionary history. As outlined introductorily, social exchange for mutual benefit is considered an important factor with respect to the enhancement of *inclusive fitness*. Secondly, many theoretical aspects of social exchange have been explicated in detail and, as Cosmides and Tooby (1992) argue, in an unambiguous manner (see, for example, the work of Axelrod & Hamilton (1981) on the evolution of cooperation or the work of Trivers (1971) on reciprocation). Thus, in building their hypotheses on information-processing mechanisms involved in social exchange, the authors relied on certain features of the functional logic of social exchange (Cosmides & Tooby, 1992). Thirdly, the authors argue that theories about reasoning and rationality have been central issues to cognitive as well as social sciences. Consequently, they postulate that research on the issue of reasoning about social exchange might provide a basis for testing traditional postulates made within these scientific approaches.

Based on the work of Axelrod and Hamilton (1981), who analyzed strategies of social exchange within the scope of game theoretic models, Cosmides defined a number of constraints on which to build her theory of social exchange (Cosmides & Tooby, 1989). Referring to findings from studies with Prisoner's Dilemma games[9] (Axelrod & Hamilton, 1981), Cosmides and Tooby (1989) point out that social exchange in terms of mutual cooperation cannot emerge unless the following requirements are fulfilled (note that all these constraints may be attributed to one critical aspect of social exchange, namely the possibility to cheat). First, the probability of the cooperating parties meeting again must be sufficiently high. Considering the characteristics of our ancestors' ecological and social environment allowing for cooperation, one may think of few occasions in which cooperation might have taken place simultaneously—take, for example, cooperative hunting. However, in many instances, cooperation cannot take place concurrently and, as a consequence, reciprocation must occur belatedly (Cosmides, 1989). For example, assisting another individual in distress does not likely provide an opportunity for direct reciprocation, since the savior is unlikely to get into trouble shortly after and need to be saved in return (Cosmides & Tooby, 1989; Trivers, 1971). Yet this gives rise to the question why the rescued individual should bother to reciprocate at all. Even if the cost of a supportive act to the performer is, by definition, lower than the benefit to the recipient (Trivers, 1971), one may be tempted to cheat on paying this cost if one has already received the benefit. Thus, nearly any exchange that is not simultaneous enhances the possibility of defection. Second, none of the cooperating individuals should know the exact number of future encounters. It is plausible that an act of cheating will be perceived as innocuous from the viewpoint of the cheater if he knows that there will be no occasion for the deceived to get back at him. Finally, none of the involved should value later benefits by too much less than earlier benefits.

To cooperate under circumstances that allow for cheating will, as likely as not, turn out to be what Axelrod and Hamilton (1981) call an unstable strategy that will be selected out

[9] In the Prisoner's Dilemma game, two individuals strive for a certain payoff which may be defined in terms of the effect on each player's fitness in that they either cooperate or defect. It is a simple fact that the selfish choice of defection renders higher payoffs than cooperation. However, if both parties cheat, each of the players gets a lesser payoff than in the case of mutual cooperation. Hence, when two individuals are sure enough not to meet again, as is the case in a one-move Prisoner's Dilemma, defection turns out to be the most successful strategy to solve the game. Instead, in an iterated Prisoner's Dilemma, a system of incentives for cooperation and disincentives for defection can emerge (Axelrod & Hamilton, 1981).

quickly. Taking into account that any transgression of the above-mentioned constraints paves the way for acts of cheating, one may obviously be better off to carefully select one's interactants. This, in turn, requires solving a number of specific information processing problems. Cosmides (Cosmides & Tooby, 1989) incorporated some of what she thought were the most specific of these problems into a "grammar of social contracts," that is, a set of assumptions about which rules may govern a particular social exchange. In her opinion, these rules must be embodied by psychological mechanisms in the sense of Darwinian algorithms, or modules, as outlined above. Consequently, testable hypotheses on the computational outputs of such social-exchange algorithms may be derived from her grammar, thus providing a possibility to shed light on the nature and functioning of the human mind.

3.1 A Grammar of Social Contracts

As mentioned above, Cosmides defined her grammar of social contracts in terms of a set of rules governing social exchange. In this regard, the grammar is assumed to provide a defining framework from which one may understand the conditions that characterize well-formed social contracts and, in turn, the deviations of these conditions leading to ill-formed social contracts. This is done by specifying the cost-benefit relations that are thought to underlie any social exchange[10] (Cosmides & Tooby, 1989).

Within the cost-benefit terminology used by Cosmides (Cosmides & Tooby, 1989), a social contract is expressed as the formal rule "If *P* then *Q*"[11]. One basic presumption with respect to the grammar is that, at the time an offer of the type "If *P* then *Q*" is made, any of the potential interactants have a certain level of well-being, accompanied by certain expectations about the future, the so-called *zero level utility*. Based on this assumption, it is further assumed that one will accept an offer of the above-mentioned type only if this will raise one's utility level over zero. Since not-*P* and not-*Q* are part of one's zero level baseline[12], this requires that *Q* is perceived as a benefit. Moreover, since *P* is something one would possibly not do in the absence of the offer, *P* is likely to be perceived as a cost, thus decreasing the utility level below zero.

---

[10] Cosmides argues that social exchange could not have evolved if the items that our hominid ancestors valued had not been correlated with costs and benefits in their *inclusive fitness* (Cosmides & Tooby, 1989).

[11] As an example, Cosmides gives the contract offer: "If you walk my dog, then I'll give you a million dollars" (Cosmides & Tooby, 1989, p. 74). The term *P* refers to "you walk my dog", *Q* stands for "I'll give you a million dollars".

[12] It is assumed that in the absence of the offer, no expectations with respect to the non-occurrence of *P* and *Q* should exist, and therefore these incidents should not affect one's utility level (Cosmides & Tooby, 1989).

The acceptance of an offer does, in sum, only raise one's utility level above zero if the perceived benefit from *Q* exceeds the expected costs from *P*. Equivalently, from the contractor's point of view, *P* must increase the contractor's utility level, namely to a lesser degree than *Q* may decrease it, to make him perceive the offer worthwhile. That is, the same offer does address different value systems, since contractor and acceptor variably perceive costs and benefits. A well-formed social contract is defined in that both interactants perceive the received benefits higher than the costs they paid, that is, both benefit from the interaction. In contrast, what are the defining features of an ill-formed social contract? Clearly, this denotes circumstances under which at least one of the interactants cheats on the other by purposely failing to fulfill the requirements set within the contract (Cosmides & Tooby, 1989).

3.2 Detecting Cheaters

The game-theoretic structure of the natural selection process (Axelrod & Hamilton, 1981) prescribes that social exchange in terms of cooperation for mutual benefit can evolve only if it is governed by strategies of reciprocation that rule out cheating (Cosmides & Tooby, 1989). In the context of Cosmides' theoretical framework, cheating is defined as a violation of the conditions of a social contract. That is, cheating refers to the failure to pay a cost to which one has committed by accepting a related benefit. In this regard, it does not make any difference whether a social contract is explicitly or implicitly concluded, a private agreement or a social rule. Table 3-1 illustrates the possible outcomes of a social contract in case the contract is sincerely fulfilled compared to circumstances under which one of the interactants cheats on the other. Note that Cosmides differentiates between two structural components of a social contract. The *surface structure* refers to the way in which the offer is actually made. The *deep structure*, however, corresponds to the cost-benefit description of the *surface structure* from the viewpoint of both interactants.

The central proposition within *Social Contract Theory* (Cosmides, 1989; Cosmides & Tooby, 1989) alludes to information processing algorithms, or modules, that are functionally specialized for the detection of cheaters. Such modules are assumed to involve various procedures that enable an individual to draw inferences on whether someone has cheated in prior exchanges or is about to cheat in future interactions. Yet what does this denote with respect to the grammar of social contracts? As outlined above, any social contract may be evaluated from two different viewpoints which imply different value systems. Thus, the

Table 3-1

The deep structure of a social-contract offer phrased „If you give me *P*, then I'll give you *Q*", adopted from Cosmides and Tooby (1989, p. 85). *B* stands for benefit, *C* represents costs, zero is equivalent to the baseline of well-being by the time the offer is proposed. *Me* refers to the contractor of the offer, *you* stands for the acceptor.

	I cheat you		You cheat me		Contract fulfilled	
You give me *P*	B_{me}	C_{you}	---	---	B_{me}	C_{you}
You do not give me *P*	---	---	0_{me}	0_{you}	---	---
I give you *Q*	---	---	C_{me}	B_{you}	C_{me}	B_{you}
I do not give you *Q*	0_{me}	0_{you}	---	---	---	---
My payoff	B_{me}		C_{me}		B_{me}- C_{me}	
Your payoff	C_{you}		B_{you}		B_{you}- C_{you}	

detection of cheating requires the computation of two different descriptions of each item that is part of the social contract (that is, *P* and *Q*). In a sincerely fulfilled social contract, *P* should be described as a cost for the acceptor (C_{you} in Table 3-1), and a benefit for the contractor (B_{me} in Table 3-1). In turn, *Q* should be described as a benefit for the acceptor (B_{you} in Table 3-1), and a cost for the contractor (C_{me} in Table 3-1). As Cosmides argues, inference procedures designed for the purpose of cheater detection must operate on cost-benefit descriptions of social contracts from the potential cheater's viewpoint (Cosmides & Tooby, 1989).

Given that a transaction defined by a social contract has not yet been completed, these "look-for-cheaters" procedures should lead to one of the following inferences:

(1) Ignore individual *X* in case *X* has *not* accepted the benefit (B_X).

(2) Ignore individual *X* in case *X* has paid the cost (C_X).

(3) Watch out for individual *X* in case *X* has accepted the benefit (B_X).

(4) Watch out for individual *X* in case *X* has *not* paid the cost (C_X).

Obviously, in situations (1) and (2), individual *X* cannot have cheated, since paying the cost and passing on the benefit is just to the disadvantage of individual *X* himself. However, situations (3) and (4) hold the possibility to cheat, since individual *X* might not pay the cost although having accepted the benefit (3) or might take the benefit although not having paid the cost (4).

As Cosmides posits, the innate algorithms assumed to operate on social exchange perform these computations not only reliably but also automatically (Cosmides & Tooby, 1989). This, in turn, insulates us from insight in the complex structure underlying any social exchange (an assumption that is in line with one of the basic principles of evolutionary psychology). Following this rationale, the ease with which people engage in social exchange—the complexity of the underlying information-processing problems notwithstanding—is taken as evidence for the operation of Darwinian algorithms. Still, to elaborately test for the functioning of the hypothesized modules specialized for social exchange in general, and the detection of cheaters in particular, Cosmides (1989) conducted a series of experiments that will be discussed in the following.

3.3 Empirical Evidence: Studies With the *Wason Selection Task*

As mentioned above, Cosmides based her studies on the so-called *Wason Selection Task* (Wason, 1966, 1968). Before her findings will be elucidated, the standard abstract form of the *Wason Selection Task* will be outlined and findings concerning the performance on variant modifications of the standard task will be summarized, since they built the basis for Cosmides' (1989) research.

3.3.1 The Standard *Wason Selection Task*

The *Wason Selection Task*, which was first proposed by Wason in the 1960s (Wason, 1966, 1968), was designed as a test for logical reasoning in which subjects are required to decide whether or not a given rule is violated. In the standard abstract version of the task, subjects are presented a formal rule of the form "If *P*, then *Q*," and four cards holding instances of, respectively, *P*, *Q*, not-*P*, and not-*Q* on their front. Furthermore, subjects are instructed that instances of either *P*, *Q*, not-*P*, and not-*Q* are printed on the back side of each card as well, and that they are to turn as few cards as possible to determine whether the cards fulfill the given rule, namely "If there is *P* on the front of any card, then there is *Q* on its back." Figure 3-1 illustrates an abstract *Wason Selection Task* as used by Wason (1968). In his Experiment 1, subjects were presented the conditional rule: "If there is a D on one side of any

The rule: "If there is a D on one side of any card, then there is a 3 on its other side."

The task: Cards with letters on their front have numbers on their back and vice versa. Select all those cards—but only those—that have to be turned over so as to determine whether the rule is violated.

D	3	B	7
(*P*)	(*Q*)	(not-*P*)	(not-*Q*)

Figure 3-1: The abstract form of the *Wason Selection Task*, adapted from Wason (1968, Exp. 1). Note that the logical categories (*P* and *Q*) marked on the cards are there only for the reader's benefit but never appear on problems given to subjects.

card, then there is a 3 on its other side," with D representing the logical antecedent *P*, and 3 standing for the logical consequence *Q*, and four cards showing the following letters and numbers on their front and (in brackets) on their back side: D (3), 3 (K), B (5), and 7 (D). With respect to the propositional calculus, there is only one correct conclusion to be drawn in testing for a violation of the rule, namely selecting D (that is, *P*) and 7 (that is, not-*Q*). This results from the fact that for a rule of the form "If *P*, then *Q*", there exist three combinations of antecedent (*P*) and consequent (*Q*) that may be counted as true with respect to the rule: (*P*, *Q*), (not-*P*, *Q*), and (not-*P*, not-*Q*). In contrast, there is only this one combination (*P*, not-*Q*) which breaks the rule, and therefore "it follows that only values of P and values of $\overline{\mathrm{Q}}$ allow for a valid inference. […] $\overline{\mathrm{P}}$ comes out true whether it is associated with Q or $\overline{\mathrm{Q}}$, and Q comes out true whether it is associated with P or $\overline{\mathrm{P}}$" (Wason, 1968, p. 273)[13].

However, one result reliably found throughout a capacious number of studies is that subjects show enormous difficulties in making this particular type of inference mentioned above. Generally, subjects tend to make two errors, namely failing to select the not-*Q* card, which could falsify the rule, and selecting the *Q* card instead, which could not falsify the rule (Wason & Johnson-Laird, 1972). Accordingly, it is frequently reported that less than

[13] In his original work, Wason (1968) used to refer to the conditional terms not-*P* and not-*Q*, by writing $\overline{P}$ and $\overline{Q}$, respectively.

10% of the subjects produce the correct solution in the standard abstract version of the task (Bracewell & Hidi, 1974; Cox & Griggs, 1982; Gilhooly & Falconer, 1974; Griggs & Cox, 1983; Johnson-Laird, Legrenzi, & Legrenzi, 1972; Lunzer, Harrison, & Davey, 1972; van Duyne, 1974; Wason, 1968, 1969; Wason & Shapiro, 1971). For example, Wason (1968, Exp. 1) found that 10 out of 16 participants (62.5 per cent) chose (*P*, *Q*), whereas only one person (0.06 per cent) made the correct choice (*P*, not-*Q*) in the standard abstract task[14]. In another study conducted by Wason, none of the 32 participants made the correct decision, whereas 13 subjects (40.6 per cent) chose (*P*, *Q*)[15] (Wason, 1969).

Within the last decades, a multitude of research has been done on what might influence performance on the *Wason Selection Task*, and a number of attempts have been made to find an explanation for the above-mentioned phenomenon. Wason (1966, 1968) interpreted the fallacy of choosing the affirmation of the consequent instead of its denial as the result of a confirmation bias whereat "the need to establish the 'truth' of the statement predominates over the instruction" (Wason, 1966, p. 147). He attributed this bias to the observation that "individuals are biased, through a long learning process, to expect a relation of truth, correspondence or match to hold between sentences and states of affairs," and that "in adult experience truth is encountered more frequently than falsity, and we seldom use a proposition or judgment that something is false in order to make a deduction" (Wason, 1968, p. 274).

An alternative explanation was proposed by Evans (1972a, 1972b; Evans & Lynch, 1973; Manktelow & Evans, 1979) who pointed out that "subjects were responding in a manner consistent with a 'matching bias': preferentially selecting those cards which are mentioned in the rule, i.e. the p and q cards" (Manktelow & Evans, 1979, p. 477). There is empirical evidence compatible with both approaches (see, for example, Reich & Ruth, 1982; Yachanin & Tweney, 1982). However, a more detailed examination would go beyond the scope of this work. Instead, a closer look shall be taken on a well-investigated, yet controversially discussed phenomenon that was a determining factor for Cosmides' research (Cosmides, 1989)—the so-called *thematic-materials effect* or *content effect* in the *Wason Selection Task*.

[14] Note that these data refer to so-called initial choices made by a control group which performed the standard task as compared to an experimental group which performed a slightly modified task.
[15] These data also refer to initial choices in a modified task as compared to the standard abstract task.

3.3.2 The *Thematic-Materials Effect* in the *Wason Selection Task*

The tendency to draw the wrong conclusions in the *Wason Selection Task* has been proven to be resistant to a number of interventions, such as projecting falsities (Wason, 1968), making the task strictly binary and introducing contradictions to sensitize subjects for the propositional calculus (Wason, 1969), inducing former experience with the problem's logical structure (Lunzer et al., 1972; Wason & Shapiro, 1971), or giving visual aids (Goodwin & Wason, 1972). However, a variety of studies conducted in the early 1970s suggested that the performance in the *Wason Selection Task* may be drastically improved by using natural as compared to abstract materials (Bracewell & Hidi, 1974; Gilhooly & Falconer, 1974; Johnson-Laird et al., 1972; Lunzer et al., 1972; van Duyne, 1974; Wason & Shapiro, 1971). For example, Wason and Shapiro (1971) compared the standard abstract form of the *Wason Selection Task* [16] with another, so-called thematic form, in which they presented the following rule: "Every time I go to Manchester, I travel by car" together with four cards showing different towns (Manchester, Leeds) and modes of transportation (car, train). They found that 10 out of 16 subjects (62.5 per cent) dealing with the thematic material produced the correct answer (i.e., they chose Manchester and train), whereas only 2 persons (12.5 per cent) chose the correct cards in the abstract task (D and 7, respectively). Referring to findings of an enhanced memory for concrete as compared to abstract material, the authors argued that the terms used in the concrete version of the task—namely, names of towns and modes of transportation—had been much more comprehendible than abstract terms such as letters and numbers and therefore must have been easier to deal with. Furthermore, they pointed out that not just the concreteness of the terms per se, but the concreteness of the relations between the terms might have considerably influenced the subjects' performance. Obviously, it seems easier to link names of towns to different modes of transportation thinking of traveling, than to relate letters to numbers by means of artificial relations as "the other side of the card." Finally, the authors referred to the higher coherence of concrete as compared to abstract rules.

In another often cited and just as often disputed study, Johnson-Laird et al. (1972) instructed their subjects to imagine themselves being postal workers who were to investigate whether some letters had been sufficiently stamped. Therefore, they presented the rule: "If a letter is sealed, then it has a 50 lire stamp on it" together with five envelopes, which are

[16] In line with Wason (1968, Exp. 1), the rule "If there is a D on one side of any card, then there is a 3 on its other side" was presented together with the following four cards: D (3), 3 (K), B (5), and 7 (D). Items in brackets refer to what was printed on the back side of the cards.

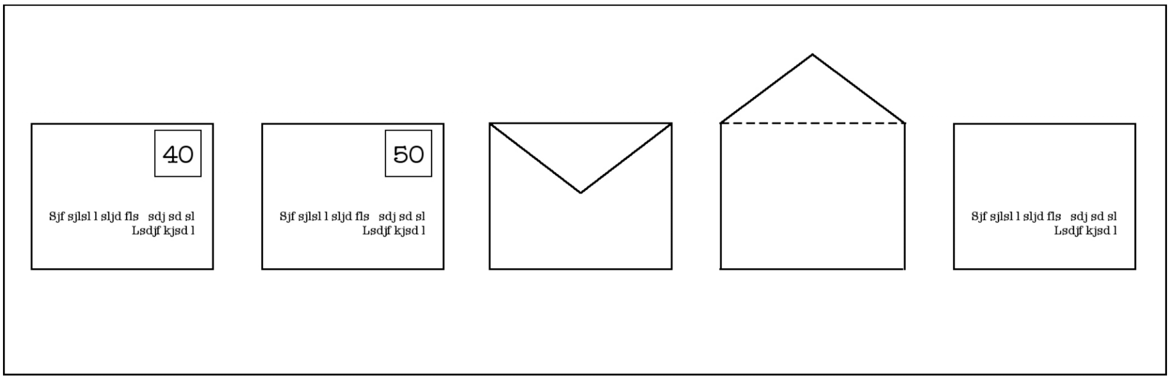

Figure 3-2: The envelopes used in the letter-stamp task, adapted from Johnson-Laird et al. (1972). Differing from other experiments, five instead of four instances of the conditional rule "If a letter is sealed, then it has a 50 lire stamp on it" were presented. Obviously, the third envelope from the left was sealed, just as the envelope right next to it was unsealed. For the other envelopes, one could impossibly know whether or not they were sealed.

illustrated in Figure 3-2. Working as his own control, each subject additionally handled a second abstract task with the rule "If a letter has an A on one side, then it has a 3 on the other side", and four cards showing A, B, 3, and 2. What was found was that 87.5 per cent of the subjects generated the correct solution with respect to the concrete rule, whereas only 8.3 per cent were able to solve the abstract problem. The authors concluded that the realistic nature of the concrete task had led to considerable insight in the problem structure and thus to an obviously enhanced performance. As an interesting side effect, Johnson-Laird et al. (1972) reported that in spite of the within-subject design realized in their experiment, no transfer effects could be found in the way that subjects performed better on the abstract task after having solved the concrete task. Thus, the authors reasoned that "it is the content of the problems which is crucial rather than their structural identity" (Johnson-Laird et al., 1972, p. 399)[17]. Furthermore, they pointed out that with thematic material it might be easier to detect the reversibility of the terms mentioned in the rule, that is, "that the cards are identical apart from their orientation" (Johnson-Laird et al., 1972, p. 399), an aspect subjects seemingly were not aware of when dealing with the abstract version of the task. Finally, just as Wason and Shapiro (1971), the authors emphasized the realistic relationship between the terms in the concrete rule as the crucial factor for the increased performance in solving these rules.

At first view, the explanations offered by Wason and Shapiro (1971) as well as Johnson-Laird et al. (1972) seem plausible. However, others have raised objections, pointing out that

[17] Note that this lack of transfer was repeatedly replicated (see, for example, Griggs & Cox, 1982; Lunzer et al., 1972).

a serious confounding of the concreteness of the terms of the rule per se and the concreteness of the relationship of the terms may have been causative to the content effects found in both of the above-mentioned studies (see, for example, Bracewell & Hidi, 1974; Gilhooly & Falconer, 1974). Furthermore, Bracewell and Hidi (1974) found that the enhanced performance in the concrete version of the *Wason Selection Task* also varied as a function of the order of presentation of *P* and *Q*, in that the thematic-materials effects was only found for rules in which *P* preceded *Q* but not vice versa.

However, though consensus was not reached as for an explanation of the thematic-materials effect, it still appeared to be reliably found within numerous different studies and experimental designs. Still, there have also been a number of cases in which no such effect emerged (C. A. Brown, Keats, Keats, & Seggie, 1980; Cox & Griggs, 1982; Griggs & Cox, 1982, 1983; Manktelow & Evans, 1979; Pollard, 1981; Yachanin & Tweney, 1982). For example, in a series of five experiments conducted by Manktelow and Evans (1979), no performance enhancement could be found with thematic materials, not even in their Experiment 5 that was designed as an exact replication of the Wason and Shapiro (1971) study. Likewise, Griggs and Cox (1982) failed to replicate the findings both of Wason and Shapiro (1971) and of Johnson-Laird et al. (1972). However, arguing that “the impressive results of Johnson-Laird et al. (1972) may have been due to long-term memory cues” with the falsifying instance being “available immediately from a subject’s past experience” (Griggs & Cox, 1982, p. 414), they designed their third experiment close to the letter-stamp paradigm by Johnson-Laird et al. (1972) while using an implication rule closer to their subjects’ specific experience[18]. According to a law in the state of Florida with regard to the legal age for drinking alcoholic beverages which they assumed their subjects might be quite familiar with, they presented the conditional rule “If a person is drinking beer, then the person must be over 19 years of age” together with four cards labeled “drinking a beer,” “drinking a coke,” “16 years of age,” and “22 years of age.” Subjects’ performance on this so-called drinking-age problem was compared to the outcomes found for an abstract version of the task[19]. Contrary to their first experiments, the authors found a considerable thematic-materials effect in that 73 per cent of the subjects generated the correct solution for

[18] The authors pointed out that the letter-stamp problem applied in the study of Johnson-Laird et al. (1972) enabled their British subjects to benefit from their own experience, since they were to deal with postal regulations concerning the amount of postage in everyday life. However, no such rule existed in the United States, therefore “such thematic material would not cue the falsifying instance in our [American] subjects’ memories”(Griggs & Cox, 1982, p. 414).

[19] Following Johnson-Laird et al. (1972), the abstract rule was “If a card has an «A» on one side, then it has a »3« on the other side,” and the response cards were labeled “A,” “B,” “2,” and “3,” respectively.

for the drinking-age problem whereas no one solved the abstract task. They pointed out that this finding was fully in line with their memory-cueing hypothesis, and that "[w]ith the drinking-age problem, American subjects were given a problem that cued both knowledge of a rule and its counter-examples and the detective-set strategy" (Griggs & Cox, 1982, p. 418), that is, "searching for a violator of the rule" (Griggs & Cox, 1982, p. 417).

At this point, it seems appropriate to sum up the issue of content effects on the *Wason Selection Task* and, even more important, to revert to the actual matter of interest, namely empirical evidence for Cosmides' (Cosmides & Tooby, 1989) *Social Contract Theory*. Taken together, the findings concerning the thematic-materials effect are highly inconsistent, as well as the numerous rival explanatory approaches to give consideration to what exactly induces or hinders this characteristic performance enhancement—e.g., task understanding (Wason & Shapiro, 1971), concreteness of thematic terms (Gilhooly & Falconer, 1974), goal-means relationships (Bracewell & Hidi, 1974) or memory cueing (Griggs & Cox, 1982), to name just a few. With regard to this state of affairs, Tooby (1987, p. 72) argues:

> Research on so-called content effects in logical reasoning has been bogged down in a quagmire of conflicting results and interpretations, and none of the prevailing hypotheses have demonstrated any predictive power. Cosmides (1985)[20] has productively reorganized this confused literature through the application of the evolutionary approach. The content effects become very orderly when they are scrutinized for the presence of evolutionary significant content themes.

Thus, in line with her rationale that the human mind is comprised of functionally specialized information processing modules and referring to her theory of social exchange, Cosmides (1989) claims that there is indeed a common denominator to all of the findings concerning the thematic-materials effect: an innate looking-for-cheaters procedure. Reviewing a multitude of studies concerned with the thematic-materials effect, Cosmides detected an interesting coherence (Cosmides, 1989, p. 199-200):

> Robust and replicable content effects are found only for rules that relate terms that are recognizable as benefits and costs in the format of a standard social contract. No thematic rule that is not a social contract (e.g., rules about food[21], transportation or school[22]) has ever produced a content effect that is both robust and replicable. For thematic content areas that do not express social contracts, either no content effect is found (e.g., food prob-

---

[20] Doctoral dissertation, cited from Tooby (1987).

[21] Note that in their first four experiments, Manktelow and Evans (1979) used a number of conditional rules referring to food and beverages such as "If I eat haddock, then I drink gin" together with four cards showing pictures of what the experimenter allegedly ate and drank at separate meals.

[22] Note that van Duyne (1974) presented thematic rules relating academic subjects and universities such as "If a student studies philosophy, then he is at Cambridge" together with four cards supposedly taken from a students' register and indicating different departments and places of study, respectively.

> lems), or there are at least as many studies that do not find content effects as there are studies that do (e.g., transportation and school problems).

At first sight, and with respect to the above-mentioned sample of studies, this statement seems plausible and justified. For example, substantial effects of thematic material were found in studies using the drinking-age problem (Cox & Griggs, 1982; Griggs & Cox, 1982, 1983), demanding that the subjects select persons who cheated against a national law, namely by drinking alcohol although being younger than 19. According to Cosmides, the drinking-age problem clearly has the cost-benefit structure of a social contract, since it may be translated as "If you take the benefit, then you pay the cost" (Cosmides, 1989; Cosmides & Tooby, 1989). The benefit in that case is the permission to drink alcohol, whereas the cost to be paid is to fulfill the requirement of having a certain age. Similarly, in the letter-stamp task that yielded considerable content effects (Johnson-Laird et al., 1972), this cost-benefit structure is given in that a letter is delivered only in case the cost, that is, the correct amount of postage, has been paid.

According to Cosmides, with this sort of conditional rules, subjects behaved as if they were looking for cheaters, persons who accepted the benefit without paying the cost (Cosmides & Tooby, 1989). To substantiate her assumptions she designed a number of experiments in which she compared what she calls social-contract rules to a variety of other thematic versions of the *Wason Selection Task*. This work will be outlined in the following paragraph.

3.3.3 A Logic of Social Exchange?

Why should a looking-for-cheaters strategy as proposed by Cosmides (Cosmides & Tooby, 1989) enhance performance on conditional reasoning? As she argues, the crucial factor is *coincidence*. In her opinion, it is *not* that subjects actually *reason* more logically when solving social-contract rules. It is rather that for standard social-contract rules as illustrated in Figure 3-3, the "Benefit Accepted" card *happens* to correspond to the logical antecedent *P*, just as the "Cost Not Paid" card comes up to the logical consequent not-*Q*. Therefore, Cosmides (1989) hypothesizes that whenever a subject is looking for a cheater, he will accidentally choose the two cards that are in accordance with the logically correct response.

In her first two experiments, Cosmides (1989) compared subjects' performance on thematic rules which were expressed either in the form of a social contract or not (to these problems, the author refers to as *descriptive*). Both problems were designed as unfamiliar in content to test *Social Contract Theory* against a rival explanatory approach, namely *availability theories*.

The rule: "If you take the benefit, then you pay the cost."

The task: The cards below have information on four people. Each card represents one person. One side of a card tells whether a person accepted the benefit, and the other side of the card tells whether that person paid the cost. Indicate those card(s) you definitely need to turn over to see if any of these people are breaking this law.

Benefit Accepted	Cost Paid	Benefit NOT Accepted	Cost NOT Paid
(*P*)	(*Q*)	(not-*P*)	(not-*Q*)

Figure 3-3: The cost-benefit structure of a standard social-contract rule, adapted from Cosmides (1989). According to Cosmides, a looking-for-cheaters procedure should lead subjects to choose only those cards that indicate potential cheaters and in return, ignore any other card that characterizes a person who could not possibly have cheated. Thus, unaffected by formal logic subjects should select the "Cost Not Paid" card and the "Benefit Accepted" card. Note that the logical categories (*P* and *Q*) marked on the cards are there only for the reader's benefit but never appear on problems given to subjects.

This class of explanatory approaches derived from Tversky and Kahnemann's (1973) *availability heuristic* and provides explanations that, albeit differing within their formulations, attribute thematic-materials effects to the amount of experience that subjects may have had with a given problem content (see, for example, Gigerenzer & Hug, 1992; Manktelow & Evans, 1979; Pollard, 1982).

As opposed to *Social Contract Theory*, availability theories may be applied to any problem context since they postulate that subjects' reasoning on thematic tasks is not influenced by the problem content per se. Instead, it is exclusively the familiarity with a content that is regarded as the determining factor. Additionally, subjects were to deal with a standard abstract version of the *Wason Selection Task* as well as a familiar descriptive problem[23]. All of the rules were embedded in a short cover story phrased so as to activate a "detective set" (Cosmides, 1989)[24]. The unfamiliar rules (both social-contract and descriptive, varying only with respect to the surrounding context story[25]) were: "If a man eats cassava root, then he must have a tattoo on his face," and "If you eat duiker meat, then you have found an os-

[23] Performance on the abstract task was thought to somehow represent a baseline since abstract problems are commonly used as a standard for assessing availability (Cosmides, 1989).
[24] For a full outline of the materials used, please refer to the extensive appendix of Cosmides (1989).
[25] Note that social-contract and descriptive versions of each rule were counterbalanced across subjects.

trich eggshell." In the context story that implied a social contract to the "cassava-root problem", subjects were cued into the role of a member of the Kaluame (Polynesian culture providing for a highly unfamiliar problem context) who is charged with the enforcement of a series of laws. One of these laws is concerned with the privilege of eating the cassava root, a powerful aphrodisiac which is only given to married men. It is an accepted custom that a man gets a tattoo on his face when he marries. Thus, having a tattoo on his face normally should indicate a man's marriage and therefore, his authorization to receive the cassava root. However, some men might cheat on the law, since the cassava root is greatly in demand. Subjects were instructed to see if any of the Kaluame men were breaking the law by turning over as few of the following cards as possible: "eats cassava root," "no tattoo," "eats molo nuts," "tattoo." As for the descriptive rule, subjects were cued into the perspective of an anthropologist who—in the case of the "duiker-meat problem"—studied a hunter-gatherer culture in the African desert and wanted to figure out whether a common saying (namely, the "duiker-meat rule") was metaphorical or reflecting a real relationship. Just as for the social-contract rule, subjects were instructed to look for violations of the rule by choosing from the following cards: "eats some duiker meat," "has never found an ostrich shell," "does not eat any duiker meat," "has found an ostrich shell." As a familiar but descriptive (non-social-contract) rule, a transportation problem known from Wason and Shapiro (1971) was used, since "the transportation problem had been the most successful non-social-contract problem in the literature" (Cosmides, 1989, p. 211). Finally, the abstract version of the task was closely adapted from Wason (1968) yet embedded in a context story as were the other rules.

Table 3-2 illustrates the predictions made with respect to *Social Contract Theory* and *availability theories* and the actual results[26]. Cosmides (1989) interpreted the results as clearly in line with her assumption that content effects in the *Wason Selection Task* are due to a functionally specialized cheater-detection mechanism. In fact, in both experiments the highest percentage of logically correct responses was found for social-contract problems. However, as evident from Table 3-2, the familiar descriptive problem also elicited response rates that were about twice as high as for unfamiliar descriptive and abstract problems, which was contrary to Cosmides' prediction. Accordingly, Gigerenzer and Hug (1992), who slightly varied the experimental design proposed by Cosmides (1989), noted that the findings obtained by Cosmides (1989) as well as their own results did not imply that *Social Contract Theory* does

[26] Note that Experiment 2 was designed as a replication of the first experiment except for the slight variation that private exchange (rather than social laws) was indicated by the given rules.

Table 3-2

Predictions and results obtained for Cosmides' (1989) Experiments 1 and 2. Both predictions and results refer to the percentage of (logically correct) *P* & not-*Q* responses. *Social Contract Theory* predicts a high performance level only for the social-contract rule, independent of familiarity aspects. Availability theories predict performance enhancement exclusively for familiar problem content.

	Predictions		Results	
	Social Contract	Availability	Exp. 1	Exp. 2
Unfamiliar social contract (cassava-root problem)	High	Low	75%	71%
Unfamiliar descriptive (duiker-meat problem)	Low	Low	21%	25%
Abstract problem	Low	Low	25%	29%
Familiar descriptive (transportation problem)	Low	Middling to low	46%	38%

better than availability for *familiar* rules. Conceding this limitation to Cosmides' rationale, the authors still lent support to her work, reporting even higher rates of logically correct *P* & not-*Q* choices (94%) for unfamiliar social-contract rules.

In another two experiments that followed the same design and procedure as the above-mentioned, Cosmides (1989) compared subjects' performance on social-contract rules and non-social-contract rules involving what she called a social-purpose permission[27]. These experiments were conducted to test *Social Contract Theory* against another competing approach, *Permission Schema Theory* (Cheng & Holyoak, 1985, 1989; Cheng, Holyoak, Nisbett, & Oliver, 1986). Table 3-3 illustrates the relevant results. According to *Permission Schema Theory*, the key to any content effects on the *Wason Selection Task* is that the content of a conditional rule functions as a cue for so-called *pragmatic reasoning schemas*, abstract knowledge structures that are induced from ordinary life experiences such as permissions or obligations (Cheng & Holyoak, 1985). A single pragmatic reasoning schema is assumed to consist of a set of highly generalized and abstracted rules which are nonetheless defined in terms of classes of goals and types of relationships (Cheng et al., 1986). *Permission schemas* are regarded as one example of these clustered rules, referring to situations in which taking a particular action requires to satisfy a certain precondition (Cheng & Holyoak, 1985). The permission schema is defined as consisting of the following four production rules that are

[27] For a detailed description of the materials please refer to the appendix of Cosmides (1989).

Table 3-3

Predictions and results in terms of the percentage of (logically correct) *P* & not-*Q* responses obtained for Cosmides' (1989) Experiment 5 and 6. *Social Contract Theory* predicts a high performance level only for social-contract rules, but not for non social-contract permission rules. *Pragmatic Reasoning Schema Theory* predicts the same high performance level for both types of rules.

	Predictions		Results	
	Social Contract Theory	*Pragmatic Reasoning Schema Theory*	Exp. 5	Exp. 6
Social-contract problem	High	High	75%	80%
Non-social-contract permission problem	Low	High	30%	45%

assumed to be activated whenever a conditional rule has the action-precondition structure of one of the rules inherent in the permission schema: (1) If the action is to be taken, then the precondition must be satisfied. (2) If the action is not to be taken, then the precondition need not be satisfied. (3) If the precondition is satisfied, then the action must be taken. (4) If the precondition is not satisfied, then the action must not be taken (Cheng & Holyoak, 1985). Obviously, there is considerable similarity to the cost-benefit structure of social-contract rules. Accordingly, Cheng and Holyoak (1989) claim that social contracts are nothing but a subset of permission (or obligation) rules.

At first sight, the results illustrated in Table 3-3 confirm the hypothesis that social exchange is a necessary precondition for an enhanced performance due to the problem content, and this is exactly what Cosmides (1989) argues. However, even if the percentage of logically correct answers obtained for the non-social-contract permission rules was lower than for the social-contract rules, performance on these rules was still better than for abstract rules or other, non-permission rules (remember previous findings in which subjects' performance on such tasks was barely as good as 20% of correct responses).

As the most convincing evidence for her assumption that subjects' performance on the *Wason Selection Task* is guided by a cheater-detection algorithm, Cosmides (1989) reports findings from experiments in which she compared standard social-contract rules to so-called *switched* social-contract rules. In a switched rule, the terms *P* and *Q* are exchanged. Thus, instead of reading "If you take the benefit, then you pay the cost," a switched social-contract rule reads "If you pay the cost, then you take the benefit." Cosmides (1989) argues that a "look-for-cheaters" procedure should always lead subjects to pick the "Cost Not Paid" and the "Benefit Accepted" card, independent from formal logic. Therefore, in a switched social-contract rule, a high percentage of logically incorrect not-*P* and *Q* responses should be found[28]. As a matter of fact, Cosmides (1989, Exp. 3 & 4) found increased rates of not-*P* and *Q* responses for switched social-contract rules (namely, 67% and 75%). Moreover, these were accompanied by only marginal rates of logically correct *P* and not-*Q* responses (namely, 4% and 0%). Cosmides (1989) concluded that this pattern of results was completely in line with *Social Contract Theory* and clearly ruled out the possibility that social-contract content might just enhance subjects' ability for logical reasoning.

At this point, a more detailed description of evidence provided by Cosmides to support her theory of social exchange shall be skipped, since her experiments are basically the same as for their design, procedure, and results. Instead, a résumé shall be drawn on the explanatory power of her findings in general and with respect to rival approaches in particular.

3.3.4 Résumé

Aside from the above-outlined evidence, Cosmides reports a variety of similar findings to substantiate her proposition that innate mental algorithms, functionally specialized for operating on social exchange, are the key to any performance enhancement found for thematic (social-contract) versions of the *Wason Selection Task* (see, for example, Cosmides, 1989, Exp. 7-9; Stone, Cosmides, Tooby, Kroll, & Knight, 2002; Sugiyama, Tooby, & Cosmides, 2002). In line with her postulate of a "logic of social exchange" (Cosmides, 1989), she devoutly believes that her findings sufficiently rule out any competing explanation. However, as noted previously, her findings may not always be interpreted as unambiguously as she suggests. Instead, other explanatory approaches seem to hold for most, if not almost all, of the findings as well, as shall be discussed in the following.

3.3.4.1 *Social Contract Theory* versus Availability Theories

In contrast to Cosmides' social-contract argumentation, availability theories share the view that in the course of the subjects' past experience, associations between the propositions of conditional rules (*P* and *Q*) are created and strengthened with any further exposure that the subjects may have to these propositions. Following this, it is assumed that the ease with which *P* and *Q* come to mind while solving a thematic *Wason Selection Task*, their availability as a response, varies as a function of the strength of these associative conjunctions (Giger-

---

[28] Note that in a switched social-contract rule, not-*P* refers to the "Cost Not Paid" card, whereas *Q* refers to the "Benefit Accepted" card.

enzer & Hug, 1992). Having looked at the above-mentioned studies, availability theories do indeed hold for at least most of the reported findings of content effects (e.g., Griggs & Cox, 1982; Johnson-Laird et al., 1972; Pollard, 1982; Wason & Shapiro, 1971) and do, on the other hand, provide a plausible explanation for a variety of failures to replicate the typical performance enhancement (e.g., Griggs & Cox, 1982; Pollard & Gubbins, 1982). Thus, their explanatory power is not less convincing than that of *Social Contract Theory*. Moreover, against the background of Cosmides' (1989) first two experiments, *Social Contract Theory* does not seem warranted to rule out any availability explanation. In fact, Gigerenzer and Hug (1992, p. 140)—otherwise strong proponents of *Social Contract Theory*, admit that "availability might be the (only) decisive cognitive process in familiar rules." Yet their argumentation in favor of social-contract explanations whereupon entirely unfamiliar rules may activate other processes than availability at least requires consolidation by further empirical evidence.

3.3.4.2 *Social Contract Theory* versus *Pragmatic Reasoning Schema Theory*

As suggested above, proponents of both *Social Contract Theory* and *Pragmatic Reasoning Schema Theory* are concordant in their argumentation that people often reason using neither syntactic, context-free rules of inference nor memory of specific experiences (Cheng & Holyoak, 1985). Instead, within both approaches it is assumed that the underlying structure of a given problem content does, by accident, yield the same solutions as reasoning by formal logic (Cheng & Holyoak, 1985; Cosmides, 1989). However, there is deep disagreement with respect to Cheng and Holyoak's proposition that social contracts are nothing but instances of permission rules. Yet going into the heated debate that has taken place between the two parties in more detail would go beyond the scope of this work (see, however, Cheng & Holyoak, 1989; Fiddick, Cosmides, & Tooby, 2000).

Still, Cosmides' (1989) results supporting *Social Contract Theory* notwithstanding, viewing social contracts as a subset of permission schemas seems highly plausible, especially with regard to the rationale proposed by Cheng and Holyoak (1989, p. 288):

> Cosmides faces a dilemma. On the one hand, she clearly needs to broaden her definition of an exchange to include situations in which no cost is paid in order to account for the many non-exchange contexts (e.g., the drinking age rule). On the other hand, by stretching her definition in this way she is left with either an incoherent concept (if she retains her definition as is), or a concept that includes non-social-exchange permissions (if she replaces the concept of a cost in her theory by the more general concept of a requirement), thus abandoning her claim that social exchange is crucial.

Clearly, the crucial difference between the two theories is in the semantic interpretation of a given rule as either a social contract or a permission (or obligation) rule (Gigerenzer & Hug, 1992). As Gigerenzer and Hug (1992) point out, the precise difference is fairly blurred, and the underlying concepts of costs and benefits, and actions and preconditions are ambiguous. Yet this seems enough of a reason to doubt the unconfined validity of Cosmides' postulate (Cosmides, 1989; Cosmides & Tooby, 1989) unless further evidence is provided.

3.3.4.3 Further Criticism

Apart from the objections to *Social Contract Theory* that have been raised by proponents of availability theories and *Pragmatic Reasoning Schema Theory* (see Pollard, 1990, for further comments), additional criticism has been postulated with respect to the work of Cosmides (1989), focusing on her methodological approach. For example, Lawson (2002) argues that any task failures occurring on a *Wason Selection Task* may be due to misunderstandings of the given rule. Such misunderstandings, in turn, may considerably vary from task to task and from individual to individual. From a series of five experiments in which he compared a variety of social-contract rules and non-social-contract rules differing with regard to their perspicuity, Lawson provides support for his misunderstanding hypothesis, concluding that his results do not only fail to support *Social Contract Theory* but also the broader theory that conditional logic has its origin in the evolution of social exchange.

Similarly, Liberman and Klar (1996) doubt the validity of *Social Contract Theory*. Pointing out serious confoundings in Gigerenzer and Hug's (1992) materials, the authors argue that these may have led to a misunderstanding of the given rules, thus inducing the observed pattern of results[29]. Based on further task analyses, they propose three aspects to affect performance on the *Wason Selection Task*: first, the clarity of a given rule; second, the nature of the alternative to the tested rule and the falsifying instances entailed; third, the perceived relevance of a "looking-for-violation strategy". From a series of experiments in which, as they propose, they compared the confounded original cheating versions of the *Wason Selection Task* (adopted from Gigerenzer and Hug, 1992) to unconfounded cheating and non-cheating versions, Liberman and Klar (1996) conclude that rather than cheating content (Cosmides, 1989) and perspective (Gigerenzer & Hug, 1992), cognitive features related to

[29] As mentioned above, Gigerenzer and Hug (1992) strongly support Cosmides' proposal of innate cheater-detection algorithms, yet amend *Social Contract Theory* by stressing the importance of the perspective in which subjects are cued in dealing with social-contract rules (the perspective of the to-be-cheated).

task interpretation facilitate performance on the *Wason Selection Task*. Moreover, in line with Lawson (2002), the authors argue in terms of a general logic mechanism which they propose might be just as adaptive and just as high in its evolutionary value as the specified cheater-detection algorithm proposed by Cosmides.

Furthermore, Fodor (2000) strongly argues against a cheater-detection module as providing a clear example of natural selection's impact on the human cognitive architecture. Instead, he proposes that the alleged cheater-detection effects on the *Wason Selection Task* are nothing but materials artifacts. As Fodor (2000) criticizes, in her cheater-detection interpretation of content effects on the *Wason Selection Task*, Cosmides (1989) disregarded that different inferential routes (rather than different mental modules) are sufficient to account for performance differences on separate versions of the task[30].

Even Gigerenzer and Hug (1992), although basically proponents of *Social Contract Theory*, have raised criticism with respect to Cosmides' (1989) experiments based on switched social-contract rules. As they point out, and as Cosmides (1989) herself already conceded, by the way, not all thematic rules may be switched like abstract rules. For example, the accurately switched cassava root problem would read "If a man *must* have a tattoo on his face, then he eats cassava root"[31]. However, for linguistic reasons, Cosmides (1989) left out the deontic *must* from the original rule, thus presenting a switched rule that was obviously not the exact converse of the original one (Gigerenzer & Hug, 1992; Manktelow & Over, 1987).

3.3.5 Conclusions

It is clear that the above summary of empirical evidence derived from the *Wason Selection Task* and its implications for the validity of *Social Contract Theory* is by far not exhaustive. On any account, however, a far more detailed examination would have gone beyond the scope of this work. Still, one thing should have become clear: there is an abundant number of empirical findings that may be interpreted in terms of *Social Contract Theory*, thus supporting the notion of a domain-specific, content-dependent "Darwinian algorithm" leading to domain-appropriate inferences, judgments, and choices within the area of social exchange (Cosmides, 1989; Gigerenzer & Hug, 1992; Stone et al., 2002; Sugiyama et al., 2002).

[30] For a more detailed exemplification of Fodor's rationale, see his debate with Beaman on the issue of cheater detection (Beaman, 2002; Fodor, 2000, 2002).

[31] Italics added. Remember that the problem in its unswitched form was phrased "If a man eats cassava root, then he must have a tattoo on his face" (Cosmides, 1989).

However, there have been just as many conflicting results and critiques (Cheng & Holyoak, 1989; Lawson, 2002; Liberman & Klar, 1996; Manktelow & Over, 1990; Pollard, 1990; Shapiro & Epstein, 1998; Sperber & Girotto, 2002). Thus, even after more than two decades of research, there is still ambiguity regarding the role that evolutionary approaches in the sense of domain-specific mental modules may play with respect to human reasoning and, as a consequence, human behavior. It seems that in the course of all the contentions on what factors may or may not influence subjects' performance on the *Wason Selection Task*, some researchers lost sight of their actual intention. In the case of Cosmides, this was originally the test of assumptions proposed by evolutionary psychology concerning the human cognitive architecture. Sperber and Girotto (2002, p. 278f) call a spade a spade in pointing out that "what seems to drive the continuous production of selection task experiments is that their interpretations can be endlessly contested by means of further experiments with the task," and that "[t]his in itself would be reason enough to question the reasonableness of the proliferation of research based on the selection task of which the work of Cosmides, Tooby and their collaborators is a striking example." In line with Sperber and Girotto, it shall be called into question why there has been such an adherence to the *Wason Selection Task* as the one and only paradigm for examining social-exchange hypotheses. It seems astonishing that apparently no attempt has been made to investigate the issue of "Darwinian algorithms" in general, or the cheater-detection mechanism in particular, with methods other than the *Wason Selection Task*. With respect to the enormous importance ascribed to such mechanisms by evolutionary psychologists, it appears even more inscrutable to test their functioning by means of only a single methodic approach and thus to constrain the interpretation and generalization of findings that support the "evolutionary" point of view. If the proposed cheater-detection mechanism is de facto activated within social interactions and if it enables us to reliably identify cheaters, then it should reasonably have considerable impact on tasks that require the identification of *real* persons who *actually* cheated, rather than just cards *representing* persons who *potentially* cheated.

It appears highly plausible that if a cheater-detection mechanism as proposed by Cosmides (1989; Cosmides & Tooby, 1989) exists, functioning to prevent us from deception perpetrated by others in the course of social interactions, then this mechanism should at least meet two claims: sensitizing our perception with respect to others within situations of social exchange and moreover, enhancing our memory for persons we interacted with, our memory for who they were, and how they behaved, and which implications we made concerning further interactions with these persons. This, in turn, implies two things: a distinct memory for faces and good source-monitoring skills which enable us to distinguish faces of cheaters from any other faces that appear familiar to us. It is known that we do very well with respect to the first (E. Brown, Deffenbacher, & Sturgill, 1977; Bruce, 1988; Burton, Bruce, & Dench, 1993; Neath & Surprenant, 2003) but have considerable problems as for the latter (E. Brown et al., 1977). For example, E. Brown et al. (1977) found that facial photos were correctly identified as old in a recognition test at a rate of 96%. However, when subjects were to indicate in which of two places a given face had been presented to them before, only 58% of correct judgments emerged. Although subjects knew that they had encountered certain faces before, they were not as reliably able to remember the context of the encounter. Many factors have been discussed to influence face recognition (Bruce, 1988; Ellis, 1975; Geary, 1998; Rodin, 1987), for example stimulus features (such as feature saliency, facial attractiveness, facial expression), subject factors (such as sex, age, intentions), methodological aspects (such as exposure duration, retention interval, training effects), and, as outlined above, design features of the human mind. However, full consensus on exactly what accounts for our excellent face recognition ability has not been reached so far (Bruce et al., 1993)[32].

Still, to recognize a person per se is obviously not a sufficient precondition to prevent oneself from future fraud. Instead, on the basis of information on previous social interactions retrieved from memory, one must be able to evaluate others persons' "threat potential." To the best of my knowledge, only one attempt to account for a cheater-detection mechanism with respect to such considerations has been made (Mealey et al., 1996). This is even more astounding taking into account that Cosmides herself emphasizes the importance of what she calls "associative cognitive capacities" (Cosmides & Tooby, 1989, p. 60)[33]. In their experiment, Mealey et al. (1996) tested for the impact of behavioral information that did or did not characterize target persons as cheaters on subjects performance in a face-recognition test. This work will be attended to in more detail in the following.

---

[32] For a detailed examination of the issue of face recognition, you may refer to the extensive work of Bruce and colleagues (e.g., Bruce, 1988; Bruce, Burton, & Craw, 1992; Bruce & Young, 1986) or Ellis (e.g., 1975).

[33] As Cosmides points out, these associate cognitive capacities are not only important with respect to successfully participating in social exchange; moreover, they are assumed to be of major significance in a wide range of other evolutionarily crucial social interactions, such as mating, pair-bonding, and parenting. As for the issue of social exchange, five capacities are specified. These are, amongst others, the ability to recognize many different individuals and the ability to remember aspects of one's history of interaction with different individuals (Cosmides & Tooby, 1989).

3.3.6 Enhanced Memory for Faces of Cheaters?

Originally, Mealey et al. (1996) focused on the investigation of processes underlying face recognition. Arguing from an evolutionary perspective and pleading for a notion of the human mind as a complex entity of specialized information-processing devices, they proposed that specialized adaptive features should be built into the individual face-recognition mechanism. In fact, there is considerable support for the existence of specialized recognition systems, for example, with regard to facial identity (e.g., Duchaine, 2000; Duchaine et al., 2001; Kanwisher, 2000) or facial expressions of emotion (e.g., Adolphs et al., 1999; Blair & Curran, 1999). Probably the strongest evidence for the impact of specialized mechanisms on the recognition of faces comes from neuropsychological literature, mainly on the issue of prosopagnosia. Prosopagnosia refers to the phenomenon of a failure in discriminating and recognizing familiar faces despite sufficient perception of the relevant facial details. However, most of the patients suffering from prosopagnosia do not exhibit impaired object detection (Goldenberg, 2003; Orgass, 1989) which is commonly interpreted in terms of different processing mechanisms responsible for face and object related information processing (Kanwisher, 2000). Within their proposition of specialized face-recognition mechanisms, Mealey et al. (1996) especially pleaded for an innate, adaptive threat-detection feature. They based their assumption on findings of a particularly high ability to process angry faces and, even more important, on Cosmides' work (1989; Cosmides & Tooby, 1989, 1992), suggesting the impact of specialized cheater-detection algorithms on human reasoning. With respect to Cosmides' findings, the authors concluded that the evaluation of other persons' threat potential was the most critical element of one's subjective perception of unknown faces[34].

Furthermore, the authors took into account that social status has been found to play a major role in social interactions. They hypothesized that there may be some concurrence of an interactant's perceived social status and threat potential in affecting later cognition (Mealey et al., 1996, p. 121):

> We postulated that faces presented as high status and threatening (likely to cheat) might be attended to, encoded, and recognized even more readily than those presented as low status, because high-status individuals would be perceived as having more power, and therefore be more likely to act in line with their character.

[34] Note that threat potential was defined in terms of *likeliness to cheat*; the higher the likeliness of being cheated by a certain person, the higher is this person's threat potential.

In their experiment, subjects were to judge the attractiveness of 36 male faces, presented in black-and-white pictures under the disguise of a study concerning the test-retest reliability of attractiveness measures. With each of the photos a fictional behavioral description was presented, giving information on behavior of cheating, behavior of trustworthiness or irrelevant behavior. Additionally, each sentence contained information on the depicted person's social status which was induced by means of information on the to-be-judged person's profession. Overall, there were six categories of descriptive sentences, with three levels of threat potential (information on cheating, information on trustworthiness, and irrelevant information) and two levels of social status (low and high social status). After a retention interval of one week, subjects were presented 72 photos, half of which were old (presented the week before) and half of which were new. However, no behavioral descriptions were presented additionally. As in Week One, subjects were to rate the depicted persons' attractiveness. Furthermore, they were instructed to judge whether or not they remembered a given photo from the week before.

The results were in line with the authors' assumptions in that the expected bias for remembering faces formerly associated with cheating (with potential threat) was found. However, as evident from Figure 3-4, this was only true for faces of persons associated with low social status. In the high-status condition, obviously no differences in the recognition performance occurred as a function of the threat-potential variable. As Mealey et al. (1996) pointed out, this finding was contrary to their hypothesis that cheating of high-status persons might be perceived as even more threatening than cheating of low-status persons. Based on the result that subjects rated the attractiveness of high-status persons clearly higher compared to low-status persons, the authors accounted for this finding with the activation of a cognitive mechanism related to the perception of attractiveness rather than the perception of faces in general. Thus they interpreted their findings as "a mitigation of the threat-detection effect in the high-status condition, mediated by the enhanced attractiveness of high-status individuals" (Mealey et al., 1996, p. 123). Furthermore, Mealey et al. (1996) found that the effect of threat potential on the recognition-test data was more distinctive for male than for female participants. Additionally, they pointed out that for female subjects the two conditions with the highest recognition performance were found (low-status cheating and high-status trustworthiness). The authors concluded that to-be-judged persons' social status and potential threat value were important social environmental features. Furthermore, supporting the view of Cosmides (Cosmides, 1989; Cosmides & Tooby, 1989), they reasoned that, in the course of evolution, humans must have developed non-conscious biases in their per-

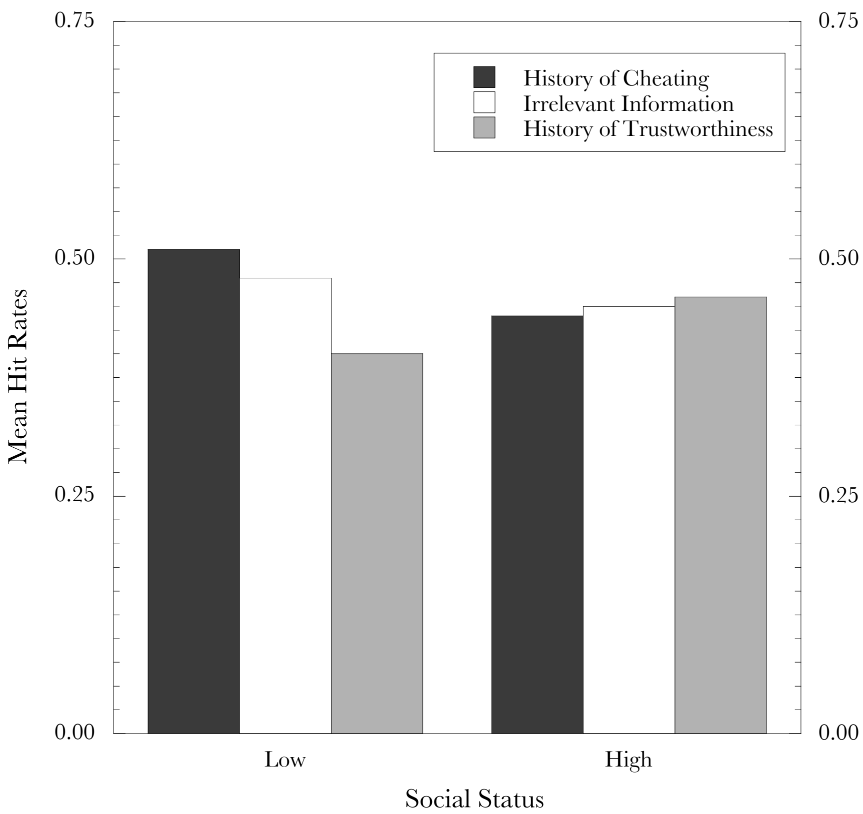

Figure 3-4: Mean hit rates calculated from the recognition-test data reported by Mealey et al. (1996). The highest hit rates were obtained for the recognition of low-status cheaters. No standard deviations or standard errors were reported.

ceptual and cognitive processes so as to be especially attuned to individuals who are perceived as potentially threatening.

At first view, the findings of Mealey et al. (1996) seem to provide a validation of what Cosmides (1989) concluded from her studies with the *Wason Selection Task*, namely that there must be some kind of mechanism inherent in the human mind that is concerned with tracking cheaters. However, a closer look at the data reveals that the size of the cheater-detection effect found by Mealey et al. (1996) is only marginal. The effect size resulting from the *F*-statistic reported for the main effect of threat potential was as small as $\eta^2 = .05$[35]. Therefore, the chapter of mental algorithms, functionally specialized for the detection of cheaters, may still not be closed. Aside from the negligibly small effect size, there are a series of aspects one may criticize with respect to the approach of Mealey et al. (1996).

[35] Note that the effect size measure η^2 refers to the proportion of variance explained by a given factor relative to the variance not explained by any other factor (Cohen, 1977). Obviously, 5% of variance explained by the threat potential factor is by no means sufficient, taking into account the enormous importance that has been ascribed to the hypothesized cheater-detection mechanism.

First of all, the authors did not give any information concerning the time frame for the attractiveness ratings in the first phase of the experiment. However, this seems important considering that faces associated with high threat potential presumably attracted the subjects' attention to a greater extend than faces presented together with descriptions of irrelevant or positive (unoffending) behavior. As a consequence, subjects may have spent more time examining faces of persons depicted as cheaters than faces of persons described as trustworthy or irrelevant. Accordingly, this might have led to a better encoding and therefore an enhancement of the recognition of faces of cheaters in the second phase of the experiment.

Secondly, although Mealey et al. (1996) noted that "people are more proficient at recognizing faces of individuals who had previously been perceived as important," they did not provide any evidence that this was true for the faces presented in their experiment. That is, subjects were not asked for their appraisal of the importance or, in other words, the personal relevance of the depicted persons and, in particular, the behavior associated with them. It seems reasonable that the memory of persons whose behavior strongly affects me as a participant of such an experiment might be better than the reminiscence of persons whose behavior does not concern me at all. Similarly, the authors did not ensure that subjects perceived the social status of the applied job titles as intended, and therefore, that the manipulation of the social-status variable was successful[36].

Thirdly, though the authors reported the finding that subjects chose lower attractiveness ratings with respect to faces of cheaters compared to faces of persons depicted as trustworthy or irrelevant, it remains unclear whether this was observable in Week Two as well as in the Week One attractiveness ratings. However, this is an important aspect with regard to the interpretation of the recognition-test results. As for Week One attractiveness ratings, it is obvious that the observed bias was directly evoked by the written descriptions presented together with the to-be-judged faces. However, in Week Two, no behavioral descriptions were presented. If the same bias as found in Week One still had been observed, this must have been due to the implicit memory of the character information given in the previous week. In contrast, if no such differences in the Week Two attractiveness ratings had been found, this would have acted as an indicator that judging a face as old did not function as a cue for the recollection of the character information. However, as was outlined previously,

[36] Of course, the findings for the attractiveness ratings strongly suggest this was the case. However, a straightforward assessment of the social status ascribed to the given names would have been apt to resolve any doubts.

this is what should be expected, keeping in mind that the proposed cheater-detection mechanism was thought to enable us to protect ourselves from being cheated.

Finally, as a side note, due to the authors' parsimonious way of reporting their data, some information that might have been of further interest is simply not given. For example, neither the mean Week One attractiveness ratings nor the mean values as for the Week Two ratings have been reported. Similarly, Mealey et al. (1996) did not exemplify the rating scales used for the attractiveness ratings. However, there is evidence that the nature of the response scale may have considerable impact on the results (Strack, 1994)[37]. These pieces of information, albeit not mandatory for interpreting the recognition-test results, may have been useful cues with respect to the design of a replication.

Taken together, the approach of Mealey et al. (1996) may seriously be considered as an alternative to more studies based on the *Wason Selection Task*. Besides, focusing on face recognition compared to the more or less abstract *Wason Selection Task* paradigm seems to represent an increase of the ecological validity of the obtained results. Therefore, it appeared advantageous to embark on the strategy proposed by Mealey et al. (1996) to further investigate the issue of cheater detection. Accordingly, the ensuing experiments were designed as extensions of the work of Mealey et al. (1996) with the intention of testing for the replicability of a biased recognition of faces of cheaters and of gaining further insight in the underlying processes. Furthermore, some methodological and procedural improvements were made with respect to the aspects criticized above.

[37] For example, Strack (1994) reports findings that subjects' ratings varied as a function of the provided scale values and alternatives.

4 Experiment 1

4.1 Overview

The first experiment was designed as a conceptual replication of the work of Mealey et al. (1996). Subjects were told to evaluate the attractiveness of a variety of facial photos that were presented together with short descriptive sentences giving information on the social status and threat potential of the people depicted. One week later, in a recognition test, Week One faces were presented together with an equal number of new faces. Again, subjects were told to rate the attractiveness of the people shown and also to estimate whether a picture was old (already shown in Week One) or new. As an extension of the work of Mealey et al. (1996), the recognition-test phase was followed by a series of ratings that will be described in more detail below.

4.2 Hypotheses

Acting on the assumption that a cheater-detection module as postulated by Cosmides (Cosmides, 1989; Cosmides & Tooby, 1989) might in fact exist, and in view of the findings of Mealey et al. (1996), the following hypotheses were stated.

4.2.1 Recognition-Test Data

As a matter of course, the core assumption with respect to a test of *Social Contract Theory* (Cosmides & Tooby, 1989) was concerned with the cheater-detection effect that had been found by Mealey et al. (1996). Accordingly, faces of people characterized as cheaters were expected to be more reliably recognized than faces presented together with information on a history of trustworthiness or irrelevant information. Note that no specific predictions may be derived from *Social Contract Theory* regarding an impact of information on trustworthiness and irrelevant information on the recognition-test data. Since both types of information do not imply an act of cheating, *Social Contract Theory* simply prescribes to ignore any individual that is characterized in terms of trustworthiness or irrelevance. However, with respect to findings reported by Rodin (1987) both types of information were expected to affect face-recognition performance differently. She found face-recognition abilities to vary as a function of the perceived likelihood of a face to be of later significance[38]. Therefore, it was assumed that people characterized as trustworthy and supportive should be perceived as more important with respect to future encounters than persons whose behavior was neither of advantage nor causing any harm to the observer.

Similarly, *Social Contract Theory* does not make any specific predictions with regard to the social status of interactants. On the one hand, the assumption initially held by Mealey et al. (1996) that subjects might perceive persons of high social status as powerful and, consequently, as particularly threatening seems plausible; yet the question is whether this assumption fits the narrow definition of threat potential in terms of likeliness to cheat. On the other hand, however, Cummins (1999) suggests the exact opposite. In the context of her *Dominance Theory*[39] she argues that primarily low-ranking individuals may perform acts of cheating since they might attempt to improve their access to competitive resources that are otherwise accessible for individuals of higher rank in particular. Thus, even if persons of high social status might be more powerful, having more means to conduct an act of cheating, the actual probability of cheating might be even higher for low-status persons for reasons specified by Cummins. Therefore, with regard to the present experiment an enhanced memory for faces of low-status cheaters was expected to be found, replicating the result obtained by Mealey et al. (1996).

4.2.2 Attractiveness Ratings

With respect to the attractiveness ratings, it was expected to replicate the effect of threat potential on the Week One data as it was found by Mealey et al. (1996). Thus, persons described as cheaters were supposed to get clearly lower ratings of attractiveness than people described as trustworthy or irrelevant. Moreover, this effect was supposed to be found for the attractiveness ratings assessed in the recognition-test phase. Even if there were no descriptive sentences presented in the recognition-test phase giving information on the threat potential of the people shown, it was assumed that an enhanced memory for faces of cheaters and for characteristics associated with these faces should result in the above-mentioned effect. That is, to reencounter the face of a cheater should, by means of the proposed

---

[38] More precisely, from a series of experiments Rodin (1987) concluded that ignoring others is largely determined by the purposes of the observer that she proposes to be embodied in the disregard criteria.

[39] The basic notion of *Dominance Theory* is that of a set of implicit social norms reflecting behaviors that are permitted, prohibited, or obligated given an individual's rank. Dominance in this context is defined in terms of priority of access to resources in competitive situations. A direct relationship between dominance and reproductive success is suggested to exist for most species, affirming ambitions of low-ranking individuals to get access to otherwise unapproachable resources by means of cheating (Cummins, 1999).

cheater-detection algorithm, somehow trigger the relevant source information (the behavioral information given in Week One).

In contrast, regarding the social-status factor it was expected to find an impact on the attractiveness ratings only with respect to the Week One ratings. Considering that *Social Contract Theory* does not propose a functionally specialized mechanism designed for the purpose of reasoning on social status per se (detached from information on cheating), forgetting was assumed to occur regarding to specific information on a given person's social status. Thus, in line with the findings of Mealey et al. (1996) subjects were assumed to rate the attractiveness of high-status persons higher compared to low-status persons, yet only in Week One.

4.3 Method

4.3.1 Participants

Participants were 64 female and 32 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. They were paid for participating in the experiment. Their age ranged from 19 to 53 (M = 26.13, SD = 5.86).

4.3.2 Materials

4.3.2.1 Pictures

Facial photos of 116 Caucasian males were taken from the Internet. Permission for usage was obtained via e-mail. To make sure that subsequent effects of face recognition might not be ascribed to color features, all pictures were converted into 256 bit grayscales. The size was adjusted to 116 × 164 pixels. Based on judgments of a group of experts (N = 8), particularly salient pictures were excluded. Finally, 72 pictures were randomly assigned to two sets of 36 pictures each.

4.3.2.2 Descriptive Sentences

According to Mealey et al. (1996), photos were to be presented together with a fictional descriptive sentence typed below, giving information on the depicted individual's threat potential and status. There were three types of information implying three different types of threat potential: information containing a history of cheating (e.g., "R. O. is a cashier.

Again and again, he would shortchange and keep the rest of the money for himself."), information containing a history of trustworthiness (e.g., "E. K. is an architect. Any structural damages he might have caused he would get repaired at his own expense."), and irrelevant information (e.g., "J. M. is a teacher. He teaches mathematics and biology at a high school."). Social status was varied by means of information on the profession of the people depicted, using high-status jobs such as "notary," "dentist" or "business manager," and low-status jobs such as "waiter," "scavenger" or "cashier." In a pretest, subjects (N = 24) had rated 82 job titles in terms of their social status using a scale ranging from 1 ("status low") to 5 ("status high"). The 20 titles with extreme values scoring above and below the median of Z = 3.00 were selected as "low-status jobs" (M = 1.83, SD = .40) and "high-status jobs" (M = 4.01, SD = .33), respectively[40]. Subsequently, 40 sentences describing behavior of cheating, trustworthiness or irrelevant behavior were made up as stated above. As in the examples presented above, the behavior described was always directly linked to the job title. Subsequently, another group of subjects (N = 21) was asked to rate these descriptive sentences with respect to the valence of the specified behavior, using a scale ranging from −3 ("negative") to +3 ("positive"). This was to make sure that instances of behavior of cheating were perceived as negative (as having a high threat potential), whereas instances of trustworthiness were supposed to be sensed as positive. Furthermore, subjects were thought to judge the instances of irrelevant behavior as neutral, corresponding to valence ratings at zero. Finally, six sentences with high- and low-status job titles and behavioral descriptions of cheating (M = −2.35, SD = .72), trustworthiness (M = 1.75, SD = .73), and irrelevant behavior (M = 0.22, SD = .28), respectively, were selected for the experiment[41].

4.3.2.3 Rating Scales

To make sure that the social-status variable was successfully manipulated, subjects were asked to evaluate the social status of the job titles used. The same scale as applied in the pretest was used, with ratings ranging from 1 ("status low") to 5 ("status high"). This rating will be referred to as the *status rating*. Furthermore, as a manipulation check of the threat potential variable, participants were requested to rate the descriptive sentences alluding to a history of cheating in regard to how grave they thought the behavior depicted was. They used a scale ranging from 1 ("not grave at all") to 6 ("extremely grave"). This rating will be indicated as the *severity rating*. Analogically, subjects were asked to rate descriptions of trust-

[40] See Appendix A for a full list of the selected job titles.
[41] See Appendix B for a complete list of the behavioral descriptions.

worthiness with respect to how commendable they thought the behavior was. The rating scale ranged from 1 ("not commendable at all") to 6 ("extremely commendable"). This rating will be termed *commendation rating*. Finally, participants were to estimate how relevant each of the described behaviors (cheating as well as trustworthiness) was for them. The rating scale used ranged from 1 ("does not affect me at all") to 6 ("strongly affects me"). This rating will be referred to as the *relevance rating*.

4.3.3 Procedure

The experiment took place in a laboratory at the Heinrich-Heine-Universität Düsseldorf. It was run on Apple iMac computers that controlled a standard 15-in. color display. Up to five participants were able to take part simultaneously. To be protected from distracting noise, subjects sat in sound-attenuated booths and wore sound-attenuating headphones.

Under the disguise of a study concerning the test-retest reliability of attractiveness measures, participants were at first asked to rate the attractiveness of 36 facial photos (i.e., picture set 1 or 2, respectively), using a scale ranging from 1 ("not attractive at all") to 6 ("extremely attractive"). The pictures were presented in random order, and the assignment of the picture set was counterbalanced across subjects. Each photo was also randomly assigned to one of 36 descriptive sentences characterizing the depicted person as either cheating, trustworthy or none of both and as holding an occupation of either high or low social status. Each trial started with the display of a headline ("How attractive do you think is the depicted person?") and a photo together with the social-status information (i.e., the job information) written below. After a fixed time frame of 2 seconds the behavior description was shown additionally. To make sure that the participants read the description carefully, the rating scale for the attractiveness rating appeared not until 4.5 seconds later. Subjects then were supposed to rate the depicted person's attractiveness by choosing the appropriate value on the scale with the computer mouse. As long as they did not click a *continue* button, subjects were able to change their decision and choose a different scale value. Pressing the *continue* button initiated the next trial after a delay of 500 ms. This first part of the experiment lasted approximately 10 minutes.

One week later, the same subjects were randomly presented 72 photos (i.e., picture set 1 and 2), half of which had been presented the week before and half of which were new. Subjects were told to rate the depicted persons' attractiveness and, afterwards, to decide whether a picture was old or new. Just like in Week One, each trial started with the display of a headline ("How attractive do you think is the depicted person?") and of a photo but no descriptive sentence. After a fixed time frame of 1.5 seconds the rating scale appeared, and subjects were able to make their choice by selecting the appropriate scale value with the computer mouse. Again, subjects were informed that they might change their decision as long as they did not press the *continue* button. Clicking the *continue* button removed the rating scale and a new headline was shown ("Is this face old or new?"). 500 ms later two check boxes appeared labeled "old" and "new", and subjects were able to make their decision by clicking into one of the boxes with the computer mouse. Similar to the attractiveness ratings, participants were able to change their decision as long as they did not click the *continue* button, which initiated the next trial.

Following the recognition test, subjects were randomly presented the 36 job titles used before and were asked to evaluate their social status. As before, each trial began with the display of a headline ("In your opinion, how reputable is the following job?") and a job title. The rating scale appeared immediately. Participants were to choose a value on the scale and were again able to change their decision as long as they did not click the *continue* button to initiate the next trial. Subsequent to the status ratings, the severity ratings, the commendation ratings, and the relevance ratings were to be made. At the beginning of each trial a headline was displayed ("How do you evaluate the following behavior?") and one of the 24 descriptive sentences containing information on cheating or trustworthiness was presented at random. After a fixed time frame of 4.5 seconds two rating scales appeared: the scales for the severity rating and the relevance rating (in case the sentence displayed contained information on cheating) or the scales for the commendation rating and the relevance rating (in case the sentence displayed contained information on trustworthiness). Subjects then had to choose the appropriate value on each of the scales, and just like before they were able to revise their decision as long as they did not select the *continue* button. After the final rating subjects were informed about the purpose of the experiment and thanked for their participation. This second part of the experiment lasted approximately 25 minutes.

4.3.4 Design

The within-subject independent variables were threat potential (history of cheating, irrelevant information, history of trustworthiness) and social status (low, high). The Week One dependent measures were the attractiveness ratings and, to control for effects of picture encoding, response latencies measured from the picture onset to the confirmation of the rating by clicking the *continue* button. For the second part of the experiment in Week Two the main dependent measures were the old-new ratings. In addition, the time needed for the old-new decisions was recorded (this measure will be referred to as the *recognition response latency*). As a measure of subjects' recognition performance, discrimination indices $P_r = H - FA$ (Snodgrass & Corwin, 1988; Macmillan & Creelman, 1991) were computed, where H and FA refer to the rates of, respectively, hits and false alarms obtained in a recognition test[42]. Attractiveness ratings and corresponding response latencies were measured as in Week One. Finally, Week Two dependent measures also included the status ratings, the severity ratings, the commendation ratings, and the relevance ratings. Following Mealey et al. (1996), attractiveness ratings as well as response latencies and recognition-test data were analyzed in a full factor 3 × 2 × 2 design, with the third factor participant sex included. This was done on the grounds that "females tend to be better than males at face recognition tasks" (Mealey et al., 1996, p. 122)[43].

Given a total sample size of $N = 96$, $\alpha = .05$, and the assumption that the average population correlation between the levels of the repeated measures factor threat potential is $\rho = .60$ (estimated from pilot data), effects of size $f = .15$ (that is, $\eta^2 = .14$; small to medium effects as defined by Cohen, 1977) could be detected for the threat-potential variable with a probability of $1 - \beta = .95$. Note that all power calculations reported in the following were conducted using the *G*-Power program (Buchner, Faul, & Erdfelder, 1996; Erdfelder, Faul, & Buchner, 1996).

In all experiments reported in the following a multivariate approach was used for all within-subject comparisons. All multivariate test criteria correspond to the same (exact) *F*-statistic, which is reported. The level of α was set to .05 for all analyses. For post-hoc tests, the significance level was Bonferoni-Holm corrected (Holm, 1979). This sequential method was preferred to the better-known Bonferoni correction, since it is less restrictive and there-

---

[42] The discrimination index P_r is used quite frequently in recognition-memory studies (Snodgrass & Corwin, 1988). In the context of chapter five of the present work, attention will be drawn to its underlying memory model, the so-called two-high-threshold model for recognition memory. Taking into account that subjects in a recognition test may misjudge new distractor items as old due to processes of guessing, Snodgrass and Corwin (1988) also defined a corresponding bias index B_r. However, since no hypotheses were stated with respect to subjects' guessing strategies in the present experiment, no bias measures will be reported in the following.

[43] Note, however, that this assumption has not been satisfactorily verified. While it has indeed sometimes been reported that women perform better on face-recognition tasks than men, this has often been observed to interact with the target sex in that women did especially well on recognizing female faces (e.g., Going & Read, 1974; Lewin & Herlitz, 2002; for a review, see Shepherd, 1981). Moreover, sometimes no sex differences were observed at all (e.g., Grimshaw, Bulman-Fleming, & Ngo, 2004) or statistically significant effects were only small in size (Bruce, 1988).

fore minimizes the risk of mistakenly accepting the null hypothesis. Partial η^2's are reported as a measure of the size of an effect.

4.4 Results

In the following, the results obtained for Experiment 1 will be reported in line with the order of data assessment. Note that for reasons of lucidity, this will be done with respect to the following experiments as well.

As mentioned above, all data analyses were first conducted including the participant sex variable. However, since no statistically significant main effects or interactions with other variables could be detected, 3 × 2 analyses without this variable were performed additionally and will be reported in the following. To illustrate statistically significant effects, means and standard errors will be parenthesized as far as they are not illustrated in figures.

4.4.1 Week One Data

4.4.1.1 Attractiveness Ratings

Figure 4-1 illustrates the mean attractiveness ratings as a function of threat potential and social status. The 3 × 2 multivariate analysis of variance (MANOVA) performed on the ratings revealed the expected main effect of threat potential, $F(2, 94) = 42.08$, $p < .001$, $\eta^2 = .47$. On a scale from 1 to 6, faces that had been presented together with information on cheating were rated significantly less attractive than faces of people described as trustworthy, $t(95) = -8.96$, $p < .001$, $\eta^2 = .46$ (*M*s = 2.09 vs. 2.97, *SE*s = .08 vs. .08) or irrelevant, $t(95) = -8.72$, $p < .001$, $\eta^2 = .44$ (*M*s = 2.09 vs. 2.83, *SE*s = .08 vs. .08). Furthermore, faces that had been combined with irrelevant information were rated significantly less attractive than faces of people characterized as trustworthy, $t(95) = 2.72$, $p = .008$, $\eta^2 = .07$ (*M*s = 2.83 vs. 2.97, *SE*s = .08 vs. .08). However, it seems noteworthy that subjects generally tended to choose comparatively low ratings of attractiveness. As expected, and in line with the results of Mealey et al. (1996), the analysis also showed a statistically significant main effect of social status, with high-status information coming along with higher attractiveness ratings than low-status information, $F(1, 95) = 9.78$, $p = .002$, $\eta^2 = .10$ (*M*s = 2.68 vs. 2.58, *SE*s = .07 vs. .07). There was no statistically significant interaction between the threat-potential and social-status variables, $F(2, 94) < 1$, $p = .685$, $\eta^2 = .01$.

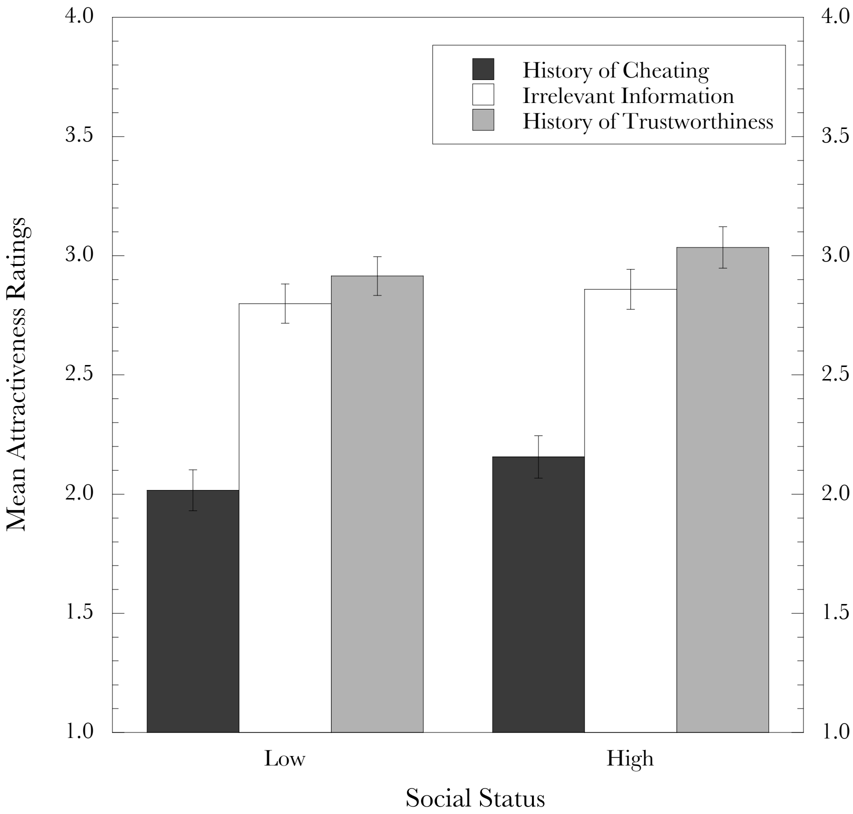

Figure 4-1: Week One mean attractiveness ratings depending on threat potential and social status. On a scale ranging from 1 to 6, subjects rated faces of cheaters significantly lower in attractiveness than faces of people depicted as trustworthy or irrelevant. The error bars represent the standard errors of the means.

4.4.1.2 Response Latencies

The 3 × 2 MANOVA conducted for the response latencies revealed a statistically significant main effect of threat potential, $F(2, 94) = 3.62$, $p = .031$, $\eta^2 = .07$. As post-hoc tests showed, it took subjects significantly longer to judge the attractiveness of people described as cheaters compared to people described by irrelevant information, $t(95) = 2.48$, $p = .015$, $\eta^2 = .06$. The mean differences between the conditions of irrelevant information and information on trustworthiness missed the Bonferoni-Holm corrected significance level, as well as the mean differences between the conditions of information on cheating and trustworthiness, all $|t|(95) \leq 2.05$. However, this main effect was qualified by a statistically significant interaction between the threat-potential and the social-status variables, $F(2, 94) = 8.88$, $p < .001$, $\eta^2 = .16$. Figure 4-2 illustrates the corresponding mean reaction times. Post-hoc analyses showed that in the high-status condition it took subjects significantly longer to rate the attractiveness of people depicted as cheaters compared to people described as trustworthy, $t(95) = 2.72$, $p = .008$, $\eta^2 = .07$ or irrelevant, $t(95) = 3.63$, $p < .001$, $\eta^2 = .12$.

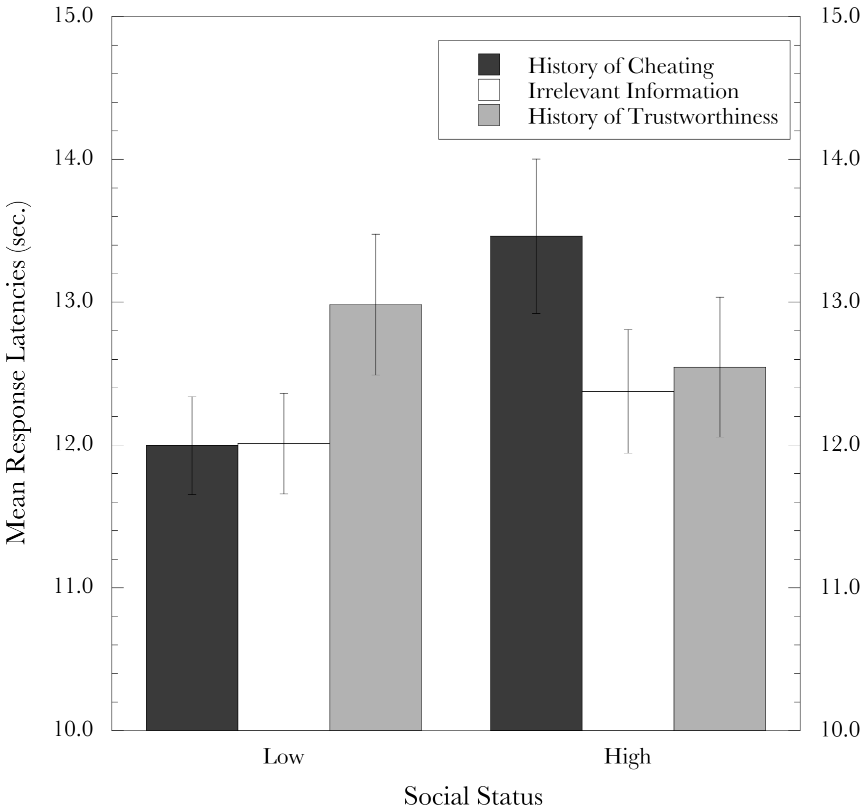

Figure 4-2: Week One mean response latencies depending on threat potential and social status. Apparently, behavior of cheating performed by high-status persons attracted the subjects' attention in particular, as well as positive behavior of low-status persons. The error bars represent the standard errors of the means.

In contrast, in the low-status condition subjects needed significantly more time to judge the attractiveness of people pictured as trustworthy compared to people characterized by information on cheating, $t(95) = -2.71$, $p = .008$, $\eta^2 = .07$ or irrelevant information, $t(95) = 2.68$, $p = .009$, $\eta^2 = .07$, all other $|t|(95) < 1$. Finally, the main effect of social status just missed statistical significance, $F(1, 95) = 3.74$, $p = .056$, $\eta^2 = .04$. This result may lead to the assumption that for stereotypically unusual instances of behavior subjects might have needed additional time to adjust the judgment they had drawn before just on the basis of the photo displayed[44].

[44] It appears verisimilar that persons of high social status are perceived as more respectable (i.e., as more socially approved) than persons of low social status. Moreover, one may expect that persons who are socially disapproved are more likely to be associated with negative behavior (e.g., acts of cheating) than persons who are prestigious. In turn, generous acts of behavior seem likely to be ascribed to persons who are highly respectable but not to persons who are socially disapproved. Consequently, subjects may have perceived the association of negative behavior (cheating) and high social status just as startling (or unusual) as the association of positive behavior (trustworthiness) and low social status.

4.4.2 Week Two Data

4.4.2.1 Attractiveness Ratings

Figure 4-3 illustrates the mean attractiveness ratings for faces from Week One as a function of threat potential and social status. In contrast to the Week One data, the 3 × 2 MANOVA performed on the Week Two attractiveness measures did neither reveal a statistically significant main effect of threat potential, $F(94) < 1$, $p = .929$, $\eta^2 < .01$, nor of social status, $F(95) = 2.10$, $p = .15$, $\eta^2 = .02$. The interaction also missed statistical significance, $F(94) < 1$, $p = .896$, $\eta^2 < .01$. That is, no differences in the attractiveness ratings could be found with respect to the information on threat potential or social status with which the to-be-rated pictures had been combined in Week One.

Overall, subjects chose significantly lower attractiveness ratings in Week Two than in Week One, $t(95) = -3.00$, $p = .003$, $\eta^2 = .09$. As Figure 4-4 shows, this was due to an increase in attractiveness ratings for faces of cheaters, $t(95) = -6.07$. $p < .001$, $\eta^2 = .28$, whereas there was a decrease in the judgments for people described as trustworthy, $t(95) = 7.13$. $p < .001$,

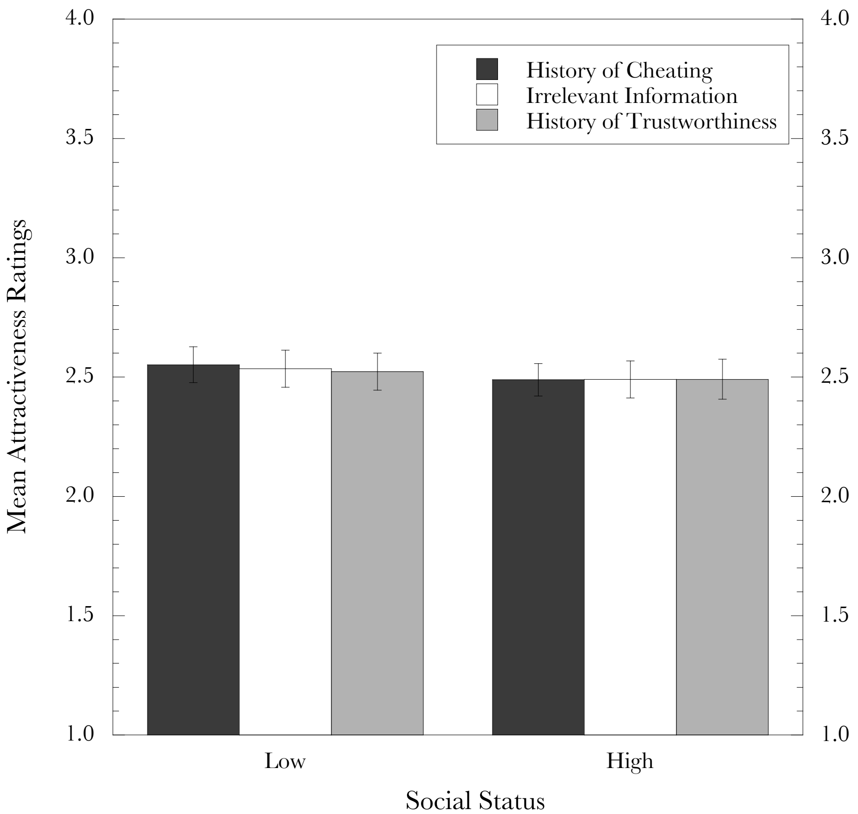

Figure 4-3: Week Two mean attractiveness ratings for faces from Week One depending on threat potential and social status. On a scale ranging from 1 to 6, subjects rated faces of cheaters just as high in attractiveness as faces of people depicted as trustworthy or irrelevant. The error bars represent the standard errors of the means.

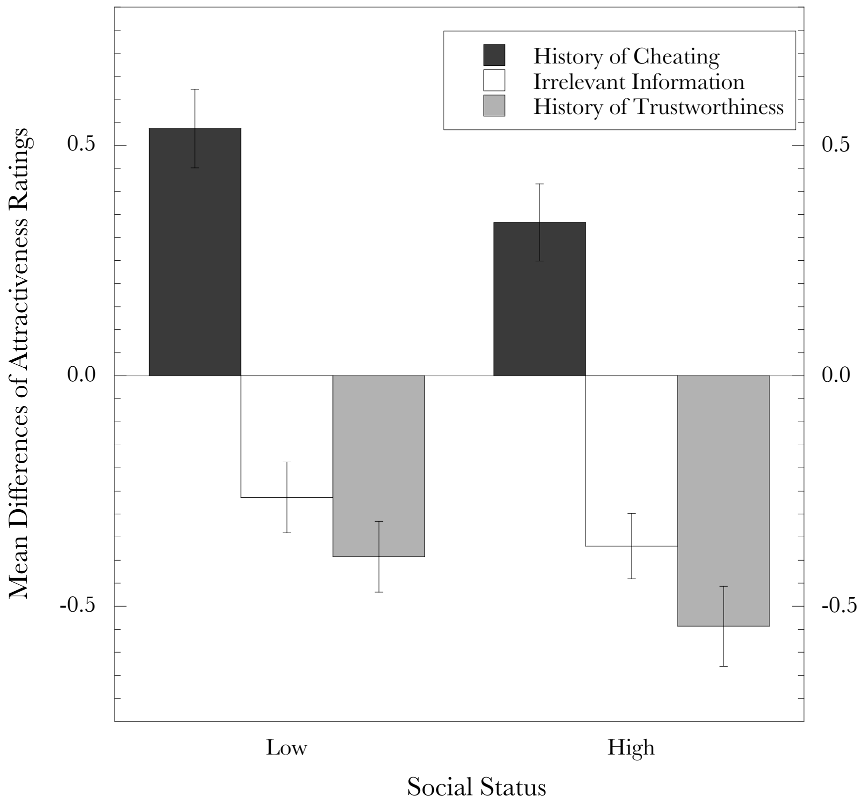

Figure 4-4: Mean differences of attractiveness ratings for Week One faces between Week Two and Week One, depending on threat potential and social status. In Week Two, subjects rated the attractiveness of faces of cheaters higher than in Week One, as the positive difference scores indicate. In contrast, faces of people depicted as trustworthy or irrelevant were perceived as less attractive in Week Two compared to Week One, which is indicated by the negative difference scores. The error bars represent the standard errors of the means.

$\eta^2 = .35$ and irrelevant, $t(95) = 6.15$. $p < .001$, $\eta^2 = .28$. Finally, it seems noteworthy that, on an average, old faces (faces that had been presented in Week One) were judged just as attractive as new faces (faces that had not been presented in Week One), $t(95) < 1$, $p = 987$, $\eta^2 < .01$ (*M*s = 2.52 vs. 2.52, *SE*s = .08 vs. .07 on a scale from 1 to 6).

4.4.2.2 Recognition-Test Data

Figure 4-5 illustrates the discrimination indices P_r depending on threat potential and social status for Week One faces. A 3 × 2 MANOVA did not confirm the expected bias for remembering faces that had been presented with a description indicating potential threat. The main effect of threat potential did not reach statistical significance, $F(2, 94) < 1$, $p = .811$, $\eta^2 < .01$, and neither did the main effect of social status, $F(1, 95) = 1.85$, $p = .177$, $\eta^2 = .02$. There was a statistically significant interaction between threat potential and social status, $F(2, 94) = 3.26$, $p = .043$, $\eta^2 = .07$, but subsequent tests did not reveal any significant differences, $|t|(95) \leq 2.02$. However, it seems noteworthy that the values of P_r in the cheat-

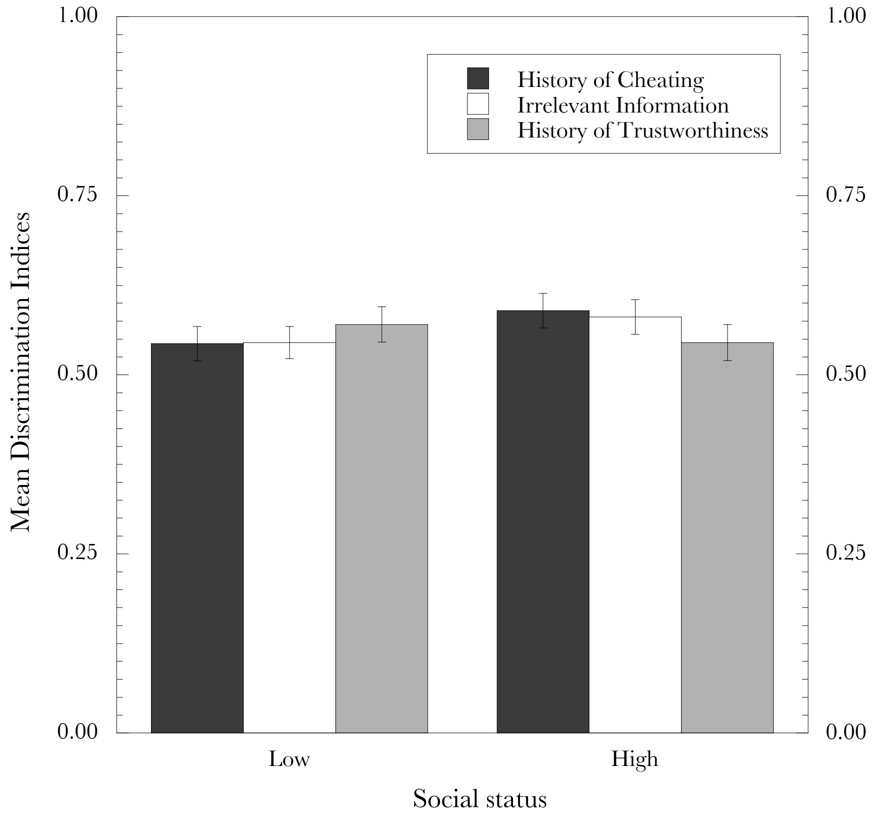

Figure 4-5: Mean discrimination indices P_r for faces from Week One as a function of threat potential and social status. The highest value of P_r was found for the detection of faces that had formerly been presented together with information on high social status and cheating. However, post-hoc tests did not reveal any statistically significant effects. The error bars represent the standard errors of the means.

ing condition as well as in the irrelevant information condition were lower for low-status than for high-status faces, whereas the reverse pattern emerged for hit rates in the trustworthiness condition. Apart from that, a test of whether P_r was different from zero for all cells of the design simultaneously confirmed that recognition performance was above chance, $F(1, 95) = 817.39, p < .001, \eta^2 = .90$.

The pattern of results resembles that of the Week One response latencies. As a reminder, the highest response latencies for Week One attractiveness ratings had been assessed with respect to judgments of high-status cheaters and low-status trustworthy persons, yielding the assumption that the unusualness of a given behavior may have governed how the faces on display were perceived. Comparably, the highest value of P_r was found with respect to high-status cheaters, and a high value of P_r also emerged with regard to low-status trustworthy persons. This may strengthen the assumption that the unusualness of the behavior associated with the to-be-remembered faces has an impact on the perception, encoding, and later retrieval of these faces from memory.

4.4.2.3 Response Latencies

A similar pattern of results as was found for the attractiveness ratings emerged for the response-time data. The 3 × 2 MANOVA on the response latencies measured for the attractiveness ratings neither showed a statistically significant main effect of threat potential, $F(2, 94) < 1$, $p = .675$, $\eta^2 < .01$, nor of social status, $F(1, 95) < 1$, $p = .624$, $\eta^2 < .01$, nor did the interaction reach statistical significance, $F(2, 94) < 1$, $p = .678$, $\eta^2 < .01$.

Testing for the recognition response latencies, however, revealed a statistically significant main effect of threat potential, $F(2, 94) = 3.75$, $p = .027$, $\eta^2 = .07$. As subsequent analyses showed and as evident from Figure 4-6, the old-new decision took participants significantly longer for faces that had been presented together with irrelevant information than with information of trustworthiness, $t(95) = -2.69$, $p = .008$, $\eta^2 = .07$ (*M*s = 2.68 vs. 2.53 sec., *SE*s = .07 vs. .06). The response-time difference between faces paired with information on cheating and irrelevant information missed the Bonferoni-Holm corrected significance level, and the same was found with respect to faces associated with information on cheating

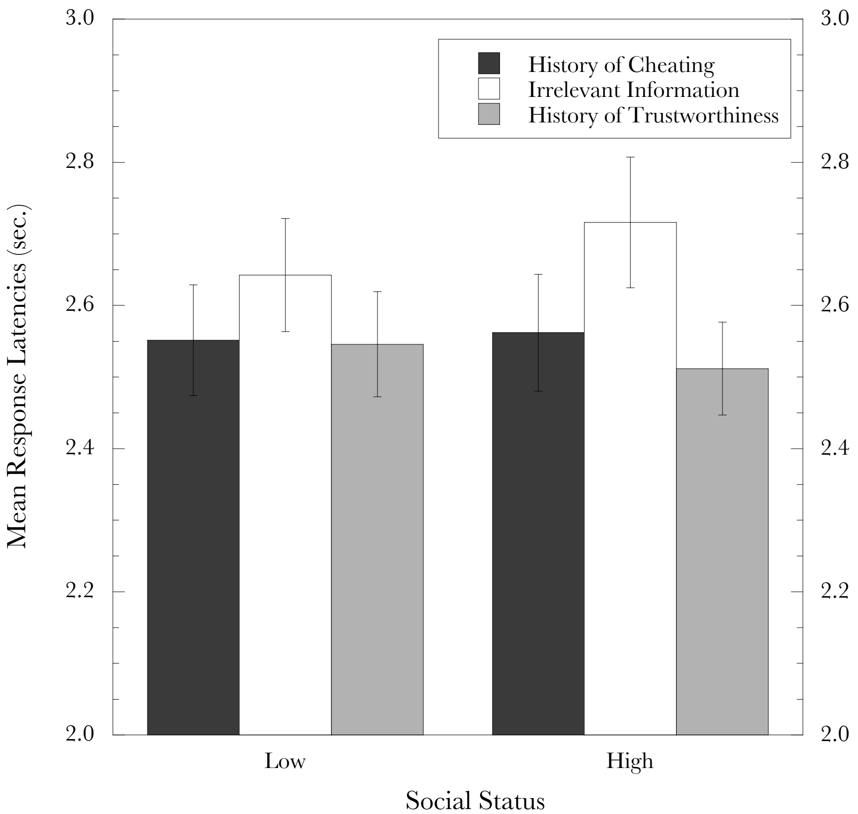

Figure 4-6: Mean recognition response latencies for faces from Week One depending on threat potential and social status. Subjects needed more time to make their old-new decision for faces of irrelevant persons compared to faces of persons described as cheating or trustworthy. The error bars represent the standard errors of the means.

and trustworthiness, all $|t|(95) \le 2.08$. Neither the main effect of social status was statistically significant, $F(1, 95) < 1$, $p = .742$, $\eta^2 < .01$, nor the interaction of the threat-potential and social-status variables, $F(2, 94) < 1$, $p = .684$, $\eta^2 < .01$.

4.4.2.4 Additional Ratings

Status Ratings

The analysis of the status ratings confirmed the successful manipulation of the social-status variable using high- and low-status job titles as defined by means of the above-mentioned pretest. Subjects rated the social status of job titles that were supposed to be of low status significantly lower than the status of job titles that were planned to be of high status, $t(95) = -28.03$, $p < .001$, $\eta^2 = .89$ (Ms = 1.91 vs. 4.01, SEs = .05 vs. .05 on a scale from 1 to 5). As illustrated in Appendix C, all low-status job titles were rated significantly below the median ($Z = 3.00$), while the ratings for all high-status job titles were significantly above the median.

Ratings of Severity and Commendation

Figure 4-7 displays the ratings of severity and commendation on a scale from 1 to 6 as a function of social status. As intended, subjects perceived the instances of cheating as notably grave ($M = 4.91$, $SE = .08$). This result supports the assumption that such behavior descriptions may indeed have had a high threat potential, and that the manipulation of this variable was successful. Analogously, participants judged the instances of trustworthiness as highly commendable ($M = 4.76$, $SE = .10$). Two one-way MANOVAs performed on the ratings of severity and commendation, respectively, revealed a statistically significant main effect of social status for the commendation ratings, $F(1, 95) = 14.15$, $p < .001$, $\eta^2 = .13$, but not for the severity ratings, $F(1, 95) = 1.75$, $p = .189$, $\eta^2 = .02$. Ratings of commendation were significantly higher for the high-status condition than for the low-status condition, whereas the status variable had no effect on the severity ratings.

Relevance Ratings

Figure 4-8 illustrates the relevance ratings depending on threat potential and social status. A 2 × 2 MANOVA performed on the ratings showed that the perceptions of relevance varied as a function of the social status associated with the to-be-judged behavior, $F(1, 95) = 14.11$, $p < .001$, $\eta^2 = .13$. Subjects perceived behavior of persons of low social status as more personally relevant than behavior of high-status persons (Ms = 3.61 vs. 3.38, SEs = .11 vs. .11 on a scale from 1 to 6). No statistically significant main effect of threat potential

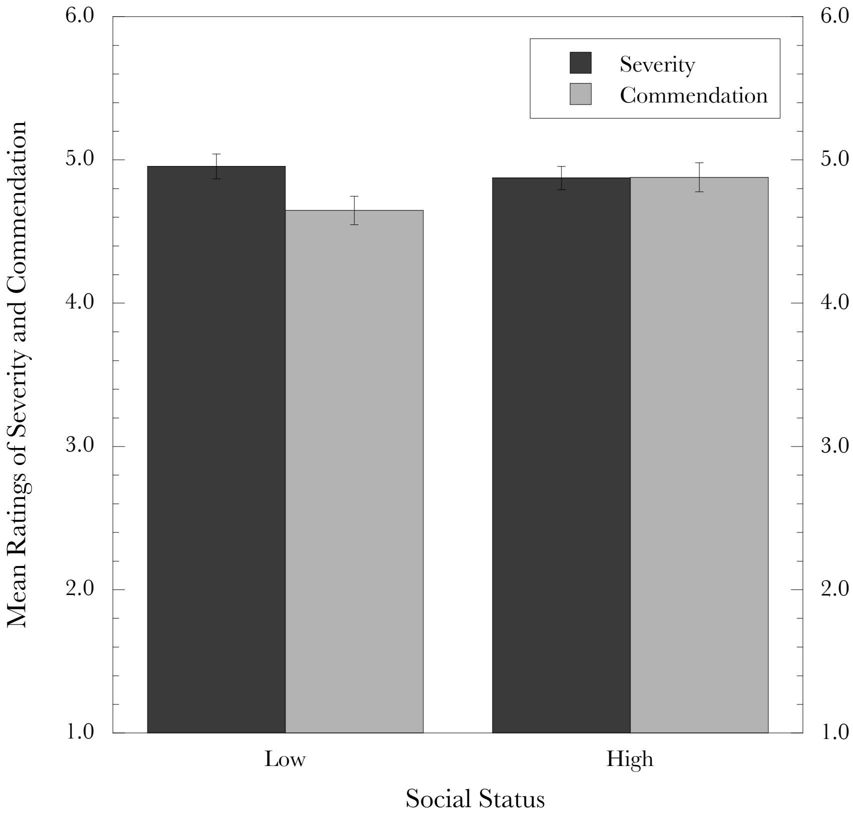

Figure 4-7: Mean ratings of severity and commendation depending on social status. On a scale ranging from 1 to 6, subjects judged instances of behavior of trustworthiness associated with high-status persons as more commendable than trustworthiness of low-status persons. The error bars represent the standard errors of the means.

could be detected, $F(1, 95) < 1$, $p = .660$, $\eta^2 < .01$. Instead, the interaction between threat potential and social status was statistically significant, $F(1, 95) = 57.89$, $p < .001$, $\eta^2 = .38$. In the low-status condition, subjects judged the personal relevance of incidents of cheating to be clearly higher than in the high-status condition, $t(95) = -7.14$, $p < .001$, $\eta^2 = .35$. In contrast, the personal relevance of positive behavior (trustworthiness) was rated significantly higher in the high-status condition than in the low-status condition, $t(95) = 3.28$, $p = .001$, $\eta^2 = .10$.

4.5 Discussion

Experiment 1 was designed as a replication of the work of Mealey et al. (1996). However, contrary to the findings reported by Mealey and colleagues, no evidence for an enhanced memory for faces of cheaters emerged. First of all, no effect of the threat-potential variable on the recognition-test data emerged. Likewise, the analysis of the Week Two attractiveness ratings did not reveal an (implicit) effect of cheater detection in that faces of cheaters were

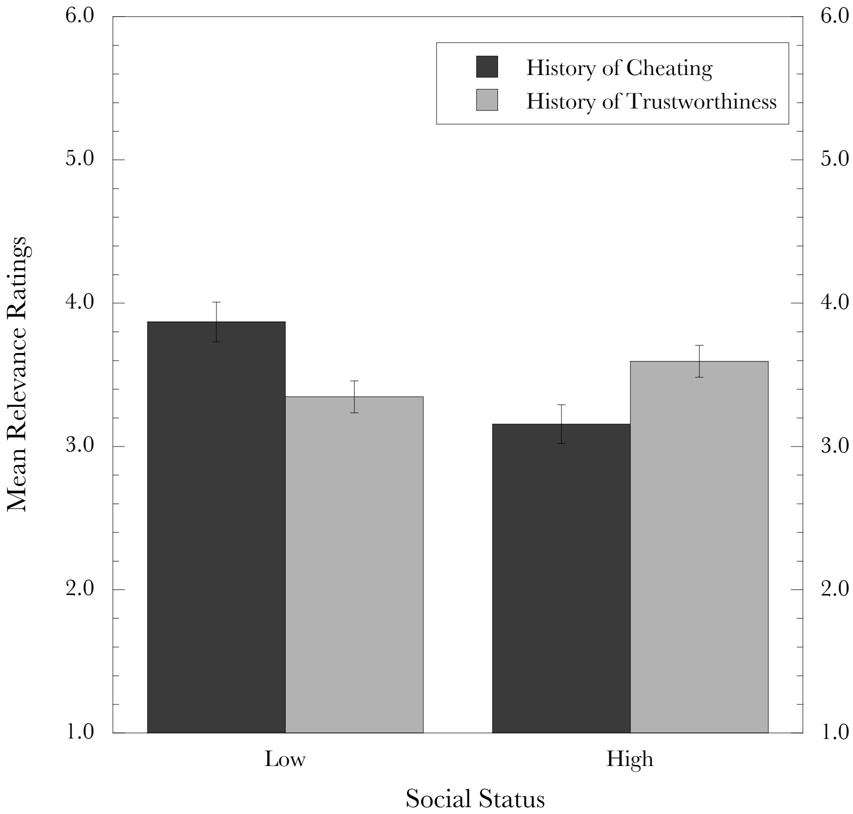

Figure 4-8: Mean relevance ratings depending on threat potential and social status. On a scale ranging from 1 to 6, subjects estimated behavior of cheating performed by persons of low status as even more relevant for themselves than behavior of cheating by high-status persons. In contrast, trustworthiness of high-status persons was judged as more relevant than of low-status persons. The error bars represent the standard errors of the means.

rated lower in attractiveness than faces associated with information on trustworthiness or irrelevant information. Yet for the Week One attractiveness ratings such an effect of threat potential was found, replicating the finding of Mealey et al. (1996) and suggesting that the manipulation of threat potential using instances of behavior of cheating, trustworthiness, and irrelevant behavior was successful[45]. Thus, since the hypothesis of an enhanced memory for faces of cheaters derives directly from *Social Contract Theory* (Cosmides & Tooby, 1989), the present results conflict with Cosmides' conceptualization of a functionally specialized cheater-detection module.

Nevertheless, there are some results that are in line with the findings of Mealey et al. (1996) and deserve closer attention. For example, Mealey and her colleagues found that "the effect of the perception of threat appeared when the person in the photo was depicted as being of low status rather than of high status" and that this was "however, in the opposite

[45] Note that the analysis of the severity ratings also substantiated this assumption, since subjects obviously perceived the instances of cheating as sufficiently grave to have an adequate degree of threat potential.

direction from the anticipated" (Mealey et al., 1996, p. 122). As outlined above, Mealey et al. (1996) expected high-status persons to be perceived as having more power than low-status individuals and hence as having a greater threat potential. In line with their hypothesis, the present recognition-test data showed, at least on a descriptive level, the highest values of P_r for faces of persons described as high-status cheaters. In contrast, subjects assigned the highest severity to behavior of cheating associated with low-status persons. Thus, the impact of the social-status variable on the perception of cheating remains ambiguous. There are some findings (e.g., regarding the Week One response latencies) that may strengthen the assumption that cheating of high-status persons was experienced as somewhat exceptional, maybe since it does not fit stereotypic ideas of high-status behavior. The same seems to be true for behavior of trustworthiness relating to persons of low social status. Possibly, high-status cheaters were perceived as more salient in contrast to low-status cheaters whose behavior may certainly be evaluated as more stereotypical. As a consequence, faces of high-status cheaters might have been more distinctive and therefore easier to keep in mind.

However, regardless of the possible implications of social status the data clearly do not support the idea of a cheater-detection module as proposed by Cosmides (Cosmides, 1989; Cosmides & Tooby, 1989). Aside from the fact that the threat-potential variable did not show the expected effect on the recognition-test data, there are a series of other findings that also fail to meet the expectations. As outlined above, starting from the existence of a cheater-detection mechanism in the sense of Cosmides, the same data pattern as for the Week One attractiveness ratings could have been expected for the Week Two data. Considering an enhanced memory for cheaters due to a cheater-detection module, the (conscious or unconscious) recognition of a cheater's face should in a way have reactivated the information of negative behavior associated with the face before and plausibly have resulted in the same biased perception of that person's attractiveness as was found in Week One. However, the threat-potential variable did not have any effect on the Week Two attractiveness ratings. Apparently, devoid of the Week One descriptive sentences, subjects did not identify faces of cheaters as such. As a consequence, their attractiveness ratings were not influenced by any information on the persons' behavior and were therefore solely based on the visual information given by the pictures shown. Similarly, contemplating that the main purpose of any cheater-detection mechanism should be to prevent an individual from deception in the future, the recognition of a person who is known to be a cheater should happen efficiently, that is, accurately and fast. Thus, recognition response latencies should have been clearly lower for faces of cheaters than for faces of persons depicted as trustworthy or irrelevant. In fact, the recognition response latencies assessed with respect to faces of cheaters were lower compared to the irrelevant information condition. However, there was evidently no difference in the promptness of responding to faces of cheaters and faces of persons depicted as trustworthy. It simply seems that not the information on cheating per se accelerated subjects' decision, but the fact that faces associated with somewhat unusual information had been more closely examined in Week One. Plausibly, those faces may have appeared more familiar than faces previously associated with irrelevant information, provoking faster responding.

Given these results, it seemed interesting to focus more strongly on behavior of cheating and trustworthiness as opposite instances of somewhat unusual behavior. Consequently, Experiment 2 was designed as a replication of Experiment 1, yet omitting the irrelevant information condition.

5 Experiment 2

5.1 Overview and Hypotheses

Experiment 2 was equivalent to Experiment 1 with the exception that there were only two conditions of threat potential varied, that is, information on cheating and information on trustworthiness. The same hypotheses as with respect to Experiment 1 were tested.

5.2 Method

5.2.1 Participants

Participants were 84 female and 39 male persons who were paid for participation. Most of them were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 18 to 57 (*M* = 24.89, *SD* = 7.11). None of them had already taken part in Experiment 1.

5.2.2 Materials

The materials used were identical to those of Experiment 1, except for the following modifications. Since only two conditions of threat potential were to be manipulated, 48 facial photos were randomly chosen out of the pictures used in Experiment 1 and assigned to two sets of 24 pictures each. Analogically, 24 descriptive sentences used in Experiment 1 were chosen, namely 12 sentences containing information on cheating and 12 sentences containing information on trustworthiness. Furthermore, half of these descriptions implied information on low and high social status, respectively, as elucidated further above. As for the rating scales, all scales applied were identical to those in Experiment 1.

5.2.3 Procedure

The procedure was exactly the same as in Experiment 1 with the only exception that the duration was reduced because of the smaller number of to-be-judged stimuli. Thus, the first phase of the experiment lasted about 5 minutes, whereas the recognition-test phase lasted 15 minutes, approximately.

5.2.4 Design

The within-subject independent variables were threat potential (history of cheating, history of trustworthiness) and social status (low, high). The dependent measures were the same as in Experiment 1. Differing from Experiment 1, attractiveness ratings as well as response latencies and recognition-test data were analyzed in a 2 (threat potential) × 2 (social status) factorial design, disregarding participant sex as an additional factor. This was done for the following reasons: first of all, the data analyses conducted for Experiment 1 did not reveal any statistically significant effects of the sex-of-respondents factor at all. Secondly, contrary to their assumption that women might perform better on the recognition task than men, Mealey et al. (1996) did not find a statistically significant main effect of participant sex. Instead, they detected a statistically significant interaction of threat potential and participant sex yet, on closer examination, the effect size turned out to be diminutive, namely $\eta^2 = .03$. Thus, analyzing the data of Experiment 2 it seemed justifiable to leave the participant sex variable aside.

Given a total sample size of $N = 123$, $\alpha = .05$, and the assumption that the average population correlation between the levels of the repeated measures factor threat potential is $\rho = .60$ (estimated from pilot data), effects of size $f = .15$ ($\eta^2 = .10$; small to medium effects as defined by Cohen, 1977) could be detected for the threat-potential variable with a probability of $1 - \beta = .96$.

5.3 Results

Parallel to Experiment 1, the results of the present experiment are reported in line with the order of data assessment. Means and standard errors referring to statistically significant effects are parenthesized as far as they are not illustrated in figures.

5.3.1 Week One Data

5.3.1.1 Attractiveness Ratings

Figure 5-1 illustrates the mean attractiveness ratings depending on threat potential and social status. As expected, and replicating the finding of Experiment 1, a 2 × 2 MANOVA revealed that faces of people depicted as cheaters were judged clearly lower in attractiveness than faces of people described as trustworthy, $F(1, 122) = 148.70$, $p < .001$, $\eta^2 = .55$

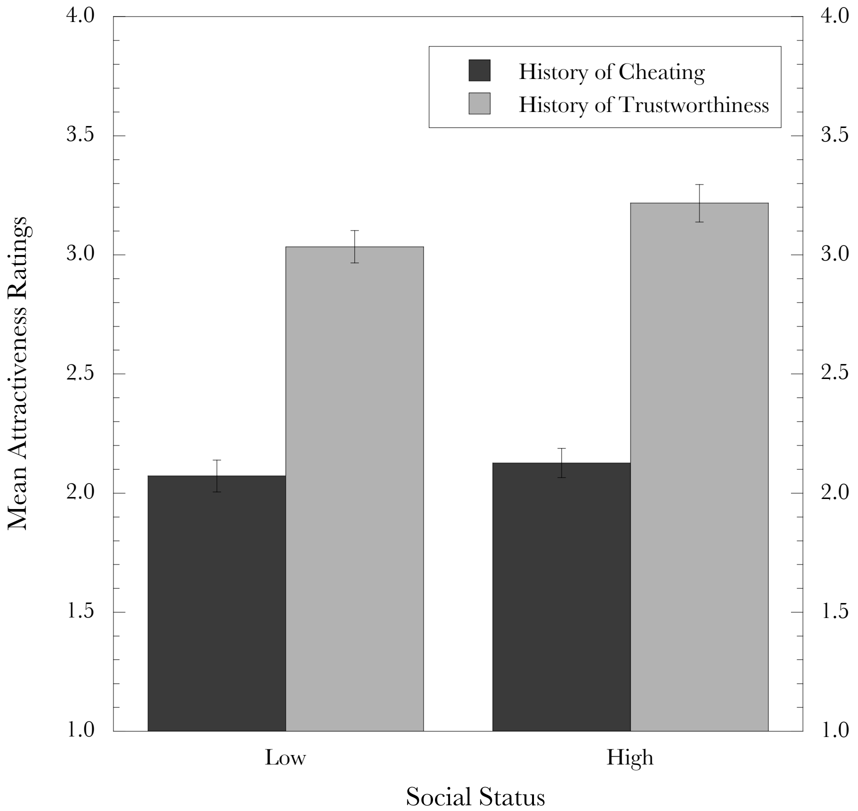

Figure 5-1: Week One mean attractiveness ratings depending on threat potential and social status. On a scale ranging from 1 to 6, subjects rated faces of people described as cheaters significantly lower in attractiveness than faces of people depicted as trustworthy. The error bars represent the standard errors of the means.

(*M*s = 2.10 vs. 3.13, *SE*s = .06 vs. .07 on a scale from 1 to 6.). The expected main effect of social status was also statistically significant, indicating that high-status information was coming along with higher ratings of attractiveness than low-status information, $F(1, 122) = 9.51$, $p = 003$, $\eta^2 = .07$ (*M*s = 2.67 vs. 2.55 *SE*s = .05 vs. .05). As in Experiment 1, there was no statistically significant interaction of threat potential and social status, $F(1, 122) = 2.57$, $p = .112$, $\eta^2 = .02$. Besides, subjects tended to choose relatively low ratings of attractiveness, also replicating the finding of Experiment 1.

5.3.1.2 Response Latencies

Figure 5-2 displays the mean response latencies for the attractiveness ratings. No statistically significant main effect of threat potential could be detected, resembling the finding of Experiment 1 for the response latencies concerning the evaluation of faces of cheaters and of trustworthy persons, $F(1, 122) = 2.93$, $p = .090$, $\eta^2 = .02$. However, the 2 × 2 MANOVA performed on the data revealed a statistically significant main effect of social status, denoting that attractiveness ratings were performed faster in the low-status condition than in the

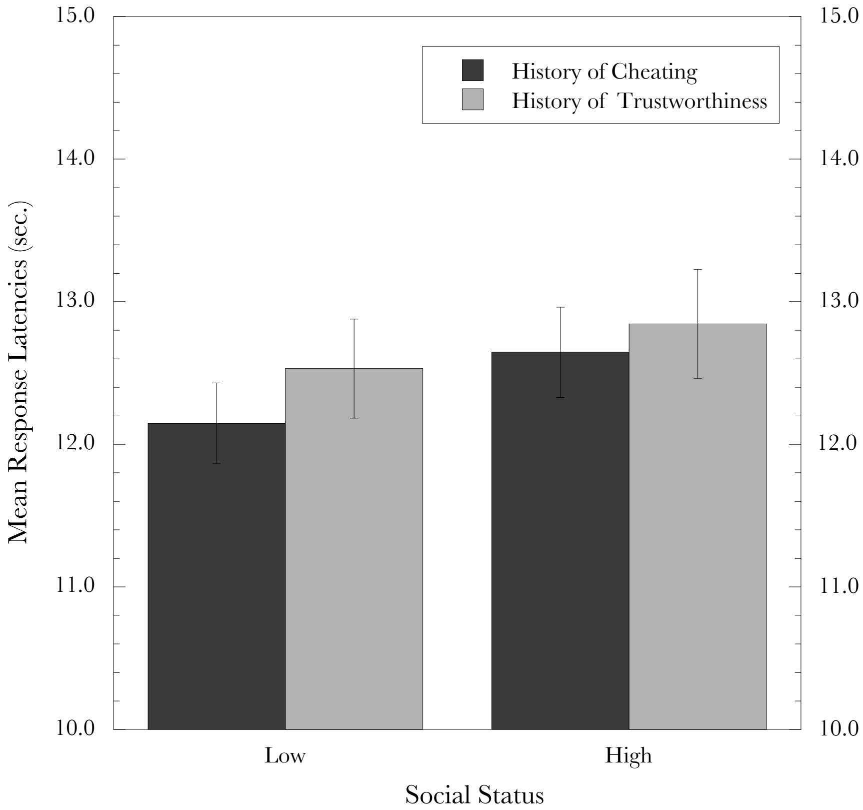

Figure 5-2: Week One mean response latencies depending on threat potential and social status. On average, subjects made their judgments faster in the low-status condition than for the high-status condition. The error bars represent the standard errors of the means.

high-status condition, $F(1, 122) = 4.81$, $p = .030$, $\eta^2 = .04$ (*M*s = 12.34 vs. 12.75 sec., *SE*s = .30 vs. .33 sec.). The interaction between threat potential and social status was not statistically significant, $F(1, 121) < 1$, $p = .581$, $\eta^2 < .01$.

5.3.2 Week Two Data

5.3.2.1 Attractiveness Ratings

Figure 5-3 illustrates the mean attractiveness ratings for Week One faces as a function of threat potential and social status. Replicating the results of Experiment 1, no statistically significant main effects of threat potential and social status emerged, $F(1, 122) = 1.67$, $p = .283$, $\eta^2 < .01$ and $F(1, 122) = 1.57$, $p = .213$, $\eta^2 = .01$, respectively. The interaction of the threat-potential and social-status variables also missed statistical significance, $F(1, 122) < 1$, $p = .539$, $\eta^2 < .01$. Thus, subjects rated the attractiveness of persons depicted as cheaters just as high as they judged the attractiveness of persons described as trustworthy. As Figure 5-4 suggests, this may be ascribed to an increase in the attractiveness ratings for faces of cheaters, whereas there was a decrease in the judgments for people formerly depicted as

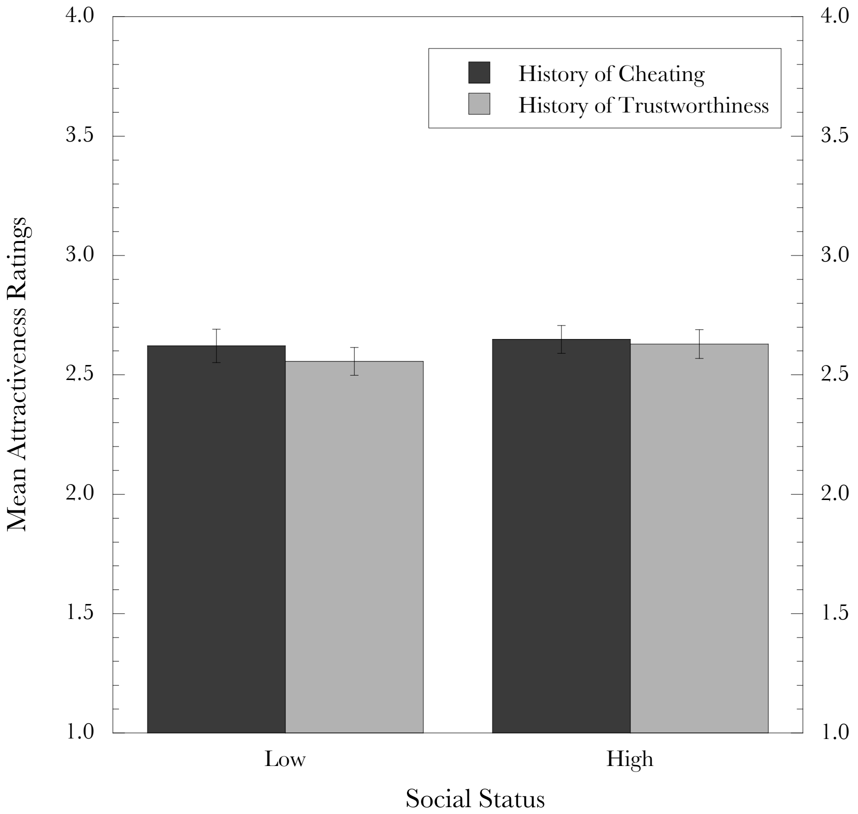

Figure 5-3: Week Two mean attractiveness ratings for faces from Week One depending on threat potential and social status. On a scale ranging from 1 to 6, subjects rated faces of cheaters just as high in attractiveness as faces of people depicted as trustworthy. The error bars represent the standard errors of the means.

trustworthy. Furthermore, a one-way MANOVA performed on the data did not show a statistically significant difference in the judgments of old faces and new distractor faces, $F(1, 122) < 1$, $p = .958$, $\eta^2 < .01$ (*M*s = 2.61 vs. 2.61, *SE*s = .05 vs. .05 on a scale from 1 to 6), replicating the finding of Experiment 1.

5.3.2.2 Recognition-Test Data

Figure 5-5 illustrates the discrimination indices P_r for faces from Week One depending on threat potential and social status. Again, the expected bias for remembering faces of cheaters could not be detected. Replicating the finding of Experiment 1, the 2 × 2 MANOVA performed on the data revealed neither a statistically significant main effect of threat potential, $F(1, 122) < 1$, $p = .545$, $\eta^2 < .01$, nor of social status, $F(1, 122) < 1$, $p = .619$, $\eta^2 < .01$. The interaction between both variables was just statistically significant, $F(1, 122) = 3.93$, $p = .050$, $\eta^2 = .03$. It seems noteworthy that, with respect to the threat potential by social status interaction, the data pattern closely matches that of Experiment 1 in that the highest values of P_r were found for the detection of faces that had formerly been presented together with

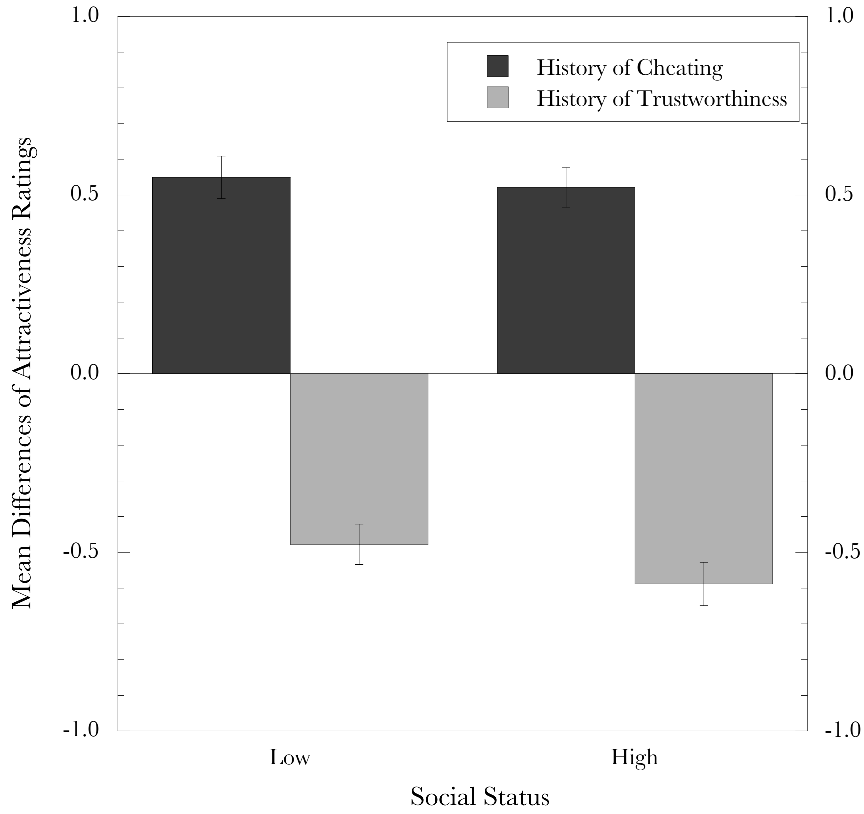

Figure 5-4: Mean differences of attractiveness ratings for Week One faces between Week Two and Week One as a function of threat potential and social status. While attractiveness ratings increased with respect to faces of cheaters, as indicated by positive difference scores, they decreased for faces of persons described as trustworthy, as indicated by negative difference scores. The error bars represent the standard errors of the means.

somewhat unusual behavioral descriptions (i.e., high-status cheating and low-status trustworthiness). Parallel to Experiment 1, a test of whether P_r was different from zero for all cells of the design simultaneously confirmed that recognition performance was above chance, $F(1, 122) = 2494.16$, $p < .001$, $\eta^2 = .95$.

5.3.2.3 Response Latencies

Similar to Experiment 1, the 2 × 2 MANOVA performed on the reaction-time data assessed for the attractiveness ratings yielded neither a statistically significant main effect of threat potential, $F(1, 122) = 3.42$, $p = .067$, $\eta^2 = .03$, nor of social status, $F(1, 122) < 1$, $p = .909$, $\eta^2 < .01$, nor was there a statistically significant interaction, $F(1, 122) = 2.60$, $p = .109$, $\eta^2 = .02$. Likewise, the 2 × 2 MANOVA performed on the recognition response latencies did not reveal a statistically significant main effect of threat potential, $F(1, 122) < 1$, $p = .369$, $\eta^2 < .01$, or of social status, $F(1, 122) < 1$, $p = .916$, $\eta^2 < .01$. Finally, the interaction of

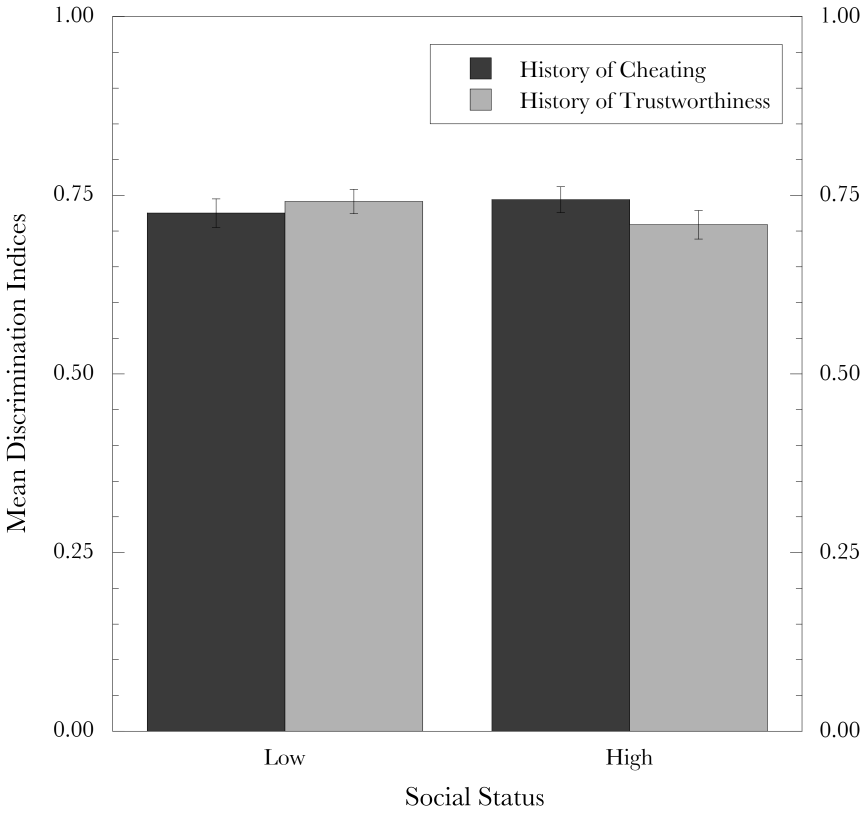

Figure 5-5: Mean discrimination indices P_r for faces from Week One depending on threat potential and social status. The highest values of P_r emerged with respect to high-status cheaters and low-status trustworthy persons. The error bars represent the standard errors of the means.

threat potential and social status just missed the level for statistical significance, $F(1, 122) = 3.67$, $p = .058$, $\eta^2 = .03$. The results are displayed in Figure 5-6.

5.3.2.4 Additional Ratings

Status Ratings

In line with the results of Experiment 1, subjects rated the social status of job titles previously chosen to stand for high-status jobs significantly higher than that of job titles supposed to be of low status, $t(122) = -26.56$, $p < .001$ (*M*s = 2.10 vs. 4.07, *SE*s = .04 vs. .05 on a scale from 1 to 5). Again, all low-status job titles were rated significantly below the median ($Z = 3.03$). Accordingly, all high-status jobs were judged to be significantly higher than the median. Obviously, the manipulation of social status using job titles was successful[46].

[46] See Appendix D for a detailed description of the results.

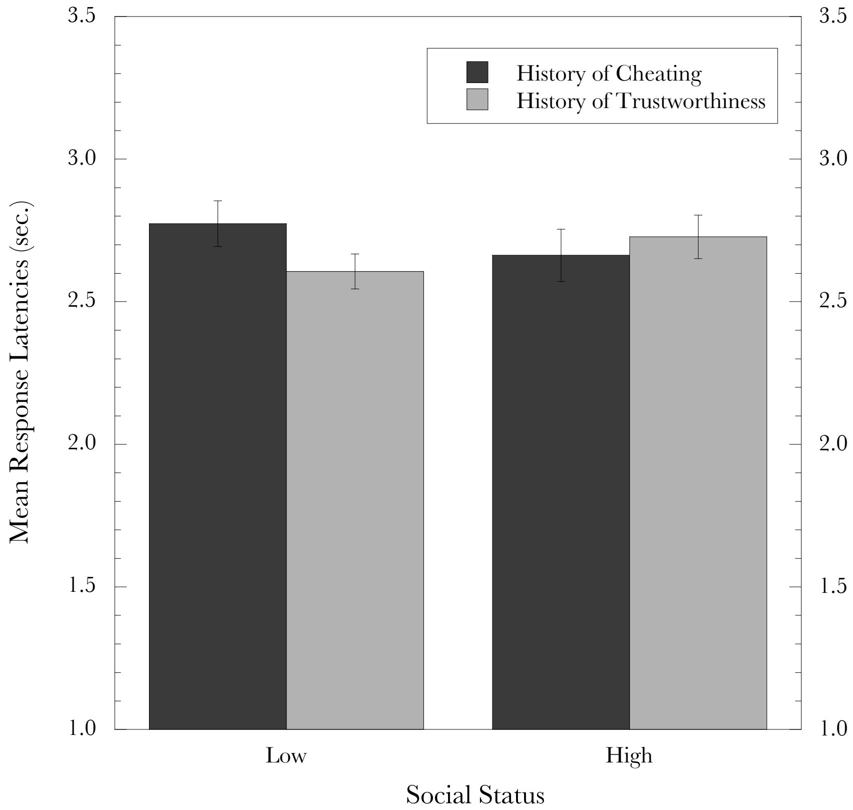

Figure 5-6: Mean recognition response latencies for faces from Week One as a function of threat potential and social status. No statistically significant main effects or interaction were found. The error bars represent the standard errors of the means.

Ratings of Severity and Commendation

As Figure 5-7 displays, subjects perceived the instances of behavior of cheating as notably grave (M = 4.95, SE = .07) and the instances of trustworthiness as comparably commendable (M = 4.91, SE = .08). According to the findings of Experiment 1, the one-way MANOVAs performed on the data revealed a statistically significant main effect of social status for the commendation ratings, $F(1, 122) = 21.32$, $p < .001$, $\eta^2 = .15$, but not for the severity ratings, $F(1, 122) = 3.28$, $p = .072$, $\eta^2 = .03$. Thus, ratings of commendation were significantly higher for the high-status condition than for the low-status condition, whereas the status variable had no effect on the severity ratings.

Relevance Ratings

Figure 5-8 illustrates the mean relevance ratings as a function of threat potential and social status. The 2 × 2 MANOVA performed on the ratings revealed a statistically significant main effect of social status, indicating that subjects perceived behavior of low-status persons as more personally relevant than behavior of high-status persons, $F(1, 122) = 35.47$, p <

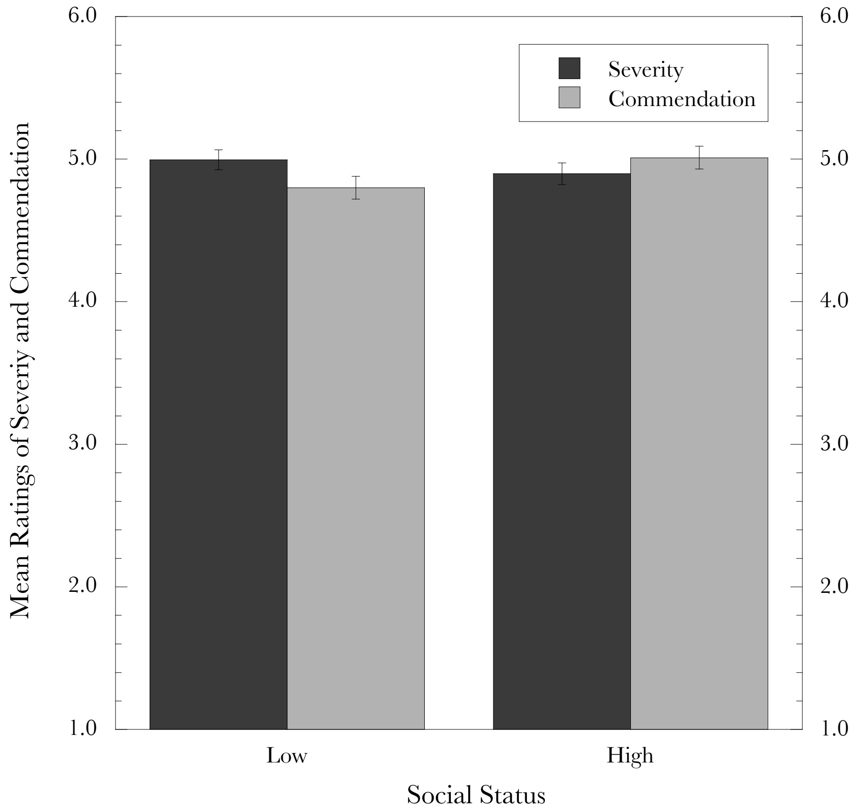

Figure 5-7: Mean ratings of severity and commendation depending on social status. On a scale ranging from 1 to 6, subjects judged instances of behavior of trustworthiness associated with high-status persons as more commendable than trustworthiness of low-status persons. The error bars represent the standard errors of the means.

.001, $\eta^2 = .23$. The main effect of threat potential was not statistically significant, $F(1, 122) < 1$, $p = .823$, $\eta^2 < .01$. However, there was a statistically significant interaction of threat potential and social status, $F(1, 122) = 120.24$, $p < .001$, $\eta^2 = .50$. As evident from Figure 5-8, subjects judged the personal relevance of incidents of cheating to be clearly higher in the low-status condition than in the high-status condition, $t(122) = -10.49$, $p < .001$, $\eta^2 = .47$. In contrast, the personal relevance of instances of trustworthiness was rated significantly higher in the high-status condition compared to the low-status condition, $t(122) = 3.83$, $p < .001$, $\eta^2 = .11$.

5.4 Discussion

The present experiment was designed as a replication of Experiment 1, yet with a stronger focus on behavior of cheating and trustworthiness as instances of threat potential. The results closely match the findings obtained in the first experiment. First of all, the analysis of the recognition-test data did not reveal a cheater-detection effect in that faces of cheaters

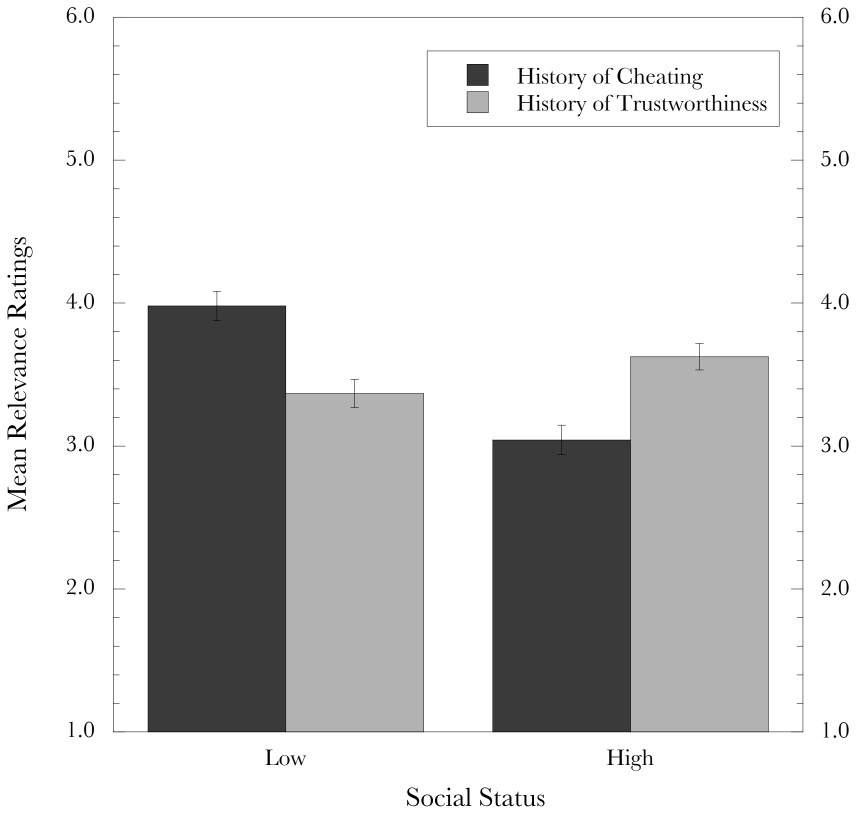

Figure 5-8: Mean relevance ratings as a function of threat potential and social status. On a scale ranging from 1 to 6, subjects estimated behavior of cheating performed by low-status persons as even more personally relevant than behavior of cheating by persons of high social status. The reverse occurred concerning instances of trustworthiness. The error bars represent the standard errors of the means.

were recognized at a higher performance level than faces previously associated with trustworthiness. Likewise, no (implicit) effect of an enhanced memory for faces of cheaters was found with respect to the Week Two attractiveness ratings, also replicating the finding of Experiment 1. However, as in the first experiment, it seems that the manipulation of the threat-potential and social-status variables was successful. For example, the analysis of the Week One attractiveness ratings yielded a noticeable effect of threat potential in that faces of cheaters were judged clearly less attractive than faces of trustworthy persons, replicating Mealey et al. (1996). Moreover, as evident from the severity ratings, subjects evaluated the instances of cheating as notably grave, further justifying the assumption that the threat-potential variable was manipulated appropriately.

In Experiment 1, some effects of threat potential emerged with respect to response latency measures, indicating that subjects distinguished between the conditions of threat potential at least at an implicit level. Yet these effects seemed to be due to the differentiation between common and uncommon behavior rather than to the impact of information on cheating compared to information on non-cheating. The results of the present experiment substantiate this assumption. Omitting the irrelevant information condition, no effects of the threat-potential variable were detected with respect to the response latencies assessed in Week One and, even more important, in Week Two (i.e., the recognition response latencies). Thus, the results obtained so far fail to substantiate the hypothesis of a functionally specialized cheater-detection module affecting subjects' face-recognition performance such that faces of cheaters are more reliably recognized than faces of non-cheaters.

Of course, one may wonder whether this failure to replicate the findings of Mealey et al. (1996) might be ascribed to a lack of statistical power. Yet in both the first and the second experiment the probability of revealing a cheater-detection effect of small to medium size (as defined by Cohen, 1977) was as high as $1 - \beta = .95$. Given the supposedly enormous importance of the proposed cheater-detection mechanism, it appeared reasonable to expect an effect *at least* of small size in terms of Cohen to emerge. However, on closer examination the cheater-detection effect reported by Mealey et al. (1996) turned out to be marginal compared to what Cohen defined as a small effect. Consequently, given that no evidence for an enhanced memory for faces of cheaters was found in the present experiment and in view of the negligible cheater-detection effect observed by Mealey and colleagues, it seems adequate to challenge the pertinence of the underlying cheater-detection hypothesis. Mealey et al. (1996) deduced their assumption of an enhanced memory for faces of cheaters directly from *Social Contract Theory* (Cosmides, 1989; Cosmides & Tooby, 1989). Yet does the conception of a cognitive mechanism functionally specialized for the detection of cheaters actually implicate an enhanced memory for faces of cheaters?

As discussed in Chapter 3, one may call into question whether to simply recognize a person is a sufficient precondition of preventing oneself from future fraud. Reasonably, apart from identifying a given person as a former interactant one should be able to remember specifics of previous encounters with this person. This, in turn, may allow one to arrive at valid conclusions about this person's threat potential with respect to future interactions. Therefore, a modified cheater-detection hypothesis might predict an enhanced memory for information on prior incidents of cheating rather than an enhanced memory for cheaters per se with no concomitant recollection of their previous acts of cheating.

To test this modified hypothesis, Experiment 3 was designed as an extension of Experiment 2, providing additional data with respect to subjects' awareness of source information (i.e., the behavioral information previously characterizing the to-be-recognized faces). Based on findings of E. Brown et al. (1977) whose subjects performed almost perfectly in a face recognition task but poorly in a subsequent test of their memory for source information, it was expected that, in case the modified cheater-detection hypothesis was valid, subjects should notably benefit from a potential cheater-detection module and exhibit reliable source memory for information on behavior of cheating but not on other behavioral information.

Before turning to the detailed report of the method and results of Experiment 3, an outline of the issue of source memory shall be provided, with particular focus on so-called multinomial models of source monitoring that have built the basis for subsequent data analyses.

6 Excursus: Source Memory and Multinomial Models of Source Monitoring

The purpose of the following excursus is to impart essential knowledge on the theoretical background that has been constitutive to the design and analysis of the following experiments. At first, a short definition of source monitoring will be provided as well as an exemplification of the source-monitoring task with respect to Experiment 3. Subsequently, a multinomial approach to the modeling of cognitive processes will be illustrated, providing an alternative to more traditional methods of investigating cognitive processes. As will be shown, the source-monitoring task lends itself naturally to multinomial modeling (Batchelder & Riefer, 1990). Multinomial models of source monitoring will be discussed with respect to a general class of models that have been developed for simple detection and recognition paradigms. Finally, selected multinomial models of source monitoring will be described with respect to Experiments 3 and 4.

6.1 A Definition of Source Monitoring

Source monitoring, or source memory, means memory for contextual information acquired during the encoding of a given item (or fact), indicating the origin of the item (or fact) knowledge (Meiser & Bröder, 2002). The term *source* implies various characteristics that specify the conditions under which memory was acquired such as the spatial, temporal, and social context of an event or the media and modalities through which it was perceived (Johnson, Hashtroudi, & Lindsay, 1993). In the typical source-monitoring experiment subjects are presented with items that come from two different sources. This is followed by a memory test in which subjects are presented with old (target) items and new (distractor) items with the instruction to identify not only which items were originally presented (i.e., targets) but also the source of those items (Riefer, Hu, & Batchelder, 1994). In a special case of the source-monitoring task, *reality monitoring*, subjects have to differentiate between information that was actually presented and information that was internally generated by thought or imagination (Meiser & Bröder, 2002).

Johnson et al. (1993) have provided a theoretical source-monitoring framework based on the central claim that people do not typically directly retrieve an abstract tag or label specifying a memory's source. Instead, it is assumed that activated memory records are evaluated and attributed to particular sources through decision processes performed during remembering. This implies, in turn, that memory for source information can be distinguished from memory for certain items (or facts) per se (Meiser & Bröder, 2002). Therefore, irrespective of the memory for an item, source monitoring refers to a set of cognitive processes involved in making attributions about the origins of memories, knowledge, and beliefs (Johnson et al., 1993).

Over the past 20 years, experimental paradigms using source-monitoring tasks have gained increasing popularity within various domains of psychological research, including developmental psychology, social psychology, and neuropsychology (Bayen, Murnane, & Erdfelder, 1996; Meiser & Bröder, 2002), paving the way for the application of ideas about source memory to a diversity of phenomena such as bilingualism, eyewitness memory, amnesia or cognitive aging (Bayen et al., 1996; Johnson et al., 1993; Meiser & Bröder, 2002). In view of the immense variety of empirical studies concerned with the issue of source monitoring, a more detailed examination of this field of research shall be set aside, since it would clearly go beyond the scope of the present work. Instead, reference shall be made to the comprehensive review on source-monitoring research provided by Johnson et al. (1993). In the following paragraphs, attention shall be given to one facet of the issue of source monitoring that is most relevant for the present work, that is, multinomial models of source monitoring.

6.2 The Multinomial Modeling of Cognitive Processes

In general, multinomial processing tree models rest on the basic assumption that certain cognitive processes may be serial in nature; for example, it seems plausible to assume that storage takes place before retrieval (Hu, 1999). Following this proposition, multinomial models represent cognitive processes in terms of branching trees, expressing the conditional link probabilities from one stage to another stage in terms of parameters. It is presumed that in this hypothesized tree structure there are several choice points, each allowing for several possible choices associated with a certain probability. Finally, each branch of the processing tree results in an observable response category. Multinomial processing tree models have been applied to the analysis of data from a variety of paradigms in the areas of memory, reasoning, perception, and social psychology. Most advantageously, these models provide an opportunity for testing predictions concerning latent cognitive processes in that they constitute statistically based techniques to estimate hypothetical processing parameters from probabilities of events that are per se unobservable (Riefer & Batchelder, 1988).

In the past decades, models of this class have been particularly popular in scientific areas other than psychology such as statistical genetics (Riefer & Batchelder, 1988). Within the field of psychology, the growing interest in multinomial approaches may be attributed to the work of Batchelder, Hu, and Riefer (see, for example, Batchelder & Riefer, 1990; Hu, 1993, 1999; Hu & Batchelder, 1994; Riefer & Batchelder, 1988; Riefer et al., 1994) who pointed out the high potential of this methodology for a variety of productive applications in the context of cognitive psychology. Riefer and Batchelder (1988) point out some advantageous features of multinomial models due to which they may outplay various other methods that are more commonly used for the purpose of investigating cognitive processes (for example, general statistical models such as the general linear model of analysis of variance (ANOVA)). Most notably, the authors highlight the strong accordance of the structure of multinomial models with the fundamental assumption implied by many cognitive theories that human behavior is determined in that the cognitive system can be in a variety of discrete states. Furthermore, it is assumed that whether or not a given behavior falls into a certain observable category is determined by the co-occurrence or absence of these cognitive states. Resting on this notion, multinomial models allow for the direct measurement of the effects of cognitive processes on behavior. Moreover, Riefer and Batchelder (1988) emphasize that this only requires the specification of assumptions on which cognitive processes might be relevant to a given situation and on how these processes might influence behavioral data. In the following, the statistical methodology for multinomial modeling will be outlined with reference to the basic concepts necessary for using multinomial models (as proposed by Riefer & Batchelder, 1988).

Theoretical Orientation

According to the assumption that cognitive processing involves a finite set of discrete processing states and that each behavioral act to be observed in experimental paradigms involving a limited number of behavioral categories arises from exactly one of these underlying cognitive states, multinomial processing tree models formally define the probability of all C_j ($j = 1, \ldots, J$) mutually exclusive behavioral categories observable within a given experimental paradigm as a function of psychologically interpretable parameters θ_s, that is,

$$p(C_j) = f(\theta_s), \text{ with } 0 < \theta_s < 1. \quad (1)$$

θ_s (s = 1, …, S) refers to the functionally independent model parameters that are assumed to be the probabilities of various cognitive events. Thereby, it is assumed that

$$p(C_j) = p_j = \sum_{i=1}^{I} P(C_j|T_i)P(T_i) = \sum_{i=1}^{I} a_{ij}b_i, \quad (2)$$

where a_{ij} is the conditional probability of behavior category C_j given cognitive state T_i, and b_i is the unconditional probability of being in state T_i. Following from Equations (1) and (2), it is assumed that each of the a_{ij} and b_i is determined from a set of parameters θ_s. With the number of unknown parameters smaller than the number of known data probabilities ($S < J$), unique estimates for the parameters may be obtained[47].

Data Representation

The general rationale of multinomial modeling is as follows: in order to generate appropriate data for the model to be applied, an experimental paradigm should first be selected in which data can be classified into a finite number of discrete, observable categories C_j. Suppose that, in the most general case, there are N_j total data observations in C_j, and the data vector of observations for the model may be defined as $\mathbf{D} = (N_1, \ldots, N_j, \ldots, N_J)$. Given the assumption that the data observations are mutually independent and identically distributed with the probability p_j of an observation falling into category C_j, the joint distribution of the data $\mathbf{D}$ is given by the general multinomial model

$$P(\mathbf{D}; p_1, \ldots, p_J) = N! \prod_{j=1}^{J} \frac{p_j^{N_j}}{N_j!}, \quad (3)$$

where $N = \sum_{j=1}^{J} N_j$. The parameter space for this general model may be viewed as

$$\Gamma_J = \{\mathbf{p} = (p_1, \ldots, p_J) \,|\, 0 \le p_j \le 1, \ \sum_{j=1}^{J} p_j = 1\}. \quad (4)$$

[47] Note that acting on Equation (2) holds the problem that the equation contains only $J - 1$ independent known quantities (since the p_j sum to one) and an even larger number of unknown cognitive parameters (that is, $IJ - 1$). Thus, further progress is not possible without imposing on Equation (2) the restriction that the a_{ij} and b_i are determined by a set of θ_s, with their number S no larger than $J - 1$. Given this restriction, the $J - 1$ independent values of p_j are each a function of the cognitive parameters θ_s through Equation (2).

Model Development

Starting from the general model expressed in Equation (3), various substantive models may be defined as restrictions of this superordinate general model. A substantive model assigns a parameter value θ_s to each cognitive event, representing its probability of occurring. The values of p_j in Equation (4) may then be expressed in terms of these postulated parameters. Each θ_s ($1 \leq S < J$) lies within some interval of real numbers I_s such as [0, 1]. Thus, the parameter space for the substantive model is given by

$$\Omega = \left\{ \theta = (\theta_1,\ \ldots,\ \theta_s,\ \ldots,\ \theta_S) \middle| \theta_s \in I_s,\ s = 1,\ \ldots,\ S \right\}. \tag{5}$$

As mentioned above, substantive multinomial models may be derived from the general model by imposing restrictions with respect to the probabilities p_j. As soon as the parameter space of a substantive model is defined in terms of Equation (5), a set of model equations needs to be specified, expressing the probabilities of the data events as a function of the model's parameters. For this purpose, a continuous function **p** from Ω into Γ_J may be specified, giving the probability distribution over the J categories. More precisely, $\mathbf{p}: \Omega \rightarrow \Gamma_J$ specifies how the probabilities of the observable events p_j are determined by the probabilities of the cognitive processes underlying these observable events. Thus, $\mathbf{p}(\theta) = (p_1(\theta), \ldots, p_J(\theta))$ for each θ in Ω. Given the range of **p** is

$$\Omega^* = \left\{ \mathbf{p}(\theta) \middle| \theta \in \Omega \right\}, \tag{6}$$

Ω^* is a proper subset of Γ_J (that is, usually, S-dimensional). A substantive model may be considered as globally identifiable in case **p** is a one-to-one function from Ω to Ω^* (that is, $\theta \neq \theta'$ implies $\mathbf{p}(\theta) \neq \mathbf{p}(\theta')$)[48].

Parameter Estimation

As soon as the data events and a substantive model have been identified, the next step in the multinomial modeling procedure involves using the empirical data to obtain estimates of the model's parameters. Riefer and Batchelder (1988) discuss parameter estimation in terms of maximum likelihood methods. In general, to obtain maximum likelihood estimators for any model of interest, one first needs to generate the likelihood function for that model, expressing the probabilities of the data as a function of the parameter values. As

[48] Note that global identifiability is useful provided that the researcher's goal is to measure a unique value of θ from a given data set **D**.

outlined above, Equation (3) gives the likelihood function for the general multinomial model (preconditioned the equation is viewed as a function of $\mathbf{p} = (p_1, \ldots, p_J)$ given the data $\mathbf{D}$). Thus, the maximum likelihood estimators $\hat{p}_j$ are the values that maximize the likelihood function expressed in Equation (3) for any given data set $\mathbf{D}$, that is,

$$\hat{p}_j(\mathbf{D}) = N_j/N. \quad (7)$$

With respect to substantive models, the maximum likelihood estimators $\hat{\theta}$ are the values of θ in Ω that maximize the restrictive likelihood function for the particular model that may be derived from Equation (3). In case a given substantive model is true and N is sufficiently large, $\hat{\theta}$ is asymptotically unbiased, that is, $E(\hat{\theta}) = \theta$. Furthermore, $\hat{\theta}$ is efficient in that its variance is no larger than that of any other asymptotically unbiased estimator. Moreover, $\hat{\theta}$ follows an asymptotic normal distribution, if suitably standardized[49].

Goodness of Fit

Having computed parameter estimates for the model of interest, the next step in the procedure of multinomial modeling provides for testing whether the specified model adequately fits the data. Naturally, the general multinomial model always fits the data perfectly since it does not impose any restrictions on the true $\mathbf{p}$ in Γ_J. Yet the substantive model usually contains fewer parameters than the general model (that is, $S < J - 1$), requiring that $\mathbf{p}$ be contained within Ω^* (which is an S-dimensional subset of Γ_J). Therefore, the substantive model may be viewed as fitting the data only in case its description of the data is comparable to that achieved by the general model.

As Riefer and Batchelder (1988) point out, the maximum likelihood estimator $\hat{\theta}$ is the member of Ω that minimizes the "distance" between the maximum likelihood estimator $\hat{\mathbf{p}}$ of the general model (given by Equation (7)) and the "fittet" probability distribution $\mathbf{p}(\theta)$ in Ω^*, given by

$$G^2 = [\hat{\mathbf{p}}, \mathbf{p}(\theta)] = 2\sum_{j=1}^{J} N_j \log[N_j/Np_j(\theta)]. \quad (8)$$

---

[49] Riefer and Batchelder (1988) explicate in detail the determination of maximum likelihood estimators $\hat{\theta}$ for substantive models with only one parameter as well as for models with many parameters (which is true for most multinomial models), including the computation of confidence intervals and confidence regions for θ (for models with individual parameters and multiple parameters, respectively). However, for reasons of straightforwardness, a more detailed description of the statistical background shall be set aside in the context of the present work.

In case a given model is true, the asymptotic distribution of $G^2[\hat{\mathbf{p}}, \mathbf{p}(\theta)]$, called the log-likelihood ratio statistic, is a chi-square with $J - S - 1$ degrees of freedom. Thus, with N being sufficiently large the hypothesis may be tested that the model fits the data by computing G^2 and seeing if it is sufficiently small.

Hypothesis Testing

As a final step in multinomial modeling one may consider the testing of various hypotheses about parameter values. For example, given empirical data that come from several independent experimental conditions, one may wish to test the hypothesis that certain parameters have a constant value across the conditions. Generally, any hypothesis for a model's parameters will constitute some restriction of the dimensionality of the parameter space Ω. There are two commonly used restrictions, namely on the one hand, a set of parameters are constrained to be equal to each other, and on the other hand, one or more parameters are set equal to specific values. If such a restriction is put on one or more parameters of a given substantive model M_1 with the parameter space Ω_1, this creates a restricted version M_2 of M_1. Given a total of R restrictions on the parameters specified by the model M_1 (with $R < S$), the restricted model M_2 will have a new parameter space Ω_2 that will be a nested subset contained in Ω_1. Moreover, Ω^*_2 will be a proper subset of Ω^*_1 (usually, an $S - R$ dimensional subset).

If the restricted version of the model is not able to capture the true parameter vector of the model, then the restriction placed on the model's parameters is not a valid one and can be rejected. Thus, assuming that θ is the true parameter vector, one basically tests the null hypothesis that θ lies in the restricted parameter space Ω_2 versus the alternative hypothesis that θ is not in Ω_2. According to the general model fit, the asymptotic test of such hypotheses is based on the statistic

$$G^2 = [\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta})] - G^2[\hat{\mathbf{p}}, \mathbf{p}(\hat{\hat{\theta}})], \tag{9}$$

where $\hat{\theta}$ is the vector of maximum likelihood estimators within Ω_1 for the full version of the model M_1, while $\hat{\hat{\theta}}$ refers to the vector of new maximum likelihood estimators for the restricted model M_2.

The proposed methodology has been implemented in various computer programs, two of which have been designed by Hu (see, for example, Hu, 1999) and Rothkegel (1999). Both programs operate on the basis of the so-called EM algorithm[50] as adapted and extended by Hu, providing an iterative method of obtaining maximum likelihood estimates (for details, see Hu & Batchelder, 1994). For more detailed information on the full statistical background of multinomial models see the work of Riefer, Hu, and Batchelder (e.g., Hu & Batchelder, 1994; Riefer & Batchelder, 1988). For a comprehensive review on multinomial modeling in cognitive psychology you may refer to Erdfelder (2000).

6.2.1 Multinomial Models of Source Monitoring

Multinomial models have been developed for a variety of source-monitoring tasks. For example, different models have been provided that apply to the analysis of data from tasks with two sources (Batchelder & Riefer, 1990; Bayen et al., 1996), three sources (Riefer et al., 1994) or an arbitrary number of sources (Meiser & Bröder, 2002). In general, multinomial models of source monitoring may be viewed as variants of the general class of threshold theories (see, for example, Macmillan & Creelman, 1991; Snodgrass & Corwin, 1988) that have been developed for simple detection and recognition paradigms. Threshold theories rest on the assumption that the decision space in a detection paradigm can be divided into an arbitrary number of discrete states. In a so-called one-high-threshold theory, one threshold defines two memory states, namely recognition and non-recognition. That is, an old (target) item will be correctly identified as old only if crossing a subject's memory threshold. Otherwise, if a target item fails to pass the threshold, it may solely be identified as old or new due to a subject's response bias. By definition, new (distractor) items can never exceed the threshold in the one-high-threshold model. They may only be misidentified as old by guessing from the non-recognition state, depending on a subject's response bias (Snodgrass & Corwin, 1988). In contrast, so-called two-high-threshold models propose two memory thresholds, one for old (target) items and one for new (distractor) items. According to this assumption, a target item will always be recognized as old if crossing the old-recognition threshold, whereas a distractor item will always be identified as new if crossing the new-item threshold. Comparable to the one-high-threshold model, distractor items can by no means cross the old-recognition threshold, and target items can never exceed the new-item threshold. However, items may fall into a state of uncertainty in case they fail to pass either threshold. In this case, an item will be classified as old or new only

[50] Note that »EM« denotes »expectation-maximization« (Hu & Batchelder, 1994).

depending on a subject's response bias, that is, due to guessing processes (Snodgrass & Corwin, 1988).

In the following paragraphs, various multinomial models of source monitoring will be introduced. Introductorily, a one-high-threshold model will be exemplarily outlined. Due to its comparatively simple structure the one-high-threshold model seems appropriate to establish a fundamental understanding of the process of multinomial modeling of source monitoring. Subsequently, more complex two-high-threshold models of source monitoring will be introduced that have been applied to the data analyses in the context of Experiments 3 and 4.

6.2.1.1 A One-High-Threshold Multinomial Model of Source Monitoring

Batchelder and Riefer (1990) presented a multinomial model of source monitoring including a one-high-threshold model of item detection. Their model refers to the standard source-recognition task with two source categories. In this task subjects are asked to distinguish between target items (previously presented in a learning phase) and new distractor items and then discriminate between two different sources of target items. As illustrated in Figure 6-1, separate processing trees are proposed for items from Source A, Source B, and new distractor items. Models of this type are referred to as *joint multinomial models* since they separate the experimental observations into two or more disjoint data sets based on the experimental design. Thus, when there is more than one set of data a possibly different multinomial model applies to each set independently, and the likelihood function is given by the product of the separate multinomial likelihood functions (Riefer & Batchelder, 1988).

Batchelder and Riefer's (1990) model describes subjects' responses to items in a source-monitoring task as a function of a series of hypothetical cognitive processes, that is, target identification, source discrimination, and various response biases. As for the target identification, a target item may, on test trials, either be recognized as old within memory or remain undetected. The parameters D_1 and D_2 define the probabilities of correctly identifying old A and B items, respectively. D_1 and D_2 may differ since experimental factors may create different detection rates. Conditional on the correct detection of an old item, a subject either is or is not able to correctly identify the appropriate item source. The parameters d_1 and d_2 refer to the probabilities of discriminating the source of detected A and B items, respectively. Again, these parameters are allowed to differ. The remaining parameters (that

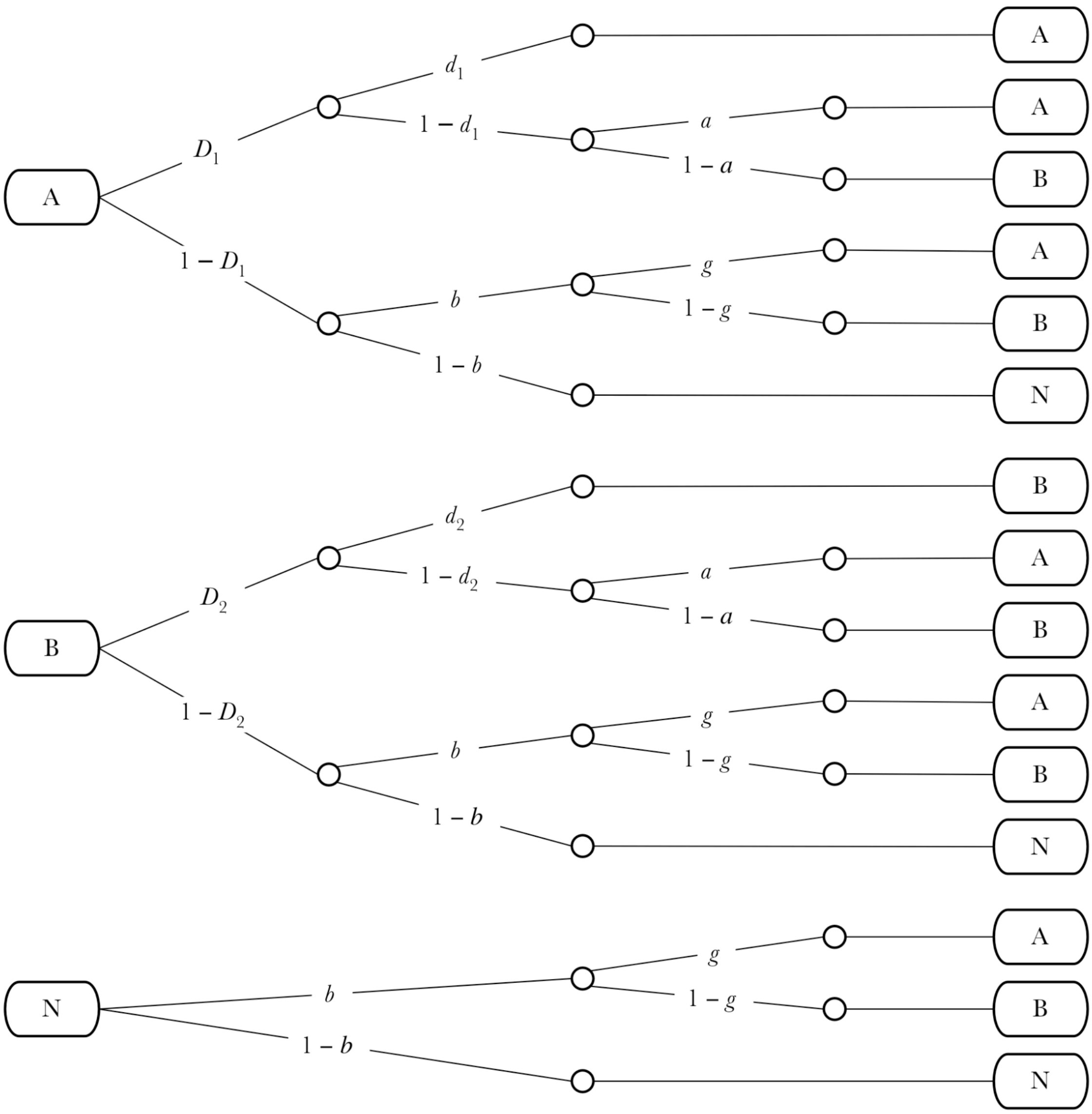

Figure 6-1: Seven-parameter, one-high-threshold joint multinomial model of source monitoring, adapted from Batchelder and Riefer (1990). A = Source A item; B = Source B item; N = distractor item; D_1 = probability of detecting an item from Source A; D_2 = probability of detecting an item from Source B; d_1 = probability of correctly discriminating the source of an item from Source A; d_2 = probability of correctly discriminating the source of an item from Source B; *a* = probability of guessing that a detected item is from Source A; *g* = probability of guessing that an undetected item is from source A, *b* = probability of guessing that a distractor item is old.

is, *a*, *b*, and *g*) denote various response biases. Both parameter *a* and *g* define probabilities of guessing that an item belongs to Source A. However, parameter *a* is conditional on the correct identification of a target as old, whereas parameter *g* is related to undetected items that a subject has guessed as old, which is represented by parameter *b*. Parameters *a* and *g* may differ, since detected and non-detected items might be biased differently.

Thus, the model specified in Figure 6-1 has a total of seven parameters in the parameter space

$$\Omega_7 = \{(D_1, D_2, d_1, d_2, a, b, g)\}. \quad (10)$$

Data Representation

As outlined above, in a source-monitoring experiment with two sources (A and B) subjects are presented a mix of old (target) items from Sources A and B together with some new distractor items (N) in the recognition test. Each item presented then has to be classified as an item of the category A, B or N. Individual subject data can fully be described by a frequency table **T** given as

$$\mathbf{T} = \mathbf{D} = \text{Source} \begin{matrix} \\ \\ \text{A} \\ \text{B} \\ \text{C} \end{matrix} \begin{matrix} & \text{Response} & \\ \text{A} & \text{B} & \text{N} \\ \end{matrix} \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \end{bmatrix} \begin{matrix} Y_1 \\ Y_2 \\ Y_3 \end{matrix}, \quad (11)$$

where Y_{ij} is the frequency of *j*-type responses to *i*-type items. The marginal frequency $Y_i = Y_{i1} + Y_{i2} + Y_{i3}$ is the total number of *i*-type items on the memory test, and usually it is $Y_1 = Y_2$ and $Y_3 = Y_1 + Y_2$ (Batchelder & Riefer, 1990).

Model Equations and Identifiability

As a reminder, multinomial models of source monitoring are proposed to serve the purpose of measuring (unobservable) cognitive processes that underlie observable responses (i.e., source judgments in the present example) falling into certain response categories. As evident from Figure 6-1, there are 15 branches of the tree that combine into nine observable events E_{ij}, corresponding to the Y_{ij} frequencies of the data table **T** in equation (11). Equations for the $p(E_{ij})$ may be written by summing up the branch probabilities over the combined classes. Thus, according to the notation provided in Equation (11), the model equations are, for Source A items,

$$p_{11} = D_1 d_1 + D_1(1 - d_1)a + (1 - D_1)bg, \quad (12a)$$

$$p_{12} = D_1(1 - d_1)(1 - a) + (1 - D_1)b(1 - g), \quad (12b)$$

$$p_{13} = (1 - D_1)(1 - b), \quad (12c)$$

for Source B items,

$$p_{21} = D_2(1 - d_2)a + (1 - D_2)bg, \quad (13a)$$

$$p_{22} = D_2d_2 + D_2(1 - d_2)(1 - a) + (1 - D_2)b(1 - g) \quad (13b)$$

$$p_{23} = (1 - D_2)(1 - b), \quad (13c)$$

and for distractor items,

$$p_{31} = bg, \quad (14a)$$

$$p_{32} = b(1 - g), \quad (14b)$$

$$p_{33} = (1 - b). \quad (14c)$$

The model as depicted in Figure 6-1 and specified by Equations (12) to (14) is not technically identifiably, since it has seven free parameters, but there are only six independent model equations. In other words, the general model has only 6 degrees of freedom, that is, a given **p** might be generated by multiple sets of parameters in Ω_7, complicating the analysis of the model because the classical methods of parameter estimation and hypothesis testing may fail to yield unique estimates of the parameters (Batchelder & Riefer, 1990). However, the non-identifiability problem may be handled considering special cases of Ω_7 by imposing restrictions on the parameters. Batchelder and Riefer (1990) propose three kinds of restrictions on their model that will be exemplified in the following to establish a basic understanding of model development before turning towards more complex models.

As a first restrictive assumption one may consider the target-detection parameters being equal, that is $D_1 = D_2$. Furthermore, the discrimination parameters may be assumed to be equal as well, therefore $d_1 = d_2$. Finally, the guessing rates may also be considered being equal, thus $a = g$. Depending on the combination of the presence or absence of each of the suggested restrictions, eight possible models can be specified, as illustrated in Figure 6-2. Starting from the original seven-parameter model, three restricted models with six parameters each can be specified, containing one of the above-outlined restrictions. Another three models may then be defined by combining two of the three restrictions proposed, resulting in five parameters. Finally, at the bottom of Figure 6-2 there is a four-parameter version of the model, resulting from the combination of all three restrictions suggested above. Note that, statistically, Figure 6-2 represents a nested hierarchy of processing-tree models for

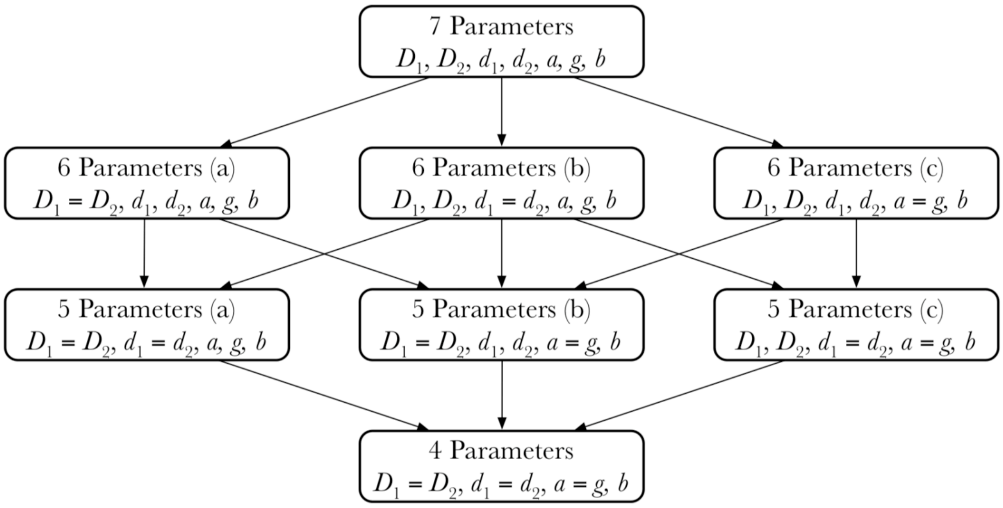

Figure 6-2: Nested hierarchy for the eight versions of the multinomial model depicted in Figure 6-1, adapted from Batchelder and Riefer (1990). Directed arrows indicate subset relations between models.

source monitoring, each of which corresponds to a joint multinomial model for the data structure in Equation (11). Thus, for example, Model 5a is a nested subset of Model 6a or 6b in that any $\mathbf{p} \in \Gamma$ that satisfies Model 5a also satisfies Models 6a and 6b[51].

The two-source model proposed by Batchelder and Riefer (1990) may easily be adapted for an extended number of sources by adding additional processing-tree diagrams for items from each source adjoined. For example, Riefer et al. (1994) reported the corresponding three-source multinomial model, containing four separate processing-tree diagrams for items from Source A, Source B, Source C, and new distractors. However, there has been criticism with respect to one-high-threshold multinomial models of source monitoring, mainly based on general reviews of one-high-threshold theories for detection and recognition paradigms (Bayen et al., 1996). For example, Kinchla (1994) and Batchelder, Riefer, and Hu (1994) discussed the value of one-high-threshold models. Thereby, Kinchla (1994, p. 166) argued against the above-outlined model proposed by Batchelder and Riefer (1990), pointing out the model's ambiguous relationship "to a previously discredited model of signal detection and recognition memory." Indeed, Snodgrass and Corwin (1988, p. 38) already refrained from dwelling on one-high-threshold theories, succinctly pointing out that

[51] Batchelder and Riefer (1990) report in detail additional facts and relationships about the models in Figure 6-2 that are, however, not of proximate relevance with respect to the present work. Moreover, they portray various empirical examples in which they applied their model to data from different studies, comparing the results of the model's analysis with the conclusions of the original investigators. For details, see Batchelder and Riefer (1990).

"the one-high-threshold model is easily falsified by data". However, Bayen et al. (1996, p. 211) pointed out that "it is unclear whether criticisms that can be brought to bear against one-high-threshold models of simple detection carry equal weight when brought to bear against threshold models of decision in the more complex decision environment imposed by the typical source-monitoring task." The authors empirically tested three multinomial models of source monitoring based on different models of decision in a simple detection paradigm (i.e., one-high-threshold, low-threshold, and two-high-threshold models). They concluded that only the two-high-threshold source-monitoring model "provided accurate measures of both item detection and source discrimination across all […] cells of the experimental design" (Bayen et al., 1996, p. 211). Based on this finding, two-high-threshold models of source monitoring were preferred to one-high-threshold models with respect to the data analyses in the context of the following experiments. The next section focuses on the two-high-threshold source-memory model applied to the data analysis of Experiment 3.

6.2.1.2 A Two-High-Threshold Multinomial Model of Source Monitoring

Basically, a two-high-threshold multinomial model of source monitoring may be derived from a one-high-threshold model as illustrated in Figure 6-1 by adding a detection parameter, indicating the probability of a new item being detected as new. Then, any predictions may be derived in the same manner as for the one-high-threshold model. For example, Bayen et al. (1996) presented a two-high-threshold analog to Batchelder and Riefer's (1990) above-outlined one-high-threshold model[52]. Meiser and Bröder (2002) provided a more complex two-high-threshold multinomial model for crossed dimensions of source information. Since their model accurately fits the structure of the frequency data arising from the source-monitoring task of Experiment 3, as will be outlined in detail in Chapter 7, it shall be explicated in the following.

Data Representation

Table 6-1 illustrates the structure of the frequency data that arise from a source-memory experiment with crossed binary source dimensions, with one dimension concerning the discrimination between two sources A and B and the other pertaining to the discrimination between two other sources X and Y. As evident from the table, the source-memory task for crossed binary source dimensions allows for responses in five mutually exclusive response

[52] Note that a modified version of their model has been applied to analyzing the data from Experiment 4 and will be taken up later in the present work.

Table 6-1

Frequency data (*f*) arising from a source-memory task for crossed binary source dimensions, adapted from Meiser and Bröder (2002). Bold values indicate correct responses.

True category		Response category				
		A		B		
		X	Y	X	Y	New
A	X	$\boldsymbol{f_{AX/AX}}$	$f_{AY/AX}$	$f_{BX/AX}$	$f_{BY/AX}$	$f_{New/AX}$
	Y	$f_{AX/AY}$	$\boldsymbol{f_{AY/AY}}$	$f_{BX/AY}$	$f_{BY/AY}$	$f_{New/AY}$
B	X	$f_{AX/BX}$	$f_{AY/BX}$	$\boldsymbol{f_{BX/BX}}$	$f_{BY/BX}$	$f_{New/BX}$
	Y	$f_{AX/BY}$	$f_{AY/BY}$	$f_{BX/BY}$	$\boldsymbol{f_{BY/BY}}$	$f_{New/BY}$
New		$f_{AX/New}$	$f_{AY/New}$	$f_{BX/New}$	$f_{BY/New}$	$\boldsymbol{f_{New/New}}$

categories to items from each of the four source combinations (AX, AY, BX, and BY) as well as to new distractor items. This exactly corresponds to the response paths in the source-monitoring task of Experiment 3, as will be illustrated in the next chapter.

$f_{ij|ij}$ refers to the frequency of correct assignments of items from source combination (i, j) to source i on the first dimension and to source j on the second dimension, with $i \in \{A, B\}$ and $j \in \{X, Y\}$. For $k \neq i$ and $l \neq j$, $f_{il|ij}$ and $f_{kj|ij}$ represent the frequencies of correct source assignments on only one dimension, and $f_{kl|ij}$ denotes incorrect source assignments on both dimensions. Furthermore, $f_{New|ij}$ refers to the frequency of misses of target items from source combination (i, j). With respect to distractor items, $f_{kl|New}$ denotes the frequency of false alarms with assignment to source k on the first dimension and source l on the second dimension, while $f_{New|New}$ refers to the number of correct rejections of new items.

Figure 6-3 illustrates the processing tree for the data structure depicted in Table 6-1. D_{ij} refers to the high-threshold parameter of item detection for target items from source combination (i, j), while D_N denotes the high-threshold parameter of distractor identification. Given that targets from source combination (i, j) are recognized as old, d^1_{ij} indicates the probability of correctly recollecting the first dimension, resulting in correct assignments of

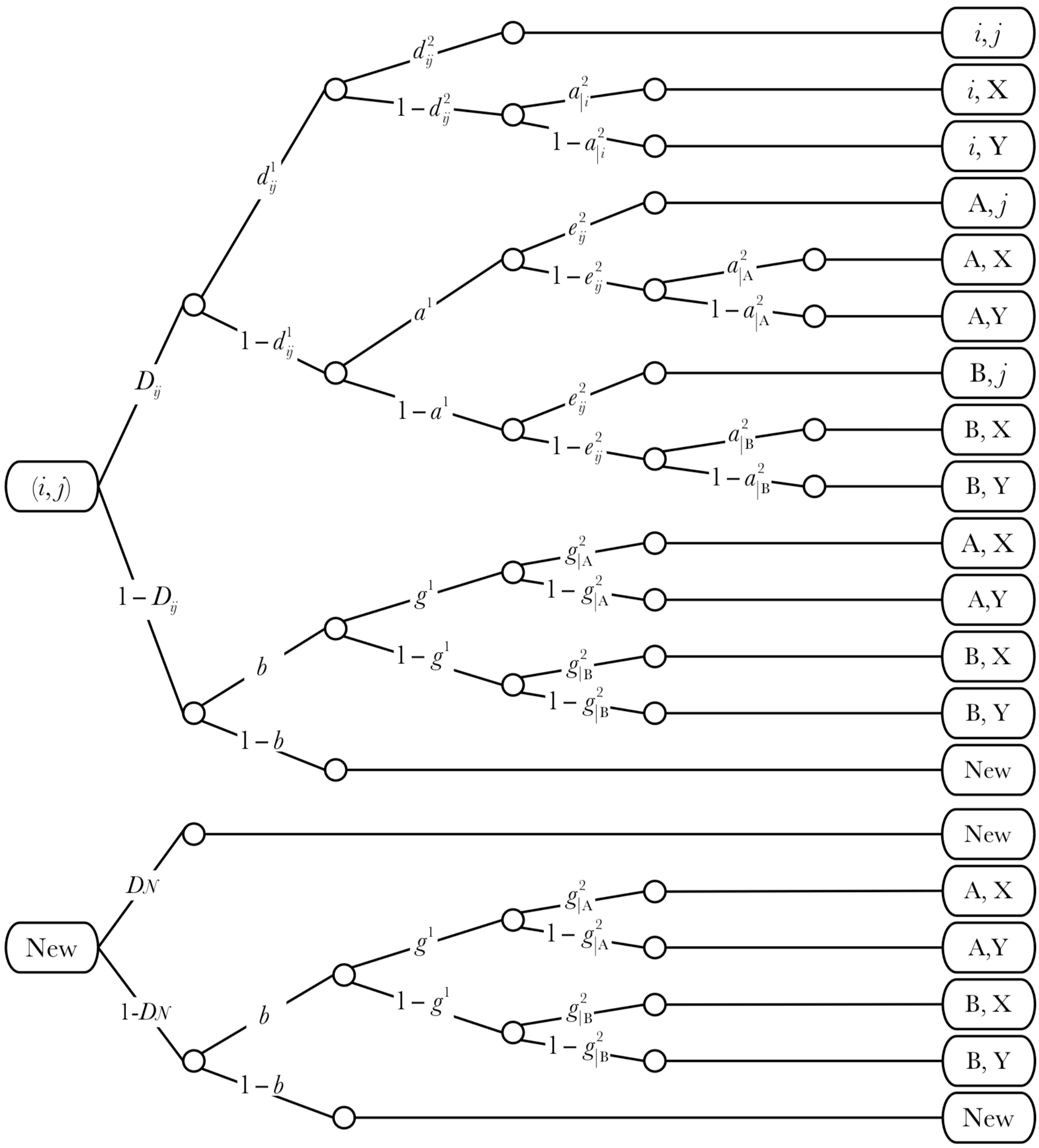

Figure 6-3: Two-high-threshold joint multinomial model of source monitoring for crossed source information, adapted from Meiser and Bröder (2002). (i, j) indicates items from source i of the first dimension and source j of the second dimension, $i \in \{A, B\}$ and $j \in \{X, Y\}$; New denotes distractor items. D_{ij}, D_N = high-threshold parameters of item detection; d_{ij}^1, d_{ij}^2, e_{ij}^2 = parameters of source recollection for recognized target items; a^1, $a_{|A}^2$, $a_{|B}^2$ = parameters of guessing sources for recognized target items; g^1, $g_{|A}^2$, $g_{|B}^2$ = parameters of guessing sources for unrecognized target items and unidentified distractors; b = probability of guessing that an item is old.

source i. However, if source recollection fails on the first dimension with probability $1 - d_{ij}^1$, Source A is selected with guessing probability a^1, and Source B is selected with the complementary guessing probability $1 - a^1$.

Provided that the source category of the first dimension is recollected, the source pertaining to the second dimension is recollected with probability d^2_{ij}. Yet given no recollection of the source category of the first dimension, the source on the second dimension may still be recollected with probability e^2_{ij}, resulting in correct assignments to source *j*. If recollection fails on the second dimension with probability $1-d^2_{ij}$ or $1-e^2_{ij}$, either Source X or Source Y is selected on the basis of a guessing process that may be influenced by the source assignment on the first dimension. Thus, the model contains two conditional guessing proportions for Source X, one given assignment of Source A on the first dimension, $a^2_{|\mathrm{A}}$, and one given assignment of Source B on the first dimension, $a^2_{|\mathrm{B}}$. Source Y, in turn, is selected with probabilities $1-a^2_{|\mathrm{A}}$ and $1-a^2_{|\mathrm{B}}$, respectively. Given that target items are not recognized with probability $1 - D_{ij}$, they may be considered old with guessing probability *b*. Then, items are assigned to either Source A and B on the first dimension with guessing probabilities g^1 and $1 - g^1$, respectively. Note that the guessing proportions for source assignments on the second dimension are specified conditional on source assignments on the first dimension, with $g^2_{|\mathrm{A}}$ denoting the guessing proportion of Source X given assignment to Source A, and $g^2_{|\mathrm{B}}$ referring to the guessing proportion of Source X given assignment to Source B on the first dimension. Consequently, Source Y is selected with probabilities $1-g^2_{|\mathrm{A}}$ and $1-g^2_{|\mathrm{B}}$, respectively. Finally, responses to distractors that are not identified as new with probability $1 - D_{\mathrm{N}}$ are evoked by the same guessing processes as in the case of unrecognized target items.

Meiser and Bröder (2002) assume that the probability of source recollection on the second dimension may be influenced by the source assignment on the first dimension, which is specified by the different parameters d^2_{ij} and e^2_{ij}. That is, the model allows for *stochastic dependence* of source memory for the two dimensions[53]. Complete source recollection of the combination (*i*, *j*) is modeled in terms of the product of the parameters d^1_{ij} and d^2_{ij} (as the intersection of recollecting the individual sources *i* and *j*) rather than as memory for the combination itself. Thus, despite of the assumption of stochastic dependence of the recollection processes for source information on the two dimensions, no single process is specified that reflects memory for the conjunction of the source information. Moreover, the

[53] As Meiser and Bröder (2002) point out, the assumption of stochastic dependence reflects the hypothesis that recollection of the episodic context information on one dimension may facilitate (or cue) the retrieval of information on a different dimension, which one may predict from the principle of encoding specificity (see, for example, Tulving & Thomson, 1973).

model implies that the guessing parameters that correspond to the second dimension are conditional on the outcome of the assignment process on the first dimension[54]. By specifying three *a* and three *g* parameters for the four source combinations, the guessing distributions are saturated in the model (since they may differentially affect assignments to the four source combinations). Furthermore, Meiser and Bröder (2002) point out that the distinction between *a* and *g* parameters is justified, taking into account that metacognitive strategies may result in different guessing proportions for recognized target items compared to non-recognized target items or unidentified distractors.

Of course, starting from the general model illustrated in Figure 6-3, submodels may be specified that reflect alternative assumptions on the representation and retrieval of multidimensional source information. For example, with respect to the premise of stochastic dependence of source memory for the two dimensions, the alternative view of stochastically independent source recollection may be embodied within a submodel that is specified by the parameter restriction $d_{ij}^2 = e_{ij}^2$. Similarly, an all-or-none principle of source memory may be incorporated into a submodel by the parameter fixations $d_{ij}^2 = 1$ and $e_{ij}^2 = 0$ for all source combinations (i, j), representing the assumption that source memory refers to the combination of context dimensions. This, in turn, implies that source information is either recollected jointly on both dimensions or on neither dimension[55].

Model Equations and Identifiability

As outlined on the basis of the one-high-threshold model illustrated in Figure 6-1, the multinomial model for crossed source dimensions can be expressed in terms of various model equations. With $a_{\mathrm{A}}^1 = a^1, a_{\mathrm{B}}^1 = 1 - a^1, a_{\mathrm{X}|i}^2 = a_{|i}^2, a_{\mathrm{Y}|i}^2 = 1 - a_{|i}^2$, and $g_{\mathrm{A}}^1 = g^1, g_{\mathrm{B}}^1 = 1 - g^1, g_{\mathrm{X}|i}^2 = g_{|i}^2$, $g_{\mathrm{Y}|i}^2 = 1 - g_{|i}^2$, the probability of assigning a target of source combination (i, j) to source k on the first dimension and to source l on the second dimension, $i, k \in \{\mathrm{A, B}\}$ and $j, l \in \{\mathrm{X, Y}\}$, can be written as expressed in Equation (15):

---

[54] According to Meiser and Bröder (2002), this assumption accounts for empirical evidence that guessing proportions in source assignments can reflect perceived or expected correlations between item contents and source categories (see, for example, Klauer & Wegener, 1998), and that they may mirror the function of metacognitive strategies tending to minimize assignment errors (see, for example, Hoffman, 1997).

[55] Note that the models of stochastic independence and of an all-or-none principle of source memory represent the extremes on a continuum of gradual dependence among memories for different aspects of the encoding episode (Meiser & Bröder, 2002).

$$p_{kl|ij} = \begin{cases} D_{ij}\left\{d^1_{ij}\left[d^2_{ij} + \left(1-d^2_{ij}\right)a^2_{j|i}\right] + \left(1-d^1_{ij}\right)a^1_i\left[e^2_{ij} + \left(1-e^2_{ij}\right)a^2_{j|i}\right]\right\} + (1-D_{ij})bg^1_i g^2_{j|i} & \text{for } k=i,\, l=j; \\ D_{ij}\left[d^1_{ij}\left(1-d^2_{ij}\right)a^2_{l|i} + \left(1-d^1_{ij}\right)a^1_i\left(1-e^2_{ij}\right)a^2_{l|i}\right] + (1-D_{ij})bg^1_i g^2_{l|i} & \text{for } k=i,\, l\neq j; \\ D_{ij}\left(1-d^1_{ij}\right)a^1_k\left[e^2_{ij} + \left(1-e^2_{ij}\right)a^2_{j|k}\right] + (1-D_{ij})bg^1_k g^2_{j|k} & \text{for } k\neq i,\, l=j; \\ D_{ij}\left(1-d^1_{ij}\right)a^1_k\left(1-e^2_{ij}\right)a^2_{l|k} + (1-D_{ij})bg^1_k g^2_{l|k} & \text{for } k\neq i,\, l\neq j. \end{cases}$$

The probability of missing a target item from source combination (i, j) is given by

$$p_{New|ij} = \left(1-D_{ij}\right)\left(1-b\right). \tag{16}$$

Furthermore, distractors elicit false alarms and are assigned to sources k and l with probability

$$p_{kl|New} = (1-D_{\text{N}})bg^1_k g^2_{l|k}, \tag{17}$$

whereas distractors are correctly rejected as new with probability

$$p_{New|New} = D_{\text{N}} + \left(1-D_{\text{N}}\right)\left(1-b\right). \tag{18}$$

Taken together, the model specified in Figure 6-3 has a total of 24 parameters in the parameter space

$$\Omega_{24} = \{D_{\text{AX}}, D_{\text{AY}}, D_{\text{BX}}, D_{\text{BY}}, D_{\text{N}}, d^1_{\text{AX}}, d^1_{\text{AY}}, d^1_{\text{BX}}, d^1_{\text{BY}}, d^2_{\text{AX}}, d^2_{\text{AY}}, d^2_{\text{BX}}, d^2_{\text{BY}}, e^2_{\text{AX}}, e^2_{\text{AY}}, e^2_{\text{BX}}, e^2_{\text{BY}}, a^1, a^2_{|\text{A}}, a^2_{|\text{B}}, g^1, g^2_{|\text{A}}, g^2_{|\text{B}}, b\} \tag{19}$$

and thus violates the necessary condition for identifiability: there are more parameters than unrestricted probabilities, or frequencies, in the unconstrained model. Therefore, the parameter space needs to be reduced by appropriate restrictions on the model's parameters, resulting in identifiable submodels. Meiser and Bröder (2002) discuss numerous restrictions that shall be briefly outlined in the following.

At first, the model may be simplified by imposing the restrictive assumption that the probability of distractor identification equals the probability of target item recognition, that is, $D_{\text{N}} = D_{ij}$ for one or more source combinations (i, j). As the authors point out, this assumption has been suggested as a standard restriction in two-high-threshold models and it has proved to be empirically tenable. Note, however, that a conceptual decision may be necessary for the specification of D_{N} since the model contains four high-threshold parameters of item recognition and, as a consequence, it may be difficult to determine which of these parameters should be equated with D_N. Therefore, Meiser and Bröder (2002) suggest to initially test for the equality of item recognition parameters D_{ij}. Based on this preliminary analysis, restrictions including D_N may be specified. The authors argue that, in case D_N is set to be equal to at least one item recognition parameter, the remaining model parameters corresponding to the old-new decision are identifiable irrespective of the further model specifications[56].

The authors discuss additional restrictions with respect to the parameters that correspond to source assignments of recognized target items, that is, d^1_{ij} and a^1 (for the first source dimension) as well as d^2_{ij}, e^2_{ij}, $a^2_{|A}$, and a^2_B (for the second source dimension). Figure 6-4 illustrates several sets of such restrictions, yielding globally identifiable submodels. As previously outlined with regard to Batchelder and Riefer's (1990) one-high-threshold multinomial model, arrows between different sets of restrictions indicate hierarchical relations among the resulting submodels (that is, relations in which one model entails the other model as a special case).

Empirical Validation

Meiser and Bröder (2002) report two experiments in which they applied their multinomial model for crossed source information. For the present purposes it suffices to outline some of the conclusions that the authors drew from their empirical validation (Meiser & Bröder, 2002, p. 133):

> […] the model shows an excellent fit to empirical data and […] the source recollection parameters for different context dimensions are sensitive to experimental manipulations and dissociable from each other. Thus, by indicating that the multinomial memory model for crossed source information meets the criteria for empirical and psychological validity, the results lend strong support to the veridicality of the model as a measurement tool for multidimensional source memory.

Moreover, the results supported the authors' assumption of differential guessing processes in source assignments for recognized and unrecognized stimuli. Finally, the analyses ruled out an all-or-none principle of source recollection, yet leaving the case of stochastically dependent versus independent memory for various aspects of an encoding episode ambigu-

[56] That is, the memory parameters D and the guessing parameters b, as well as the guessing parameters for source assignments of unrecognized items g^1, $g^2_{|A}$ and $g^2_{|B}$. For more detailed information on the identifiably of the model, see the appendix of Meiser and Bröder (2002).

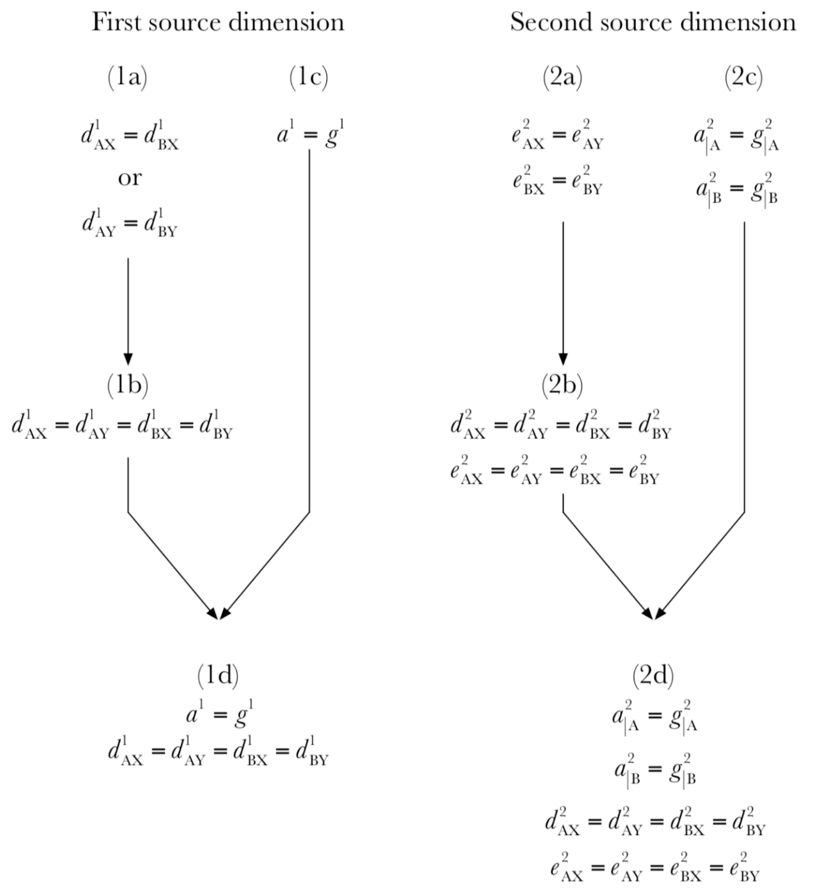

Figure 6-4: Restrictions on the source recollection and source guessing parameters that render globally identifiable submodels of the multinomial source-memory model for crossed source information, adapted from Meiser and Bröder (2002). Arrows indicate hierarchical relations among the resulting submodels.

ous. With respect to the model specification, Meiser and Bröder found that the order of the source dimensions in the model did not have an effect on the outcome of the model selection procedure. Taken together, Meiser and Bröder's multinomial model for crossed source information provided a solid basis for own analyses, which will be reported in the following.

7 Experiment 3

7.1 Overview and Hypotheses

Experiment 3 was similar to Experiment 2 with few procedural differences that will be reported in the following. First of all, the retention interval was minimized, that is, the recognition-test phase immediately followed the acquisition phase. This was done to make sure that, in case the hypothesized cheater-detection effect on the recognition-test data repeatedly failed to appear, this might not be due to the duration of the retention interval.

Moreover, since the threat-potential variable had no effect on the Week Two attractiveness ratings in any of the previous experiments, no ratings of attractiveness were assessed in the recognition-test phase. Instead, only old-new judgments were obtained but specified by means of source-monitoring data. As discussed at the end of Chapter 5, the results of Experiments 1 and 2 failed to substantiate Mealey et al.'s (1996) hypothesis of an enhanced memory for faces of cheaters due to a functionally specialized mental module. This gave reason to consider the actual pertinence of the cheater-detection hypothesis as deduced from *Social Contract Theory* (Cosmides, 1989; Cosmides & Tooby, 1989) by Mealey and colleagues.

Consequently, the present experiment was designed to test a modified cheater-detection hypothesis. It was expected that subjects might rather show an enhanced memory for information on prior cheating than an enhanced memory for cheaters per se (without a concomitant recollection of their previous behavior). More precisely, it was hypothesized that, given the cheater-detection hypothesis as proposed by Mealey et al. (1996) was valid, the multinomial analysis of the source-monitoring data should yield the highest item-detection parameters D_{ij} for faces of cheaters, which should exceed the remaining detection parameters. However, if the modified hypothesis was true, a cheater-detection effect should be reflected in an enhanced memory for source information on cheating, represented by high source-recollection parameters d_{ij}^2 (and possibly e_{ij}^2) with respect to information on cheating, which should exceed the remaining source-recollection parameters for source assignments on the second dimension (threat potential).

7.2 Method

7.2.1 Participants

Participants were 41 female and 23 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their age ranged from 20 to 58 (M = 26.73, SD = 6.39). None of them had already taken part in Experiments 1 or 2. As in the previous experiments, they were paid for participation.

7.2.2 Materials

The materials used were mostly identical to those of Experiment 1, except for the following modifications. The number of to-be-judged faces was increased up to ten per Week One condition (instead of six as in the previous experiments). Thus, the 72 facial photos used for Experiment 1 and another eight pictures out of the picture set primarily selected from the Internet were randomly assigned to two sets of 40 pictures each. Accordingly, the 24 descriptive sentences used for Experiment 2 were chosen, and 16 new behavioral descriptions were added. Previously, a group of subjects (N = 33) had judged the job titles and behavioral descriptions with respect to their social status and valence[57]. All rating scales applied were identical to those in Experiments 1 and 2.

7.2.3 Procedure

In principle, the procedure was identical to Experiments 1 and 2. Yet, as mentioned previously, the following modifications were implemented. First, the recognition-test phase immediately followed the acquisition phase. Second, no ratings of attractiveness were assessed in the recognition-test phase. Third, the old-new judgments obtained in the recognition test were specified by means of source-monitoring data. Thus, in the recognition-test phase, subjects were randomly presented 80 photos (i.e., picture set 1 and 2), half of which had been presented before and half of which were new. As in Experiments 1 and 2, two check boxes labeled "old" and "new" appeared 1000 ms after the picture onset, and subjects were instructed to make their decision by clicking into one of the boxes with the computer mouse. In case of a "new" judgment, clicking the *continue* button initiated the next trial after a fixed time frame of 500 ms. In contrast, when evaluating a photo as old, clicking the *con-*

[57] See Appendix A for a full list of the selected job titles and Appendix B for a complete list of the behavioral descriptions.

tinue button induced the assessment of source-monitoring data. Subjects had been informed that, in the first part of the experiment, some of the to-be-judged persons held high-status jobs and some held low-status jobs and were described as either cheating or trustworthy independently from their social status. Furthermore, they were instructed to recollect a person's social status and to denote whether a given picture had been associated with behavior of cheating or trustworthiness during the acquisition phase. Therefore, following an "old" judgment, two check boxes labeled "high status job" and "low status job" were presented. Immediately after choosing one of these options and clicking the *continue* button, two other check boxes labeled "behavior of cheating" and "behavior of trustworthiness" appeared. Subsequently, clicking the *continue* button initiated the next trial after an intertrial interval of 500 ms.

7.2.4 Design

In line with Experiment 2, the within-subject independent variables were threat potential (history of cheating, history of trustworthiness) and social status (low, high). The Phase One dependent measures were attractiveness ratings and response latencies. As elucidated above, the main Phase Two dependent measures were old-new ratings aside from source-monitoring data and reaction times assessed for all of the afore-mentioned ratings. Furthermore, and in line with Experiments 1 and 2, status ratings, ratings of severity and commendation as well as relevance ratings were acquired.

According to Experiment 2, attractiveness ratings as well as response latencies and recognition-test data were analyzed in a 2 (threat potential) × 2 (social status) factorial design. Source-monitoring data were analyzed using Meiser and Bröder's (2002) multinomial model of source memory described in the previous chapter. Given a total sample size of $N = 64$, $\alpha = .05$, and the assumption that the average population correlation between the levels of the repeated measures factor threat potential is $\rho = .60$ (estimated from previous experiments), effects of size $f = .20$ ($\eta^2 = .17$) could be detected for the threat-potential variable with a probability of $1 - \beta = .94$.

7.3 Results

In line with Experiments 1 and 2, means and standard errors referring to statistically significant effects are parenthesized as far as they are not illustrated in figures.

7.3.1 Phase One Data

7.3.1.1 Attractiveness Ratings

Figure 7-1 illustrates the mean attractiveness ratings depending on threat potential and social status. In line with the results of Experiments 1 and 2, subjects rated the attractiveness of persons depicted as trustworthy clearly higher than the attractiveness of persons described as cheaters. Accordingly, the 2 × 2 MANOVA showed a statistically significant main effect of threat potential, $F(1, 63) = 92.40$, $p < .001$, $\eta^2 = .60$ (*M*s = 3.36 vs. 2.17, *SE*s = .11 vs. .08). Surprisingly, no main effect of social status was found, $F(1, 63) = 1.93$, $p = .169$, $\eta^2 = .03$, as well as no interaction of the threat-potential and social-status variables, $F(1, 63) = 1.60$, $p = .210$, $\eta^2 = .03$.

7.3.1.2 Response Latencies

Figure 7-2 displays the mean response latencies as a function of threat potential and social status. Although the data pattern looks similar to that found in Experiment 1, the 2 × 2

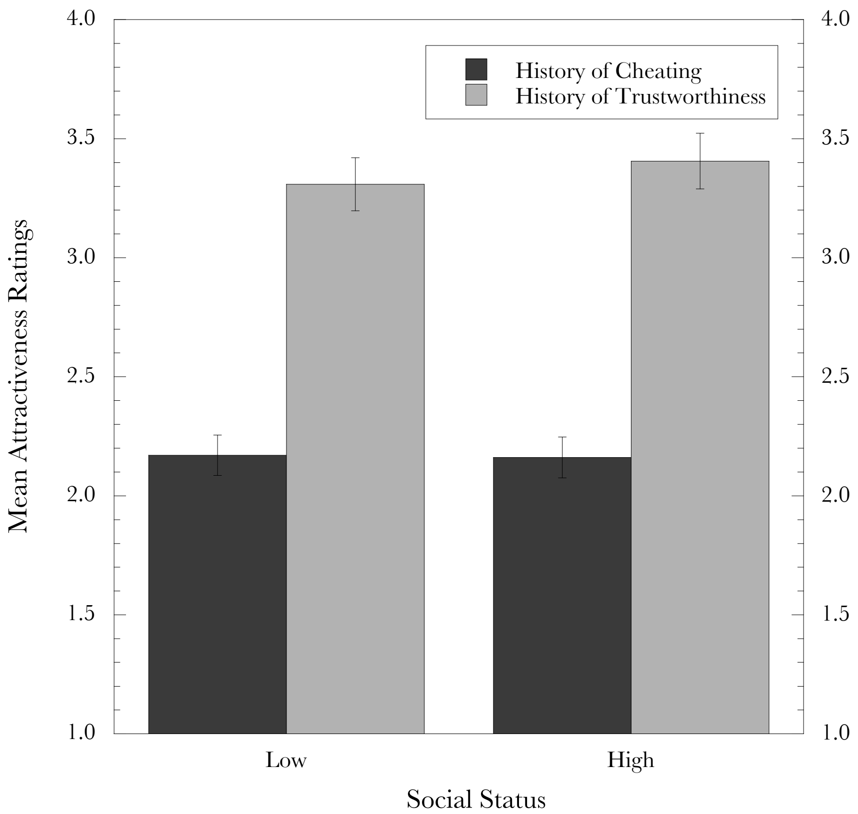

Figure 7-1: Phase One mean attractiveness ratings as a function of threat potential and social status. On a scale ranging from 1 to 6, subjects rated persons described as cheaters clearly lower in attractiveness than persons depicted as trustworthy. The error bars represent the standard errors of the means.

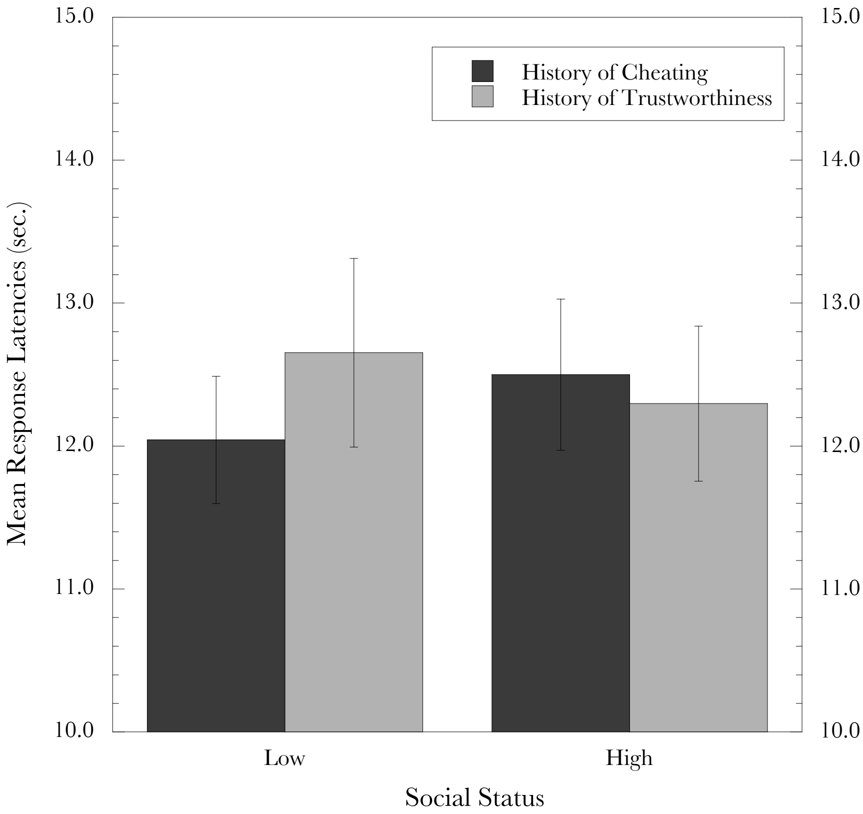

Figure 7-2: Phase One mean response latencies depending on threat potential and social status. No statistically significant main effects or interactions emerged. The error bars represent the standard errors of the means.

MANOVA conducted revealed neither a statistically significant main effect of threat potential nor of social status, $F(1, 63) < 1$, $p = .419$, $\eta^2 = .01$, and $F(1, 63) < 1$, $p = .750$, $\eta^2 < .01$, respectively. There was also no statistically significant interaction of threat potential and social status, $F(1, 63) = 1.92$, $p = .171$, $\eta^2 = .03$. Thus, the time spent for the attractiveness ratings did not vary between the conditions of threat potential and social status, indicating that possible differences concerning the recognition-test data may not be due to effects of encoding.

7.3.2 Phase Two Data

7.3.2.1 Recognition-Test Data

Figure 7-3 illustrates the discrimination indices P_r for faces from Phase One depending on threat potential and social status. As evident from the figure, the highest value of P_r was obtained for the detection of faces of high status cheaters. However, replicating the results of the previous experiments, the 2 × 2 MANOVA performed on the data did not reveal a statistically significant main effect of threat potential, $F(1, 63) < 1$, $p = .762$, $\eta^2 < .01$. There

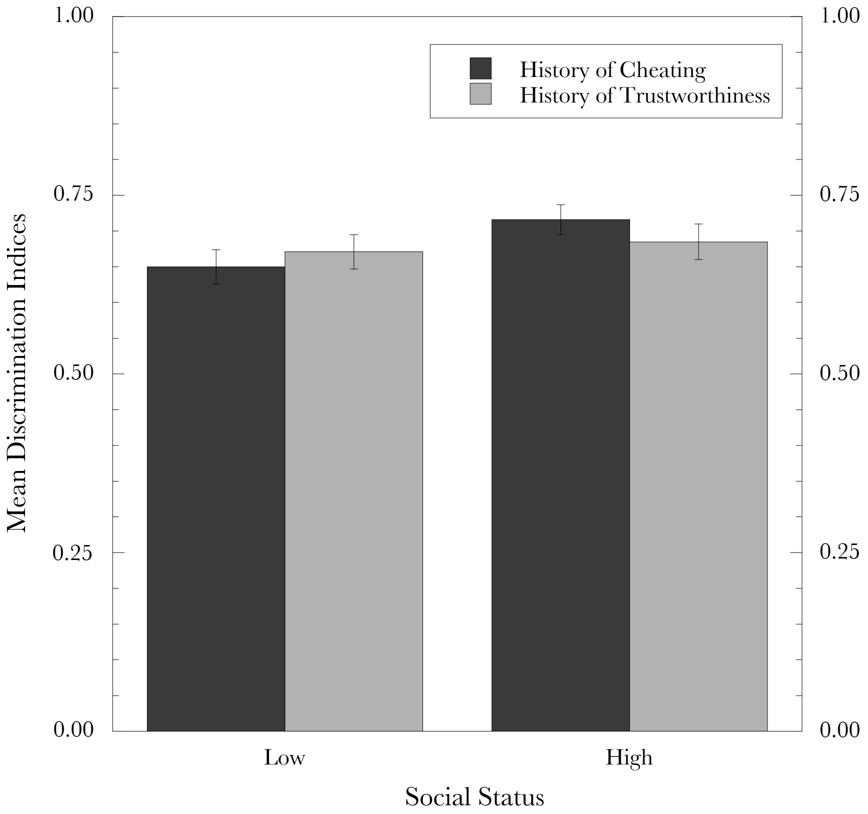

Figure 7-3: Mean discrimination indices P_r for faces from Phase One depending on threat potential and social status. Apparently, subjects performed slightly better in recognizing faces of persons associated with high compared to low social status. The error bars represent the standard errors of the means.

was also no statistically significant interaction between the threat-potential and social-status variables, $F(1, 63) = 2.91$, $p = .093$, $\eta^2 = .04$. Yet the analysis revealed a statistically significant main effect of social status, indicating that high-status persons were identified as old more likely than low-status persons, $F(1, 63) = 4.96$, $p = .030$, $\eta^2 = .07$ (*M*s = .70 vs. .66, *SE*s = .02 vs. .02). It seems noteworthy that the data pattern closely matches that of Experiments 1 and 2. Moreover, replicating the findings of Experiments 1 and 2, the test of whether P_r was different from zero for all cells of the design simultaneously revealed that recognition performance was clearly above chance, $F(1, 63) = 1313.07$, $p < .001$, $\eta^2 = .95$.

7.3.2.2 Source-Monitoring Data

Table 7-1 displays the response frequencies in the source-monitoring task with source dimensions threat potential and social status. The data were analyzed using the multinomial memory model for crossed source information of Meiser and Bröder (2002). The order of the source dimensions in the model was specified according to the temporal order of source decisions in the memory task. Thus, social status was considered the first dimension while

Table 7-1

Response frequencies in the source-monitoring task for threat potential and social status. Bold values indicate correct source assignments.

		Response category				
		Cheating		Trustworthiness		
True category		High Status	Low Status	High Status	Low Status	New
Cheating	High Status	**204**	115	111	81	32
	Low Status	109	**157**	61	93	20
Trustworthiness	High Status	98	76	**197**	106	30
	Low Status	77	98	99	**179**	37
New		152	194	172	181	**2441**

threat potential represented the second dimension. Note, however, that the outcome of the model-selection procedure reported in the following was not at all affected by reversing the order of dimensions in the model, which has also been observed by Meiser and Bröder (2002). Apart from this, note that no temporal order of the cognitive processes yielding source judgments on either of the two source dimensions is implicated per se by the model's tree structure. The criterion of statistical significance for tests of goodness-of-fit was set to $\alpha = .05$.

For lack of space, the processing tree representation of the model for source dimensions social status and threat potential is illustrated in extenso in Appendix E. The model defines 62 response paths by means of parameters for target item detection, distractor identification, source recollection, and guessing processes. The multinomial analyses were carried out using the AppleTree program (Rothkegel, 1998). Following Meiser and Bröder (2002), it was initially examined whether the high-threshold parameters of target item recognition were equal across the four source combinations. The test yielded a statistically non-significant result, $G^2(3) = 7.41$, $p = .060$, supporting the assumption of equal high-threshold parameters. Batchelder and Riefer (1990) suggested a different test for equality of the detection rates; with respect to their two-source multinomial model, the authors point out that the restriction of equal detection rates for target items leads naturally to the statistical null hypothesis that $p_{13} = p_{23}$, as evident from Equations (12c) and (13c). This hypothesis can be tested by a chi-square test of equality of independent proportions defined from **T** in Equation (11). An equivalent test was calculated for the present data and yielded a non-significant result, $\chi^2(3) = 5.13$ ($\chi^2_{crit.} = 7.81$). Therefore, the hypothesis of equal high-threshold parameters for target items could be maintained, and following Meiser and Bröder (2002) the equality constraints $D_{HC} = D_{HT} = D_{LC} = D_{LT} = D_N$ were specified in the multinomial model.

According to Meiser and Bröder (2002), the most parsimonious submodel was specified at first by imposing on the model restrictions (1d) and (2d) from Figure 6-4. The restrictions implicate that the guessing proportions of source assignments for recognized items equal those for unrecognized items, that is,

$$a^1 = g^1, \tag{20}$$

$$a^2_{|H} = g^2_{|H}, \tag{21}$$

$$a^2_{|L} = g^2_{|L}, \tag{22}$$

and that the source-recollection parameters are invariant across source combinations for both dimensions, that is,

$$d^1_{HC} = d^1_{HT} = d^1_{LC} = d^1_{LT} = d^1, \tag{23}$$

$$d^2_{HC} = d^2_{HT} = d^2_{LC} = d^2_{LT} = d^2, \tag{24}$$

$$e^2_{HC} = e^2_{HT} = e^2_{LC} = e^2_{LT} = e^2. \tag{25}$$

Conveniently, this most restrictive submodel with the fewest unconstrained model parameters comprises all restrictions that may be imposed on the model to test the cheater-detection hypotheses. First, the restriction that all target item detection parameters D_{ij} are equal implies the null hypothesis that the probability of identifying a target item as old is independent of an item's source. In contrast, the alternative hypothesis of an enhanced memory for faces of cheaters implies that the target item detection parameters differ among source combinations, with higher detection parameters for the detection of faces of cheaters compared to faces of trustworthy persons. Second, the restriction that the source-recollection parameters d^2_{ij} and e^2_{ij} are equal for the second source dimension (threat potential) implies the null hypothesis that the threat-potential variable has no impact on source assignments. In contrast, the alternative hypothesis of an enhanced memory for source information on cheating implies that the source-recollection parameters for source dimension two (threat potential) differ among source combinations, with higher source-recollection parameters for source assignments of information on cheating compared to information on trustworthiness.

Despite of these demanding restrictions, the selected submodel yielded an excellent fit to the data, $G^2(12) = 13.87$, $p = .309$. Table 7-2 shows the parameter estimates in the specified submodel for crossed source information together with their 95% asymptotic confidence intervals. In line with the results of the conventional data analysis of the discrimination indices P_r, the item detection parameter $D = .68$ suggested that recognition performance was above chance. In fact, fixing the detection parameter to $D = .50$ (that is, chance level) yielded a significant misfit of the model to the data, $G^2(13) = 292.99$, $p < .001$. Most notably, however, the equality restrictions of the source-recollection parameters d^2_{ij} and e^2_{ij} for source dimension two (threat potential) in no way affected the model fit. Apparently, source recollection was not enhanced for information on cheating compared to information on trustworthiness. Taken together, these results fail to support Mealey et al.'s (1996) hypothesis of an enhanced memory for faces of cheaters as well as the modified hypothesis of an enhanced memory for source information on cheating due to a functionally specialized cheater-detection module.

7.3.2.3 Response Latencies

In line with Experiment 2, the 2 × 2 MANOVA conducted for the recognition response latencies revealed neither a statistically significant main effect of threat potential, $F(1, 63) < 1$, $p = .624$, $\eta^2 < 1$, nor of social status, $F(1, 63) < 1$, $p = .797$, $\eta^2 < 1$, nor did the interaction reach statistical significance, $F(1, 63) < 1$, $p = .699$, $\eta^2 < 1$. Thus, independent of the threat potential or social status formerly associated with the to-be-judged faces, it took subjects equally long to evaluate the presented faces as old or new, respectively.

The same 2 × 2 MANOVA as mentioned above was additionally performed on the recognition response latencies coming along with perfect recollection, that is, with the correct identification of old faces as old and, furthermore, with the correct assignment of high or low social status and high or low threat potential to the to-be-judged face. Still, no response

Table 7-2

Parameter estimates and 95% asymptotic confidence intervals (CI) in the multinomial memory model for the crossed source dimensions of social status and threat potential.

Parameter	Estimate	CI
D	.68	[.66; .70]
b	.15	[.12; .17]
d^1	.24	[.19; .28]
d^2	.48	[.27; .69]
e^2	.21	[.12; .28]
g^1	.53	[.51; .56]
$g^2_{\mid H}$	.52	[.48; .56]
$g^2_{\mid L}$	.46	[.42; .50]

Note. D = probability of recognizing target items as old and identifying distractor items as new; b = probability of guessing that an item is old; d^1 = probability of recollecting the social status of recognized target items; d^2 = probability of recollecting the threat potential of recognized target items given recollection of social status; e^2 = probability of recollecting the threat potential of recognized target items given no recollection of social status; g^1 = proportion of guessing "high social status" for recognized and unrecognized target items and unidentified distractor items; $g^2_{\mid H}$ = proportion of guessing "behavior of cheating" for recognized and unrecognized target items and unidentified distractor items given assignment to high-status category; $g^2_{\mid L}$ = proportion of guessing "behavior of cheating" for recognized and unrecognized target items and unidentified distractor items given assignment to low-status category.

time differences could be detected as a function of threat potential and social status, neither concerning the recognition response latencies, nor the reaction times collected for the assessment of the source-monitoring data.

7.3.2.4 Additional Ratings

Status Ratings

The analysis of the status ratings confirmed the successful manipulation of the status variable. A one-way MANOVA performed on the ratings showed that subjects rated the social status of job titles that were supposed to be of low status significantly lower than the social status of job titles that were planned to be of high status, $F(1, 63) = 724.14$, $p < .001$, $\eta^2 = .92$ (*M*s = 2.07 vs. 4.10, *SE*s = .05 vs. .05). As illustrated in Appendix F, all high-status job titles were rated significantly above the median ($Z = 3.00$). Furthermore, 17 out of 20 alleged low-status job titles were rated significantly below the median.

Ratings of Severity and Commendation

Figure 7-4 displays the mean ratings of severity and commendation depending on social status. Obviously, subjects evaluated the instances of behavior of cheating as notably grave (M = 4.80, SE = .10) and the instances of trustworthiness as comparably commendable (M = 4.86, SE = .10). In line with the findings of Experiments 1 and 2, subjects judged behavior of trustworthiness associated with persons of high social status as more commendable than trustworthiness of low-status persons. Accordingly, the one-way MANOVA performed on the commendation ratings revealed a statistically significant main effect of social status, $F(1, 63) = 7.11$, $p = .010$, $\eta^2 = .10$. As for the severity ratings, no differences could be found as a function of social status, $F(1, 63) < 1$, $p = .369$, $\eta^2 = .01$, which also resembles the findings of the previous experiments.

Relevance Ratings

Figure 7-5 illustrates the mean relevance ratings depending on threat potential and social status. Evidently, the data pattern is similar to that found in the previous experiments. The 2 × 2 MANOVA performed on the data revealed a statistically significant main effect of threat potential, $F(1, 63) = 10.79$, $p = .002$, $\eta^2 = .15$., as well as a statistically significant main effect of social status, $F(1, 63) = 22.94$, $p < .001$, $\eta^2 = .27$. Thus, subjects evaluated instances of cheating as more personally relevant than instances of trustworthiness (Ms = 3.21 vs. 2.94, SEs = .12 vs. .11). Furthermore, they judged the personal relevance of behavior associated with persons of high social status as minor compared to behavior associated with low-status persons (Ms = 2.89 vs. 3.26, SEs = .12 vs. .11), which is in line with the findings of Experiments 1 and 2. There was also a statistically significant interaction of threat potential and social status, $F(1, 63) = 131.42$, $p < .001$, $\eta^2 = .68$. As post-hoc tests indicated, subjects judged the personal relevance of instances of cheating higher in the low-status condition than in the high-status condition, $t(63) = -9.17$, $p < .001$, $\eta^2 = .57$. In contrast, the personal relevance of incidents of trustworthiness was rated significantly higher in the high-status condition than in the low-status condition, $t(63) = 5.26$, $p < .001$, $\eta^2 = .31$. Again, this is in line with the findings of the previous experiments.

7.4 Discussion

Experiment 3 was designed as an extension of Experiment 2, focusing on a more detailed investigation of the cognitive processes operating in the recognition test. Replicating the

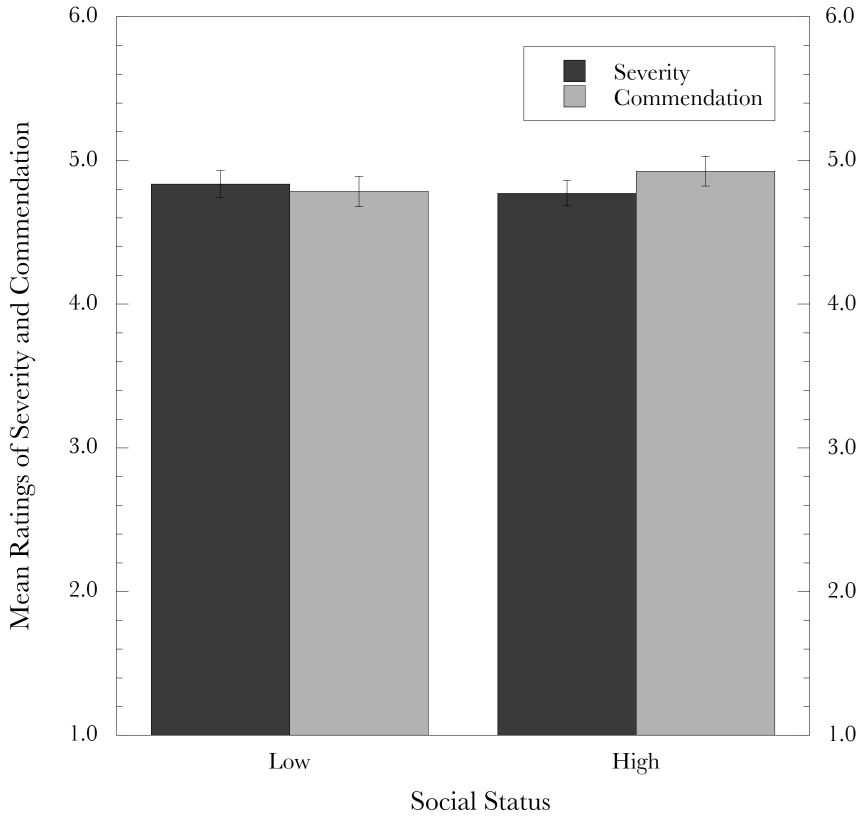

Figure 7-4: Mean ratings of severity and commendation as a function of social status. On a scale ranging from 1 to 6, subjects evaluated instances of trustworthiness associated with low-status persons as more commendable than trustworthiness of high-status persons. The error bars represent the standard errors of the means.

findings of the previous experiments, no results emerged in favor of a cheater-detection hypothesis in the sense of Mealey et al. (1996). The analysis of the recognition-test data did not reveal a cheater-detection effect in terms of an enhanced memory for faces of cheaters. Likewise, the analysis of the recognition response latencies did not provide an indication of an enhanced responsiveness to faces previously associated with information on cheating. Along with the findings of Experiment 2, the results regarding response latency measures substantiated the conclusion that had been drawn from the results of the first experiment, namely that the effects of the threat-potential variable on response latencies that were found in Experiment 1 may have been due to the differentiation between ordinary and somewhat exceptional behavior rather than to the impact of information on cheating.

Apart from that, the present results failed to confirm the modified cheater-detection hypothesis of an enhanced memory for source information on cheating. The rationale underlying the design of the present experiment was that an advanced memory for faces of cheaters might not be an essential but only a sufficient implication of cheater detection. In contrast, accurate memory for source information on prior cheating was considered a nec-

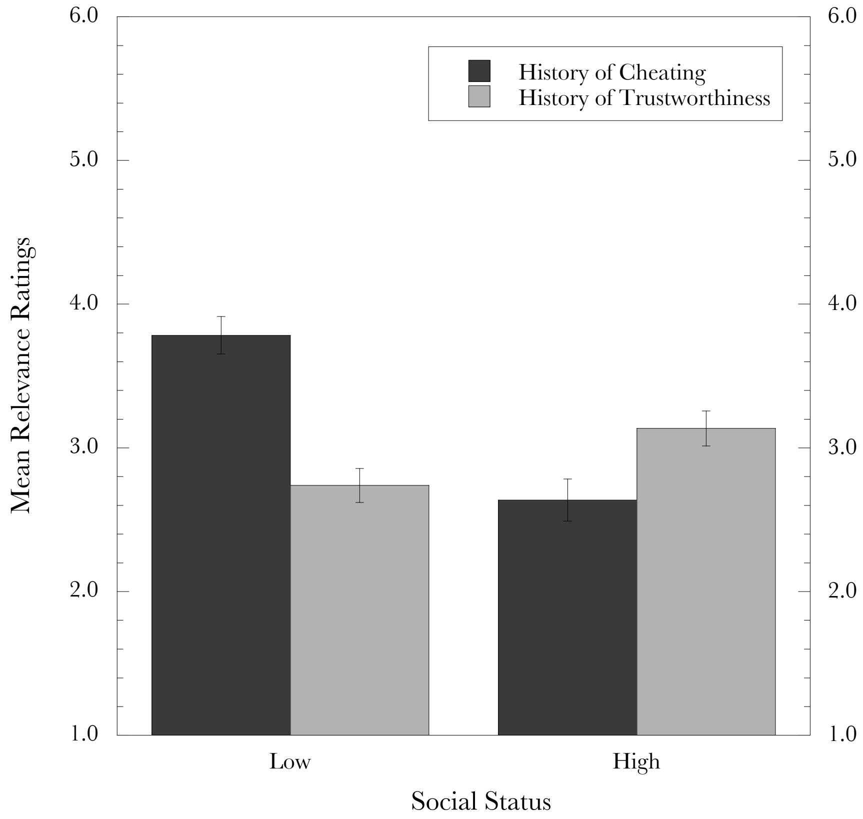

Figure 7-5: Mean relevance ratings as a function of threat potential and social status. On a scale ranging from 1 to 6, subjects estimated behavior of cheating performed by persons of low social status as more personally relevant than behavior of cheating by persons of high social status. The reverse was found concerning instances of trustworthiness. The error bars represent the standard errors of the means.

essary precondition for evaluating a potential interactant's threat potential. Consequently, it was assumed that the possible impact of a mental module as proposed by Cosmides (Cosmides, 1989; Cosmides & Tooby, 1989) might be reflected in a higher amount of correct source assignments with respect to information on cheating compared to information on trustworthiness. However, the application of Meiser and Bröder's (2002) multinomial memory model for crossed source information revealed an excellent model fit given the restriction of equal source-detection parameters for information on cheating and trustworthiness, which is incompatible with the modified cheater-detection hypothesis.

Obviously, the failure of any cheater-detection effect to appear may not be ascribed to a deficient manipulation of the experimental variables. Parallel to the previous experiments, the analysis of the Phase One attractiveness ratings confirmed that persons associated with information on cheating were perceived as more negative than persons associated with information on trustworthiness, indicating that the threat-potential variable was successfully manipulated. Moreover, the instances of cheating were evaluated as sufficiently grave to imply a high threat potential. Finally, a higher personal relevance was ascribed to behavior of cheating compared to behavior of trustworthiness, indicating that the threat potential associated with instances of cheating exceeded the potential convenience associated with trustworthiness. So far, the findings of Experiment 3 are fully in line with the previous experiments.

One might argue that the failure to detect an enhanced source memory for information on cheating could be ascribed to the conditions of information encoding. Johnson et al. (1993) point out that source judgments on certain events highly depend on the information available from activated memory records and therefore fundamentally rely on the quality of the information that was initially recorded about these events. Consequently, there are several characteristics of the encoding episode that may have an impact on the quality of source judgments. That is, anything that prevents a person from fully contextualizing information at acquisition will reduce encoding of potentially relevant source information. As examples, Johnson et al. (1993) refer to stress or divided attention as factors to potentially disrupt normal perceptual and reflective processes. This may, in turn, result in incompletely encoded information from which source could later be derived. As for the present experiment, none of the above-mentioned factors may have contributed to an insufficient encoding of the relevant information; at least, not particularly with respect to information on cheating.

However, Dodson and Johnson (1993) indicate that source attributions are not only influenced by characteristics of the encoding episode but also by certain decision processes used at the time of test, which may be due to characteristics of the test format. In their experiment, they compared subjects' performance in source-memory tests with either a binary question response format or a four alternative forced-choice (4AFC) source-listing format. The authors found that such test features influenced the rate of false source attributions; they observed fewer source confusions with the 4AFC source format than with the binary question format. Similarly, Marsh and Hicks (1998) found that source-monitoring accuracy was influenced by whether a single source or multiple sources were the focus of the source-monitoring question or by the combination of sources tested. Thus, it is unclear whether the mode of test possibly influenced subjects' source assignments in the present experiment. For example, the order of dimensions in the model applied to the present data had no impact on the outcome of the model-selection procedure (and the same was reported by Meiser & Bröder, 2002). Still, it appears possible that subjects' performance might change in a source-memory test with only one dimension in the focus of the source-monitoring question (namely, threat potential).

Taken together, the previous three experiments yielded a consistent pattern of results conflicting with the proposition of a functionally specialized mental module operating on the detection of cheaters. Although successfully manipulated, the threat-potential variable obviously had no impact on subjects' recognition performance. Yet the results of the first experiment suggested that the unusualness of the behavioral descriptions—rather than the implied threat potential—might be the variable of actual interest, potentially influencing the recognition-test data as well as other measures indicating how the facial stimuli may have been perceived. Unfortunately, the behavioral descriptions used by Mealey et al. (1996) for the manipulation of threat potential were unavailable. It is possible that there has been some confounding in their materials. Beyond doubt, behavior of cheating is exceptional to a greater or lesser extent, and such deviant behavior hardly qualifies as ordinary. Therefore, it was expected to gain further insight into the ongoing processes and relevant variables in the context of another experiment, allowing for a direct test of possible effects of unusualness.

Experiment 4 was designed as a modification of Experiments 1 and 3, where the type of behavior implying various specifications of threat potential was additionally varied as either somewhat exceptional or ordinary. Apart from that, with regard to the above-mentioned findings that the test format may affect subjects' performance in a source-memory test, source-monitoring data were assessed only with respect to source dimension threat potential.

8 Experiment 4

8.1 Overview and Hypotheses

In its main features, Experiment 4 was adapted from Experiment 1 and, with respect to the recognition-test phase, from Experiment 3 as follows. Parallel to Experiment 1, subjects at first evaluated a variety of facial photos that were presented together with behavioral descriptions of cheating, trustworthiness, and irrelevant behavior. Yet in contrast to the previous experiments, sympathy ratings were assessed instead of attractiveness ratings for the following reason. The previous experiments consistently yielded the result that the attractiveness ratings assessed in Week One (or, as for Experiment 3, in Phase One) varied as a function of the threat potential implied by the behavioral descriptions presented together with the to-be-judged faces. This was taken as evidence for a successful manipulation of the threat-potential variable. No such effect was found with respect to the attractiveness ratings acquired in Week Two (Experiments 1 and 2), where no behavioral information was presented along with the facial photos. This was interpreted such that the Week Two ratings were no longer influenced by the behavioral information previously associated with the Week One faces. Yet given that in Week One (Phase One in Experiment 3) subjects obviously based their ratings on the behavioral information rather than on the photos, it was assumed that they probably judged the depicted persons' *sympathy* rather than their actual attractiveness. In contrast, with no clues to sympathy available in Week Two, subjects possibly focused more strongly on the pictures and actually evaluated the depicted persons attractiveness. Therefore, it was considered that by explicitly instructing subjects to judge the sympathy of the persons presented, in Week Two they might rather tend to recollect the behavioral information previously associated with a given face and to base their judgments on this information. This, in turn, might enhance the probability of Week Two ratings to vary as a function of the threat-potential variable as it was consistently found with respect to the Week One ratings. According to this rationale, sympathy ratings were assessed in the present experiment both in Week One and in Week Two.

As inferable from the previous paragraph, the one-week retention interval originally proposed by Mealey et al. (1996) was realized again in the present experiment in contrast to Experiment 3. The reduction of the retention interval that had been implemented in the third experiment obviously had no effect on the results and therefore could be abolished. Moreover, in contrast to the previous experiments, the social-status variable was not varied in the present experiment. This was done since the only experimental condition on which a significantly enhanced memory for faces of cheaters could be detected before was Mealey et al.'s (1996) low-status condition. While the social status was held constant across the behavioral descriptions, the type of behavior was varied in terms of exceptionality. That is, the type of behavior implying various specifications of threat potential was additionally varied as either somewhat exceptional or ordinary. As outlined previously, this was done to test whether potential effects of cheater detection as reported by Mealey et al. (1996) may have been due to confoundings of threat potential and exceptionality.

Following Experiment 3, in the recognition-test phase faces from Week One and new distractor faces were presented and recognition judgments as well as source assignments were assessed. However, with respect to the above-mentioned findings that the test format may affect subjects' performance in a source-memory test (Dodson & Johnson, 1993; Marsh & Hicks, 1998), source judgments were surveyed only with respect to source dimension threat potential in a three alternative forced choice format. Based on the work of Dodson and Johnson (1993) it was assumed that more accurate source monitoring might occur when all of the potential sources for a memory were simultaneously salient to the subjects.

Based on the previous experiments, the following hypotheses were tested in the present experiment. First, given Mealey et al.'s (1996) original cheater-detection hypothesis was valid, a cheater-detection effect in terms of an enhanced memory for faces of cheaters should be detected. In contrast, if the modified cheater-detection hypothesis was true, then an enhanced memory for source information on cheating should be observed. Furthermore, assumed that the type of behavior influenced the recognition-test data as discussed in the previous chapter, recognition performance should be superior for faces associated with exceptional compared to ordinary behavior, independent of threat potential. Moreover, similar effects might be observed with respect to the source-monitoring data. No specific hypotheses were postulated with respect to an impact of the type-of-behavior variable on the sympathy ratings.

8.2 Method

8.2.1 Participants

Participants were 51 female and male 21 persons who were all students at the Heinrich-Heine-Universität Düsseldorf. As in the aforementioned experiments, they were paid for participation. Their age ranged from 20 to 35 (*M* = 23.01, *SD* = 2.84). None of them had already taken part in one of the previous experiments.

8.2.2 Materials

The same 72 facial photos as in Experiment 1 were used, half of them as targets in the acquisition phase and half of them as distractors in the recognition-test phase. Additionally, 36 descriptive sentences were used implying three different types of threat potential by giving information on cheating, irrelevant information or information on trustworthiness. Equivalent to the previous experiments, the behavioral information followed information on the depicted person's profession, indicating the person's social status. As mentioned above, only low-status job titles were used in the present experiment. In a pretest, subjects (*N* = 36) had rated 200 job titles with respect to their social status using a scale ranging from 1 ("status low") to 5 ("status high"), and 36 job titles with low ratings were chosen (*M* = 1.88, *SD* = .33)[58]. To assure for a successful manipulation of behavior of cheating, trustworthiness, and irrelevant behavior, a different group of subjects (*N* = 22) had rated 72 behavioral descriptions with respect to the depicted person's character, using a scale ranging from −3 ("cheating") to +3 ("trustworthy"). Additionally, and differing from the previous experiments, half of the descriptive sentences were thought to specify somewhat exceptional behavior, whereas the other half were supposed to represent instances of ordinary behavior. Therefore, subjects were also asked to rate the 72 statements with respect to the type of behavior, using a scale raging from −3 ("very exceptional") to +3 ("very ordinary"). Finally, six sentences with descriptions of exceptional (*M* = −1.30, *SD* = .61) and ordinary (*M* = 0.76, *SD* = .54) behavior and behavioral descriptions of cheating (*M* = −2.50, *SD* = .31), trustworthiness (*M* = 2.32, *SD* = .39), and irrelevant behavior (*M* = 0.26, *SD* = .34), respectively, were selected for the experiment[59]. Finally, the rating scales applied were identical to those in the previous experiments, yet no status ratings were obtained.

8.2.3 Procedure

The procedure was similar to the previous Experiments, except for the following modifications mentioned above: first, in line with Experiments 1 and 2, there was a retention interval of one week. Since the modification of the retention interval in Experiment 3 did not

---

[58] See Appendix G for details.
[59] See Appendix H for details.

affect the recognition-test results compared to Experiments 1 and 2, the one-week interval was chosen to bring the experimental design more in line with the study of Mealey et al. (1996).

Second, sympathy ratings were assessed instead of attractiveness ratings in both parts of the experiment. Thus, comparable to the previous experiments, each Week One trial started with the display of a headline ("How likable do you think is the depicted person?"), and after the random presentation of a picture and descriptive sentence subjects made their judgments using a scale ranging from 1 ("not likable at all") to 6 ("extremely likable"). Analogically, each Week Two trial was induced by the sympathy assessment as elucidated above.

Third, similar to Experiment 3 subjects were to give source judgments in the recognition-test phase. However, since the status variable was not manipulated they only made assignments with respect to source dimension threat potential. Thus, subjects were randomly presented the 72 photos of picture sets 1 and 2, and right after the sympathy-ratings another headline ("Is this face old or new?") and two check boxes labeled "old" and "new" appeared. In case of a "new" judgment, clicking the *continue* button initiated the next trial after a fixed time frame of 500 ms. In contrast, when evaluating a photo as old clicking the *continue* button induced the assessment of source-monitoring data. Subjects were asked to choose one of the following options displayed as check boxes: "What was the behavior of this person?"… "cheating?", "trustworthy?" or "neither cheating nor trustworthy?". Choosing one of these options and clicking the *continue* button initiated the next trial after an inter-trial interval of 500 ms.

8.2.4 Design

The within-subject independent variables were threat potential (history of cheating, irrelevant information, history of trustworthiness) and type of behavior (exceptional, ordinary). The Week One dependent measures were the sympathy ratings and corresponding response latencies. As elucidated above, the main Week Two dependent measures were ratings of sympathy and old-new ratings aside from source-monitoring data and reaction times assessed for all of the aforementioned ratings. Furthermore, and parallel to the previous experiments, ratings of severity and commendation as well as relevance ratings were acquired.

Sympathy ratings as well as response latencies and recognition-test data were analyzed in a 3 (threat potential) × 2 (type of behavior) factorial design. An a priori power analysis revealed that—given an assumed total sample size of $N = 70$, $\alpha = .05$, and the assumption that the average population correlation between the levels of the repeated measures factor threat potential is $\rho = .60$ (estimated from previous studies)—effects of size $f = .20$ ($\eta^2 = .23$) could be detected for the threat-potential variable with a probability of $1 - \beta = .99$.

8.3 Results

In line with the previous experiments, means and standard errors referring to statistically significant effects are parenthesized as far as they are not illustrated in figures.

8.3.1 Week One Data

8.3.1.1 Sympathy Ratings

Figure 8-1 illustrates the mean sympathy ratings depending on threat potential and type of behavior. As expected, the 3 × 2 MANOVA performed on the data revealed a statistically significant main effect of threat potential, $F(2, 70) = 155.70$, $p < .001$, $\eta^2 = .82$. As post-hoc tests showed, people depicted as cheaters were judged as significantly less likable than people described by irrelevant information, $t(71) = -14.36$, $p < .001$, $\eta^2 = .74$. (Ms = 1.97 vs. 3.48, SEs = .07 vs. .07) or by information on trustworthiness, $t(71) = -17.51$, $p < .001$, $\eta^2 = .81$. (Ms = 1.97 vs. 4.36, SEs = .07 vs. .09). Furthermore, persons associated with irrelevant behavior were also rated as significantly less likable than persons described as trustworthy, $t(71) = 13.97$, $p < .001$, $\eta^2 = .73$. (Ms = 3.48 vs. 4.36, SEs = .07 vs. .09). The analysis of the data also showed a statistically significant main effect of type of behavior, indicating that persons associated with somewhat exceptional behavior were perceived as more likable than persons described by more ordinary behavior, $F(1, 71) = 13.95$, $p < .001$, $\eta^2 = .16$ (Ms = 3.35 vs. 3.20, SEs = .05 vs. .05). The interaction of threat potential and type of behavior was also statistically significant, $F(2, 70) = 15.75$, $p < .001$, $\eta^2 = .31$. Post-hoc tests showed that alleged cheaters were judged as less likable when associated with exceptional acts of cheating compared to more ordinary frauds, $t(71) = -2.40$, $p = .019$, $\eta^2 = .08$ (Ms = 1.89 vs. 2.06, SEs = .09 vs. .07). In contrast, supposedly trustworthy persons were perceived as more likable when accredited with exceptional compared to common behavior, $t(71) = 6.17$, $p < .001$, $\eta^2 = .35$ (Ms = 4.59 vs. 4.13, SEs = .10 vs. .09).

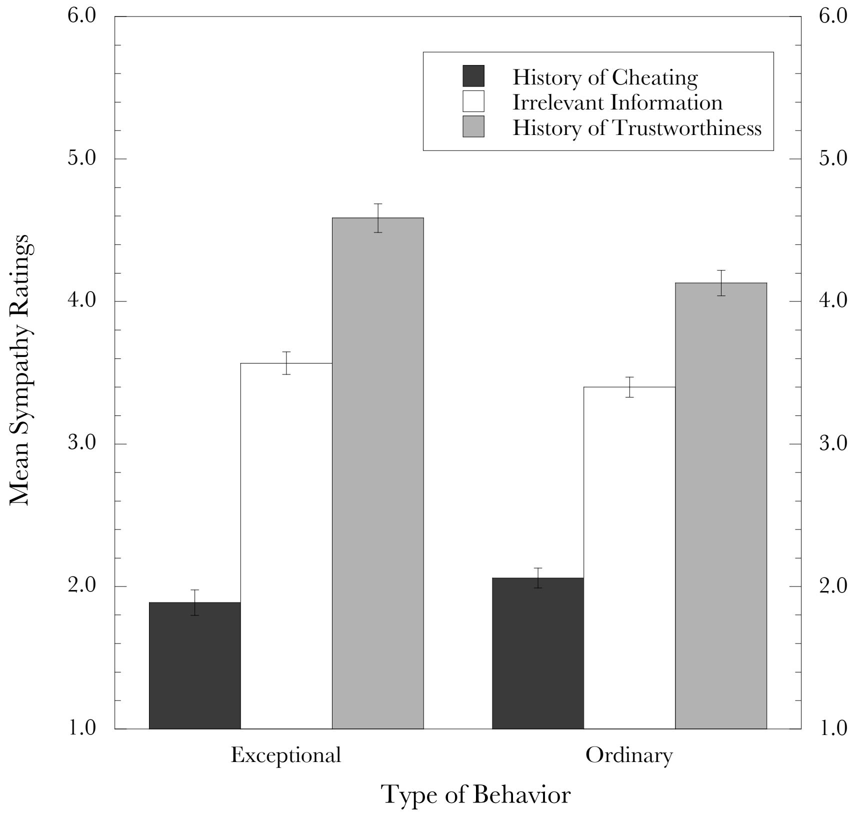

Figure 8-1: Week One mean sympathy ratings depending on threat potential and type of behavior. On a scale ranging from 1 to 6, subjects rated people described as cheaters as significantly less likable than people depicted as trustworthy or irrelevant. The error bars represent the standard errors of the means.

8.3.1.2 Response Latencies

Figure 8-2 displays the mean Week One response latencies as a function of threat potential and type of behavior. The 3 × 2 MANOVA performed on the data revealed a statistically significant main effect of threat potential, $F(2, 70) = 4.00$, $p = .023$, $\eta^2 = .10$. As post hoc *t*-tests showed, it took subjects significantly longer to make their judgments concerning cheaters than persons depicted as trustworthy, $t(71) = 2.85$, $p = .006$, $\eta^2 = .10$. (*M*s = 11.57 vs. 10.99 sec., *SE*s = .29 vs. .23). The response time differences between the other conditions of threat potential missed the Bonferoni-Holm corrected significance level, $|t|(71) \leq 1.91$. Moreover, the main effect of type of behavior just missed the level of statistical significance, $F(1, 71) = 3.55$, $p = .064$, $\eta^2 = .05$. However, there was a statistically significant interaction, $F(2, 70) = 3.99$, $p = .023$, $\eta^2 = .10$.

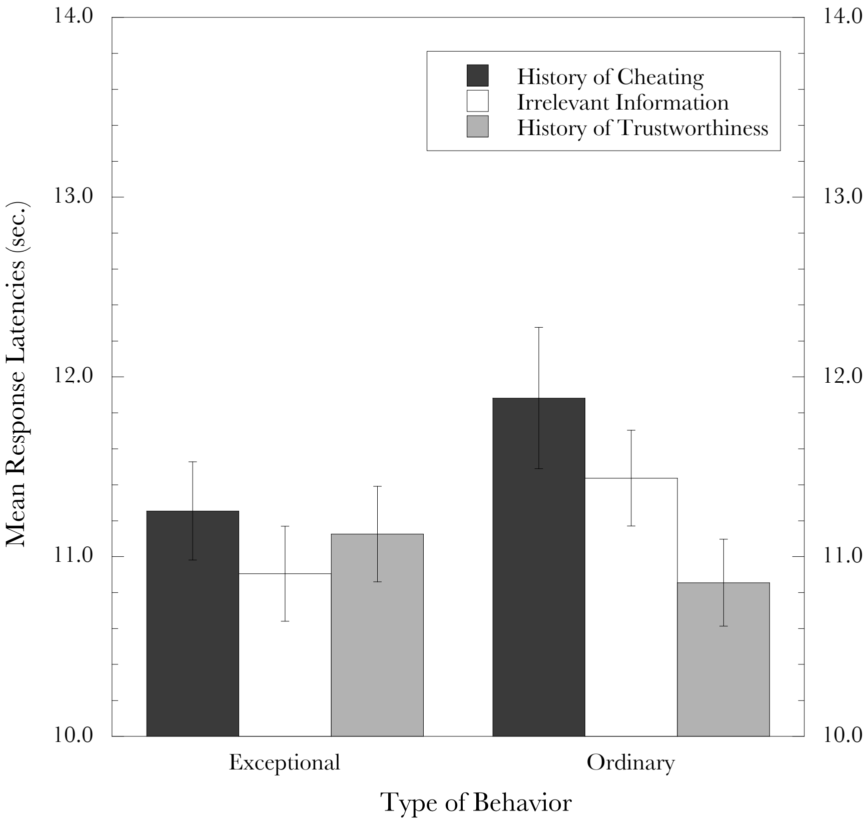

Figure 8-2: Week One mean response latencies depending on threat potential and type of behavior. The longest reaction times were found for judgments concerning cheaters. The error bars represent the standard errors of the means.

8.3.2 Week Two Data

8.3.2.1 Sympathy Ratings

Figure 8-3 displays the mean sympathy ratings for faces from Week One depending on threat potential and type of behavior. Contrary to the findings of Experiments 1 and 2, a 3 × 2 MANOVA yielded a statistically significant main effect of threat potential, $F(2, 70) = 3.93$, $p = .024$, $\eta^2 = .10$. Taking into account the Bonferoni-Holm corrected significance level, post hoc tests revealed that subjects rated cheaters as significantly less likable than persons formerly associated with irrelevant information, $t(71) = -2.82$, $p = .006$, $\eta^2 = .10$ (*M*s = 3.06 vs. 3.22, *SE*s = .05 vs. .04), all other $|t|(71) \leq 1.54$. Neither the main effect of type of behavior, nor the interaction of the threat-potential and type-of-behavior variables reached statistical significance, $F(1, 71) < 1$, $p = .434$, $\eta^2 < .01$, and $F(2, 70) < 1$, $p = .413$, $\eta^2 = .03$, respectively.

A one-way MANOVA performed on the mean sympathy ratings for Week One faces showed that, overall, subjects chose significantly lower ratings in Week Two than in Week

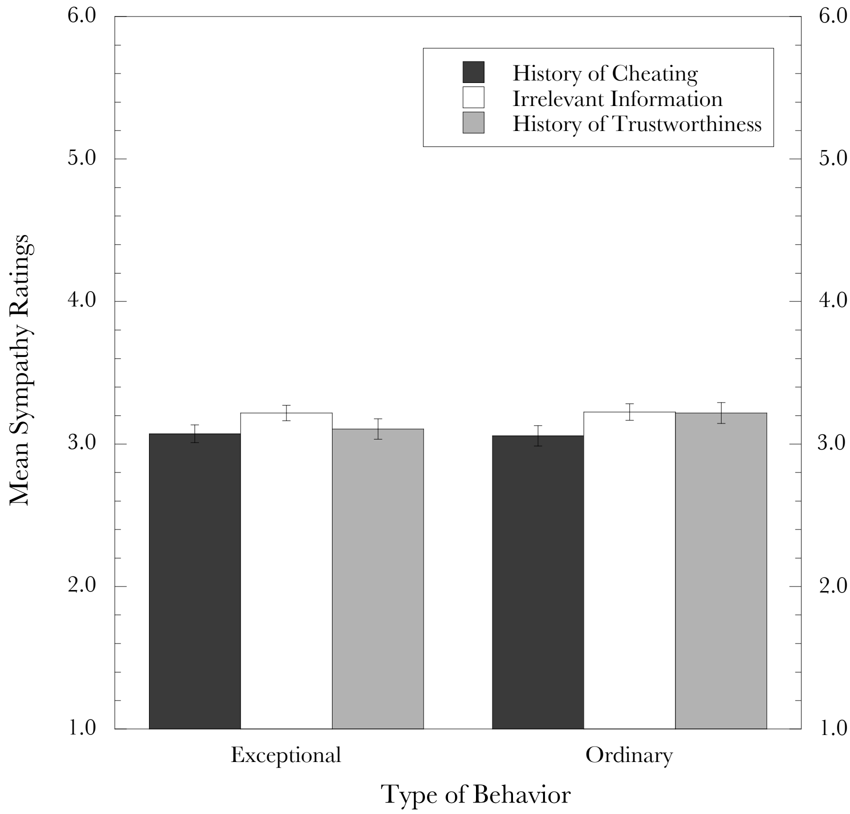

Figure 8-3: Mean Week Two ratings of sympathy for faces from Week One as a function of threat potential and type of behavior. On a scale ranging from 1 to 6, persons formerly described as cheaters were rated as slightly less likable than persons associated with irrelevant behavior. The error bars represent the standard errors of the means.

One, replicating the findings of Experiments 1 and 2, $F(1, 71) = 10.16$, $p = .002$, $\eta^2 = .13$. As illustrated in Figure 8-4, this was due to an increase in the sympathy ratings for faces of cheaters, whereas there was a decrease in the judgments concerning persons associated with trustworthiness or irrelevant behavior. Finally, in line with findings of Experiments 1 and 2, a one-way MANOVA performed on the mean sympathy ratings for old and new faces assessed in Week Two did not reveal a statistically significant difference, $F(1, 71) < 1$, $p = .430$, $\eta^2 < .01$ (*M*s = 3.15 vs. 3.18, *SE*s = .04 vs. .04). Thus, on an average, Week One faces were judged just as likable as new distractor faces.

8.3.2.2 Recognition-Test Data

Figure 8-5 illustrates the discrimination indices P_r as a function of threat potential and type of behavior. The 3 × 2 MANOVA performed on the data did not reveal a statistically significant main effect of threat potential, $F(2, 70) = 2.73$, $p = .072$ $\eta^2 = .07$, and also a statistically significant main effect of type of behavior was not found, $F(1, 71) < 1$, $p = .497$ $\eta^2 < .01$. The interaction between the threat-potential and the type-of-behavior variables also

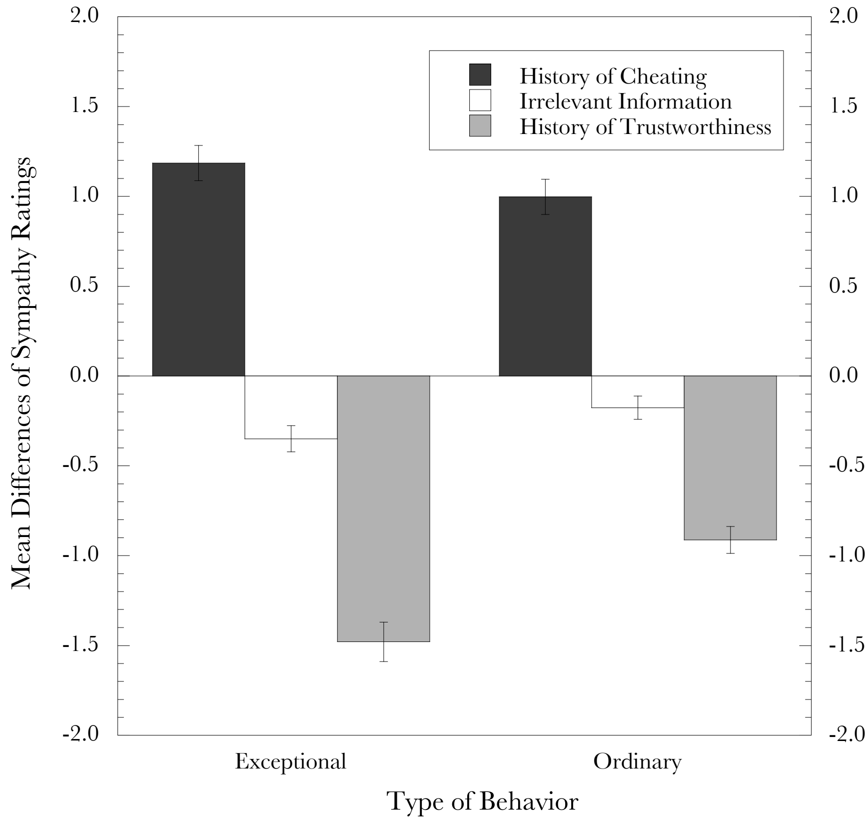

Figure 8-4: Mean differences of sympathy ratings for faces from Week One between Week Two and Week One depending on threat potential and type of behavior. In Week Two, subjects perceived cheaters as more likable than in Week Two, as the positive difference scores indicate. In contrast, persons formerly associated with trustworthiness or irrelevant behavior were judged as less likable in Week Two compared to Week One, which is indicated by the negative difference scores. The error bars represent the standard errors of the means.

did not reach statistical significance, $F(2, 70) < 1$, $p = .492$ $\eta^2 = .02$. However, it seems noteworthy that the main effect of threat potential just missed significance. Apart from that, and replicating the results of the previous experiments, a test of whether P_r was different from zero for all cells of the design simultaneously confirmed that recognition performance was above chance, $F(1, 71) = 591.90$, $p < .001$, $\eta^2 = .89$.

8.3.2.3 Source-Monitoring Data

Table 8-1 summarizes the response frequencies from the source-monitoring task of the present experiment. In contrast to Experiment 3, source judgments had been assessed only with respect to source dimension threat potential. The data were analyzed through a three-source adaptation of the multinomial source-memory model proposed by Bayen et al. (1996). Bayen and colleagues presented the two-high-threshold analog to Batchelder and Riefer's (1990) one-high-threshold model that has been introduced in detail in Chapter 6.

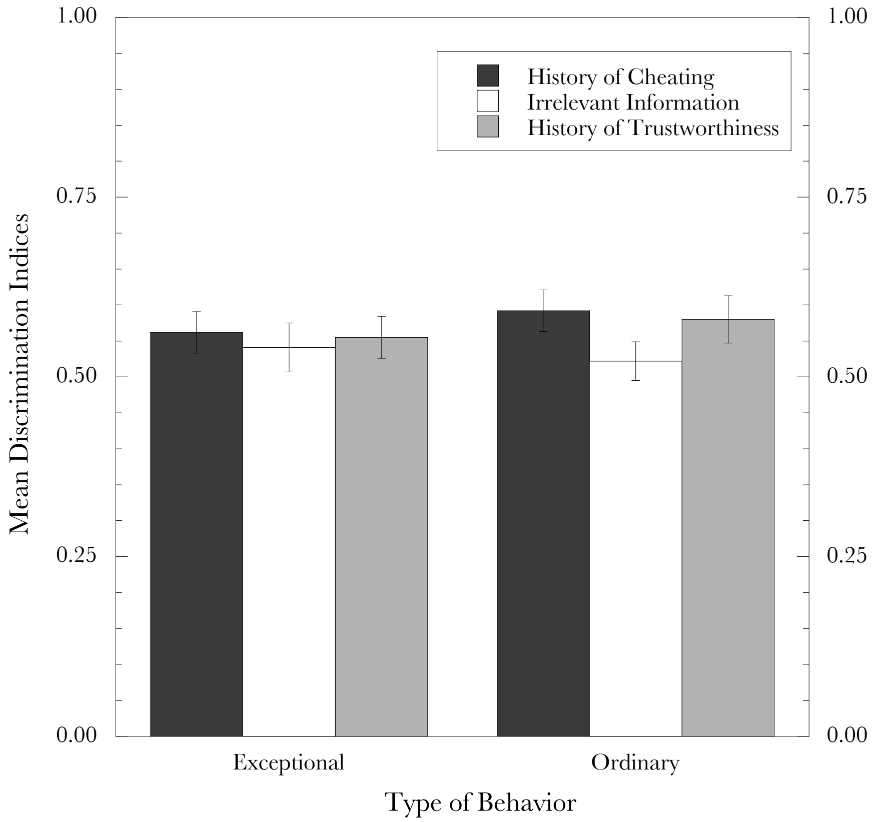

Figure 8-5: Mean discrimination indices P_r for faces from Week One depending on threat potential and type of behavior. The highest value of P_r was found for faces of cheaters, yet this was only on a descriptive level. The error bars represent the standard errors of the means.

Figure 8-6 illustrates the model as adapted from Bayen et al. (1996). The detection parameters D_C, D_I, and D_T refer to the probability of detecting an item as old that belongs to

Table 8-1

Response frequencies in the source-monitoring task for threat potential broken down by type of behavior. Note that the response category "neither… nor" refers to target items previously associated with irrelevant information. Bold values indicate correct source judgments.

		Response			
	True category	Cheating	Neither… nor	Trust-worthiness	New
Exceptional	Cheating	**143**	81	72	136
	Irrelevant	101	**86**	106	139
	Trustworthiness	111	76	**100**	145
Ordinary	Cheating	**145**	92	72	123
	Irrelevant	111	**92**	101	128
	Trustworthiness	93	84	**102**	153
	New	93	133	94	**2272**

Source C (i.e., Week One faces associated with information on cheating), to Source I (i.e., Week One faces associated with irrelevant information), and to Source T (i.e., Week One faces associated with information on trustworthiness), respectively. Distractor items are identified as new with probability D_N. Given no target item detection or distractor identification, parameter *b* denotes the probability of guessing that an item is old. The source-recollection parameters d_C, d_I, and d_T refer to the probability of correct source assignments to target items from Sources C, I, and T. Moreover, a_C, a_I, and a_T represent the proportions of guessing Sources C, I, and T for recognized target items. Analogously, parameters g_C, g_I, and g_T denote the proportions of guessing Sources C, I, and T for unrecognized target items and unidentified distractors. Note that the model implies identical guessing parameters a_C, a_I, and a_T and g_C, g_I, and g_T for all item-categories.

Unfortunately, conventional analyzers such as the AppleTree program (Rothkegel, 1999), the MBT program (Hu, 1993), and the SOURCE program (Hu, 1993) have been designed for analyzing only binary-link multinomial processing tree models. Since the precondition of a binary tree structure is violated for the three-source model depicted in Figure 8-6 due to the guessing parameters *a* and *g*, the above-mentioned computer programs are not applicable to the analysis of the model in its original definition. Therefore, the model was reparameterized as a binary-tree model in that the bias parameters *a* and *g* were redefined in terms of conditional link probabilities.

The amended model is illustrated in a general notation in Figure 8-7. For lack of space, the complete seven-tree model with separate processing trees for target items from source categories *i* depending on the two conditions of type of behavior (i.e., exceptional, ordinary) is depicted in Appendix I. In the reparameterized model, items from source *i* are recognized as old with probability D_i, while distractor items are identified as new with probability D_N. Parameter *b* denotes the probability of guessing that an unrecognized target item or an unidentified distractor is old. Furthermore, parameters d_i indicate correct source recollection. In case recollection of Source *i* fails with probability $(1 - d_i)$, the model implies that either of the response categories C and T is selected by guessing with probabilities a_{CT} (given prior target recognition) or g_{CT} (given no target recognition and distractor identification, respectively), where the final selection of Source C is specified by guessing probabilities a_C and g_C, respectively. The third bias parameter from the original model, referring to the selection of Source I by guessing, is represented in the model as, respectively, $(1 - a_{CT})$ and $(1 - g_{CT})$, that is the probabilities of *not* selecting either of the sources C and T.

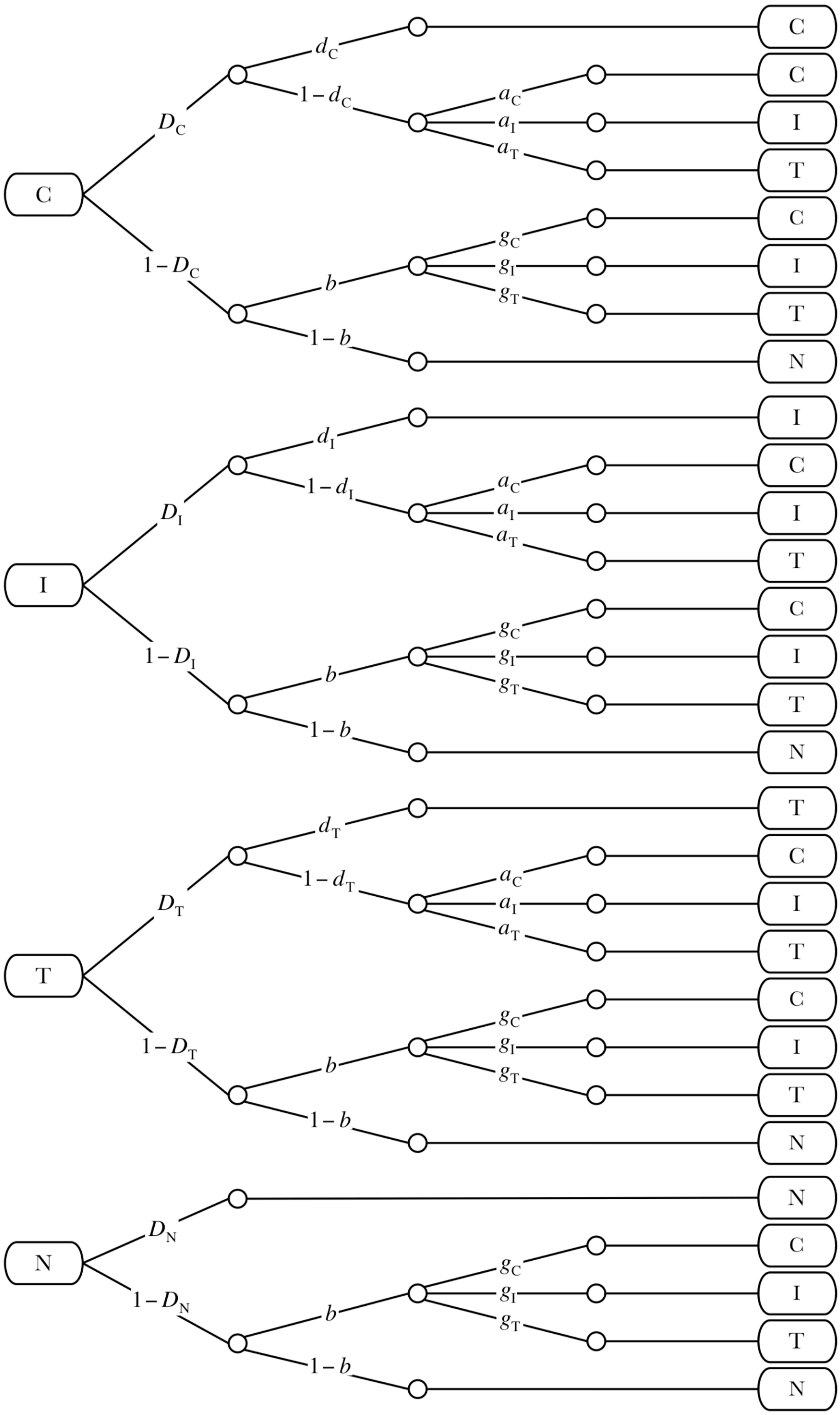

Figure 8-6: Two-high-threshold joint multinomial model of source monitoring for three source categories, adapted from Bayen et al. (1996). For lack of space, annotations with respect to the parameters are provided in the full text.

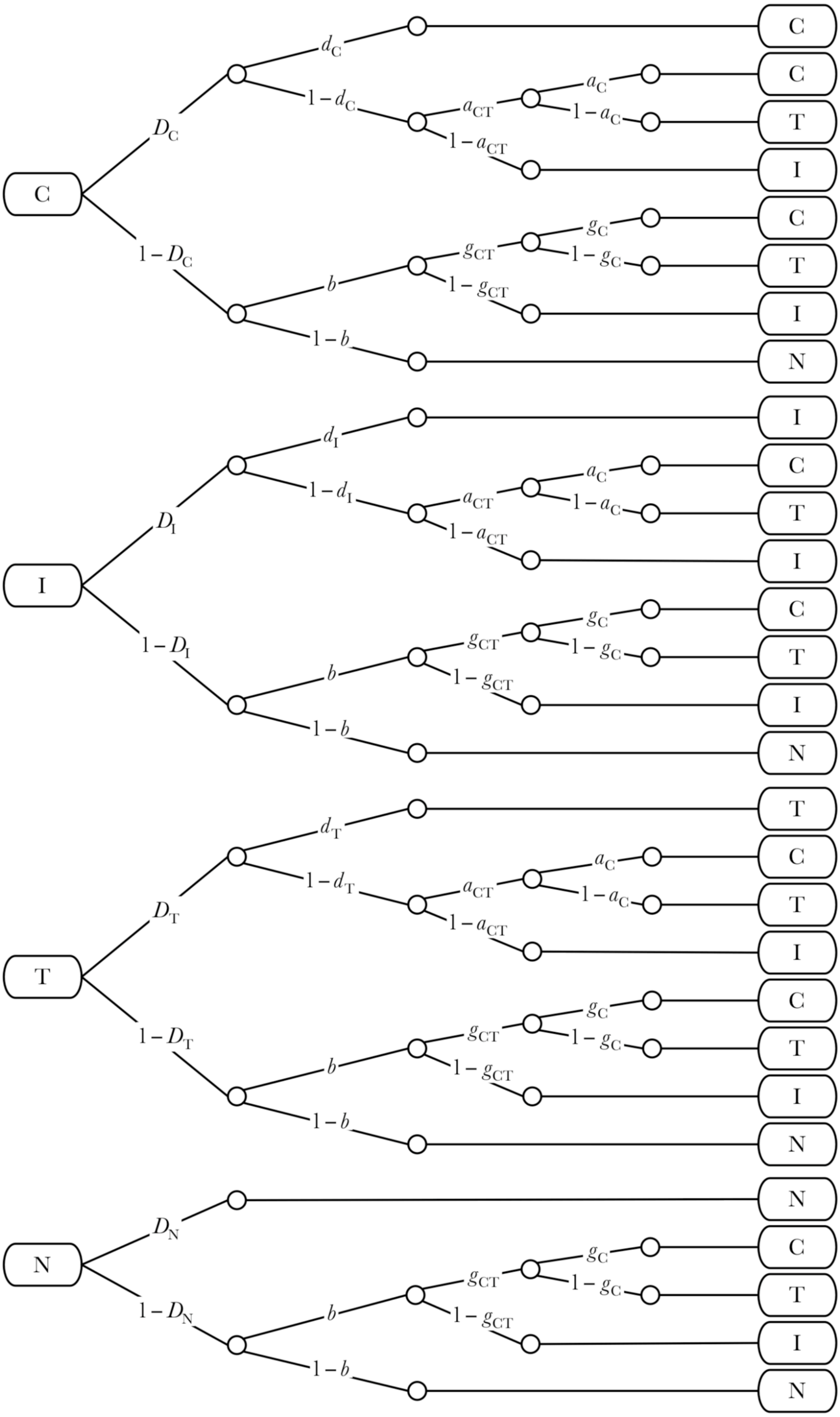

Figure 8-7: Reparameterized version of the two-high-threshold joint multinomial model of source monitoring for three source categories, adapted from Bayen et al. (1996). For lack of space, annotations with respect to the parameters are provided in the full text.

Table 8-2

Values of the Likelihood Ratio Statistic G^2 for identifiable submodels of the multinomial memory model for three source categories of threat potential.

Restrictions	G^2	*df*	*p*
(1) $D_i = D_N$	14.40	9	.109
(2) $D_i = D_N$; $d_{Ce} = d_{Co}$ (= d_C); $d_{Ie} = d_{Io}$ (= d_I); $d_{Te} = d_{To}$ (= d_T)	14.72	12	.257
(3) $D_i = D_N$; $d_C = d_I = d_T$	27.61	14	.016*

Note. D_i = probability of recognizing target items as old, $i \in$ {Ce, Co, Ie, Io, Te, To}; D_N = probability of identifying distractor items as new, d_i = probability of correct source recollection, $i \in$ {Ce, Co, Ie, Io, Te, To}.

The multinomial analyses were carried out using the AppleTree program (Rothkegel, 1999), and the criterion of statistical significance was set to α = .05 for all goodness-of-fit tests. Table 8-2 displays the goodness-of-fit statistics for various identifiable submodels of the seven-tree multinomial memory model. Along the lines of Experiment 3, it was initially examined whether the high-threshold parameters of target item recognition D_i were equal across the six combinations of threat potential and type of behavior, that is, $D_{Ce} = D_{Ie} = D_{He} = D_{Co} = D_{Io} = D_{Ho}$[60]. The test yielded a statistically non-significant result, $G^2(5) = 6.42$, $p = .267$. A chi-square test of equality of independent proportions as suggested by Batchelder and Riefer (1990) substantiated this result, $\chi^2(5) = 4.38$ ($\chi^2_{crit.} = 11.07$). Therefore, the assumption of equal target-detection parameters could be maintained, and the corresponding equality constraints were specified in the model. Note that equality of the target-detection parameters is incompatible with the original cheater-detection hypothesis of an enhanced memory for faces of cheaters (Mealey et al., 1996). Following Bayen et al. (1996), the additional restriction of *all* high-threshold parameters being equal was imposed on the model in terms of Restrictions (1) in Table 8-2. The resulting submodel fitted the data satisfactorily and was selected as the base model for all subsequent tests[61].

[60] Note that parameters D_{Ce}, D_{Ie}, and D_{He} denote target identification in the experimental condition of exceptional behavior, while parameters D_{Co}, D_{Io}, and D_{Ho} refer to face recognition in the experimental condition of ordinary behavior.

[61] Actually, the less restrictive submodel requiring invariant item-detection parameters D_i while leaving the high-threshold parameter for distractor identification D_N unrestricted fitted the data superiorly, $\Delta G^2(1) = 5.86$ ($G^2_{crit.} = 3.84$). However, this submodel turned out to be inapplicable as the base model for subsequent hypothesis testing, since specifying the additional restrictions necessary with respect to testing the modified cheater-detection hypothesis resulted in a non-identifiable submodel.

Starting from this base model, source-recollection parameters d_i were set equal across the conditions of type of behavior for each of the three source categories of threat potential by moving from Restrictions (1) to (2). This was done to test whether the type of behavior implied in source information influenced subjects' source-recollection performance. As evident from Table 8-2, imposing these additional restrictions did not affect the model fit, and the equality constraint could be maintained. Obviously, the type of behavior was irrelevant with respect to recollecting source information on behavior of cheating, trustworthiness or irrelevant behavior. Finally, to test the modified cheater-detection hypothesis, Submodel (3) was specified implying the additional restriction of *all* source-recollection parameters d_i being equal. Note that the modified cheater-detection hypothesis predicts the value of the source-recollection parameter d_C to exceed the values of the remaining parameters d_I and d_T, given a cheater-detection effect in terms of an enhanced source-judgment performance for information on cheating.

Most surprisingly facing the findings from the previous experiments, moving from Restrictions (2) to (3) resulted in a significant misfit to the data. Thus, the equality constraint on the source-recollection parameters had to be rejected, lending support to the modified

Table 8-3

Parameter estimates and 95% asymptotic confidence intervals (CI) in the multinomial memory model for three source categories of threat potential.

Parameter	Estimate	CI
D	.56	[.53; .58]
b	.28	[.26; .30]
d_C	.21	[.12; .31]
d_I	< .01	[< –.01; .08]
d_T	.06	[< –.01; .15]
a_C	.54	[.49; .59]
g_C	.50	[.43; .57]
a_{CT}	.71	[.67; .75]
g_{CT}	.58	[.53; .64]

Note. D = probability of target-item detection and distractor identification; b = probability of guessing that an item is old; d_C = probability of recollecting source information on cheating; d_I = probability of recollecting source information on irrelevant behavior; d_T = probability of recollecting source information on trustworthiness; a_C = probability of selecting "history of cheating" by guessing given target item detection; g_C = probability of selecting "history of cheating" by guessing given no target item detection or distractor identification; $(1 - a_{CT})$ = probability of selecting "irrelevant information" by guessing given target item detection; $(1 - g_{CT})$ = probability of selecting "irrelevant information" by guessing given no target item detection or distractor identification.

cheater-detection hypothesis for the first time. Table 8-3 illustrates the parameter estimates obtained from Submodel (2) together with their 95% asymptotic confidence intervals. As predicted by the modified cheater-detection hypothesis, the source-recollection parameter d_C evidently turned out to be distinct from the parameters d_I and d_T. Actually, while the probability of correct source recollection was close to zero for irrelevant information and information on trustworthiness, indicating that such sources could not be consciously remembered, the probability of recollecting source information on cheating was clearly above zero.

8.3.2.4 Response Latencies

Comparable to the findings of Experiments 1 and 2, the 3 × 2 MANOVA performed on the response latencies assessed for the sympathy ratings did not reveal a statistically significant main effect of threat potential, $F(2, 70) < 1$, $p = .825$, $\eta^2 < .01$. There was also neither a statistically significant main effect of type of behavior, $F(1, 71) < 1$, $p = .781$, $\eta^2 < .01$, nor a statistically significant interaction of the threat-potential and type-of-behavior variables, $F(2, 70) < 1$, $p = .390$, $\eta^2 = .03$. Testing for the recognition response latencies yielded the same pattern of results: neither the main effect of threat potential, $F(2, 70) < 1$, $p = .605$, $\eta^2 = .01$, nor of type of behavior reached statistical significance, $F(1, 71) < 1$, $p = .842$, $\eta^2 < .01$, nor was there a statistically significant interaction of threat potential and type of behavior, $F(2, 70) < 1$, $p = .390$, $\eta^2 = .03$. Finally, the analysis of the response latencies assessed for the source-monitoring judgments also failed to reveal any statistically significant main effects, $F(2, 70) < 1$, $p = .869$, $\eta^2 < .01$ and $F(1, 71) < 1$, $p = .583$, $\eta^2 < .01$, and there was no statistically significant interaction of the threat-potential and type-of-behavior variables either, $F(2, 70) < 1$, $p = .563$, $\eta^2 = .02$.

8.3.2.5 Additional Ratings

Ratings of Severity and Commendation

Figure 8-8 displays the mean ratings of severity and commendation as a function of type of behavior. As intended and as found in the previous experiments, subjects perceived the instances of cheating as notably grave ($M = 4.53$, $SE = .08$). Accordingly, instances of trustworthiness were judged as highly commendable ($M = 4.88$, $SE = .07$). Two one-way MANOVAs performed on the ratings revealed a statistically significant main effect of type of behavior for the severity ratings, $F(1, 71) = 208.90$, $p < .001$ $\eta^2 = .75$, as well as for the

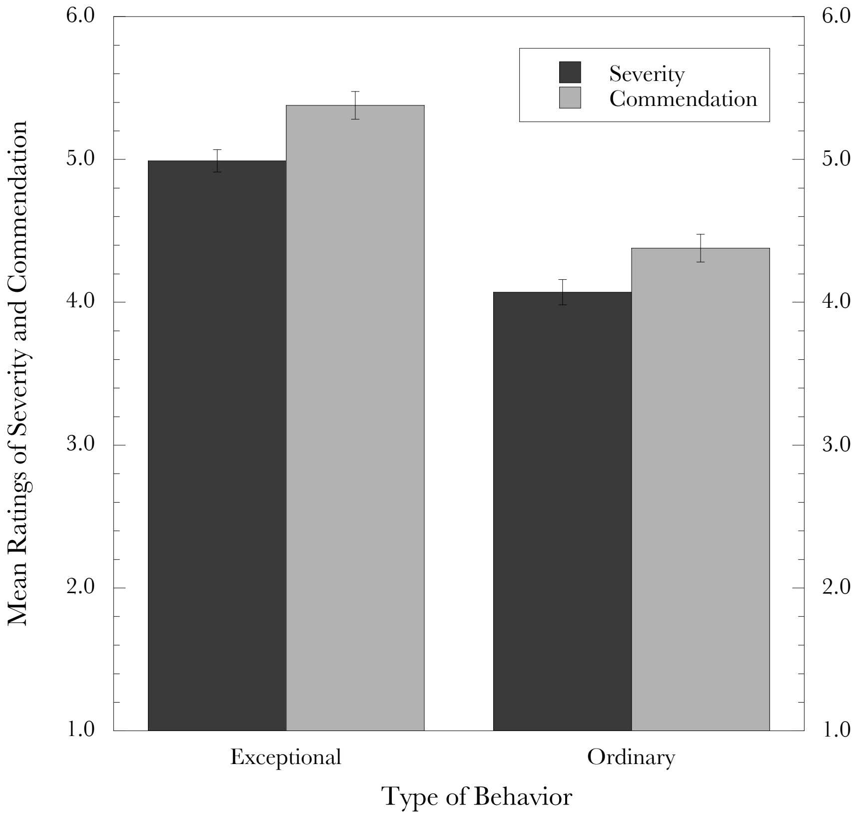

Figure 8-8: Mean ratings of severity and commendation as a function of type of behavior. Subjects chose higher ratings for instances of somewhat exceptional behavior compared to more common behavior. The error bars represent the standard errors of the means.

commendation ratings, $F(1, 71) = 164.25$, $p < .001$ $\eta^2 = .70$. Thus, subjects perceived behavior of cheating as clearly more severe when the behavior was very exceptional compared to more ordinary behavior. In the same way, instances of trustworthiness were judged as even more commendable when the described behavior was unusual compared to everyday behavior.

Relevance Ratings

Figure 8-9 illustrates the mean relevance ratings depending on threat potential and type of behavior. A 2 × 2 MANOVA performed on the data revealed a statistically significant main effect of threat potential, indicating that subjects found instances of trustworthiness to be more relevant than behavior of cheating, $F(1, 71) = 6.66$, $p = .012$ $\eta^2 = .09$. Note that this result is contrary to the findings of the previous experiments. In all of the present experiments so far, subjects evaluated the cases of cheating as more personally relevant than the acts of trustworthiness, and this effect was especially obvious in the low-status condition. Given that in Experiment 4 low-status descriptions had been used exclusively, the present finding appears even more astounding. Apart from that, the main effect of type of behavior

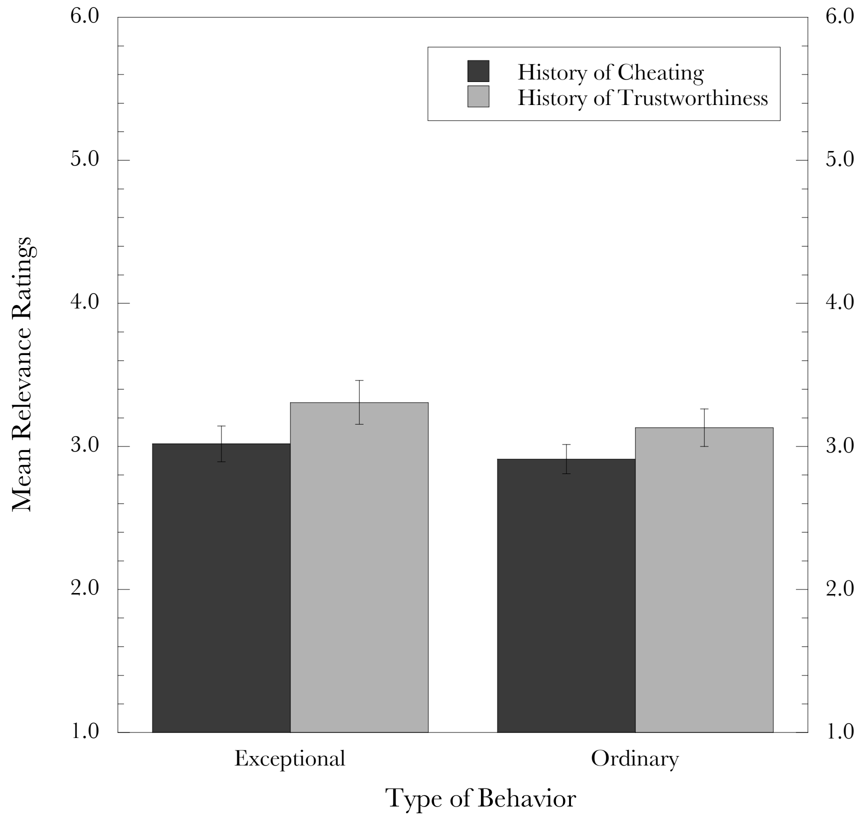

Figure 8-9: Mean relevance ratings depending on threat potential and type of behavior. Independent of the type of behavior, subjects perceived instances of trustworthiness as more personally relevant than instances of cheating. The error bars represent the standard errors of the means.

just missed statistical significance, $F(1, 71) = 3.41$, $p = .069$ $\eta^2 = .05$, and there was no statistically significant interaction of the threat-potential and type-of-behavior variables either, $F(1, 71) < 1$, $p = .684$ $\eta^2 < .01$.

8.4 Discussion

The present results once more confirm what has already become apparent from the previous experiments. Parallel to Experiments 1 to 3, no evidence emerged in favor of a cheater-detection hypothesis in the sense of Mealey et al. (1996). The analysis of the recognition-test data (i.e., the discrimination indices P_r) did not reveal a cheater-detection effect in terms of an enhanced memory for faces of cheaters. This finding was supported by the results from the multinomial analysis of the source-monitoring data, yielding satisfactory model fit in spite of the restriction of equal target-detection parameters. Likewise, the recognition response latencies did not reflect an enhanced responsiveness to faces previously associated with information on cheating.

Parallel to the previous experiments, this failure of a cheater-detection effect to appear may obviously not be ascribed to a deficient manipulation of the threat-potential variable. The analysis of the Week One sympathy ratings showed that persons associated with information on cheating were perceived as less likable than persons associated with irrelevant information or information on trustworthiness, indicating that the threat-potential variable was successfully manipulated. Furthermore, the instances of cheating were judged as sufficiently grave to imply a high threat potential. So far, the results of Experiment 4 are fully in line with the previous experiments.

However, in contrast to Experiment 3, the present results lend support to a modified cheater-detection hypothesis of an enhanced memory for source information on cheating, implying that the possible impact of a mental module as proposed by Cosmides (1989; Cosmides & Tooby, 1989) might be reflected in a higher amount of correct source assignments with respect to information on cheating compared to irrelevant information or information on trustworthiness. The multinomial analysis of the source judgments revealed that the only submodels that fitted the data well did not impose an equality restriction on the source-recollection parameters with respect to the threat-potential variable. In other words, the hypothesis that the probabilities of choosing any of the three possible source categories be equal could not be maintained. Instead, the estimated value of the source-recollection parameter denoting the probability of selecting source category "information on cheating" clearly exceeded the values of the remaining source-recollection parameters indicating the probabilities of selecting source categories "information on trustworthiness" and "irrelevant information". Apparently, subjects reliably remembered source information associated with behavior of cheating, while virtually no source information was remembered with respect to the other source categories, as indicated by parameter values near zero for the respective parameters.

There are numerous factors that may account for the fact that a cheater-detection effect in terms of an enhanced source memory was observed in the present experiment but not in Experiment 3. Of course, the findings of Experiments 3 and 4 may not be directly compared since the experiments differed in many respects. Most notably, there have been considerable changes in the test format of the source-memory test, yielding a different data structure and thus necessitating the application of completely different multinomial models to the data analysis. Following from the fact that the multinomial memory models used define source-recollection parameters as conditional link probabilities, the parameter estimates resulting from various models bear a different meaning contingent on the model structure and, as a consequence, are incommensurable. Apart from this, it is conceivable that in Experiment 3 more false source attributions occurred due to the more complicated test format, potentially masking a cheater-detection effect. This appears reasonable given the results of Marsh and Hicks (1998) who found that source-monitoring accuracy was influenced by whether a single source or multiple sources were the focus of the source-monitoring question as well as by the combination of sources tested. As for Experiment 4, it follows from the data that subjects' memory for source information *not* associated with cheating (or threat) obviously fully decayed over the one-week retention interval. This failure to remember source information is in line with the results of E. Brown et al. (1977) whose subjects performed almost perfectly in a face-recognition task but poorly in a subsequent source-memory test. However, in contrast to the findings of Brown and colleagues, in the present experiment subjects' source memory was not affected in general. Instead, source information on cheating was very well recalled even after one-week. It is conceivable that initially *all* behavioral information presented to subjects was encoded and stored equivalently. Yet subsequently, forgetting may have taken place to a stronger degree for information *not* associated with cheating compared to information implying potential threat. Possibly, depending on the content of source information different consolidation processes operated, resulting in a deeper processing and, consequently, a more durable storage of information on cheating. However, this assumption is highly speculative and needs to be empirically tested. To substantiate this supposition, future research should, as a start, reveal that *all* source information is remembered equally well irrespective of its connotation, given immediate test. In Chapter 10, such future perspectives will be further discussed.

Seeing that in a series of three experiments no cheater-detection effect in terms of an enhanced memory for faces of cheaters could be observed, Experiment 4 was designed to test whether the unusualness of the behavioral descriptions indicating threat potential—rather than their actual content of meaning—might affect subjects' judgments in the recognition test. The present results do not support this assumption. Obviously, the type of behavior was irrelevant with respect to the recollection of Week One faces in the recognition test. Moreover, the multinomial data analysis revealed that the type-of-behavior variable had no impact on subjects' source-recollection performance either. However, it appears that this failure of such effects to emerge may not be ascribed to a deficient manipulation of the type-of-behavior variable. In fact, interesting effects of type of behavior occurred with respect to how subjects judged the severity and commendation of instances of cheating and trustworthiness, respectively. With increasing unusualness, subjects perceived acts of cheating as more severe (more negative), but acts of trustworthiness as more commendable (more positive). This was also reflected by the Week One sympathy ratings, indicating that the degree to which alleged cheaters were perceived as likable *decreased* with increasing unusualness of their behavior, while the reverse was found with respect to supposedly trustworthy persons. Given these results, for the nonce it remains unclear why the type of behavior did in no way influence subjects' face-recognition or source-judgment performance.

Based on the findings of Mealey et al. (1996) the present experiments were designed with the intention to test for the replicability of a biased recognition of faces of cheaters. Yet in a series of four experiments no evidence was found supporting the hypothesis of an enhanced memory for cheaters. At this point, it seems appropriate to consider whether a cheater-detection effect in the sense of Mealey et al. (1996) might have been observable in the present experiments if the statistical power underlying the relevant data analyses had been larger. In fact, as has been reported, a priori power analyses had been conducted for all four experiments, continuously yielding satisfactory probabilities of detecting the effect of interest. These a priori analyses were based upon the claim that a potential cheater-detection effect should at least be of small to medium size as defined by Cohen (1977) to measure up to the great importance that is attached to the cheater-detection module and its functionality by Cosmides (1989; Cosmides & Tooby). However, Mealey et al. (1996) already reported an only diminutive effect and, as well, in all of the previous experiments the effect sizes that were calculated for the statistically non-significant main effects of threat potential on the recognition-test data were virtually negligible. Therefore, a combined analysis of the recognition-test data from Experiments 1 to 4 was conducted to rule out the possibility that the failure to observe a cheater-detection effect as observed by Mealey et al. (1996) was simply a matter of statistical power. In the following, the results of this combined analysis will be reported.

9 Combined Analysis of the Present Experiments

As mentioned above, the following combined analysis of the data from Experiments 1 to 4 was conducted to investigate whether the failure to reveal a cheater-detection effect as observed by Mealey et al. (1996) might have been due to an insufficient statistical power underlying the data analyses that have been made in the context of each of the previously reported experiments. This assumption derived from the finding that the study of Mealey et al. (1996) as well as the present four experiments yielded diminutive effect sizes with respect to the potential cheater-detection effect.

Table 9-1 provides a survey of the effect sizes calculated for the cheater-detection effect—that is, the main effect of threat potential on the recognition-test data—with respect to the study of Mealey et al. (1996) and the present experiments. Evidently, the size of the effect is persistently very small—or, to put it bluntly, completely negligible. To find an effect of such small size requires a very high statistical power which may, in turn, be achieved by an enormous sample size. Therefore, to enlarge the statistical power underlying the analysis of the potential cheater-detection effect, the recognition-test data derived from all of the present experiments were conjointly analyzed[62].

Given the total sample size of $N = 355$, $\alpha = .05$, and the assumption that the average population correlation between the levels of the repeated measures factor threat potential is $\rho = .60$ (estimated from the previous experiments), effects of size $f = .09$ (that is, $\eta^2 = .04$) could be detected for the threat-potential variable with a probability of $1 - \beta = 0.96$, approximately.

However, the one-way MANOVA performed on the discrimination indices P_r did not reveal a statistically significant main effect of threat potential, $F(1, 354) = 3.33$, $p = .069$, $\eta^2 < .01$, where P_r for faces of cheaters only negligibly exceeded P_r for faces of trustworthy persons (Ms = .61 vs. .60, SEs = .01 vs. .01). Thus, even with an immense statistical power, no evidence could be detected for an enhanced memory for faces of cheaters. Of course, adherents of the cheater-detection hypothesis might argue that in the combined analysis—as

62 Note that, with respect to the differences in the four experimental designs, this combined analysis was confined to comparing the mean discrimination indices P_r obtained for faces associated with information on cheating and trustworthiness.

Table 9-1

Overview on the effect size measures η^2 as well as the corresponding *F*-statistics and sample sizes with respect to the main effect of threat potential derived from the pilot study of Mealey et al. (1996) and the four experiments that constitute the present work.

Data source	*N*	*F*-statistic	Empirical η^2
Mealey et al. (1996)	124	$F(2, 244) = 6.63$*	$\eta^2 = .05$
Experiment 1	96	$F(2, 94) < 1$	$\eta^2 < .01$
Experiment 2	123	$F(1, 122) < 1$	$\eta^2 < .01$
Experiment 3	64	$F(1, 63) < 1$	$\eta^2 < .01$
Experiment 4	72	$F(2, 70) = 2.73$	$\eta^2 = .07$

* $p < .05$

well as in Experiment 4—the critical level for statistical significance was just missed. Maybe, reverting to an even larger sample size *might* in the end be sufficient for detecting a statistically significant effect. However, aside from an effect's statistical significance, the practical significance needs to be concerned as well. Given the great importance that is attached to the cheater-detection effect by proponents of *Social Contract Theory*, effect sizes as small as found in the previous experiments may hardly be conceived as practically significant.

10 General Discussion

The General Discussion first provides a résumé of the present work. Subsequently, with respect to the present results, implications of *Social Contract Theory* will be discussed. In the following, some of the central propositions of evolutionary psychology which are constitutional to *Social Contract Theory*, will be challenged. Finally, future perspectives will be elaborated with respect to an empirical validation of *Social Contract Theory* and its implications.

10.1 Résumé

The present work focused on the central claim of Leda Cosmides' so called *Social Contract Theory* (Cosmides, 1989; Cosmides & Tooby, 1989), comprising the existence of neural circuits inherent in the human brain that are functionally specialized for the detection of cheaters. Since *Social Contract Theory* reverts to a Darwinist approach to psychology, the fundamental principles of this evolutionary psychology were presented and the formation, basic assumptions, and predictions of *Social Contract Theory* were explicated. An outline of research on the issue of cheater detection, solely based on Wason's (1966, 1968) selection task, was provided and findings both supporting and challenging *Social Contract Theory* were analyzed. A face-recognition paradigm was introduced as an alternative methodological approach to the investigation of cheater-detection effects. Furthermore, findings previously achieved by means of this method were reported, suggesting an enhanced memory for faces of cheaters compared to non-cheaters (Mealey et al., 1996). Finally, results from a series of four experiments based on the afore-mentioned face-recognition paradigm were illustrated.

Experiment 1 was designed as a close replication of the work of Mealey et al. (1996) but failed to find evidence in favor of an enhanced memory for faces of cheaters. Obviously, the threat-potential variable had no impact on subjects' recognition performance. Apart from that, an (implicit) cheater-detection effect was not observable in the Week Two attractiveness ratings either. Experiment 2 was derived from Experiment 1 and focused more strongly on behavior of cheating and trustworthiness as opposite instances of threat potential. The results closely matched the findings of the first experiment. Again, no cheater-detection effect in terms of an enhanced memory for faces of cheaters was found. In Experiment 3, the one-week retention interval realized in the first two experiments following Mealey et al. (1996) was omitted to rule out the possibility that the failure to observe a cheater-detection effect so far was simply a matter of time. Yet replicating the findings of the previous experiments, no evidence emerged in favor of a cheater detection effect in the sense of Mealey et al. (1996). Finally, Experiment 4 was designed as a modification of Experiment 3 but once more failed to confirm the hypothesis of an enhanced memory for faces of cheaters. In a nutshell, the results from all four experiments were not in line with the idea of a functionally specialized cheater-detection module improving subjects' recognition performance with respect to faces of cheaters.

In Experiments 3 and 4, for the first time, a source-memory approach to the investigation of a potential cheater-detection effect was realized. With respect to the failure to observe such an effect in Experiments 1 and 2, it seemed adequate to challenge the pertinence of the underlying cheater-detection hypothesis. Based on the assumption that an advanced memory for faces of cheaters might not be an essential but only a sufficient implication of cheater detection, a modified cheater-detection hypothesis was deduced. According to this modified hypothesis, accurate memory for source information on prior cheating was considered a necessary precondition for evaluating a potential interactant's threat potential. Consequently, it was assumed that the possible impact of a mental module as proposed by Cosmides (Cosmides, 1989; Cosmides & Tooby, 1989) might be reflected in a higher amount of correct source assignments with respect to information on cheating compared to other source information. To test this modified cheater-detection hypothesis, in Experiment 3 as well as in Experiment 4 a source-memory test was embedded in the recognition-test phase. Multinomial models of source memory were applied to the analysis of the source-monitoring data.

Experiment 3 failed to confirm the modified cheater-detection hypothesis of an enhanced memory for source information on cheating. The analysis of the source-monitoring data using Meiser and Bröder's (2002) multinomial memory model for crossed source information revealed an excellent model fit given the restriction of equal source-detection parameters for all source information, which is incompatible with the modified cheater-detection hypothesis. Experiment 4, however, yielded evidence in favor of a cheater-detection effect in terms of an enhanced source memory for information on cheating. The analysis of the source-monitoring data using an adapted version of a source-memory model proposed by Bayen et al. (1996) revealed that, independent of prior target detection, subjects showed an improved memory for source information on cheating compared to source information not associated with cheating (or threat).

Taken together, the results of the present of experiments did in no way support the cheater-detection hypothesis in the sense of Mealey et al. (1996). Throughout a series of four experiments, an effect of an enhanced memory for faces of cheaters continuously failed to emerge. Consequently, on the basis of the present results the postulate of an enhanced memory for faces of cheaters due to a looking-for-cheaters procedure that is triggered by a functionally specialized mental module cannot be sustained. Solely the results of Experiment 4, suggesting superior source memory for information on cheating, give reason to further speculate about the possible effects of a cheater-detection algorithm as proposed by *Social Contract Theory* (Cosmides, 1989; Cosmides & Tooby, 1989).

10.2 Enhanced Memory for Faces of Cheaters—Essential or Sufficient Implication of Cheater Detection?

The results of the present experiments raise the question whether an enhanced memory for faces of cheaters is an essential or just a sufficient implication of cheater detection. Actually, in a series of four experiments no such effect was observed. Instead, the only indication for cheater detection arose from the multinomial analysis of the source-monitoring data obtained in Experiment 4. While subjects did not show an enhanced memory for faces of alleged cheaters compared to faces of non-cheaters, they were obviously able to reliably recollect source information on cheating and, at the same time, failed to remember any other source information. Given the importance ascribed to the proposed cheater-detection mechanism with respect to its potentially protective function, one may wonder whether such functionality is met by the sole recollection of information that cheating has occurred but not of the cheater per se. In other words, it appears reasonable to consider how one should be able to avoid future interactions with cheaters if one fails to identify persons who cheated in the past.

Adherents of *Social Contract Theory* might argue that the prevention of potentially harmful stimuli does not necessarily require conscious rejection. Consider, for example, food aversions. It is known that humans as well as numerous mammals avoid potentially toxic aliment even if negative consequences are long delayed (Birbaumer & Schmidt, 1991), whereas conscious recollection of prior experience with the toxin is not an essential precondition. Instead, the mere perception of, for example, displeasing odors is sufficient to elicit specific aversions, triggered by specific food selection mechanisms and independent from consciousness or awareness (Buss, 2004). Thus, one might argue that an unspecific feeling of discomfort in the presence of an earlier cheater—rather than the reminiscence of his face as well as of episodic details with respect to prior encounters—should be sufficient to affect an individual's decision on future interactions with this person who might potentially cheat repeatedly. However, with respect to excluding a known cheater from future interactions, only the conscious recollection of the circumstances under which this person committed deception before allows for balancing his motives, for weighing up the consequences arising for the deceived, and thus for drawing inferences on the actual threat potential of this person. From this point of view both an explicit memory for the cheater (that is, for the cheater's face) and for the episode of cheating may be regarded as essential implications of the proposed cheater-detection module.

Of course, one might argue that, in her primary publication of *Social Contract Theory*, Cosmides (1989) did not manifest herself at all with regard to the possible impact of the proposed looking-for-cheaters mechanism on face recognition. Closely following her narrow definition of cheating as a violation of social-contract rules of the type "If you give me *P*, then I will give you *Q*" and in accordance with the use of the *Wason Selection Task* as the only method for testing her theory, Cosmides' (1989) original cheater-detection hypothesis focused on performance in logical reasoning. Thus, strictly speaking, one may not view the present experiments as providing a proximate test of Cosmides (1989) hypotheses on cheater detection. However, such plea may be overruled since the assumption of an enhanced memory for faces of cheaters initially hypothesized by Mealey et al. (1996) simply is a corollary of the initial postulates concerning the functionality of the proposed mental module. This, in turn, suggests a revision of the original cheater-detection hypothesis and, more important, of the underlying definition of cheating.

10.3 Challenging *Social Contract Theory*

As indicated above, a general issue that should be discussed with respect to *Social Contract Theory* is the definition of cheating. Starting from the fundamental assumption that the mental modules proposed in the context of Cosmides and Tooby's theoretical framework (Cosmides, 1989; Cosmides & Tooby, 1987, 1989, 1992, 1994, 1997, 2002; Tooby & Cosmides, 1989) are designed for solving problems arising from Pleistocene environmental conditions, the narrow definition of cheating as a violation of social-contract rules of the form "If you give me *P*, then I will give you *Q*" may appear plausible. However, there are so many facets of defraud that an individual may experience that it seems justified to expand the initial definition of cheating provided by *Social Contract Theory*. For example, no one would probably dissent that committing adultery is a form of cheating even if there is no evident cost-benefit structure. Of course, one might think of matrimony as a cost-benefit-relation such as “If you are loyal to me, then I will be loyal to you”. However, this appears somewhat arranged and obviously does not commensurate with Cosmides’ (1989) primary definition of a social contract in which two individuals engage to enhance their chances of survival. Similarly, fiscal fraud, to cite another example, is beyond doubt a form of cheating. Still, it does not fit Cosmides’ (1989) original definition of cheating since it hardly emanates from a clear-cut social contract in the sense of reciprocity.

Still, one might criticize that the failure to observe a cheater-detection effect in the present experiments could be ascribed to the fact that the behavioral descriptions that were used did not represent social contracts in terms of cost-benefit-relations as defined by Cosmides (1989). Apart from the fact that, in most instances, the behavioral descriptions do very well imply social contracts—is this really a matter of interest? To put it bluntly, to those committed to a restrictive definition of cheating provided by Cosmides (1989) the question may be addressed of the actual usefulness of a functionally specialized mental module conducing to the detection not of cheaters in general, but of persons who commit an accurately defined variety of cheating. Obviously, in the course of our evolutionary history, more and more diversified means of social interactions have evolved, entailing just as manifold opportunities of committing defraud. Reverting to one of the fundamental propositions of evolutionary psychology denoting the development of mental modules operating to produce adaptive behavior, it appears justified to expect that such mental modules should have advanced in their functionality as well to satisfy the continuously changing requirements for adaptation resulting from our ecological and social environment. This conclusion, however, is only partially compatible with the constitutive principles of evolutionary psychology, as will be outlined in the following.

10.4 Challenging Fundamental Principles of Evolutionary Psychology

As suggested in the previous paragraph, at least some of the fundamental principles of evolutionary psychology appear questionable. In the following, possibly the most disputed of these principles shall be surveyed with respect to the results of the present experiments.

10.4.1 Stone Age Minds in Modern Skulls?

One of the basic assumptions held by evolutionary psychologists denotes that any cognitive mechanisms inherent in the human brain evolved for solving adaptive problems resulting from Pleistocene environmental conditions. This, however, completely abstracts away from the fact that evolution is by no means "accomplished" but rather continues persistently. Of course, it is true that, for the most part, our species' evolutionary history has taken place under Pleistocene conditions. Yet from the emergence of agriculture, at the latest, the development of the human species has been making rapid progress. Along with consistently opening up new habitats and, finally, with the onward industrial development, our ancestors have been increasingly faced with novel requirements for adaptation. After all, considering contemporary conditions of life (at least, with respect to industrial nations), it is obvious that the adaptive problems we are faced with are not at all comparable to those of our hunter-gatherer ancestors. Consequently, if we are in fact endowed with mental modules to solve these adaptive problems, then these modules should have adapted in their functionality as well. Instead, solving today's problems via outdated mental equipment, so to speak, hardly qualifies as adaptive problem solving.

Cosmides and Tooby (1997) refer to this manifest flaw in the rationale held by evolutionary psychologists by pointing out that our ability to solve a variety of present-day problems must be regarded as a *by-product* of our evolved problem-solving circuits. By-products are defined in terms of "characteristics that do not solve adaptive problems and do not have functional design. They are 'carried along' with characteristics that do have functional design because they happen to be coupled with those adaptations" (Buss, 2004, p. 40). As an example, Buss quotes the belly button as a by-product of the umbilical cord that may be regarded as an adaptation since its development helped to solve the problem of prenatal food supply. The belly button per se, in contrast, possibly did not help humans to survive and therefore may not be viewed as an independent adaptation (Buss, 2004).

While the example of the belly button entails at least some plausibility, it appears questionable how the dazzling array of abilities we are equipped with today may be defined in terms of by-products, tracing back to Pleistocene mental modules. As Buss (2004, p. 41) points out, the "hypothesis that something is a by-product of an adaptation […] requires identifying the adaptation of which it is a by-product and the reason why its existence is associated with the adaptation." At least, Buss (2004, p. 41) further concedes that the hypothesis that something is a by-product rather than an adaptation must be "subjected to rigorous standards of scientific confirmation" in that "specific empirical predictions must be derived from each by-product hypothesis and then tested using empirical methods." While his claim appears reasonable in theory, its actual implementation in practice seems far from feasible.

Apart from that, it appears interesting to consider whether the concept of by-products is applicable to *Social Contract Theory*'s central assumption of a functionally specialized mental module operating in terms of a looking-for-cheaters procedure and to the results of the present experiments. Cosmides (1989) originally defined the cheater-detection algorithm as an adaptation to Pleistocene conditions. Consequently, the proposed mechanism is apt only for solving adaptive problems from the past but not from 21st century conditions. From the perspective of evolutionary psychology, this might explain the failure to observe a cheater-detection effect in the present experiments. That is, the experimental task might have failed to adequately address the mechanism of interest since it possibly did not sufficiently resemble the kind of problem for which to solve the mechanism was developed. On the other hand, for example, subjects' superior performance in the "drinking-age problem"—which Cosmides frequently cites as evidence in favor of *Social Contract Theory*—rather supports a by-product interpretation of cheater detection, since reasoning on national laws referring to the minimum age for drinking alcoholic beverages hardly qualifies as a Pleistocene adaptive problem. This, in turn, raises the question why no unambiguous effects of cheater detection could be observed in the present series of experiments—given the assumption that there actually exist however natured by-products of the primary adaptation of a cheater-detection module. Thus, the issue of adaptations and by-products is elusive and the present results do not allow for a clear-cut conclusion in this regard.

10.4.2 Modularity of Mind?

The notion of the human mind as made up of so-called evolved psychological mechanisms, or mental modules, is definitely the most disputed principle on which evolutionary psychology is based. For a long time there has been dissension between proponents of this modular conceptualization of the human mind and adherents of an alternative view defining the mind as made up of only few general learning mechanisms. Although the present work is far from putting an end to this debate, the results of the present experiments at least allow for weighing up some of the arguments that have been brought forward by evolutionary psychologists. Altogether, the results of the present series of experiments are not in line with the assumption of a functionally specialized mental module operating for the detection of cheaters. Yet as discussed in the previous paragraphs, this may be due to a number of factors other than just a misapprehension of the human mind's functionality. Apart from that, however, the conceptualization of the mind in terms of a virtually infinite number of mental modules remains in some ways unpersuasive.

First of all, the postulation of specialized mental modules for any single adaptive problem hardly makes sense from an economical point of view. Take for example the proposition of an evolved psychological mechanism responsible for learning to fear snakes, which is frequently cited in the literature as an example for the domain specificity of the proposed mental modules. Of course, no one would possibly dispute that humans, in general, are afraid of or at least disrelish snakes. Yet the same is true for spiders and numerous other insects as well as for carnivores or various reptiles. From the view of evolutionary psychology, for each of these fears there is a distinct cognitive mechanism responsible, operating to evoke an adaptive response (e.g., fight or flight). Beyond doubt, it appears plausible to trace back specific fears to specific mental modules and, apart from that, it satisfies the demand for precisely explaining human performance (Frensch & Buchner, 1999). But just as well it seems appropriate to expect that more parsimonious adaptations have outstripped unthrift solutions in the course of evolution. Following this rationale, it appears more probable that, instead of numberless highly specified mental modules, few more general problem-solving mechanisms have evolved that somehow build the substructure for human reasoning and may be extended in their functionality by means of learning. To get back to the example of snake phobia, this implies the general ability of sensing fear plus the ability to acquire knowledge on possibly harmful stimuli (such as vipers, but not blindworms) and to retain such information in memory. Consider for example urbanites that will, due to the characteristics of their habitat, hardly be confronted with vipers or other dangerous animals. It appears like a waste of capabilities to assume that they are equipped with a great number of cognitive mechanisms to produce adaptive responses to such stimuli. Instead, the supposition of a general learning mechanism that enables us to acquire relevant knowledge that, in turn, guides our behavior, seems far more conclusive. As Lawson (2002) suggests, this does not necessarily rule out any reasoning from an evolutionary perspective since one may consider that humans are born with such fundamental logical competence. However, at least some proponents of evolutionary psychology acknowledge the central problem of defining more and more specific, yet empirically unexaminable modules to come up to the great variety of abilities humans are endued with (e.g., see Cartwright, 2000).

Apart from the aspect of economy, another questionable aspect of the notion of cognitive modules refers to the proposition of their domain specificity. For example, there is still disagreement on the question of what a domain is. Even Kanwisher (2000, p. 762) who argues in favor of the domain-specific approach—at least with respect to face recognition—concedes that "[i]t seems unlikely that all of cognition will be subserved by discrete modular mechanisms, and also unlikely that all modules that exist will be domain specific." Somewhat more liberal, she holds the view that "[a] more reasonable hypothesis is that the degree of modularity and the degree of domain specificity within modules will vary across the brain and across aspects of cognition." Similarly, Buss (2004, p. 57) points out that "the human mind cannot consist solely of isolated separate mechanisms that are entirely walled of from each other." Instead, he proposes that "[s]election favors functionally specialized mechanisms *that work well together* in various combinations and permutations" and that "humans also likely have *superordinate mechanisms*[63] that function to regulate other mechanisms." In the end, this appears to be a far more rationale concept of the human mind's functionality than the limited insistence on discrete functionally specialized cognitive mechanisms as proposed by consistent skeptics of the domain-general view (e.g., see Cosmides & Tooby, 2002). Still, as Frensch and Buchner (1999) noted, the issue of domain-generality versus domain-specificity debates continues to influence much of our thinking about how the human mind works, and an end of such debates is far from foreseeable.

10.5 Future Perspectives

The purpose of the present work was the empirical test of Cosmides' (1989) hypothesis about the existence of a specific competence to deal with social exchange using an experimental paradigm that stands out from the methodological approaches favored so far. This derived from the astounding observation that, although the cheater-detection hypothesis had been at the center of heated debates for more than 15 years, it had not yet been properly tested since almost all evidence in favor of *Social Contract Theory* was based on the *Wason Selection Task*—which has been shown to be inappropriate for this purpose (Sperber & Girotto, 2002). Based on prior work of Mealey et al. (1996), the present series of experiments comprised an elaborate empirical test of Cosmides' hypothesis and, moreover, provided various clues for future research that will be discussed in the following.

---

[63] Italics in original.

For example, as outlined previously, one might trace back the failure to observe an enhanced memory for faces of cheaters in the present experiments to the stimuli that did not entirely imply social contracts as defined by Cosmides (1989). With respect to future research, one might think of an experimental design in which two faces—instead of only one as in the present experiments—are shown in each trial of the acquisition phase, representing interactants in a social contract. Following the procedure of the present experiments, information on both persons and a social-contract rule of the type "If you give me *P*, then I will give you *Q*" might then be typed below the photos. By means of these descriptions, it might be varied whether (1) both of the depicted persons conform to the rule or (2) both violate the rule, or (3) one of them acts on the rule while the other breaks it. In a subsequent recognition test, subjects' recognition performance might then be tested with respect to faces of cheaters, non-cheaters, and new distractor faces, as in the previous experiments. The results of such an experiment might shed light on the question whether the violation of a clear-cut social contract as defined by Cosmides (1989) is an essential precondition for the alleged cheater-detection mechanism to be addressed.

Futhermore, one more general constriction of the study of Mealey et al. (1996) as well as of the present experiments refers to their ecological validity. In all experiments, subjects were not actually involved in social contracts themselves but rather had to imagine various situations in which deception takes place. There are many factors that may have influenced how exactly subjects figured the presented instances of cheating or trustworthiness. As for the present experiments, assessing the relevance ratings was concerned to shed light on whether the intended severity of behavior of cheating had successfully been conveyed by way of imagination. Still, it seems appropriate to consider an alternative experimental paradigm in which subjects actually engage in social interactions and experience defrauds by real or possibly virtual interactants.

Basically following the design of Mealey et al.'s (1996) experiment, Oda (1997) at least partially realized this conception. In her experiment, subjects were asked to engage in a series of single-shot Prisoner's Dilemma games with virtual opponents who were represented by photographic reproductions of women and men, accompanied by labels indicating whether they were cooperators or defectors. One week later, in a recognition test subjects were presented faces from Week One and new distractor faces and were to indicate which of the photographs they remembered from the previous week. It was found that recognition performance was significantly lower with respect to cooperating male opponents than with respect to defecting males or females playing either strategy. Furthermore, this effect was independent from the strategy preassigned by the subjects themselves. However, the size of the effect turned out to be negligibly small (namely, $\eta^2 = .01$), which resembles the results of Mealey et al. (1996) as well as the present findings. Apart from that, presenting a photo with defection as the strategy did not necessarily mean that the depicted person was a cheater, which was due to the situation used in Oda's (1997) experiment. As outlined in Chapter 3, in the one-move Prisoner's Dilemma game defection turns out to be the most successful strategy. Cooperation, however, can only be initiated when the same players repeat the game or, in other words, when reciprocity can take place. Yet Oda (1997) argues that although a defector in the single-shot version of the game does not equal a cheater, a defector is a *potential* cheater who threatens to cheat in future interactions. Thus, subjects still had to imagine being cheated in possible future interactions, which results in the same problem as outlined with respect to the present experiments, namely that subjects' responses in the recognition test rely on uncontrollable imaginations of cheating rather than on actual defraud sustained.

This raises the question whether the results reported by Oda (1997) may be replicated in a slightly modified version of her experiment in which subjects acquire knowledge about their interactants in the course of repeated instead of single-shot Prisoner's Dilemma games. In repeated games, the degree to which subjects are cheated by their interactants could be proximately varied, and instead of just imagining that any of the persons presented *might* possibly cheat, subjects could revert to their own experience of being defrauded. Compared to the marginal effect sizes observed by Oda (1997) and Mealey et al. (1996) as well as in the present experiments, this should plausibly result in a more robust effect on subjects' recognition performance—given the cheater-detection hypothesis was true. Still, having subjects play the Prisoner's Dilemma game equals a somewhat artificial situation. Thus, for future research one might think of experimental conditions in which subjects engage in interactions that fit 21st century everyday life such as online auctions, for example.

Maybe the most auspicious indication for future research on the empirical validation of *Social Contract Theory* consists in the multinomial modeling of source memory. The present results denote that multinomial models of source monitoring are appropriate to uncover effects that may not be detected by means of more traditional analyses. Of course, the inhomogeneous results of Experiments 3 and 4 demand for a close examination of potential influencing factors. At present it is unclear why a cheater-detection effect in terms of an enhanced source memory for information on cheating was found in Experiment 4 but not in Experiment 3. As discussed in Chapter 8, the experiments differed in many respects, which does not allow for a direct comparison of the results. Future research should systematically deal with the development of identifiable models that appropriately represent the cognitive processes of interest. Yet to begin with, the so far unique finding of Experiment 4 should be replicated to substantiate the conclusion that there may be effects of cheater detection affecting source memory. It was hypothesized that the cheater-detection effect observed in Experiment 4 might have been caused by differences in the encoding and storage of source information depending on its content of meaning. More precisely, it was assumed that initially *all* source information might be encoded and stored equivalently. Subsequently, however, information on cheating might be further encoded more deeply and thus stored more durably than information *not* associated with cheating (or threat). This, in turn, might enhance later source memory for information on cheating. Given this assumption was true, no such effect should emerge with source memory tested immediately after information encoding. Therefore, to test this hypothesis a series of experiments should be conducted in which, basically following the design and procedure of Experiment 4, the retention interval is systematically varied.

Apart from this, with respect to future research one might think of addressing the issue of whether the cheater-detection effect on source memory observed in the present Experiment 4 is actually an effect of *cheater* detection. Just as well, one might think of a more general negativity bias to explain the present findings. Based on the above-outlined assumption that effects on source recollection as found in Experiment 4 might be due to differences in the encoding and storage of source information depending on its connotation, it is conceivable that not only specific information on cheating but rather negatively valent information in general is processed more deeply than, for example, neutrally valent information. This assumption is substantiated by multifaceted empirical findings concerning the impact of valence of information on information processing. For example, Fox et al. (2000) found that angry facial expressions were processed more efficiently than friendly faces. Also, in Stroop tasks negative-trait adjectives have been shown to delay the naming of the color in which they are printed more than positive-trait adjectives, and valent words turned out to delay responding more than neutral words—presumably by automatically attracting attention toward events that may have undesirable consequences for the perceiver's well-being (Pratto, 1994; Pratto & John, 1991). Wentura, Rothermund, and Bak (2000) showed that such attention-grabbing effects may be particularly distinct if valent information can be characterized as "other-relevant", with other-relevant traits denoting unconditionally positive or negative consequences for persons in the social environment of the trait-holding person. Thus, for example, other-relevant negative trait adjectives may serve as highly overlearned cues to potential threats in the environment that call for attention (Buchner, Rothermund, Wentura, & Mehl, 2004). Buchner et al. (2004) found that valent adjectives presented as distractors while memorizing target words impaired recall performance more strongly than neutral distractors[64]. Moreover, negative distractors caused more disruption than positive distractors, and this was true in particular when the negative distractor words were other-relevant. Obviously, negatively valent information, especially when other-relevant, appears to hold a form of announcement effect, indicating potential threat from the social environment and affecting information processing by directing attention towards potentially threatening environmental conditions. Thus, future research might shed light on the question of whether the present finding of an enhanced memory for source information on cheating simply denotes an enhanced memory for negatively valent information, somehow or other. To test this assumption, one might think of an experimental design in which source information is presented implying various instances of potential threat, with one source of threat deriving from potentially being cheated in the sense of *Social Contract Theory* (Cosmides, 1989).

At length, future research might focus on the investigation of physiological correlatives of the looking-for-cheaters procedure as defined in the context of *Social Contract Theory*. Interestingly, while obstinately insisting on the existence of a cheater-detection module, Cosmides (1989) did not manifest herself with respect to the physiological properties of the proposed module. Geary (1998), in the context of his model of hierarchically organized mental modules, addresses that considering a one-to-one correspondence of mental modules and certain brain regions is, in all probability, oversimplified. Instead, he suggests that localized ensembles of cells that are specialized for domain-specific information processing may operate synchronously, resulting in specialized cognitive competencies. It seems appropriate to assume that if a functionally specialized cheater-detection module really exists, then one should be able to relate its functionality to identifiable cerebral regions. Future research may address this issue making use of appropriate methods of measurement. For example, the assessment of event-related brain potentials with respect to the encoding and

[64] Note that this was found even for non-words artificially associated with valence (Buchner, Mehl, Rothermund, & Wentura, in press).

later retrieval of information on cheaters compared to non-cheaters might shed light on the so far questionable character of the proposed looking-for-cheaters mechanism.

Taken together, the present work opens up various links to future research, comprising a broad variety of methodological approaches. With respect to the so far unidirectional orientation in empirically testing *Social Contract Theory*, a multi-method approach to future research seems appropriate, if not necessary, considering the importance of findings on this issue with respect to an evaluation not only of *Social Contract Theory* but also of fundamental principles of evolutionary psychology. This, in turn, gives reason to dwell on the subject of the future of evolutionary psychology. Ten years ago, the evolutionary psychologist David Buss philosophized on the future of his discipline, worrying that "a century in the future, evolutionary psychology would be seen as merely a footnote in the history of psychology, sort of like phrenology—an intriguing idea, perhaps, but one that had not panned out" (Buss, 1995, p 81). Supposably, this amounts to an overstatement. At least, as reflected by an increasing plentitude of scientific papers in this field of research, evolutionary psychology is en vogue. Yet many inconsistencies in the rationale proposed by adherents of evolutionary psychology, most of which have been addressed in the course of the present work, still give rise to resistance and skepticism to evolutionary perspectives. Thus, evolutionary psychology is far from revolutionizing psychological research although such progression is consistently predicted by its proponents—for example, Buss (1995, 1999; Buss & Reeve, 2003) suggests the dogmas of his field as the guiding metatheory for psychological science. Yet he concedes that some adherents of evolutionary psychology carry their creed to excess, prophesizing that "at some point in the future, the term *evolutionary* would be dropped entirely from *evolutionary psychology* because the entire field of psychology will be evolutionary, and the qualifier would be superfluous" (Buss, 1999, p. 26)[65]. In contrast, Shapiro and Epstein (1998) propose a far more tempered and, as it seems, more adequate view of the changes that evolutionary theory is likely to bring, in particular, to cognitive psychology. In their elaborate discussion of the central claims of Cosmides and Tooby as well as of other evolutionary psychologists concerning the future impact of evolutionary doctrines on psychological research, the authors conclude (Shapiro & Epstein, 1998, p. 192):

> Will evolutionary theory transform cognitive psychology? On the one hand, if this entails a drastic reorientation from a domain-general view of the mind to a domain-specific one, then such a transformation is unlikely. […] On the other hand, if an affiliation with evolutionary theory requires that cognitive psychology treat all behavior as the product of cog-

[65] Italics in original.

> nitive processes, then it is an affiliation cognitive psychologists should resist. [...] However, if the marriage of evolutionary theory to cognitive psychology calls only for cognitive psychologists to be more self-conscious in their employment of teleological reasoning and to consider the ultimate factors that have influenced the nature of our cognitive capacities, then it has our blessing.

Thus, one may regard evolutionary theory as making some illuminative, yet not essential contributions to the study of cognition—for example, by providing a helpful means by which to formulate hypotheses about the function of cognitive mechanisms (Shapiro & Epstein, 1998). However, unless clear-cut evidence in favor of the central propositions of evolutionary psychology is provided, evolutionary theory will, as Shapiro and Epstein (1998) bluntly put it, supposedly play a merely heuristic role in the methodology of cognitive psychology. As a matter of fact, extensive future research is necessary to clarify at least some of the numerous unanswered questions concerning the structure and functionality of the human mind.

11 References

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., Anderson, A., Lee, G. P., & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*, 1111-1117.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390-1396.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564.

Batchelder, W. H., Riefer, D. M., & Hu, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, *101*, 172-176.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197-215.

Beaman, C. (2002). Why are we good at detecting cheaters? A reply to Fodor. *Cognition*, *83*, 215-220.

Birbaumer, N. & Schmidt, R. F. (1991). *Biologische Psychologie* (2nd edition). Berlin: Springer.

Blair, R. J. R., & Curran, H. (1999). Selective impairment in the recognition of anger induced by diazepam. *Psychopharmacology*, *147*, 335-338.

Bracewell, R. J., & Hidi, S. E. (1974). The solution of an inferential problem as a function of stimulus materials. *Quarterly Journal of Experimental Psychology*, *26*, 480-488.

Brown, C. A., Keats, J. A., Keats, D. M., & Seggie, I. (1980). Reasoning about implication: A comparison of Malaysian and Australian subjects. *Journal of Cross Cultural Psychology*, *11*, 395-410.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, *62*, 311-318.

Bruce, V. (1988). *Recognising Faces*. Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.

Bruce, V., Burton, A., & Craw, I. (1992). Modelling face recognition. In A. Cowey, V. Bruce, et al. (Eds.), *Processing the facial image* (pp. 121-128). New York, NY, US: Clarendon Press/Oxford University Press.

Bruce, V., Burton, A., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception*, *22*, 131-152.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305-327.

Buchner, A., Faul, F., & Erdfelder, E. (1996). G-Power: A priori, post-hoc, and compromise analyses for the Macintosh (Version 2.1.2) [Computer program]. Düsseldorf, Germany, Heinrich-Heine-Universität (URL: http://www.psycho.uni-duesseldorf.de /aap/projects/gpower/).

Buchner, A., Mehl, B., Rothermund, K., & Wentura, D. (in press). Artificially induced valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory and Cognition*.

Buchner, A., Rothermund, K., Wentura, D., & Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. Memory and Cognition, 32, 722-731.

Burton, A., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*, 153-176.

Buss, D. M. (1995). The future of evolutionary psychology. *Psychological Inquiry, 6*, 81-87.

Buss, D. M. (1999). Evolutionary psychology: a new paradigm for psychological science. In: M. C. Luebbert & D. H. Rosen (Eds.), *Evolution of the psyche* (pp. 1-33). Westport, CT, US: Praeger.

Buss, D. M. (2004). *Evolutionary psychology: The new science of the mind* (2nd edition). Boston, Mass., US: Pearson.

Buss, D. M. & Reeve, H. (2003). Evolutionary psychology and developmental dynamics: comment on Lickliter and Honeycutt (2003). *Psychological Bulletin, 129*, 848-853.

Cartwright, J. (2000). *Evolution and human behavior*: *Darwinian perspectives on human nature.* London: Macmillan Press Ltd.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391-416.

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition*, *33*, 285-313.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, *18*, 293-328.

Chomsky, N. (1980). Rules and representations. *The Behavioral and Brain Sciences*, *3*, 1-15.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187-276.

Cosmides, L., & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupré (Ed.) *The latest on the best: Essays on evolution and optimality* (pp. 276-306). Cambridge, MA, US: The MIT Press.

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture: II. Case study: A computational theory of social exchange. *Ethology and Sociobiology*, *10*, 51-97.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York: Oxford University Press, Inc.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, *50*, 41-77.

Cosmides, L., & Tooby, J. (1997). *Evolutionary psychology: A primer*. Retrieved from http://www.psych.ucsb.edu/research/cep/primer.html, 14. 03.2005.

Cosmides, L., & Tooby, J. (2002). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In J. C. Kaufman & R. J. Sternberg (Eds.), *The evolution of intelligence* (pp. 145-198). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory and Cognition*, *10*, 496-502.

Cummins, D. D. (1999). Cheater detection is modified by social rank: The impact of dominance on the evolution of cognitive functions. *Evolution and Human Behavior*, *20*, 229-248.

Darwin, C. (1859/1998). *The origin of species by means of natural selection, or, the preservation of favored races in the struggle for life*. New York: The Modern Library (Original work published 1859).

Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology, 106,* 541-557.

Duchaine, B. (2000). Developmental prosopagnosia with normal configural processing. *Neuroreport: For Rapid Communication of Neuroscience Research*, *11*, 79-83.

Duchaine, B., Cosmides, L., & Tooby, J. (2001). Evolutionary psychology and the brain. *Current Opinion in Neurobiology*, *11*, 225-230.

Ellis, H. D. (1975). Recognizing faces. *British Journal of Psychology*, *66*, 409-426.

Erdfelder, E. (2000). *Multinomiale Modelle in der kognitiven Psychologie.* Professorial dissertation, Rheinische Friedrich-Wilhelms-Universität, Bonn.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments and Computers*, *28*, 1-11.

Evans, J. (1972a). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, *24*, 193-199.

Evans, J. (1972b). On the problems of interpreting reasoning data: Logical and psychological approaches. *Cognition*, *1*, 373-384.

Evans, J., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*, 391-397.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition*, *77*, 1-79.

Fodor, J. (1986). *The modularity of mind* (4th ed.). Cambridge, MA, US: The MIT Press.

Fodor, J. (2000). Why we are so good at catching cheaters. *Cognition*, *75*, 29-32.

Fodor, J. (2002). Reply to Beaman. *Cognition*, *83*, 221.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion, 14,* 61-92.

Frensch, P. A. & Buchner, A. (1999). Domain-generality versus domain-specificity: A binary concept and its impact on psychological theory and research. In R. J. Sternberg (Ed.). *The nature of cognition* (pp. 137-172). Cambridge, MA: MIT Press

Geary, D. C. (1998). *Male, female. The evolution of human sex differences*. Washington DC: American Psychological Association.

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*, 127-171.

Gilhooly, K. J., & Falconer, W. A. (1974). Concrete and abstract terms and relations in testing a rule. *Quarterly Journal of Experimental Psychology*, *26*, 355-359.

Going, M., & Read, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills*, *39*, 109-110.

Goldenberg, G. (2003). Visuelle Objektagnosie und Prosopagnosie. In H.-O. Karnath & P. Thier (Eds.), *Neuropsychologie* (pp. 119-131). Heidelberg: Springer.

Goodwin, R. Q., & Wason, P. C. (1972). Degrees of insight. *British Journal of Psychology*, *63*, 205-212.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*, 407-420.

Griggs, R. A., & Cox, J. R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *35*, 519-533.

Grimshaw, G. M., Bulman-Fleming, M., & Ngo, C. (2004). A signal-detection analysis of sex differences in the perception of emotional faces. *Brain and Cognition*, *54*, 248-250.

Hoffman, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 371-383.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6,* 65-70.

Hu, X. (1993). *General processing tree models. Applications in cognitive psychology.* Doctoral thesis, University of California, Irvine.

Hu, X. (1999). Multinomial processing tree models: An implementation. *Behavior Research Methods, Instruments and Computers*, *31*, 689-695.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrica*, *59*, 21-47.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, *63*, 395-400.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*, 759-763.

Kinchla, R. A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review*, *101*, 166-171.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "Who said what?" paradigm. *Journal of Personality and Social Psychology*, *75*, 1155-1178.

Lawson, A. E. (2002). The origin of conditional logic: Does a cheater detection module exist? *Journal of Genetic Psychology*, *163*, 425-444.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—women's faces make the difference. *Brain and Cognition*, *50*, 121-128.

Liberman, N., & Klar, Y. (1996). Hypothesis testing in Wason's selection task: Social exchange cheating detection or task understanding. *Cognition*, *58*, 127-156.

Lunzer, E. A., Harrison, C., & Davey, M. (1972). The four-card problem and the generality of formal reasoning. *Quarterly Journal of Experimental Psychology*, *24*, 326-339.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Manktelow, K. I., & Evans, J. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, *70*, 477-488.

Manktelow, K. I., & Over, D. E. (1987). Reasoning and rationality. *Mind and Language*, *2*, 199-219.

Manktelow, K. I., & Over, D. E. (1990). Deontic thought and the selection task. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking: reflections on the psychology of thought*. Chichester, U.K.: Wiley.

Marr, D., & Nishihara, H. K. (1978). Visual information processing: Artificial intelligence and the sensorium of sight. *Technology Review*, *October*, 28-49.

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1137- 1151.

Mazur, J. E. (1998). *Learning and behavior* (4th edition). New Jersey, U.S.: Prentice Hall.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, *17*, 119-128.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 116-137.

Neath, I., & Surprenant, A. M. (2003). *Human memory. An introduction to research, data and theory* (2. ed ed.). Australia [u.a.]: Thomson/Wadsworth.

Oda, R. (1997). Biased face recognition in the Prisoner's Dilemma game. *Evolution and Human Behavior, 18*, 309-315.

Orgass, B. (1989). Agnosie. In K. Poeck (Ed.) *Klinische Neuropsychologie*. Stuttgart: Thieme.

Pollard, P. (1981). The effect of thematic content on the "Wason selection task". *Current Psychological Research*, *1*, 21-29.

Pollard, P. (1982). Human reasoning: Some possible effects of availability. *Cognition*, *12*, 65-96.

Pollard, P. (1990). Natural selection for the selection task: Limits to social exchange theory. *Cognition*, *36*, 195-204.

Pollard, P., & Gubbins, M. (1982). Context and rule manipulations on the Wason selection task. *Current Psychological Research*, *2*, 139-149.

Pratto, F. (1994). Consciousness and automatic evaluation. In S. Kitayama & P. M. Niedenthal (Eds.), *The heart's eye: Emotional influences in perception and attention* (pp. 115-143). San Diego, CA, US: Academic Press Inc.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380-391.

Reich, S. S., & Ruth, P. (1982). Wason's selection task: Verification, falsification and matching. *British Journal of Psychology*, *73*, 395-405.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318-339.

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 680-693.

Rodin, M. J. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition*, *5*, 144-165.

Rothkegel, R. (1999). Apple Tree: A multinomial processing tree modeling program for Macintosh computers. *Behavior Research Methods, Instruments and Computers*, *31*, 696-700.

Rozin, P., & Schull, J. (1988). The adaptive-evolutionary point of view in experimental psychology. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Steven's handbook of experimental psychology* (2nd ed.) (pp. 503-546). New York: John Wiley & Sons.

Shapiro, L., & Epstein, W. (1998). Evolutionary theory meets cognitive psychology: A more selective perspective. *Mind and Language*, *13*, 171-194.

Shepherd, J. W. (1981). Social factors in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces*. London: Academic Press.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50.

Solso, R. (2001). *Cognitive psychology* (6th edition). Boston, U.S.: Allyn & Bacon.

Sperber, D., & Girotto, V. (2002). Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides, and Tooby. *Cognition*, *85*, 277-290.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002). Selective impairment of reasoninig about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences*, *99*, 11531-11536.

Strack, F. (1994). Response processes in social judgment. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition, Vol. 1: Basic processes* (2nd ed.) (pp. 287-322). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences*, *99*, 11537-11542.

Symons, D. (1987). If we're all Darwinians, what's the fuss about? In C. Crawford (Ed.) *Sociobiology and psychology* (pp. 121-146). Hillsdale, NJ: Erlbaum.

Tooby, J. (1987). The emergence of evolutionary psychology. In D. Pines (Ed.) *Emerging synthesis in science* (pp. 67-76). Addison-Wesley Publishing Company.

Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture: I. Theoretical considerations. *Ethology and Sociobiology*, *10*, 29-49.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19-136). New York: Oxford University Press, Inc.

Tooby, J., & Cosmides, L. (1998). Evolutionizing the cognitive sciences: A reply to Shapiro and Epstein. *Mind and Language*, *13*, 195-204.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35-57.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 359-380.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.

van Duyne, P. C. (1974). Realism and linguistic complexity in reasoning. *British Journal of Psychology*, *65*, 59-67.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.) *New horizons in psychology*. Harmondsworth, U.K.: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*.

Wason, P. C. (1969). Regression in reasoning? *British Journal of Psychology*, *60*, 471-480.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, *23*, 63-71.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology*, *78*, 1024-1037.

Yachanin, S. A., & Tweney, R. D. (1982). The effect of thematic content on cognitive strategies in the four-card selection task. *Bulletin of the Psychonomic Society*, *19*, 87-90.

Appendix

Appendix A: Job titles used in Experiments 1, 2, and 3

Appendix B: Behavioral descriptions used in Experiments 1, 2 and 3

Appendix C: Mean status ratings assessed in Experiment 1

Appendix D: Mean status ratings assessed in Experiment 2

Appendix E: Two-high-threshold multinomial model of source-monitoring for crossed source dimensions applied in Experiment 3

Appendix F: Mean status ratings assessed in Experiment 3

Appendix G: Job titles used in Experiment 4

Appendix H: Behavioral descriptions used in Experiment 4

Appendix I: Two-high-threshold multinomial model of source monitoring applied in Experiment 4

Appendix A

Table A-1

Low-status job titles used in Experiments 1°, 2†, and 3‡. In a pretest, subjects (N = 24°†‡ and N = 33‡) had rated the social status of 82 jobs, using a scale ranging from 1 (“status low”) to 5 (“status high”). Those titles scoring far below the median (Z = 3.00) were chosen for behavioral descriptions representing persons of low social status.

Job Titles, Low Social Status	***M***	***SD***
Anstreicher [painter]°†‡	2.13	0.61
Autolackierer [body painter]°†‡	1.96	0.75
Automechaniker [mechanic]°†‡	2.17	0.76
Bäcker [baker]‡	2.36	0.78
Bauarbeiter [construction worker]°	1.75	0.85
Bestatter [mortician]‡	2.39	0.79
Fischer [fisherman]°†‡	1.96	0.91
Fließbandarbeiter [assembly line worker]°	1.29	0.55
Gärtner [gardener]‡	1.94	0.75
Gebrauchtwagenhändler [second-hand car dealer]°†‡	1.96	0.69
Hausmeister [caretaker]°	2.00	0.93
Kassierer [cashier]°†‡	2.17	0.92
Kellner [waiter]°	1.88	0.61
Kneipenpächter [saloon-keeper]°†‡	2.13	0.61
Krankenpfleger [hospital nurse]‡	2.70	0.85
Kurierfahrer [courier]°†‡	2.08	0.65
Käsehändler [cheese monger]°†‡	2.17	0.92
Lastwagenfahrer [trucker]°†‡	1.67	0.64
Matrose [seaman]‡	1.88	0.82
Metzger [butcher]°†‡	1.88	0.61

Job Titles, Low Social Status *(continued)*	***M***	***SD***
Obstverkäufer [greengrocer]‡	1.76	0.83
Restaurantbesitzer [owner of a restaurant]‡	3.00	0.87
Schuhputzer [shoeblack]°	1.21	0.41
Straßenkehrer [scavenger]°†‡	1.38	0.58
Tankwart [filling station attendant]°	1.58	0.58
Taxifahrer [taxi driver]‡	1.58	0.61

Table A-2

High-status job titles used in Experiments 1[°], 2[†], and 3[‡]. In a pretest, subjects (N = 24[°†‡] and N = 33[‡]) had rated the social status of 82 jobs using a scale ranging from 1 (“status low”) to 5 (“status high”). Those titles scoring far above the median (Z = 3.00) were chosen for behavioral descriptions representing persons of high social status.

Job Titles, High Social Status	***M***	***SD***
Apotheker [pharmacist][‡]	3.79	0.86
Architekt [architect][°†‡]	4.42	0.58
Bankvorstand [board member of a financial institution][°]	4.17	0.96
Bundestagsabgeordneter [member of the Bundestag][°†‡]	3.96	1.16
Börsenhändler [stock market trader][°†‡]	3.75	0.99
Chirurg [surgeon][°]	4.75	0.68
Geschäftsführer [business manager][°†‡]	3.92	0.65
Hauptkommissar [superintendent][‡]	3.85	0.83
Hochschulprofessor [professor][°†‡]	4.54	1.02
Ingenieur [engineer][‡]	4.03	0.85
Journalist [journalist][°‡]	3.92	0.78
Kernphysiker [nuclear physicist][°]	4.25	1.15
Lebensmittelingenieur [food engineer][°†‡]	3.50	0.59
Lehrer [teacher][°]	3.46	0.83
Notar [notary][°†‡]	4.04	0.62
Oberstudienrat [senior assistant master][°†‡]	3.92	1.14
Priester [priest][°†‡]	3.67	1.17
Psychotherapeut [psychotherapist][‡]	3.73	0.91
Rechtsanwalt [lawyer][‡]	4.48	0.87
Steuerberater [tax advisor][°]	3.46	0.78
Tierarzt [veterinarian][°†‡]	4.21	0.59

Job Titles, High Social Status *(continued)*	***M***	***SD***
TV-Moderator [link man]‡	3.61	1.06
Vermögensverwalter [fund manager]‡	3.50	0.84
Wirtschaftsprüfer [accountant]°†‡	3.75	0.94
Zahnarzt [dentist]°†‡	4.46	0.59

Appendix B

Table B-1

Behavioral descriptions implying information on behavior of cheating and low social status used in Experiments 1°, 2†, and 3‡. In a pretest, subjects (*N* = 21°†‡ and *N* = 33‡) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Mean valence ratings and corresponding standard deviations are listed for each descriptive statement.

Behavior of Cheating, Low Social Status	***M***	***SD***
O. G. ist Anstreicher. Die angebliche Bio-Farbe, die er seinen Kunden für teures Geld verkauft, ist höchst gesundheitsschädlich. [O. G. is a painter. The supposedly organic wall paint that he sells to his customers to a dear price is highly noxious.]°†‡	−2.61	.74
F. M. ist Automechaniker. Bei Reparaturen ersetzt er meist mehr als notwendig, um den Kunden teure Ersatzteile in Rechnung stellen zu können. [F. M. is a mechanic. When repairing cars he would mostly fix more than necessary and bill the customers expensive replacement parts.]°†‡	−2.33	.73
L. E. ist Bäcker. Brot vom Vortag wärmt er immer in der Mikrowelle auf und verkauft es dann den Kunden als frisch. [L. E. is a baker. He would always warm up old bred left over from the previous day in the oven and sell it to his customers as fresh.]‡	−2.06	1.14
K. S. ist Gebrauchtwagenhändler. Regelmäßig verkauft er restaurierte Unfallwagen als angeblich unfallfrei und verschweigt den Käufern gravierende Mängel. [K. S. is a second-hand car dealer. Regularly, he sells restored crash cars as supposedly accident-free and conceals serious defects to the customers.]°†‡	−2.14	1.42
O. G. ist Käsehändler. Immer wieder hat er schimmeligen Käse verkauft, obwohl er wusste, dass er damit die Gesundheit seiner Kunden gefährdet. [O. G. is a cheese monger. Again and again, he would sell moldy cheese, still knowing that this might harm his customers’ health.]°†‡	−2.71	.56
R. O. ist Kassierer. Immer wieder gibt er Kunden zu wenig Wechselgeld und steckt den Rest selbst ein. [R. O. is a cashier. Again and again, he would short-change and keep the rest of the money for himself.]°†‡	−2.38	.92

Behavior of Cheating, Low Social Status *(continued)*	***M***	***SD***
B. D. ist Kneipenpächter. In seiner Küche werden immer wieder verdorbene Lebensmittel verwendet, deren schlechten Geschmack er mit viel starkem Gewürz überdeckt. [B. D. is a saloonkeeper. Again and again, decomposed food is used in his tavern's kitchen, and the foul taste is covered with lots of strong spices.]°†‡	−2.29	1.55
O. W. ist Obstverkäufer. Unaufmerksamen Kunden packt er oft absichtlich überreifes und manchmal fauliges Gemüse ein, das er sonst nicht loswerden würde. [O. W. is a greengrocer. To inattentive customers, he would often consciously sell overripe or even moldy vegetables, which he would otherwise not dispose.]‡	−2.64	0.49
B. D. ist Restaurantbesitzer. Die teuren Weine auf seiner Karte sind in Wirklichkeit chemisch "aufgebesserte" Billigimporte aus Osteuropa. [B. D. is owns a restaurant. The expensive vines listed in the menu are in fact adulterated imports from Eastern Europe.]‡	−2.18	0.88
M. U. ist Taxifahrer. Wenn er bemerkt, dass Fahrgäste ortsfremd sind, fährt er in der Regel große Umwege, um mehr zu verdienen. [M. U. is a taxi-driver. Whenever he gets a hint that clients are strangers, he would go a long way round to boost the fare.]‡	−2.48	0.67

Table B-2

Behavioral descriptions implying information on behavior of trustworthiness and low social status used in Experiments 1[°], 2[†], and 3[‡]. In a pretest, subjects (*N* = 21[°†‡] and *N* = 33[‡]) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Mean valence ratings and corresponding standard deviations are listed for each descriptive statement.

Behavior of Trustworthiness, Low Social Status	***M***	***SD***
A. S. ist Autolackierer. Mehrere Anfragen, gestohlene Fahrzeuge umzulackieren, hat er abgelehnt, obwohl er dringend Geld gebrauchen könnte. [A. S. is a body painter. He declined several inquiries to revarnish stolen cars, although he urgently needed money.][°†‡]	2.19	1.08
U. I. ist Bestatter. Er legt großen Wert darauf, seine Kunden einfühlsam und mit großem Respekt zu beraten. [U. I. is a mortician. He attaches particular importance to consulting his clients in an empathetic and respectful manner.][‡]	2.61	0.70
N. K. ist Fischer. Er hält sich immer strikt an die Fangzonen und verabscheut Kollegen, die sich aus Geldgier darüber hinweg setzen. [N. K. is a fisherman. He always strictly abides by the capture zones and does abominate colleagues, who override the rules due to their greed for money.][°†‡]	1.38	1.28
A. E. ist Gärtner. Auch ohne klar umrissene Aufträge führt er keine unnötigen Arbeiten aus, nur um viele Stunden berechnen zu können. [A. E. is a gardener. He does not do any work that is not necessary just for the purpose of charging more to his employer, even he does not have a clearly defined order.][‡]	0.30	2.23
M. H. ist Krankenpfleger. Da er selbst kein Trinkgeld annehmen darf, steckt er die Zuwendungen von zufriedenen Patienten stets in die Stationskasse. [M. H. is a hospital nurse. Since he is not allowed to accept any tip, he spends any monetary gifts from his patients to the ward’s savings box][‡]	1.58	1.44
A. S. ist Kurierfahrer. Bei Unfällen, die ihm beim Einparken manchmal passieren, informiert er stets die betroffenen Fahrzeughalter. [A. S. is a courier. Whenever he has an accident, which sometimes happens to him while parking, he always would inform the car owners, who are concerned.][°†‡]	1.62	1.16

Behavior of Trustworthiness, Low Social Status ***(continued)***	**M**	**SD**
N. O. ist Lastwagenfahrer. Im Gegensatz zu vielen seiner Kollegen hält er sich immer an die vorgeschriebenen Ruhezeiten. [N. O. is a trucker. Contrary to many of his colleagues he would always abide by the prescribed rest periods.][°†‡]	1.67	1.20
A. L. ist Matrose. Die Angebote mehrerer Händler, illegale Waren für sie zu schmuggeln, hat er stets abgelehnt. [A. L. is a seaman. He would declined various offers to smuggle illegal goods.][‡]	1.82	1.21
A. I. ist Metzger. Er achtet sehr darauf, dass in seiner Metzgerei nur hochwertiges und nach strengsten Kriterien geprüftes Fleisch verarbeitet wird. [A. I. is a butcher. He is very considered that only strictly certified high-quality meat is used in his butchery.][°†‡]	1.62	1.36
T. P. ist Straßenkehrer. Wenn er bei der Arbeit verlorene Gegenstände findet, gibt er sie immer gewissenhaft im nächsten zuständigen Fundbüro ab. [T. P. is a scavenger. Whenever he finds lost property, he would always faithfully deliver them to the next lost and found.] [°†‡]	1.52	1.32

Table B-3

Behavioral descriptions implying information on irrelevant behavior and low social status used in Experiment 1. In a pretest, subjects (*N* = 21) had rated each sentence with respect to its valence, using a scale ranging from −3 ("negative") to +3 ("positive"). Mean valence ratings and corresponding standard deviations are listed for each descriptive statement.

Irrelevant Behavior, Low Social Status	***M***	***SD***
G. F. ist Bauarbeiter. Zur Zeit arbeitet er auf einer Baustelle, wo mehrere Wohnhäuser und Geschäfte entstehen sollen. [G. F. is a construction worker. He is presently working at a building site, where several apartments and business houses are about to be built.]	.05	.80
R. K. ist Fließbandarbeiter. Seit Monaten arbeitet er in einer Lebensmittelfabrik, welche ihre Produkte sowohl im Inland als auch im Ausland verkauft. [R. K. is an assembly line worker. For months he as been working at a food factory, which sells its products both at home and abroad.]	.19	.51
F. L. ist Hausmeister. Er wohnt an einer Schule und sorgt dafür, dass alles in Ordnung gehalten wird. [F. L. Is a caretaker. He is living at a school and taking care that everything is put in order.]	.67	.66
H. U. ist Kellner. Bereits seit drei Jahren arbeitet er immer am Abend in einem kleinen Restaurant, wo er viele Gäste bedient. [H. U. is a waiter. He has been working in a little restaurant for three years already, where he serves many diners.]	-.10	1.09
E. V. ist Schuhputzer. Er arbeitet in einem Hotel in Berlin, wo er schon sehr vielen unterschiedlichen Personen ihre Schuhe geputzt hat. [E. V. is a shoeblack. He is working at a hotel in Berlin, where he has already cleaned many different persons' shoes.]	.57	.98
S. Z. ist Tankwart. Er muss Autos betanken, waschen, Öl wechseln und die Luft in den Reifen kontrollieren. [S. Z. is a filling station attendant. He has to fuel and to clean vehicles, to do oil changes and air pressure control.]	.23	.70

Table B-4

Behavioral descriptions implying information on behavior of cheating and high social status used in Experiments 1°, 2†, and 3‡. In a pretest, subjects (N = 21°†‡ and N = 33‡) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Mean valence ratings and corresponding standard deviations are listed for each descriptive statement.

Behavior of Cheating, High Social Status	***M***	***SD***
F. J. ist Apotheker. „Besondere Kunden“ bekommen bei ihm auch Medikamente ohne Rezept – wenn die Bezahlung stimmt. [F. J. is a pharmacist. In case it is profitable, he hands out medicine to “special clients”, even if they do not have a prescription.]‡	−1,94	1,12
H. U. ist Geschäftsführer. Er arbeitet im Exportgeschäft und vertreibt Erbgut schädigende Pflanzenschutzmittel illegal in die Dritte Welt. [H. U. is a business manager. He carries on an export trade and illegally moves genotype-affecting pesticides into Third World countries.]°†‡	−2.62	.92
P. M. ist Hochschulprofessor. Seinen Professorentitel hat er sich erkauft und seine Veröffentlichungen haben andere für ihn geschrieben. [P. M. is a professor. He obtained his title by fraud, and others wrote his publications.]°†‡	−2.14	1.28
U. T. ist Ingenieur. Weil er keine Lust hat zu arbeiten, lässt er sich ständig mit erfundenen Symptomen krank schreiben und beurlauben. [U. T. is an engineer. Whenever he does not feel like working, he would make up some symptoms and skive off work.]‡	−1.85	0.97
F. D. ist Notar. Für einige fragwürdige Kunden hat er immer wieder gegen entsprechende Bezahlung gefälschte Dokumente beglaubigt. [F.D. is a notary. Again and again, for extra money he would attest falsified documents for several dubious customers.]°†‡	−2.52	.68
R. T. ist Priester. Mehrmals hat er Teile des Geldes, das Gemeindemitglieder für wohltätige Zwecke gespendet haben, in die eigene Tasche gesteckt. [R. T. is a priest. Several times he embezzled money from charitable donations of his parishioners.]°†‡	−2.24	4.41

Behavior of Cheating, High Social Status *(continued)*	**M**	**SD**
A. K. ist Rechtsanwalt. Zwei Abende pro Woche verbringt er beim Schäferstündchen mit seiner Sekretärin und entschuldigt sich bei seiner Frau mit wichtigen Geschäftsbesprechungen. [A. K. is a lawyer. Twice a week he meets his secretary for an amorous tête-à-tête, telling his wife that he has attend to some important meeting.]‡	−2.30	0.95
L. L. ist TV-Moderator. Er gibt regelmäßig geheime Informationen seines Senders an Journalisten weiter, damit diese im Gegenzug positiv über ihn berichten. [L. L. is a link man. He regularly betrays secret information on his channel to journalists in order that they publish positive reports on him.]‡	−1.73	0.91
D. I. ist Wirtschaftsprüfer. Er verdient sich regelmäßig Geld dazu, indem er sich bei Firmenprüfungen dafür bezahlen lässt, aufgedeckte Ungereimtheiten zu vertuschen. [D. I. is an accountant. He regularly earns some extra money by covering up inconsistencies he discloses in the context of tax examinations.]°†‡	−2.38	1.16
O. M. ist Zahnarzt. Ältere wohlhabende Patienten berät er oft gezielt falsch, um sie zu teureren Eingriffen zu überreden, als nötig wären. [O. M. is a dentist. He often gives advice to predominantly elder people and argues them into expensive interventions, which might not be medically necessary.]°†‡	−1.86	1.77

Table B-5

Behavioral descriptions implying information on behavior of trustworthiness and high social status used in Experiments 1[°], 2[†], and 3[‡]. In a pretest, subjects ($\mathcal{N}$ = 21[°†‡], and $\mathcal{N}$ = 33[‡]) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Mean valence ratings and corresponding standard deviations are listed for each descriptive statement.

Behavior of Trustworthiness, High Social Status	***M***	***SD***
E. K. ist Architekt. Von ihm zu verantwortende Schäden an Gebäuden hat er auf seine Kosten reparieren lassen. [E. K. is an architect. Any structural damages he might have caused, he would get repaired at his own expense.][°†‡]	1.62	1.69
F. G. ist Börsenhändler. Wenn er berechtigte Zweifel hat, rät er seinen Kunden von Aktiengeschäften ab, auch wenn er daran verdienen könnte. [F. G. is a stock market trader. He would discourage his clients from risky share transactions even if he might make money out of their investments.][°†‡]	1.61	.74
A. J. ist Bundestagsabgeordneter. Er vertritt gewissenhaft die Interessen seiner Wähler und lässt sich nicht in «Kungeleien» verwickeln. [A. J. is a member of the Bundestag. He would consciously represent his voters’ interests and would not get implicated in shady dealings.][°†‡]	1.67	1.11
W. O. ist Hauptkommissar. Er geht stets sensible mit Zeugenaussagen um und bringt niemals Informanten in Gefahr, indem er ihre Identität preisgibt. [W. O. is a superintendent. He always deals carefully with eyewitness testimonies and never puts a risk at an attestor in abandoning his identity.][‡]	2.21	0.99
I. U. ist Journalist. Er hält sich immer an die Absprachen mit Informanten und publiziert niemals vertrauliche Informationen. [I. U. is a journalist. He never publishes secret information and always adheres to agreements he has made with his informants.][‡]	2.18	1.07
H. B. ist Lebensmittelingenieur. Bei der Qualitätsbeurteilung von Lebensmitteln hält er sich stets an die Richtlinien und lässt sich durch nichts bestechen. [H. B. is a food engineer. Concerning quality evaluations, he always abides by the directives and is bribable in no way.][°†‡]	2.04	.97

Behavior of Trustworthiness, High Social Status ***(continued)***	***M***	***SD***
I. W. ist Oberstudienrat. Wenn er Arbeiten korrigiert, verdeckt er immer die Namen der Prüflinge, damit er sie völlig unvoreingenommen bewerten kann. [I. W. is a senior assistant master. Whenever marking written tests, he would cover the students' names to get to an unbiased judgment.]°†‡	2.04	.92
L. M. ist Psychotherapeut. Für ihn ist es selbstverständlich, dass er Informationen über Klienten absolut vertraulich behandelt und nicht einmal Kollegen mitteilt. [L. M. is a psychotherapist. As a matter of course, he never betrayes any confidential information, not even to his colleagues.]‡	2.15	1.00
O. E. ist Tierarzt. Er lehnt grundsätzlich alle Behandlungen ab, die für die Tiere quälend und aussichtslos sind. [O. E. is a veterinarian. He generally dismisses any kind of unpromising treatment, which is tantalizing to the animals.]°†‡	2.05	1.12
C. A. ist Vermögensverwalter. Obwohl er Gelegenheiten gehabt hätte, hat er sich nie auf Kosten seiner Klienten bereichert. [C. A. is a fund manager. Although he had some possibilities to enrich at the expense of his clients, he never did so.]‡	2.30	1.19

Table B-6

Behavioral descriptions implying information on irrelevant behavior and high social status used in Experiment 1. In a pretest, subjects (*N* = 21) had rated each sentence with respect to its valence, using a scale ranging from −3 ("negative") to +3 ("positive"). Mean valence ratings and corresponding standard deviations are listed for each descriptive statement.

Irrelevant Behavior, High Social Status	***M***	***SD***
A. V. ist Bankvorstand. In der Mittagspause trifft er sich gelegentlich mit Geschäftspartnern in einem kleinen italienischen Restaurant. [A. V. is a board member of a financial institute. Occasionally, he meets with some associates for lunch in a little Italian restaurant.]	−.05	1.07
B. G. ist Chirurg. Er arbeitet in einem Krankenhaus am Stadtrand, wo er täglich die unterschiedlichsten Operationen durchführt. [B. G. is a surgeon. He works at a hospital on the outskirts and does several diverse surgeries per day.]	.10	.89
W. E. ist Journalist. Für eine Tageszeitung schreibt er über politisch aktuelle Themen und betreut seine eigene Kolumne mit Fragen der Leser. [W. E. is a journalist. He reports on newsworthy political topics and answers questions to interested readers within his own column.]	.33	1.06
S. D. ist Kernphysiker. Er ist in der Forschung tätig und ist an einer Reihe von Projekten zum Beweis der Quantentheorie beteiligt. [S. D. a nuclear physicist. He does research and is involved in several projects on the quantum theory.]	.47	.74
J. M. ist Lehrer. Er unterrichtet an einem Gymnasium in der ersten Sekundarstufe die Fächer Mathematik und Biologie. [J. M. is a teacher. He teaches maths and biology at a high school.]	.10	.83
K. L. ist Steuerberater. Seit Ende des Studiums ist er in der Wirtschaft tätig, wo er verschiedene Unternehmen hinsichtlich ihrer Finanzlage berät. [K. L. is a tax advisor. He has been working in commerce since he achieved his final degree, and has been doing financial counseling for several companies.]	.05	.69

Appendix C

Table C-1

Mean status ratings for low-status job titles assessed in Experiment 1. Subjects (*N* = 96) rated all job titles, which were supposed to be associated with low social status due to the pretest, significantly below the median (*Z* = 3.00).

Job Titles, Low Social Status	***M***	***SD***	***t*(95)**	***p***	**η^2**
Anstreicher [painter]	1.90	.67	−16.09	.000	.73
Autolackierer [body painter]	2.03	.69	−13.81	.000	.67
Automechaniker [mechanic]	2.36	.76	−8.24	.000	.42
Bauarbeiter [construction worker]	1.80	.71	−16.65	.000	.74
Fischer [fisherman]	2.09	.88	−10.05	.000	.52
Fließbandarbeiter [assembly line worker]	1.48	.79	−18.76	.000	.79
Gebrauchtwagenhändler [2^{nd}-hand car dealer]	2.10	.79	−11.14	.000	.57
Hausmeister [caretaker]	1.90	.72	−15.07	.000	.71
Kassierer [cashier]	1.85	.68	−16.50	.000	.74
Kellner [waiter]	2.06	.74	−12.46	.000	.62
Kneipenpächter [saloon-keeper]	2.26	.77	−9.40	.000	.48
Kurierfahrer [courier]	2.04	.70	−13.52	.000	.66
Käsehändler [cheese monger]	2.11	.82	−10.59	.000	.54
Lastwagenfahrer [trucker]	1.81	.80	−14.57	.000	.69
Metzger [butcher]	2.21	.70	−11.17	.000	.57
Schuhputzer [shoeblack]	1.29	.66	−25.22	.000	.87
Straßenkehrer [scavenger]	1.33	.63	−26.05	.000	.88
Tankwart [filling station attendant]	1.80	.80	−14.62	.000	.69

Table C-2

Mean status ratings for high-status job titles assessed in Experiment 1. Subjects (N = 96) rated all job titles, which were supposed to be associated with high social status due to the pretest, significantly above the median (Z = 3.00).

Job Titles, High Social Status	***M***	***SD***	***t*(95)**	***p***	**η^2**
Architekt [architect]	4.30	.82	15.51	.000	.72
Bankvorstand [board member of a financial institution]	4.09	1.01	10.65	.000	.54
Bundestagsabgeordneter [member of the Bundestag]	3.75	1.21	6.10	.000	.28
Börsenhändler [stock market trader]	3.82	.83	9.67	.000	.50
Chirurg [surgeon]	4.76	.61	28.22	.000	.89
Geschäftsführer [business manager]	4.14	.90	12.34	.000	.62
Hochschulprofessor [professor]	4.65	.73	22.24	.000	.84
Journalist [journalist]	3.79	.88	8.80	.000	.45
Kernphysiker [nuclear physicist]	4.40	.83	16.54	.000	.74
Lebensmittelingenieur [food engineer]	3.58	.82	7.00	.000	.34
Lehrer [teacher]	3.40	.86	4.49	.000	.17
Notar [notary]	4.08	.98	10.83	.000	.55
Oberstudienrat [senior assistant master]	3.96	.92	10.24	.000	.52
Priester [priest]	3.57	1.00	5.60	.000	.25
Steuerberater [tax advisor]	3.59	.90	6.46	.000	.30
Tierarzt [veterinarian]	4.06	.75	13.86	.000	.67
Wirtschaftsprüfer [accountant]	3.76	.97	7.68	.000	.38
Zahnarzt [dentist]	4.40	.80	17.07	.000	.75

Appendix D

Table D-1

Mean status ratings for low-status job titles assessed in Experiment 2. Subjects (N = 123) rated all job titles, which were supposed to be associated with low social status due to the pretest, significantly below the median (Z = 3.03).

Job Titles, Low Social Status	***M***	***SD***	***t*(122)**	***p***	**η^2**
Anstreicher [painter]	2.13	.76	−13.18	.000	.59
Autolackierer [body painter]	2.10	.75	−13.77	.000	.61
Automechaniker [mechanic]	2.44	.70	−9.32	.000	.42
Fischer [fisherman]	2.15	.90	−10.86	.000	.49
Gebrauchtwagenhändler [2nd-hand car dealer]	2.10	.76	−13.57	.000	.60
Kassierer [cashier]	1.93	.82	−14.86	.000	.64
Kneipenpächter [saloon-keeper]	2.24	.83	−10.61	.000	.48
Kurierfahrer [courier]	2.15	.83	−11.86	.000	.54
Käsehändler [cheese monger]	2.36	.75	−9.97	.000	.45
Lastwagenfahrer [trucker]	1.89	.75	−16.92	.000	.70
Metzger [butcher]	2.26	.75	−11.46	.000	.52
Straßenkehrer [scavenger]	1.46	.87	−20.06	.000	.77

Table D-2

Mean status ratings for high-status job titles assessed in Experiment 2. Subjects (*N* = 123) rated all job titles, which were supposed to be associated with high social status due to the pretest, significantly above the median (*Z* = 3.03).

Job Titles, High Social Status	***M***	***SD***	***t*(122)**	***p***	**η^2**
Architekt [architect]	4.45	.68	23.12	.000	.81
Bundestagsabgeordneter [member of the Bundestag]	3.93	1.15	8.68	.000	.38
Börsenhändler [stock market trader]	3.76	1.06	7.65	.000	.32
Geschäftsführer [business manager]	4.05	.80	14.16	.000	.62
Hochschulprofessor [professor]	4.70	.64	28.96	.000	.87
Lebensmittelingenieur [food engineer]	3.65	.79	8.72	.000	.38
Notar [notary]	4.21	.87	15.05	.000	.65
Oberstudienrat [senior assistant master]	4.08	.87	13.34	.000	.59
Priester [priest]	3.62	1.00	6.55	.000	.26
Tierarzt [veterinarian]	4.29	.81	17.35	.000	.71
Wirtschaftsprüfer [accountant]	3.64	.99	6.84	.000	.28
Zahnarzt [dentist]	4.51	.66	25.00	.000	.84

Appendix E

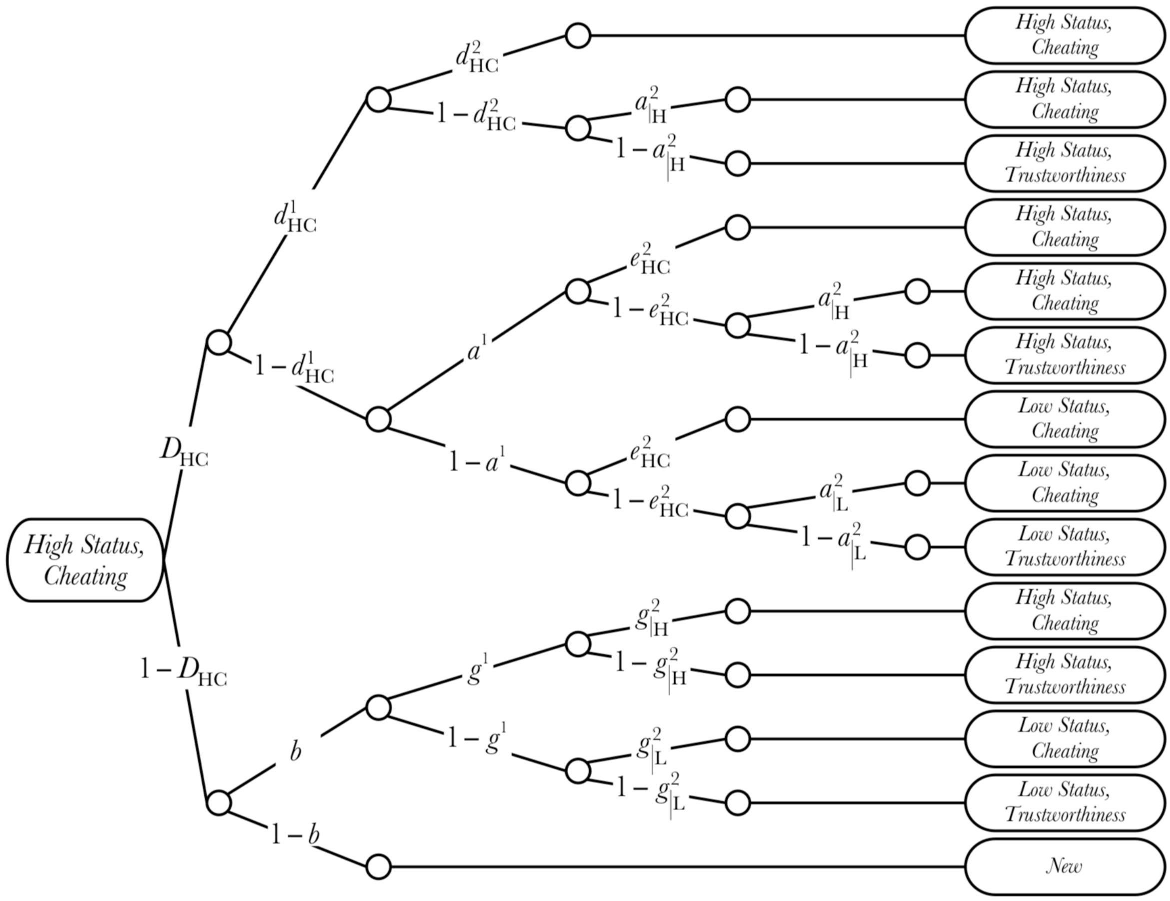

Figure E-1: Processing tree representation of the multinomial memory model for crossed source information adapted from Meiser and Bröder (2002) for target items from source combination *High Social Status*, *Information on Cheating* (High, Cheat). D_{HC} = probability of recognizing target items from source combination (High, Cheat) as old; d^1_{HC} = probability of recollecting Source (High) of the first source dimension *Social Status* for recognized target items from source combination (High, Cheat); d^2_{HC} = probability of recollecting Source (Cheat) of the second source dimension *Threat Potential* for recognized target items from source combination (High, Cheat) given recollection of Source (High) of the first dimension; e^2_{HC} = probability of recollecting Source (Cheat) of the second source dimension *Threat Potential* for recognized target items from source combination (High, Cheat) given no recollection of Source (High) of the first dimension; a^1 = proportion of guessing Source (High) for recognized target items; $a^2_{|H}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (High); $a^2_{|L}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (Low); b = probability of guessing that an item is old; g^1 = proportion of guessing Source (High) for unrecognized target items; $g^2_{|H}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (High); $g^2_{|L}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (Low).

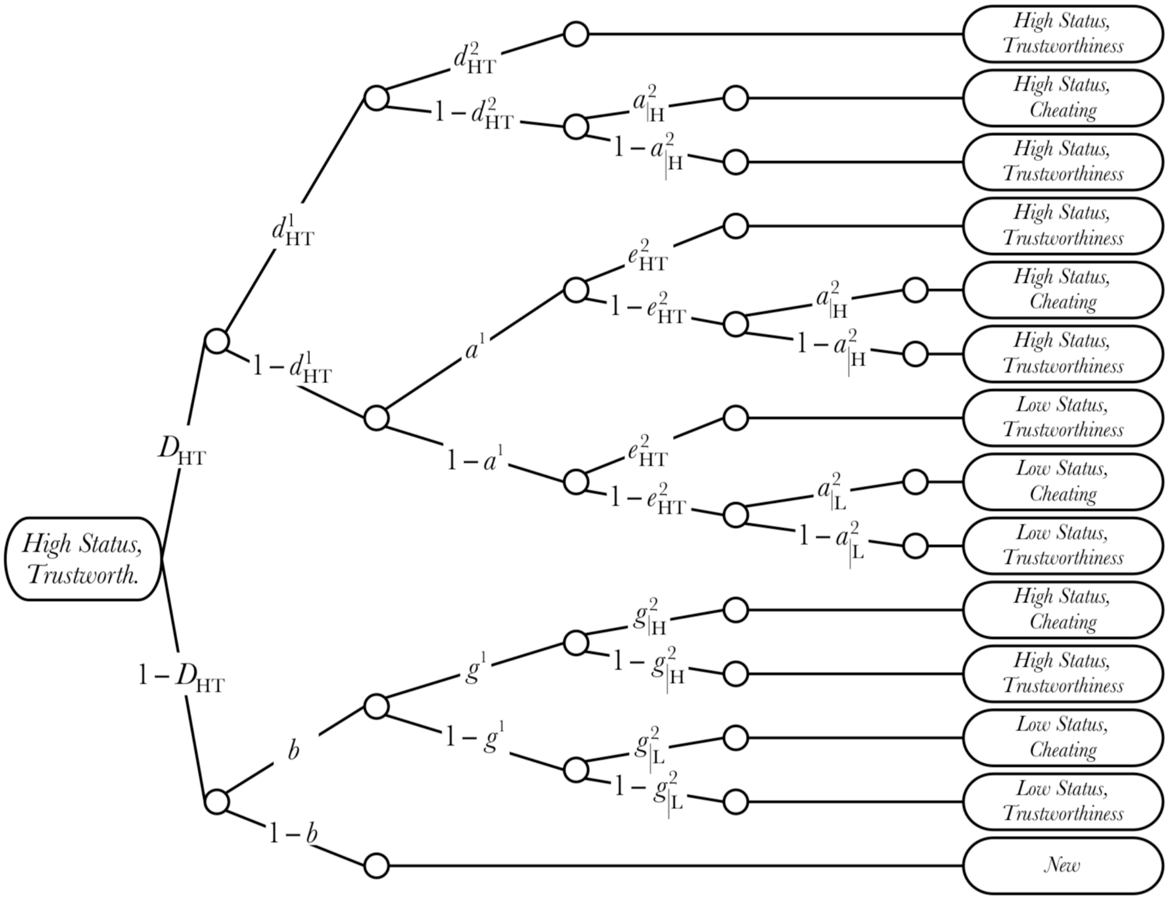

Figure E-2: Processing tree representation of the multinomial memory model for crossed source information adapted from Meiser and Bröder (2002) for target items from source combination *High Social Status*, *Information on Trustworthiness* (High, Trust). D_{HT} = probability of recognizing target items from source combination (High, Trust) as old; d^1_{HT} = probability of recollecting Source (High) of the first source dimension *Social Status* for recognized target items from source combination (High, Trust); d^2_{HT} = probability of recollecting Source (Trust) of the second source dimension *Threat Potential* for recognized target items from source combination (High, Trust) given recollection of Source (High) of the first dimension; e^2_{HT} = probability of recollecting Source (Trust) of the second source dimension *Threat Potential* for recognized target items from source combination (High, Trust) given no recollection of Source (High) of the first dimension; a^1 = proportion of guessing Source (High) for recognized target items; $a^2_{|H}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (High); $a^2_{|L}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (Low); b = probability of guessing that an item is old; g^1 = proportion of guessing Source (High) for unrecognized target items; $g^2_{|H}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (High); $g^2_{|L}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (Low).

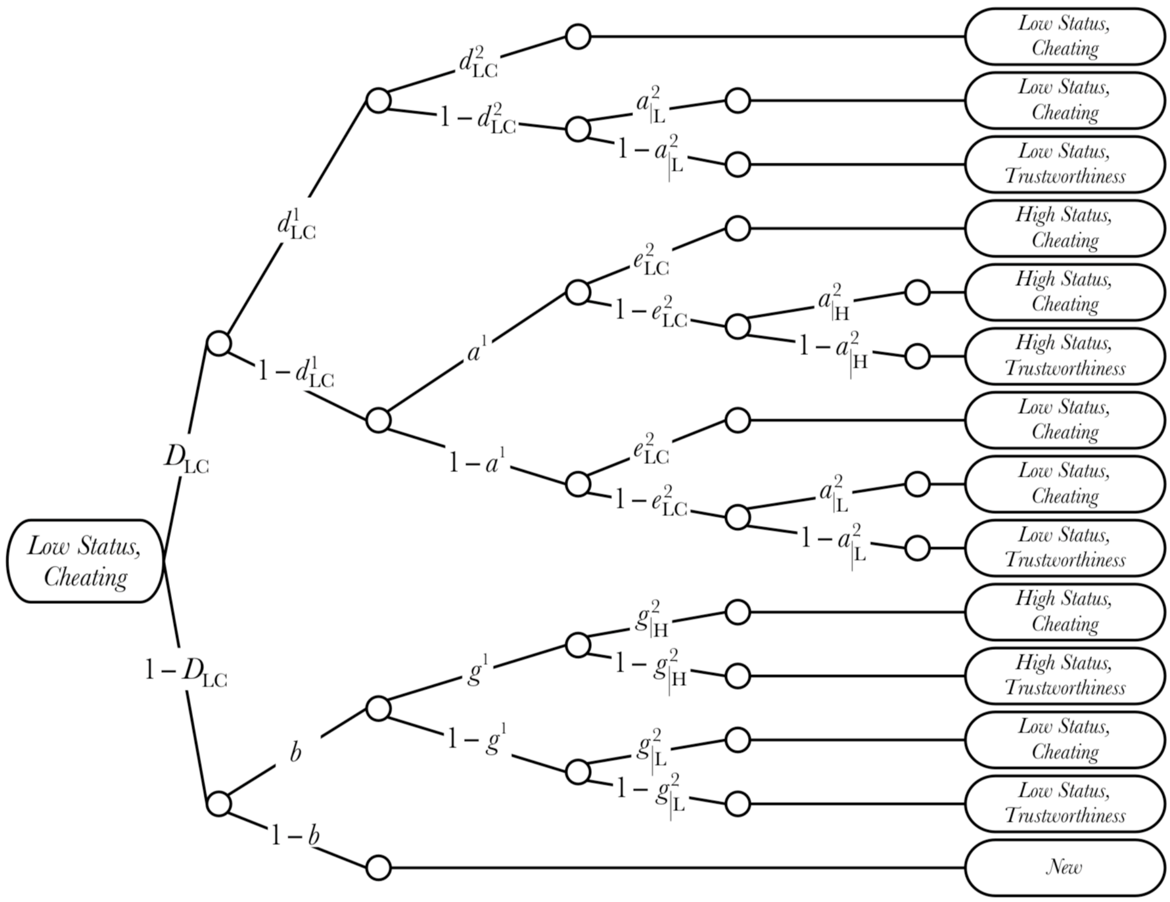

Figure E-3: Processing tree representation of the multinomial memory model for crossed source information adapted from Meiser and Bröder (2002) for target items from source combination *Low Social Status*, *Information on Cheating* (Low, Cheat). D_{LC} = probability of recognizing target items from source combination (Low, Cheat) as old; d^1_{LC} = probability of recollecting Source (Low) of the first source dimension *Social Status* for recognized target items from source combination (Low, Cheat); d^2_{LC} = probability of recollecting Source (Cheat) of the second source dimension *Threat Potential* for recognized target items from source combination (Low, Cheat) given recollection of Source (Low) of the first dimension; e^2_{LC} = probability of recollecting Source (Cheat) of the second source dimension *Threat Potential* for recognized target items from source combination (Low, Cheat) given no recollection of Source (Low) of the first dimension; a^1 = proportion of guessing Source (High) for recognized target items; $a^2_{|H}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (High); $a^2_{|L}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (Low); b = probability of guessing that an item is old; g^1 = proportion of guessing Source (High) for unrecognized target items; $g^2_{|H}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (High); $g^2_{|L}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (Low).

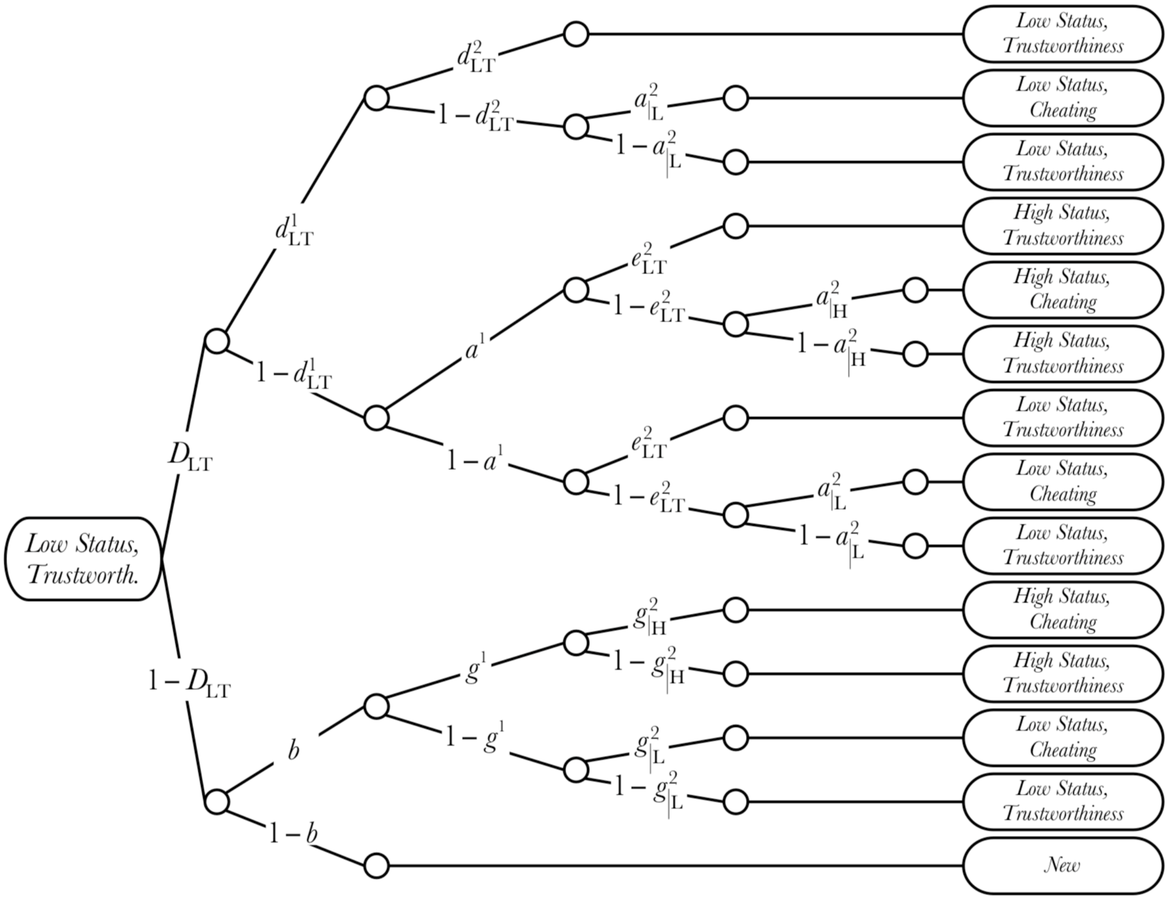

Figure E-4: Processing tree representation of the multinomial memory model for crossed source information adapted from Meiser and Bröder (2002) for target items from source combination *Low Social Status*, *Information on Trustworthiness* (Low, Trust). D_{LT} = probability of recognizing target items from source combination (Low, Trust) as old; d^1_{LT} = probability of recollecting Source (Low) of the first source dimension *Social Status* for recognized target items from source combination (Low, Trust); d^2_{LT} = probability of recollecting Source (Trust) of the second source dimension *Threat Potential* for recognized target items from source combination (Low, Trust) given recollection of Source (Low) of the first dimension; e^2_{LT} = probability of recollecting Source (Trust) of the second source dimension *Threat Potential* for recognized target items from source combination (Low, Trust) given no recollection of Source (Low) of the first dimension; a^1 = proportion of guessing Source (High) for recognized target items; $a^2_{|H}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (High); $a^2_{|L}$ = proportion of guessing Source (Cheat) for recognized target items given assignment to Source (Low); b = probability of guessing that an item is old; g^1 = proportion of guessing Source (High) for unrecognized target items; $g^2_{|H}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (High); $g^2_{|L}$ = proportion of guessing Source (Cheat) for unrecognized target items given assignment to Source (Low).

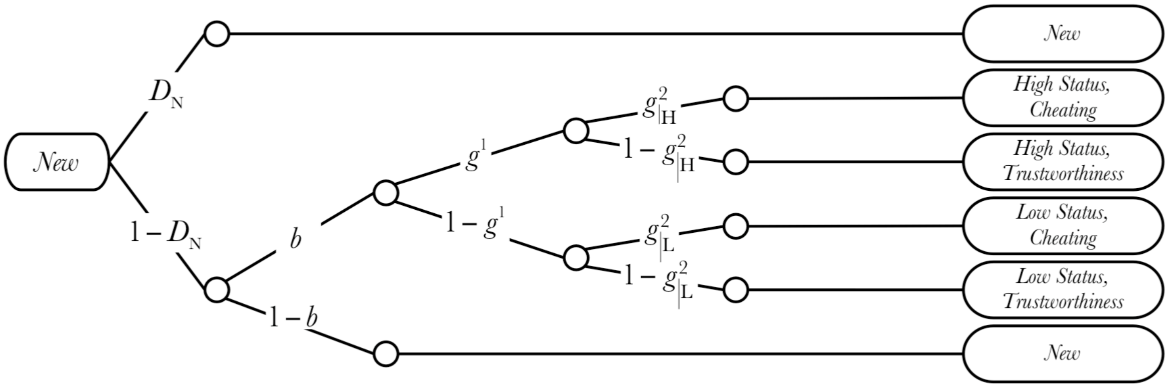

Figure E-5: Processing tree representation of the multinomial memory model for crossed source information adapted from Meiser and Bröder (2002) for new distractor items. D_N = probability of identifying distractor items as new; b = probability of guessing that an item is old; g^1 = proportion of guessing Source (High) for unidentified distractors; $g^2_{|H}$ = proportion of guessing Source (Cheat) for unidentified distractors given assignment to Source (High); $g^2_{|L}$ = proportion of guessing Source (Cheat) for unidentified distractors given assignment to Source (Low).

Appendix F

Table F-1

Mean status ratings for low-status job titles assessed in Experiment 3. Subjects (*N* = 64) rated almost all job titles, which were supposed to be associated with low social status due to the pretest, significantly below the median (*Z* = 3.00).

Job Titles, Low Social Status	***M***	***SD***	***t*(63)**	***p***	**η^2**
Anstreicher [painter]	1.69	.69	−15.28	.000	.79
Autolackierer [body painter]	1.87	.66	−13.75	.000	.75
Automechaniker [mechanic]	2.14	.69	−10.01	.000	.61
Bäcker [baker]	2.38	.75	−6.71	.000	.42
Bestatter [mortician]	2.88	.86	−1.16	.251	.02
Fischer [fisherman]	1.89	.76	−11.70	.000	.68
Gärtner [gardener]	2.25	.80	−7.53	.000	.47
Gebrauchtwagenhändler [2nd-hand car dealer]	1.97	.69	−11.97	.000	.69
Kassierer [cashier]	1.61	.63	−17.58	.000	.83
Kneipenpächter [saloon-keeper]	2.25	.84	−7.18	.000	.45
Krankenpfleger [hospital nurse]	2.78	1.06	−1.65	.104	.04
Kurierfahrer [courier]	1.86	.71	−12.86	.000	.72
Käsehändler [cheese monger]	2.11	.67	−10.64	.000	.64
Lastwagenfahrer [trucker]	1.52	.59	−20.10	.000	.87
Matrose [seaman]	2.02	.90	−8.75	.000	.55
Metzger [butcher]	1.94	.64	−13.30	.000	.74
Obstverkäufer [greengrocer]	1.88	.55	−16.39	.000	.81
Restaurantbesitzer [owner of a restaurant]	3.45	.73	4.95	.000	.28
Straßenkehrer [scavenger]	1.25	.67	−21.00	.000	.88
Taxifahrer [taxi driver]	1.69	.53	−19.78	.000	.86

Table F-2

Mean status ratings for high-status job titles assessed in Experiment 3. Subjects (*N* = 64) rated all job titles, which were supposed to be associated with high social status due to the pretest, significantly above the median (*Z* = 3.00).

Job Titles, High Social Status	***M***	***SD***	***t*(63)**	***p***	**η^2**
Apotheker [pharmacist]	4.06	.85	9.98	.000	.61
Architekt [architect]	4.59	.56	22.95	.000	.89
Bundestagsabgeordneter [member of the Bundestag]	4.14	.99	9.22	.000	.57
Börsenhändler [stock market trader]	3.84	.98	6.89	.000	.43
Geschäftsführer [business manager]	4.09	.68	12.80	.000	.72
Hauptkommissar [superintendent]	3.97	.93	8.38	.000	.53
Hochschulprofessor [professor]	4.81	.50	29.00	.000	.93
Ingenieur [engineer]	4.22	.68	14.39	.000	.77
Journalist [journalist]	3.94	.77	9.69	.000	.60
Lebensmittelingenieur [food engineer]	3.72	.68	8.48	.000	.53
Notar [notary]	4.33	.89	11.91	.000	.69
Oberstudienrat [senior assistant master]	4.17	.81	11.61	.000	.68
Priester [priest]	3.94	.89	8.44	.000	.53
Psychotherapeut [psychotherapist]	3.91	.92	7.87	.000	.50
Rechtsanwalt [lawyer]	4.39	.77	14.47	.000	.77
Tierarzt [veterinarian]	4.27	.65	15.62	.000	.79
TV-Moderator [link man]	3.63	.97	5.17	.000	.30
Vermögensverwalter [fund manager]	3.47	.91	4.13	.000	.21
Wirtschaftsprüfer [accountant]	3.94	.91	8.28	.000	.52
Zahnarzt [dentist]	4.50	.64	18.68	.000	.85

Appendix G

Table G-1

Low-status job titles used in Experiment 4. In a pretest, subjects (*N* = 36) had rated the social status of 200 titles using a scale ranging from 1 (“status low”) to 5 (“status high”). Those titles scoring far below the median (*Z* = 3.00) were chosen for behavioral descriptions representing persons of low social status.

Job Titles	***M***	***SD***
Automechaniker [mechanic]	1.97	.74
Bademeister [pool attendant]	1.69	.71
Baggerfahrer [construction worker who runs an excavator]	1.47	.56
Bauarbeiter [construction worker]	1.58	.60
Berufssoldat [lifer]	2.25	.77
Betonierer [construction worker working with concrete]	1.56	.65
Bäcker [baker]	2.22	.64
Fließbandarbeiter [assembly line worker]	1.25	.44
Färber [dipper]	1.67	.59
Gebrauchtwagenhändler [2^{nd}-hand car dealer]	2.17	.65
Gerüstbauer [rigging grip]	1.86	.64
Glaser [glazier]	2.28	.61
Gärtner [gardener]	2.22	.68
Hafenarbeiter [dock worker]	1.42	.50
Hilfsarbeiter [laborer]	1.20	.47
Holzfäller [logger]	1.61	.64
Hundezüchter [dog breeder]	2.47	.97
Installateur [installer]	2.33	.72
Inventurhelfer [inventory assistant]	1.25	.50
Kellner [waiter]	2.06	.67

Job Titles *(continued)*	***M***	***SD***
Kranführer [crane operator]	1.81	.62
Käsehändler [cheese monger]	2.06	.53
Masseur [masseur]	2.44	.61
Maurer [bricklayer]	1.81	.62
Müller [miller]	2.08	.60
Obstverkäufer [fruit merchant]	1.81	.58
Pizzabäcker [pizza baker]	1.81	.58
Postbote [postman]	2.14	.72
Prospektverteiler [brochure distributor]	1.19	.40
Taxifahrer [taxi driver]	1.86	.64
Tischler [joiner]	2.39	.77
Tätowierer [tattoo artist]	1.75	.69
Verkaufshilfe [temporary shop assistant]	1.39	.55
Weber [weaver]	2.25	.73
Zimmermann [carpenter]	2.11	.78
Zollbeamter [customs officer]	2.42	.84

Appendix H

Table H-1

Behavioral descriptions implying information on ordinary behavior of cheating used in Experiment 4. In a pretest, subjects (N = 22) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Apart from this, they judged each statement with respect to the type of behavior, using a scale ranging from −3 (“very exceptional”) to +3 (“very ordinary”). Mean valence ratings as well as mean ratings of type of behavior and the corresponding standard deviations are listed for each descriptive statement.

	Valence		***Type of Behavior***	
Behavior of Cheating, Ordinary	***M***	***SD***	***M***	***SD***
H. M. ist Betonierer. Bei Schadensersatzklagen hat er mehrmals falsche Aussagen gemacht und die Arbeitskollegen beschuldigt, um seine eigenen Fehler zu vertuschen. [H. M. is a construction worker working with concrete. Sued for damages he perjured repeatedly and blamed his colleagues to cover up his own mistakes.]	−2.64	0.49	−0.18	1.30
K. S. ist Gebrauchtwagenhändler. Regelmäßig verkauft er notdürftig wieder reparierte Unfallwagen als angeblich unfallfrei und verschweigt den Kunden gravierende Mängel der Fahrzeuge. [K. S. is a 2nd-hand car dealer. Regularly, he sells restored crash cars as supposedly accident-free and conceals serious defects to the customers.]	−2.64	0.49	0.55	1.63
R. L. ist Hilfsarbeiter. Er verkauft oft teure Werkzeuge oder Materialien an Bekannte, die er bei den Arbeiten an Baustellen entwendet hat. [R. L. is a laborer. He often sells expensive tools, which he has stolen from building lots before, to his acquaintances.]	−2.18	0.73	−0.27	1.75
P. E. ist Inventurhelfer. Bei Inventuren steckt er immer wieder Waren ein und schleust sie geschickt aus dem Laden seines ahnungslosen Vorgesetzten. [P. E. is an inventory assistant. Again and again, taking stock he rips off and takes away some goods behind his unsuspecting supervisor’s back.]	−2.18	0.66	−0.41	1.71

Behavior of Cheating, Ordinary	***Valence*** *M*	***Valence*** *SD*	***Type of Behavior*** *M*	***Type of Behavior*** *SD*
T. P. ist Maurer. Er bietet seinen Kunden oft an, Arbeiten nach Feierabend ohne Rechnung zu erledigen, um das Geld selbst einzustecken. [T. P. is a bricklayer. He often offers his customers to do a moonlight to keep the money for himself.]	−1.27	0.77	1.05	1.62
M. A. ist Obstverkäufer. Um seinen Gewinn weiter zu steigern, macht er häufig bewusst falsche Angaben über die Herkunftsländer seiner angebotenen Produkte. [M. A. is a fruit merchant. To increase his earnings he often lies about the actual countries of origin of the products he sells.]	−2.05	0.72	0.09	1.60

Table H-2

Behavioral descriptions implying information on ordinary behavior of trustworthiness used in Experiment 4. In a pretest, subjects (N = 22) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Apart from this, they judged each statement with respect to the type of behavior, using a scale ranging from −3 (“very exceptional”) to +3 (“very ordinary”). Mean valence ratings as well as mean ratings of type of behavior and the corresponding standard deviations are listed for each descriptive statement.

	Valence		***Type of Behavior***	
Behavior of Trustworthiness, Ordinary	***M***	***SD***	***M***	***SD***
N. G. ist Automechaniker. Bei Reparaturen ist er immer bestrebt, seinen Kunden möglichst günstige Ersatzteile anzubieten und die Arbeiten zügig zu erledigen. [N. G. is a mechanic. He is always eager to provide spares as cheap as possible for his clients and to fulfill his jobs efficiently.]	2.18	0.59	–0.05	1.36
J. G. ist Holzfäller. Er hat schon mehrmals verletzte Tiere, die er im Wald gefunden hat, bei sich aufgenommen und fürsorglich gepflegt. [J. G. is a logger. Several times he found harmed animals and took them home to care for them.]	2.27	0.63	0.36	1.26
L. O. ist Hundezüchter. Nach Verkäufen von Hunden besucht er die Käufer immer wieder, um sich vom Wohlergehen der Hunde zu überzeugen. [L. O. is a dog breeder. Having sold a puppy he visits the new owner again and again to make sure the dog is well off.]	1.95	0.95	0.27	1.42
A. W. ist Installateur. Er berät seine Kunden immer gerne, wie sie kostengünstige und umweltfreundliche Heizungsanlagen in ihren Häusern installieren lassen können. [A. W. is an installer. He likes to give advice to his customers and tell them how they might have installed low priced and eco friendly heating systems.]	2.00	0.76	1.27	0.98

Behavior of Trustworthiness, Ordinary *(continued)*	***Valence*** *M*	***Valence*** *SD*	***Type of Behavior*** *M*	***Type of Behavior*** *SD*
O. D. ist Käsehändler. Er achtet sehr darauf, alten Käse immer sofort auszusortieren und lässt die Kunden alle seine Produkte vorher probieren. [O. D. is a cheese monger. He strongly attends to sorting out old cheese immediately and allows his customers to try all his products.]	2.00	0.76	0.73	1.35
B. T. ist Müller. Er achtet sehr genau auf die Herkunft seiner eingekauften Produkte und verwendet nur ökologisch und qualitativ hochwertiges Getreide. [B. T. is a miller. He carefully attends to the origin of the products he buys and only uses high-quality grain.]	2.27	0.70	1.09	1.41

Table H-3

Behavioral descriptions implying information on ordinary irrelevant behavior used in Experiment 4. In a pretest, subjects (N = 22) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Apart from this, they judged each statement with respect to the type of behavior, using a scale ranging from −3 (“very exceptional”) to +3 (“very ordinary”). Mean valence ratings as well as mean ratings of type of behavior and the corresponding standard deviations are listed for each descriptive statement.

	Valence		***Type of Behavior***	
Irrelevant Behavior, Ordinary	***M***	***SD***	***M***	***SD***
J. L. ist Gärtner. Er interessiert sich sehr für Orchideen und besitzt in seiner Sammlung ein Paar sehr seltene und teure Exemplare. [J. L. is a gardener. He is interested in orchids and has a collection of some very rare exemplars.]	0.27	0.46	1.55	1.30
O. N. ist Gerüstbauer. Zur Zeit arbeitet er auf einer Baustelle im Süden Deutschlands, wo mehrere neue Wohnhäuser und Geschäfte entstehen sollen. [O. N. is a rigging grip. Presently, he works at a building site in southern Germany where several tenements and office buildings are planned to be built.	0.14	0.47	2.27	0.94
J. B. ist Glaser. In seiner Freizeit geht er regelmäßig Fallschirmspringen und spart momentan, um irgendwann später einen eigenen Flugschein zu erwerben. [J. B. is a glazier. In his spare time he regularly does parachuting, and he puts money aside for his own license.]	0.27	0.77	0.18	1.53
T. F. ist Kellner. In seiner Freizeit bastelt er gerne an seinem Motorrad und fährt jedes Wochenende sehr erfolgreich bei Motorradrennen mit. [T. F. is a waiter. In his spare time he likes to work on his motorbike, and at the weekends he usually takes part successfully in motorcycle racings.]	0.14	0.56	0.50	1.50

	Valence		*Type of Behavior*	
Irrelevant Behavior, Ordinary *(continued)*	***M***	***SD***	***M***	***SD***
O. C. ist Masseur. Er besitzt ein kleines Massagestudio in einem Vorort und behandelt dort viele Kunden, die aus der Umgebung kommen. [O. C. is a masseur. He owns a small salon in the suburbs and has many customers from his neighborhood.]	0.32	0.72	2.41	0.73
O. L. ist Prospektverteiler. An jedem Wochenende verteilt er Prospekte von Warenhäusern und Speisekarten von Lieferdiensten an viele Haushalte aus seiner Umgebung. [O. L. is a brochure distributor. On the weekends he regularly distributes prospectuses for stores and menus for delivery services in his neighborhood.]	0.14	0.47	2.27	1.12

Table H-4

Behavioral descriptions implying information on exceptional behavior of cheating used in Experiment 4. In a pretest, subjects (N = 22) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Apart from this, they judged each statement with respect to the type of behavior, using a scale ranging from −3 (“very exceptional”) to +3 (“very ordinary”). Mean valence ratings as well as mean ratings of type of behavior and the corresponding standard deviations are listed for each descriptive statement.

	Valence		***Type of Behavior***	
Behavior of Cheating, Exceptional	***M***	***SD***	***M***	***SD***
G. K. ist Berufssoldat. Er entwendet immer wieder Munition und andere Ausrüstungen aus dem Lager, um diese in zwielichtigen Kreisen zu verkaufen. [G. K. is a lifer. Again and again, he steals munitions and other equipments from the store and sells them to dubious persons.]	−2.86	0.35	−1.86	1.42
K. F. ist Postbote. Er macht häufiger Briefe auf, in denen er höhere Bargeldbeträge oder andere Wertgegenstände vermutet und steckt diese ein. [K. F. is a postman. He often opens up letters in which he assumes money or other valuables and purloins their contents.]	−2.73	0.46	−1.32	1.46
R. T. ist Tätowierer. Mehrmals hat er in seiner Wohnung Straftäter versteckt und die Polizei bei der Suche auf falsche Fährten gelockt. [R. T. is a tattoo artist. Several times he has hidden criminals in his apartment and has thrown the police off the scent.]	−2.91	0.29	−1.41	1.56
T. J. ist Taxifahrer. Er arbeitet mit einem Bankräuber zusammen und hat diesem schon einige Male bei seinen Banküberfällen zur Flucht verholfen. [T. J. is a taxi driver. He cooperates with a robber and after several robberies he helped him escape.]	−2.86	0.35	−2.73	0.55

	Valence		*Type of Behavior*	
Behavior of Cheating, Exceptional *(continued)*	***M***	***SD***	***M***	***SD***
J. F. ist Verkaufshilfe. Er nutzt den Zugang zur Kasse immer wieder dazu, um von ihm gedrucktes Falschgeld gegen echte Banknoten auszutauschen. [J. F. is a temporary shop assistant. Again and again he replaces money from the cash with bogus money that he printed himself.]	−2.77	0.43	−2.18	1.44
F. G. ist Zollbeamter. Er hilft häufig einer Verbrecherbande, LKWs mit Zigaretten über die Grenze zu schleusen und kassiert dabei hohe Bestechungsgelder. [F. G. is a customs officer. He often takes high bribes for supporting a gang of criminals in smuggling cigarettes.]	−2.86	0.35	−1.05	1.70

Table H-5

Behavioral descriptions implying information on exceptional behavior of trustworthiness used in Experiment 4. In a pretest, subjects (*N* = 22) had rated each sentence with respect to its valence, using a scale ranging from −3 (“negative”) to +3 (“positive”). Apart from this, they judged each statement with respect to the type of behavior, using a scale ranging from −3 (“very exceptional”) to +3 (“very ordinary”). Mean valence ratings as well as mean ratings of type of behavior and the corresponding standard deviations are listed for each descriptive statement.

	Valence		***Type of Behavior***	
Behavior of Trustworthiness, Exceptional	***M***	***SD***	***M***	***SD***
E. S. ist Bademeister. Er hat einmal auf dem Heimweg sein eigenes Leben riskiert, um ein im Eis eingebrochenes Kind zu retten. [E. S. is a pool attendant. Once on his way home he took a great risk and saved the life of a child threatened to drown in a frozen-up lake.]	2.55	0.60	−0.50	1.63
F. L. ist Bäcker. Obdachlose aus der Umgebung dürfen bei ihm morgens kostenlos Brötchen essen und im Winter einen warmen Kaffee trinken. [F. L. is a baker. He allows some homeless people from his neighborhood to have breakfast and, in the winter, to have some hot coffee for free.]	2.32	0.84	−1.55	1.06
T. L. ist Bauarbeiter. Um ehrenamtlich am Aufbau eines neuen Kinderdorfes in Afrika mitzuhelfen, hat er schon mehrmals auf seinen Sommerurlaub verzichtet. [T. L. is a construction worker. Several times he passed on his summer vacation to be able to assist the building of a Children’s Village in Africa.]	2.55	0.80	−1.67	1.59
O. H. ist Krankführer. Nach seiner anstrengenden Arbeit hilft er täglich in einer Organisation mit, Essen und Kleidung an Bedürftige zu verteilen. [O. H. is a crane operator. Everyday after work he supports a charitable organization by distributing food and clothing to people in need.]	2.68	0.72	−0.91	1.41

	Valence		***Type of Behavior***	
Behavior of Trustworthiness, Exceptional *(continued)*	***M***	***SD***	***M***	***SD***
M. D. ist Tischler. Einmal in der Woche besucht er ältere Personen in einem Altenheim, um ihnen Bücher und Zeitungen vorzulesen. [M. D. is a joiner. Once a week he visits elder people at a nursing home and reads books and newspapers to them.]	2.55	0.60	−0.45	1.34
G. W. ist Zimmermann. Bei einer Hausreparatur hat er einen sehr hohen Geldbetrag in der Wandverkleidung gefunden und alles dem Besitzer zurückgegeben. [G. W. is a carpenter. Doing repair work he once found a high amount of money behind a paneling and returned it to the owner.]	2.50	0.91	−1.18	1.68

Table H-6

Behavioral descriptions implying information on exceptional irrelevant behavior used in Experiment 4. In a pretest, subjects (N = 22) had rated each sentence with respect to its valence, using a scale ranging from −3 ("negative") to +3 ("positive"). Apart from this, they judged each statement with respect to the type of behavior, using a scale ranging from −3 ("very exceptional") to +3 ("very ordinary"). Mean valence ratings as well as mean ratings of type of behavior and the corresponding standard deviations are listed for each descriptive statement.

	Valence		***Type of Behavior***	
Irrelevant Behavior, Exceptional	***M***	***SD***	***M***	***SD***
C. B. ist Baggerfahrer. Bei der Arbeit an einer Baustelle hat er miterlebt, wie die Ruinen einer alten römischen Siedlung entdeckt wurden. [C. B. is a construction worker who runs an excavator. At work he witnessed the excavation of Roman ruins.]	0.14	0.47	−1.45	1.41
T. U. ist Färber. Er wohnt in einer alten Windmühle an der Nordsee und hat sich auf das Herstellen mittelalterlicher Kleidung spezialisiert. [T. U. is a dipper. He lives in an old windmill and he has specialized on the fabrication of clothing from the middle ages.]	0.32	0.57	−1.09	1.63
H. T. ist Fließbandarbeiter. Er interessiert sich sehr für den fernen Osten und meditiert als praktizierender Buddhist auch in jeder seiner Mittagspausen. [H. T. is an assembly line worker. He is very interested in the Far East and, as a practicing Buddhist, he meditates everyday even in his lunch breaks.]	0.59	0.91	−1.18	1.44
G. R. ist Hafenarbeiter. Er besitzt ein Ferienhaus in den Wäldern Alaskas und möchte im Ruhestand zusammen mit seiner Familie dort leben. [G. R. is a dockworker. He owns a house in the woods of Alaska and wants to live there with his family when retired.]	0.36	0.66	−0.59	1.68

Irrelevant Behavior, Exceptional ***(continued)***	***Valence*** *M*	***Valence*** *SD*	***Type of Behavior*** *M*	***Type of Behavior*** *SD*
P. F. ist Pizzabäcker. Er hat bereits mehrmals versucht, einen Eintrag ins Guinnessbuch für die größte gebackene Pizza der Welt zu erreichen. [P. F. is a pizza baker. For several times already he tried to receive a mention for baking the world's biggest pizza.]	0.14	0.47	–1.77	1.11
H. V. ist Weber. Er ist ein großer Liebhaber klassischer Musik und hat allen seiner drei Kinder die Namen berühmter Komponisten gegeben. [H. V. is a weaver. He loves classical music and thus he named all of his three children after famous composers.]	0.27	0.55	–0.59	1.59

Appendix I

Processing tree representation of the reparameterized two-high-threshold joint multinomial model of source monitoring for three source categories adapted from Bayen et al. (1996) for target items from Sources C (information on cheating), I (irrelevant information), and T (information on trustworthiness) broken down by type of behavior (exceptional, ordinary) and new distractor items.

Note to Figure I-1. D_{Ce} = probability of recognizing target items from Source (Cheat) as old given exceptional behavior; d_{Ce} = probability of recollecting Source (Cheat) for recognized target items from Source (Cheat) given exceptional behavior; D_{Ie} = probability of recognizing target items from Source (Neither…nor) as old given exceptional behavior; d_{Ie} = probability of recollecting Source (Neither…nor) for recognized target items from Source (Neither…nor) given exceptional behavior; D_{Te} = probability of recognizing target items from Source (Trust) as old given exceptional behavior; d_{Te} = probability of recollecting Source (Trust) for recognized target items from Source (Trust) given exceptional behavior; a_{CT} = proportion of selecting either Source (Cheat) or Source (Trust) by guessing given no source recollection for recognized target items; a_C = proportion of guessing Source (Cheat) given no source recollection for recognized target items; b = probability of guessing that an item is old; g_{CT} = proportion of selecting either Source (Cheat) or Source (Trust) by guessing for unrecognized target items; g_C = probability of guessing Soursce (Cheat) for for unrecognized target items.

Note to Figure I-2. D_{Co} = probability of recognizing target items from Source (Cheat) as old given ordinary behavior; d_{Co} = probability of recollecting Source (Cheat) for recognized target items from Source (Cheat) given ordinary behavior; D_{Io} = probability of recognizing target items from Source (Neither…nor) as old given ordinary behavior; d_{Io} = probability of recollecting Source (Neither…nor) for recognized target items from Source (Neither…nor) given ordinary behavior; D_{To} = probability of recognizing target items from Source (Trust) as old given ordinary behavior; d_{To} = probability of recollecting Source (Trust) for recognized target items from Source (Trust) given ordinary behavior; a_{CT} = proportion of selecting either Source (Cheat) or Source (Trust) by guessing given no source recollection for recognized target items; a_C = proportion of guessing Source (Cheat) given no source recollection for recognized target items; b = probability of guessing that an item is old; g_{CT} = proportion of selecting either Source (Cheat) or Source (Trust) by guessing for unrecognized target items; g_C = probability of guessing Soursce (Cheat) for for unrecognized target items.

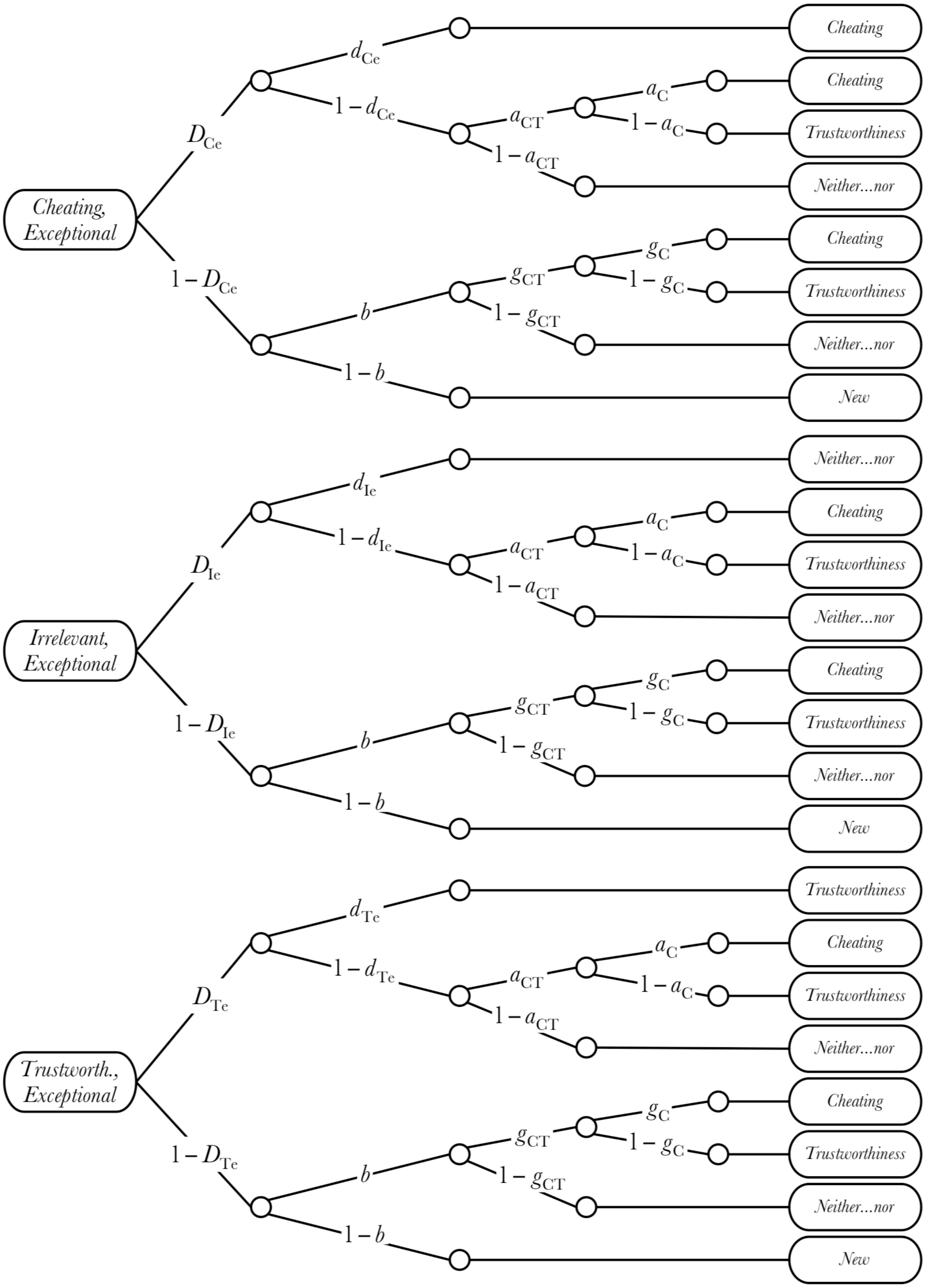

Figure I-1: Processing tree representation of the reparameterized two-high-threshold joint multinomial model of source monitoring for three source categories adapted from Bayen et al. (1996) for target items from Sources *Information on Cheating* (Cheat), *Irrelevant Information* (Neither…Nor), and *Information on Trustworthiness* (Trust), given behavioral descriptions indicating exceptional behavior.

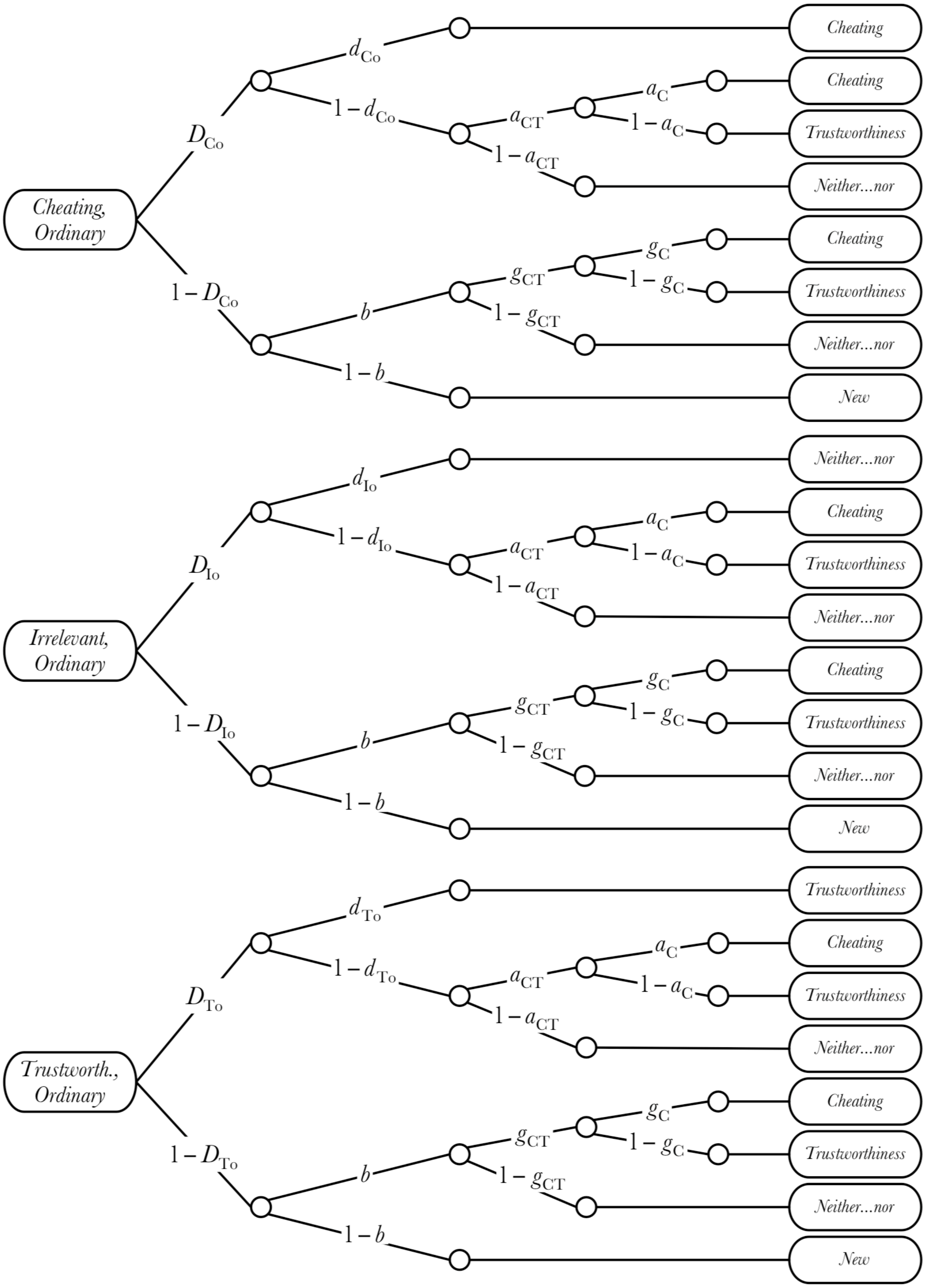

Figure I-2: Processing tree representation of the reparameterized two-high-threshold joint multinomial model of source monitoring for three source categories adapted from Bayen et al. (1996) for target items from Sources *Information on Cheating* (Cheat), *Irrelevant Information* (Neither…Nor), and *Information on Trustworthiness* (Trust), given behavioral descriptions indicating ordinary behavior.

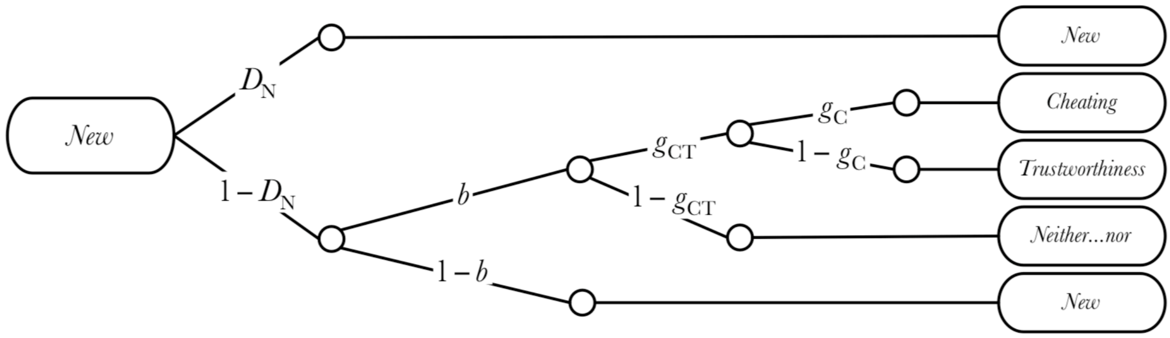

Figure I-3: Processing tree representation of the reparameterized two-high-threshold joint multinomial model of source monitoring for three source categories adapted from Bayen et al. (1996) for new distractor items. D_N = probability of identifying distractors as new; b = probability of guessing that an item is old; g_{CT} = proportion of selecting either Source (Cheat) or Source (Trust) by guessing for unidentified distractors; g_C = probability of guessing Soursce (Cheat) for unidentified distractors.

Erklärung

Hiermit versichere ich, dass ich die hier vorgelegte Dissertation eigenständig verfasst habe. Ich versichere außerdem, dass ich dabei keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen als solche kenntlich gemacht habe. Dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen. Die Dissertation wurde in der vorgelegten oder einer modifizierten Form noch keiner anderen Institution eingereicht. Ebenso versichere ich, dass bisher keine erfolglosen Promotionsversuche meinerseits stattgefunden haben.

Düsseldorf, den 09.09.2005

(Bettina Mehl)