Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 102 (2007) 179–218
COGNITION
www.elsevier.com/locate/COGNIT

Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored ☆,☆☆

Margaret C. McKinnon *, Morris Moscovitch

Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ont., Canada M6A 2E1
Department of Psychology, University of Toronto, Toronto, Ont., Canada M5S 2E1

Received 16 September 2005; revised 21 December 2005; accepted 22 December 2005

Abstract

Using older adults and dual-task interference, we examined performance on two social reasoning tasks: theory of mind (ToM) tasks and versions of the deontic selection task involving social contracts and hazardous conditions. In line with performance accounts of social reasoning (Leslie, Friedman, & German, 2004), evidence from both aging and the dual-task method suggested that domain-general resources contribute to performance of these tasks. Specifically, older adults were impaired relative to younger adults on all types of social reasoning tasks tested; performance varied as a function of the demands these tasks placed on domain-general resources. Moreover, in younger adults, simultaneous performance of a working memory task interfered with younger adults' performance on both types of social reasoning tasks; here too, the magnitude of the interference effect varied with the processing demands of each task.

☆ We are grateful to Alan Leslie, Brian Levine, Phil Zelazo, Douglas Frye, and Cheryl Grady for their comments on earlier versions of this manuscript. We thank Marilyne Ziegler for assistance in computer programming and Elizabeth Olzsewska and Jae-Jin Ryu for assistance in recruiting and testing participants. This study was supported by Natural Sciences and Engineering Research Council (NSERC) and Canadian Institutes of Health Research (CIHR) grants to M.M. and a NSERC doctoral award to M.C.M. The research reported in this manuscript was completed in partial fulfillment of the requirements for M.C.M.'s Doctoral Dissertation at the University of Toronto.

☆☆ This manuscript was accepted under the editorship of Jacques Mehler.

* Corresponding author. Tel.: +1 416 785 2500x3068; fax: +1 416 785 2862.
E-mail address: mmckinnon@rotman-baycrest.on.ca (M.C. McKinnon).

doi:10.1016/j.cognition.2005.12.011

Limits placed on social reasoning by executive functions contribute a great deal to performance, even in old age and in healthy younger adults under conditions of divided attention. The role of potentially non-modular and modular contributions to social reasoning is discussed.

Keywords: Social cognition; Modularity; Aging; Development

1. Introduction

Humans rely on a wide range of evidence to predict and understand one another's behavior. Recent work has focused on the mechanisms responsible for reasoning in social situations, such as those involved in typical 'theory of mind' tasks (ToM) and in versions of the deontic selection task describing social contracts and hazardous conditions. The ubiquitous nature of social reasoning, coupled with observed patterns of selective impairment and sparing of the mechanisms responsible for reasoning on these tasks, led to claims that, much like language, social reasoning tasks, and ToM, in particular exhibited evidence of modularity. In its original conception, this modular "component" of ToM reasoning was thought insufficient for task completion, being bolstered by executive processes (Leslie & Thaiss, 1992). Subsequent claims of dedicated and domain-specific modules for social reasoning, as in ToM (Baron-Cohen, 1995) and deontic reasoning (Cosmides, 1989), however, appeared to exclude the possibility that common domain-general resources, such as working memory, attention, and inhibition contribute additionally to task performance. By contrast, current formulations (Leslie, Friedman, & German, 2004; Leslie, German, & Polizzi, 2005) have again attempted to reconcile evidence of domain-general contributions to these tasks by positing both modular and non-modular components of social reasoning.

In the present study, we examine the contribution of domain-general processing resources to two different types of social reasoning tasks, ToM and deontic reasoning, that appear to share structural similarities and may rely on similar non-modular components for their operation. To test the influence of domain-general cognitive resources on performance on both types of tasks, we compared the performance of young adults under full attention to that of young adults under divided attention and to older adults because these resources are diminished in the latter two groups.

The debate between modular and non-modular contributions to social reasoning has played out primarily in the developmental literature, where strong arguments were made initially regarding the modularity of social reasoning tasks, such as ToM (Leslie & Thaiss, 1992). In the case of ToM, evidence for strict modularity was drawn primarily from extensive research (Baron-Cohen, 1995; Happe, 1994; Ozonoff, Pennington, & Rogers, 1991) indicating that children with autism are impaired in their ability to understand the thoughts and feelings of others. In contrast to their impaired ToM abilities, these children demonstrated relatively intact performance on matched control tasks that did not require an understanding of another's mental states (Charman & Baron-Cohen, 1992). Here, it was assumed that social reasoning modules acted as domain-specific systems that operated in an automatic and rapid fashion, producing shallow output that must be processed further by higher-order systems; such modules were said to be informationally encapsulated and impenetrable to the input of top–down processes (Fodor, 1983; Moscovitch, 1992; Moscovitch & Umilta, 1990).

More recent evidence demonstrating that children with autism are impaired across a wide range of tasks, including tests of executive function (e.g., Ozonoff et al., 1991), and correlations between executive functioning and ToM performance in children with autism (e.g., Zelazo, Jacques, Burack, & Frye, 2002), in normally developing children (e.g., Hughes, 1998), and in neurological populations (e.g., Channon & Crawford, 2000), however, has challenged the strictly modular interpretation. Indeed, most current theories of social reasoning in children, and in particular, of ToM, emphasize the contribution of potentially modular and non-modular components to task performance (Leslie et al., 2004; Moses, 2001).

In general, these revised theories form two camps, one emphasizing the role of executive functions (EFs) in the initial construction of a conceptual understanding that other people form desires and beliefs (the emergence account); these concepts, once formed, are thought to no longer require EFs for their operation (e.g., Moses, 2001). The other view emphasizes the on-line contribution of EFs to social reasoning (the performance account; e.g., Leslie et al., 2004). In this case, EFs are thought to remain involved in ToM in perpetuity, provided that the demands of the ToM task are such that they are required (e.g., in a ToM task requiring multiple levels of inhibition).

For example, in one influential formulation of the performance account, Leslie and his colleagues (Leslie et al., 2004, 2005) have suggested that ToM reasoning in young children reflects the underlying contribution of both innately specified candidate beliefs (the Theory of Mind mechanism; ToMM) and non-modular EFs (the selection process; SP), likely inhibitory in nature. In this performance account of social reasoning, the ToMM forms part of the core architecture of the human brain, and is specialized for learning about mental states, providing the underpinnings of rudimentary understandings of belief and desire observed initially in young children. Here, limits placed on ToM by developing EFs restrict the expression of belief when performance demands overwhelm the processing capacity of the developing brain (for similar formulations of cognitive capacity and task performance, see Case, 1992; Case, Kurland, & Goldberg, 1982), as when tasks require multiple levels of inhibition. In this view, progressive changes in executive functioning lead to the ability to perform increasingly complex social reasoning tasks (e.g., first- and second-order false belief tasks). "Theory of mind" EFs, however, are thought to be wholly or partly distinct from the domain-general executive functions subserving other tasks (e.g., formal reasoning; Leslie et al., 2004).

By contrast, emergence accounts of ToM reasoning (e.g., Carlson & Moses, 2001; Moses, 2001; Perner & Lang, 1999) suggest that ToM belief is constructed by children and that executive functioning plays a key role in the construction process. Proponents of these theories view developmental changes in ToM as evidence of increases in conceptual competence, with only a minor role ascribed to performance (i.e., capacity) change. Here, construction of a belief concept requires a prerequisite level of executive functioning allowing the respondent to reflect on the thoughts and actions of others, distance oneself from the immediate situation and potentially inhibit salient, but misleading, knowledge (Moses, 2001). In this view, the same executive processing mechanisms are thought to subserve "theory of mind" and "non-theory of mind" function. Once constructed, however, ToM beliefs no longer require executive functions for their expression, forming automatically in situations requiring an understanding of others' thoughts and feelings. In adults, these constructs would correspond roughly to the notion of crystallized intelligence, knowledge that is stable over long period of times, and remains impervious to declines associated with cognitive aging. In this sense, the operation of these constructs can be considered "modular", operating automatically and being impervious to domain-general processing requirements.

Although previously unexplored, the emergence and performance accounts of ToM appear applicable to other tests of social reasoning, such as the deontic selection paradigm, which, like ToM, has been subject to claims of modularity. The results of numerous investigations (Cosmides, 1989; Fiddick, Cosmides, & Tooby, 2000; Gigerenzer & Hug, 1992) reveal that, when presented with selection tasks requiring reasoning about complex social situations or hazardous conditions, participants' performance often exceeds that observed on descriptive or abstract versions of the same task. The most prominent account of these findings suggests that performance on social reasoning versions of these deontic tasks may depend on dedicated and domain-specific cognitive modules devoted to social interchange (Cosmides, 1989) and reasoning about hazardous conditions (Fiddick et al., 2000). Taken together, the results of numerous studies (e.g., Almor & Sloman, 1996; Girotto, Kemmelmeier, Sperber, & van der Henst, 2001; Griggs & Jackson, 1990) indicating that manipulation of variables, such as the structural features of conditional arguments in the reasoning task (Fodor, 2000), and the requirement to hold in mind information, results in enhanced performance on these tasks, however, suggests that participants' performance may rely in part upon the general processing mechanisms (e.g., working memory) manipulated in these experiments.

Indeed, both deontic selection tasks and ToM tasks share structure with other logical reasoning tasks that rely heavily upon domain-general processing resources, such as working memory (e.g., transitive inferencing, matrix reasoning), including a requirement to compare multiple pieces of information simultaneously. In deontic reasoning tasks (see Appendix A), participants select from four cards those which are necessary to solve a reasoning problem presented in the text of a story. Different types of information are displayed on each side of these cards (one side of which must be imagined) and participants consider simultaneously multiple pieces of information regarding societal rules and obligations when forming their judgments. When selecting from these cards, they must relate the information presented in the story to two possible outcomes inferred from reading of the story's text. Simultaneous consideration of this information requires not only that previously displayed information be recalled, but also that it be integrated to determine the appropriate response selection. Similarly, higher-order ToM tasks, such as second-order false belief tasks that require participants to hold in mind and compare the contrasting beliefs, or perspectives ('A thinks that B feels X') of different people, may draw heavily on processing resources, such as working memory, that are thought typically to mediate demands to hold in mind and integrate different pieces of information. Although putative pre-existing cognitive biases for reasoning about situations involving danger or cheating and for understanding the thoughts and feelings of others may play an important role in driving task performance, it seems unlikely that additional domain-general resources do not play a role in successful task completion where such cognitive demands exist. As for ToM, we suspect that potentially modular (e.g., cognitive reasoning biases) and non-modular components (e.g., working memory) contribute to performance on social versions of these deontic reasoning tasks.

Although many theories of social reasoning were formed from work with young children, we believe they are relevant to performance in old age and in healthy adults without disorder. Indeed, much as the developmental literature informs current theories of social reasoning in adults, we believe studies of adult performance may also speak to the existing literature on the development of ToM and social reasoning in young children. In the present set of studies, we examined the performance of older adults, with presumed declines in EF, and younger adults under conditions of divided attention that restrict executive functioning, on social reasoning tasks thought to recruit the same modular and non-modular processes as tested in young children.

In the first set of experiments, older adults completed ToM tasks and social versions of the deontic reasoning task. Older adults are known to have age-related decrements across a range of tasks tapping domain-general resources, including executive functioning (Cepeda, Kramer, & Gonzalez de Sather, 2001; Craik, 1977; Meguro et al., 2000; Salthouse, 1994). As well, the frontal lobes, implicated in working memory (e.g., Ragland et al., 2002), are among the first structures to deteriorate with age (Tisserand, van Boxtel, Gronenschild, & Jolles, 2001). Thus, correlations between older adults' ability to complete logical reasoning tasks and their performance on tests of executive function (e.g., working memory; Salthouse, 1992) suggest declines in central processing resources among this population. We consider it likely that these same resources, diminished in aging, contribute to performance of non-modular aspects of ToM and deontic selection tasks.

If performance on social reasoning tasks in older adults relies on long-established cognitive constructs that no longer require executive functions for their operation (the emergence account), performance on tasks tapping these beliefs, is likely to be relatively spared by the cognitive aging that often is associated with depletion of cognitive resources (Craik, 1994). Alternatively, however, if, as we suspect, performance on social reasoning tasks draws, at least in part, on the same cognitive resources and general abilities as does performance on other tasks that deteriorate with age, then older adults should show an age-related deficit on them. Moreover, this performance should be impacted differently by varying the load placed on EF (e.g., high and low working memory) by these social reasoning tasks.

Although our experiments with aging adults could reveal an association between declines in EF and performance on social reasoning tasks, a more direct demonstration would be required to confirm this relation. In a second set of experiments, we rely on the dual-task method to demonstrate that central processing mechanisms contribute directly to performance of social reasoning tasks. Specifically, we required young adults to complete ToM and deontic selection tasks while performing a secondary task that relied heavily upon central processing resources (i.e., working memory). A wide body of previous research demonstrates that if two concurrently performed tasks rely upon the same resource (Allport, Antonis, & Reynolds, 1972), the same hemisphere (e.g., Klein, 1976; Moscovitch, 1976; Moscovitch & Klein, 1980) or the same brain region (e.g., Moscovitch, 1994; Moscovitch, Fernandes, & Troyer, 2002), competition for shared resources will result in interference effects for one or both tasks.

Hence, we reasoned that, if performance on social reasoning tasks Leslie et al. (2004) in adulthood relies, in part, upon the same central processing resources that contribute to performance of the 2-back task, simultaneous performance of these tasks should result in interference on one or both of these tasks. Alternatively, if performance on social reasoning tasks relies strictly (or preferentially) upon long-standing constructions acquired in childhood (Moses, 2001) that no longer require EFs for their expression, little interference effects would be expected for either task.

Finally, by testing performance on two different social reasoning tasks we hoped to examine whether the same domain-general EFs contribute to performance of both tasks (the emergence account; e.g., Carlson & Moses, 2001; Perner & Wimmer, 1985), or, alternatively, if specialized EFs exist for each type of social reasoning task (the performance account; e.g., Leslie et al., 2005). Here, we predicted that performance decrements arising from executive dysfunction would be similar for both the ToM task and the deontic reasoning task. Differential patterns of impairment on social reasoning tasks involving the same level of EF demands, however, would provide an early indication that specialized EFs are required for deontic reasoning and for ToM (or, that the tasks themselves require different amounts of EF for successful performance).

2. Experiments 1–3 – Aging and social reasoning

Previous experiments reveal a conflicting pattern of findings regarding older adults' performance on social reasoning tasks. For example, whereas several studies have revealed evidence of age-related impairments on ToM tasks, other studies have failed to report any such evidence of impairment, instead demonstrating intact performance among older adults on related ToM tasks. We suspect that apparent discrepancies in performance among older adults stem from method variance across studies. Specifically, ToM deficits have been reported in older adults under conditions involving high demands to recall information (Experiment 1, Maylor, Moulson, Muncer, & Taylor, 2002). Similar age-related deficits emerge on ToM tasks involving explicit demands to integrate the conflicting perspectives of two different people (Saltzman, Strauss, Hunter, & Archibald, 2000), as in the Knower/Guesser task where participants must reconcile the differing perspectives of the experimenter and a confederate to arrive at the correct solution to a complex ToM problem. By contrast, relatively intact performance has been reported on ToM tasks that involve less explicit demands to hold in mind and compare conflicting pieces of information (Saltzman et al., 2000) and when older adults with ostensibly superior verbal and intellectual functioning (Happe, Winner, & Brownell, 1998; cf. Maylor et al., 2002) are tested. Interestingly, participants with Alzheimer's disease also fail ToM tasks when demands on domain-general resources appear high, as is the case for second-order false belief tasks; no such impairments are reported among this population for tasks involving lower-level reasoning (i.e., first-order false belief tasks; Gregory et al., 2002). Taken together, this pattern of findings suggests that task demands stemming from domain-general processing requirements may mediate the pattern of results observed in older adult on ToM tasks.

The picture appears similar for reasoning on selection tasks, where older adults demonstrate impairment on some measures of social reasoning, but appear intact on others. Thus, older adults demonstrate intact performance relative to younger control subjects on versions of these tasks involving ostensibly unfamiliar (or reportedly age atypical) scenarios such as detecting underage drinkers or impaired drivers (Overton, Yaure, & Ward, 1986). By contrast, performance is impaired on other versions of the task involving age-relevant scenarios, such as detecting individuals who cheat on collecting senior's discounts or pension plans (Overton et al., 1986). Age-related impairments are also noted on versions of the task involving affect-laden scenarios (e.g., children with AIDS, teenage abortions; Pollack, Overton, Rosenfeld, & Rosenfeld, 1995). Consistent with our view that domain-general resources, such as working memory and attention, contribute to performance on such tasks, in one experiment, use of a metacognitive strategy designed to orient older adults to the cognitive demands of the task reduced performance deficits on a version of the task involving affective reasoning (Pollack et al., 1995).

Our experiments with older adults were designed to test two hypotheses. The first concerned the question of whether older adults perform equivalently to younger adults on social reasoning tasks. Here, we hypothesized that age-related declines in executive functioning would contribute to a pattern of reduced performance in older adults (the performance account; Leslie et al., 2005), regardless of the type of social-reasoning task tested. Alternatively, if responding on social reasoning tasks is mediated preferentially by longstanding constructions that no longer rely upon EF (the emergence account; e.g., Moses, 2001; Perner & Lang, 1999), one would expect little or no difference in performance between age groups and enhanced reasoning on social versions of the reasoning tasks, unless one assumes that such constructions deteriorate with age. In addition, we were interested in whether participants' performance varied as a function of the executive functioning load of these tasks.

2.1. Experiment 1

In this experiment, older and younger adults answered a series of questions regarding the thoughts and feelings of characters described in a set of short stories typical of those used in adult investigations of ToM reasoning (e.g., Happe et al., 1998; Maylor et al., 2002). There were two different types of ToM questions. The first required participants to answer questions regarding the first-order beliefs ('A thinks or feels X') of a single character described in the story; a second set of questions addressed participants' ability to answer questions regarding the second-order beliefs ('A thinks that B thinks or feels X') of two different characters. Like other investigators (e.g., Frye, Zelazo, & Burack, 1998; Perner & Wimmer, 1985; Zelazo & Frye, 1998), we believe that second-order ToM questions involve greater cognitive demands than do first-order questions. Specifically, whereas first-order questions require recognition of a single perspective only (i.e., that of one person only), second-order belief questions require that participants not only recognize the belief and intentions of individual characters, but also that they hold in mind and integrate these perspectives to arrive at a correct solution to the ToM question posed. These requirements would appear to place high demands on domain-general resources, and in particular, on working memory, which is thought to mediate such performance demands. Hence, we made two predictions regarding participants' performance on this task. Overall, we expected that aging, which is associated with declines in cognitive resources, would lead to impaired performance, relative to younger that observed in younger adults. In addition, we predicted that the increased cognitive demands of the second-order questions would result in a heightened level of impairment among older adults on this question type, a pattern predicted by performance accounts of social reasoning (e.g., Leslie et al., 2004). Alternatively, if responding on this task mediated preferentially by longstanding constructions about others' thoughts and feelings that no longer require executive functions for their operation (Moses, 2001), there should be little or no difference between age groups or question type, unless one assumes that such constructs deteriorate with age. In that case, the deficit should be equivalent across tasks.

2.1.1. Method

2.1.1.1. Participants. Our sample included 12 older adults ($M = 78.18$ years of age; mean years of education: 15.90) and 12 undergraduates ($M = 20.16$ years of age; mean years of education: 14.74) enrolled at a large-sized Canadian university. Older adults were screened for a history of neurological or psychiatric illness; participants with a history of head injury or who were taking drugs that could alter cognitive function were excluded from the initial sample. We found no evidence of dementia-related illness amongst our sample; all scores on the MMSE were well above 23, the cut-off for impairment on this measure. All had normal or corrected-to-normal vision and hearing.

2.1.1.2. Stimulus materials. We modeled our stimuli after those used in previous investigations of ToM in adult participants. Specifically, we created a series of scenarios that described complex social situations, such as social faux pas. For example, one scenario described a case of mistaken identity; a female is carelessly identified as a male. The stories were constructed such that we were able to test participants' ability to answer first- and second-order ToM questions. A total of eight different stories were constructed, all describing complex social situations. In order to reduce memory demands, our stories were relatively brief (50–75 words) and available to inspection until a response was made.

Table 2.1. Example of short story.

Ellen's hockey team has just defeated Brad's hockey team. The teams were evenly matched, but Ellen's team worked harder and won. When Brad sees Ellen, he says, "Congratulations on your victory. It's too bad our goalie was so ill".

Each story was followed by eight questions:

(i) *First-order thought and affective questions:* These questions required consideration of one character's perspective only; questions followed the form 'A thinks X' (e.g., Why does Brad think Ellen's team won the competition?) and 'A feels X' (e.g., How does Ellen feel when Brad says what he says?). The questions were written so as to assess participants' understanding of each character's perspective. Hence, there was one first-order thought and one first-order affective question for each character.

(ii) *Second-order thought and affective questions:* These questions required comparison of two different character's perspectives; questions followed the form 'A thinks that B thinks X' (e.g., What does Brad think Ellen will think when he makes his comment?) and 'A thinks that B feels X' (e.g., How does Brad think Ellen will feel when he makes his comment?) There was one second-order thought question and one second-order affective question for each character.

2.1.1.3. Procedure. Participants were tested individually. They were told that they would read a series of stories and then answer questions about the thoughts and feelings of the characters described in the stories. They first completed a practice story and answered practice questions before proceeding to the experimental session. Participants were encouraged to read each story carefully and to take as much time as necessary to understand fully the story's contents. Upon participants' indication of their readiness, questions were presented orally following reading of the story. Participants were encouraged to refer back to the stories while forming their responses. Responses were recorded verbatim by the experimenter and on audiotape.

2.1.2. Results and discussion

Correct responses to the questions received a score of 1; incorrect responses received a score of 0. Hence, in each condition (first- and second-order questions), participants could earn a total of four points by answering correctly the four questions posed for each scenario, for a total possible score of 32. Although we found no differences between older and younger adults in accuracy of responding to the first-order questions, older adults performed more poorly than younger adults when required to adopt two perspectives simultaneously (see Fig. 1). Because responding did not differ between the thought and affective questions, scores were collapsed across these two measures.

The results were analyzed using a 2 Between (Age: old versus young) $\times$ 2 Within (Stimulus Type: first-order versus second-order) mixed-design ANOVA. The analy-

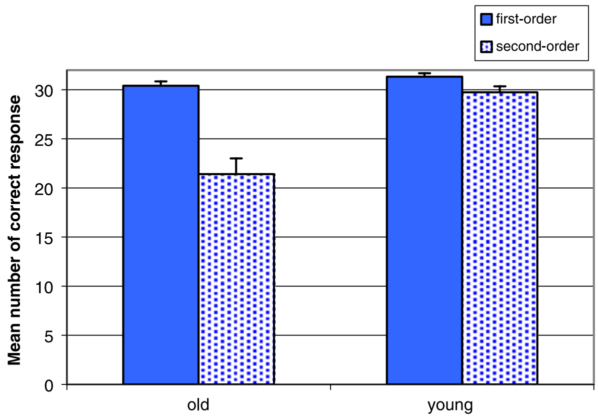

Fig. 1. Mean number of correct responses across older and younger groups.

sis revealed a main effect of age [$F(1,22) = 20.86$, $p < .001$]; the performance of younger adults exceeded that of older adults on this task. There was also a main effect of stimulus type [$F(1,22) = 50.82$, $p < .001$]; participants were more accurate in responding to first-order than to second-order ToM questions. Nonetheless, the age effects observed in this experiment appeared to be mediated by the type of question asked; the two-way interaction between age and stimulus type was highly significant, $F(1,22) = 24.96$, $p < .001$. Further contrasts revealed that whereas older adults performed more poorly relative to younger adults for second-order questions [$F(1,11) = 5.20$, $p < .001$], such differences were not apparent when responses for first-order questions were compared [$F(1,11) = 1.96$, $p > .05$].

Errors were further divided into descriptive categories in order to identify the locus of response errors on second-order questions in older adults. Five categories of errors on these questions were identified: (i) those that involved adopting the wrong perspective (that of a single person only, rather than considering both perspectives simultaneously) – 38.4% of errors in older adults (48 errors total); 7.7% of errors in young adults (2 errors total); (ii) errors that involved a failure to consider the different types of information held in mind by the two characters described in the story (e.g., known information that differed across the parties) – 29.6% of errors in older adults (37 errors total); 23.1% of errors in young adults (6 errors total); (iii) incorrect attributions of mental states – 15.2% of errors in older adults (19 errors total); 65.4% of errors in young adults (17 errors total) (iv) non-mental explanations: 11.2% of errors in older adults (14 errors); 0 percent of errors in young adults (v) failure to respond (i.e., ‘I don’t know’) – 5.6% of errors in older adults (7 errors total); 7.7% of errors in young adults (2 errors total). Hence, almost 70% of the total errors committed by older adults involved a failure to consider multiple pieces of information simultaneously either by adopting a single perspective only or by failing to consider the differing information held in mind by the two parties described in the social scenarios.

The results of this experiment indicate that whereas older adults are able readily to adopt the perspective of an individual character (first-order ToM questions), they experience difficulty when required to adopt two perspectives simultaneously (second-order ToM question), a finding mirrored in other domains, such as moral reasoning (Pratt, Diessner, Pratt, Hunsberger, & Pancer, 1996). These differences were apparent not only for questions regarding the cognitive state of the characters, but also for questions referring to their emotional states. One potential explanation for this pattern of performance is that the cognitive demands of this task, such as a need to hold in mind and to integrate two perspectives simultaneously, impacted negatively on performance among older adults, who have known deficits central processing resources. This finding is also consistent with processing accounts of ToM reasoning (e.g., Leslie et al., 2004) where older adults should be differentially influenced by varying levels of such demands. Finally, our results are at odds with emergence accounts of social reasoning which suggest that, after a critical period, performance of ToM tasks (e.g., Moses, 2001; Perner & Lang, 1999) is subserved by long-standing constructs that are impervious to domain-general processing demands. We acknowledge that older adults' pattern of performance may have stemmed from impairments to a "module" specialized for reasoning about other's thoughts and feelings. Our finding of differential performance across question types, however, argues against this conclusion.

In a second experiment, we examined performance in older adults on a different type of reasoning task, versions of the deontic task involving reasoning about social contracts and hazardous conditions. Because these tasks also appear to place high demands on participants' cognitive resources, we expected that, here too, performance deficits would emerge among older adults.

2.2. Experiments 2 and 3

We administered two different versions of the deontic selection task: an unfamiliar *social contract* version (Cosmides, 1989)and an unfamiliar *precaution* version (Fiddick et al., 2000); both versions of the task result in enhanced performance in college-aged students when compared to task versions involving descriptive or abstract reasoning. Given older adults' known deficits in central processing resources, such as working memory, and task requirements to compare multiple pieces of information, we expected them to perform poorly on deontic selection tasks, regardless of the version of the task tested. Such performance would indicate that domain-general processing resources continue to contribute to ToM performance, even in old age. This view is at odds with earlier claims of domain-specific and dedicated cognitive modules for the detection of cheaters (Cosmides, 1989) and the avoidance of hazardous conditions (Fiddick et al., 2000) and with theories of social reasoning (e.g., Moses, 2001; Perner & Lang, 1999) that stress reliance upon long-standing cognitive constructs that are impervious to the operation of domain-general resources. These views suggest that older adults should exhibit relatively spared performance on social contract and hazardous versions of these tasks, regardless of their inherent cognitive demands. Moreover, any potential deficits among older adults would stem from impairments to systems or constructs dedicated to reasoning about social interchange.

2.2.1. Experiment 2

The results of numerous studies (e.g., Cosmides, 1989; Gigerenzer & Hug, 1992) indicate that college-aged students perform better on selection tasks that imply a social contract between members of a group. Typically, this contract is described as a means of preventing members of the group from cheating (e.g., violating the social mores of a tribe). Cosmides (1989), among others, suggests that evolutionary pressures may have encouraged the development of specialized skills necessary for reasoning about cooperative interactions in a society; this model suggests that these skills are subserved by a dedicated and domain-specific cognitive module and, as such, are likely to be affected only little, if at all, by loss of the central processing resources that accompany aging. These skills may be subserved by cognitive constructs similar to those described in emergence accounts of ToM reasoning (e.g.,Moses, 2001; Perner & Lang, 1999). In this experiment, we compared the performance of older and younger adults on a social contract and descriptive version of this task. We reasoned that the cognitive demands of these tasks (e.g., the requirement to hold in mind and integrate multiple pieces of information simultaneously), coupled with reduced operational efficiency or working memory capacity in older adults, would contribute to poorer performance in this group, even on social contract tasks.

2.2.1.1. Method

2.2.1.1.1. Participants. Our sample included 24 older adults ($M = 75.39$ years of age; mean years of education: 13.63) and 24 undergraduates ($M = 21.03$ years of age; mean years of education: 14.86) enrolled at the University of Toronto. Older adults were recruited from a community-based sample of volunteers; all received nominal payment for their participation, which took approximately 10 min. These participants were screened for a history of neurological or psychiatric illness; those with a history of head injury or who were taking drugs that could potentially alter cognitive function were excluded from the initial sample. We found no evidence of dementia-related illness amongst our sample; all scores on the MMSE were well above 23, the cut-off point for impairment on this measure. Undergraduate participants received partial course credit or nominal payment for their participation in the experiment, which also took approximately 10 min. All participants had normal or corrected-to-normal vision and hearing.

2.2.1.1.2. Stimulus materials and procedure. Materials in this experiment were the same as those used in Cosmides's (1989) earlier experiments (see Appendix A). Each participant received two sheets of paper with an unfamiliar scenario on one page and four hypothetical selection cards on the following page; together these comprised the Deontic selection task (see Appendix A). There were two different unfamiliar scenarios. Half of the participants read a scenario that described an unfamiliar *social contract*; the remaining half received the descriptive version of the task. The scenario was available for inspection until the participant responded. The social contract scenario described a male member of a Polynesian culture, the Kalumae, who had been entrusted to enforce the strict laws of his community. In particular, the tribesman is charged with ensuring that unmarried men do not eat cassava root, a powerful aphrodisiac.

In the *descriptive* version of this scenario, the main character is an anthropologist who is investigating the veracity of a peculiar claim presented to him by a former colleague; this claim states that in the Kalumae culture, only men with tattoos on their face are permitted to eat cassava root. In both scenarios, tattoos are a mark of the marriage of a male member of the tribe. Both scenarios included the rule that, "If a man eats cassava root, then he must have a tattoo on his face". In the social contract version of the scenario, the Kalumae man is described as investigating whether any of four other members of his tribe have broken the rule he must enforce; to fail to catch any such cheaters would bring disgrace upon his family. By contrast, in the descriptive version of the task, the anthropologist seeks merely to verify the veracity of his colleague's claim.

Participants were tested individually; they received either the social contract or descriptive version of this task. They were encouraged to ask questions about the experiment throughout the test session; in general, responses to these questions involved restatement of the rule, "turn over only those cards that are necessary to see if what the tribesman said was true". In order to lessen the demands this task placed upon memory, participants were allowed to refer back to the scenario describing the tribesman's/anthropologist's dilemma while selecting the appropriate response card(s). Several older adults chose to look back at the stories in forming their response; few younger adults chose to do so.

2.2.1.2. Results and discussion. As predicted, younger adults outperformed older adults on the social contract version of the deontic selection task. Older adults experienced a slight improvement on the social contract version relative to their performance on the descriptive version of the task, but the benefits of social reasoning were greater for younger adults.

Because performance levels were so low amongst older adults, we were unable to use tests of proportions to analyze our results.[1] Instead, responses received a score of either −2, −1, 0, 1, 2. Participants received 1 point for each card selected correctly (turned over correctly) and a deduction of 1 point for each card incorrectly selected. For example, a score of 2 would correspond to turning over correctly the two cards necessary to verify that what the tribesman said was true and correctly leaving the two unnecessary cards unselected; a score of −2 would correspond to selection of the two unnecessary cards while the two correct cards were left unselected. This method allowed us to capture the highly variable response patterns of older adults

[1] Although the authors of at least one other similar study (Fiddick et al., 2000) have relied upon tests of the proportion of respondents (differences between two independent proportions; *z*-scores) *correctly responding* to analyze their data, use of this method would be ill-advised for these data. Specifically, because of the low number of older adults responding correctly (0 percent in one cell), these data violate the standard binomial requirement that $n(p)$ and $n(1-p)$ must be equal to or greater than 5 (Kirk, 1995); violation of this assumption renders interpretation of a test of proportions equivocal.

and avoided the use of binary scoring (i.e., 1 point for a correct answer; 0 points for an incorrect answer). For example, some participants chose to turn over all 4 cards (resulting in a score of zero: 2 cards correctly selected and 2 cards incorrectly selected); others selected one card only (score of 1 if correctly selected, score of −1 if incorrectly selected). This resulted in a conservative method of scoring, as older adults received partial credit for answers that, taken as a whole, were incorrect.

We analyzed these results using a 2 (Age: old versus young) × 2 (Stimulus Type: social contract versus descriptive) between-subjects ANOVA (see Fig. 2).

Overall, the performance of younger adults exceeded that of older adults on this task, [$F(1,44) = 21.37$, $p < .001$]. Our analysis also revealed some sparing of performance in both groups of adults on the social contract version of this task; there was a main effect of stimulus type [$F(1,44) = 11.20$, $p < .001$] in the absence of any significant interaction between age and the type of story administered ($p > .05$).

There were two main findings in Experiment 2. The first concerns the poor performance of older adults relative to that of younger adults on the social contract version of the deontic task. When older adults were asked to complete a task that required them to reason about multiple pieces of information simultaneously, performance was impaired relative to younger adults. The poor performance of older adults on this task indicates that a dedicated cognitive module for reasoning about the detection of cheaters does not spare selectively performance in these participants on a task that appears to rely heavily upon cognitive resources. As a result, we conclude that reasoning on social versions of the deontic selection task is penetrable by non-modular aspects of task performance.

Our analysis did reveal a slight sparing of performance in older adults on the social contract version of the task; performance was better for both groups of participants on the social contract as compared to descriptive version of the deontic task. Inspection of Fig. 2, however, indicates that this benefit was far greater for younger, than for older, adults. One potential explanation for this selective sparing may be that the enhanced familiarity of situations involving the detection of cheaters (e.g., misbehaved children, scheming coworkers) allows participants to schematize more readily such information, resulting in a reduction in the amount of information to be held in mind and on which to operate. This explanation would provide a partial

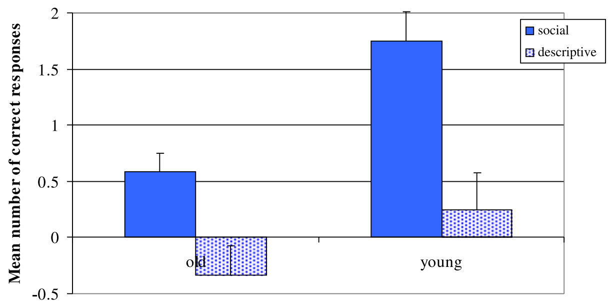

Fig. 2. Mean score across older and younger adults.

account of earlier claims of a priori reasoning biases for situations involving the detection of cheaters, in that quickly assimilating and 'chunking' familiar information would facilitate performance by reducing the amount of information to be operated upon and, hence, also reducing cognitive processing demands. This claim, however, is speculative and awaits further investigation.

In a second experiment, we sought to examine whether these performance decrements would extend to another version of the deontic selection task that has been described previously as relying on a dedicated and domain-specific cognitive system for reasoning about hazardous conditions (Fiddick et al., 2000) and that may also place similar demands on cognitive resources.

2.2.2. Experiment 3

Cosmides and her colleagues (Fiddick et al., 2000) have argued recently that in addition to a specialized mechanism for the detection of cheaters, humans also possess a dissociable mechanism dedicated to reasoning about situations involving hazardous conditions. Specifically, they found that college-age students exhibited superior performance on a *precaution* version of the deontic selection task when the test story was structured to describe a hazardous situation in which lives could be jeopardized for failing to take a precautionary measure (i.e., the wearing of bright jackets while hunting). The standard version of this task required verification of a descriptive claim only. In a second experiment, we sought to establish whether the performance decrements observed amongst older adults in Experiment 2 would extend to situations describing hazardous conditions.

2.2.2.1. Method

2.2.2.1.1. Participants. Our sample was comprised of the same older and younger adults who had participated in Experiment 2. Presentation of stimuli was counterbalanced to eliminate the possibility of order-of-administration effects.

2.2.2.1.2. Stimulus materials and procedure. The stimuli used in this experiment were similar to those used in Experiment 2, with the exception that instead of requiring participants to engage in a social contract reasoning task, the test condition contained a story that required reasoning about a hazardous situation (see Appendix A). Participants received two sheets of paper that together comprised the deontic selection task. The first sheet described an unfamiliar scenario. In this scenario, a returning tribesman notices the presence of bright orange jackets in his tribal village. Unable to understand their purpose, he asks a fellow tribesman, "What are these for?" The tribesman's conditional response is, "If you go hunting, then you wear these jackets to avoid being shot". The scenario then describes one of two versions of the selection task. In the *standard* version, the tribesman expresses his uncertainty as to whether or not his friend's claim is true. By contrast, in the *precaution* version, the tribesman expresses a different concern; he wonders whether all of his fellow tribesmen "know about the jacket" and are "needlessly endangering themselves". In both versions, participants were asked to select only those cards they would need to turn over in order to see if what the second tribesman said is true.

The procedure was identical to that in Experiment 2.

2.2.2.2. Results and discussion. The method of scoring was identical to that used in Experiment 2. There was a trend towards poorer performance in older adults on this version of the deontic selection task. Whereas differences in accuracy were apparent across the hazardous and descriptive scenarios for younger adults, these differences were diminished greatly in older adults.

We analyzed performance using a 2 (Age: old versus young) × 2 (Stimulus Type: hazard versus descriptive) between-subjects ANOVA (see Fig. 3). As in Experiment 2, there was a main effect of stimulus type [$F(1,44) = 10.08$, $p < .01$]; performance was better for the precaution condition than for the descriptive condition across all participants. Although these analyses failed to reveal a main effect of age or an interaction between the type of story presented and the age of participants (ps > .05), post hoc tests revealed a trend towards better performance in younger adults than in older adults on the precaution [$F(1,44) = 3.67$, $p = .06$] but not descriptive version ($p > .05$) of this task.

The results of Experiment 3 provided some evidence that older adults were additionally impaired relative to younger adults on the precaution version of the deontic selection task, although the findings were not as compelling as those of Experiment 2. Analysis revealed a trend towards better performance in younger than in older participants on the precaution version of the task; no differences were observed between groups on the descriptive version of the task but only because younger adults performed very poorly. This finding are consistent with the suggestion that declines in domain-general resources in older adults affect both social and descriptive versions of deontic tasks, perhaps due to structural similarities among tasks; putative "modules" or constructs for social reasoning did little to spare performance on the precaution version of the task. These results, however, should be interpreted with some caution. One possibility that cannot be dismissed is that reasoning about hazardous situations is less demanding than reasoning about social contracts. Because older adults' performance did not improve significantly on the hazardous version,

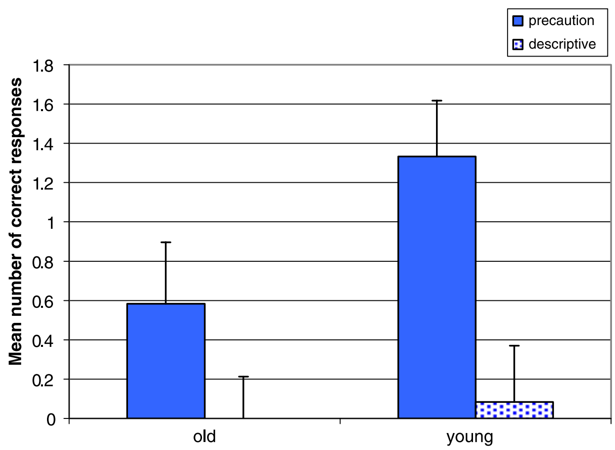

Fig. 3. Mean score across older and younger adults.

however, it is difficult to argue that our results support a view of performance that excludes the contribution of domain-general processing resources.

3. Experiments 4 and 5

Experiments 1–3 indicate that older adults demonstrate impaired performance on three different social reasoning tasks that appear to draw, at least in part, on cognitive resources for successful task completion. Although we suspect that declines in domain-general resources, such as working memory, played a substantive role in the pattern of impaired performance observed among older adults on these tasks, the evidence for this contribution remains indirect, based as it is on the purported causes of the age difference we observed. Equally plausible is that aging leads to deterioration of domain-specific modules needed to mediate performance on these tasks. To distinguish between the two interpretations, we examined performance in young adults under conditions of divided attention.

In a second set of experiments, we used the dual-task method to provide a more direct demonstration that central processing mechanisms contribute to performance of social reasoning tasks. Specifically, we required younger participants to complete the same ToM and deontic selection tasks while performing a secondary task that relies heavily upon central processing resources (i.e., working memory). A large body of previous research demonstrates that if two concurrently performed tasks rely upon the same resource (Allport et al., 1972), the same hemisphere (e.g., Klein, 1976; Moscovitch, 1976; Moscovitch & Klein, 1980) or the same brain region (e.g., Moscovitch, 1994; Moscovitch et al., 2002), competition for shared resources will result in interference effects for one or both tasks (for reviews see, Meyer & Kieras, 1997, 1997; Ruthruff, Pashler, & Hazeltine, 2003).

In the present experiments, we asked participants to complete social reasoning tasks while performing a widely used test of working memory performance, an auditory version of the 2-back task. This task, which requires participants to determine if a currently heard tone is the 'same' as, or 'different' from, a previously heard tone, relies heavily on participants' ability not only to hold information in memory, but also to compare multiple pieces of information simultaneously. We reasoned that if performance on social reasoning tasks relies at least in part upon the same working memory resources that contribute to performance of the 2-back task, simultaneous performance of these tasks should result in interference on one or both of these tasks. Moreover, consistent with performance accounts of social reasoning (e.g., Leslie et al., 2004), performance should be affected more so on the two-perspective than one-perspective task. Alternatively, if performance on social reasoning tasks relies preferentially upon pre-established constructs (or "modules") that are impenetrable to the operation of central processing resources, little interference effects would be expected for either task.

3.1. Experiment 4

Participants in this experiment were asked to complete simultaneously two different tasks: the ToM task presented in Experiment 1 and the 2-back task. We reasoned that if the working memory demands of our ToM task were indeed high, competition for working memory resources between the *n*-back and the ToM task would result in performance decrements on both tasks when performed simultaneously. Alternatively, if the ToM task relies preferentially on pre-established cognitive constructs that operate independently of domain-general resources, we would expect little competition between the ToM and working memory task, and equivalent performance, regardless of whether these tasks were performed simultaneously or performed individually.

An additional goal of our study was to compare performance on ToM tasks with high and low working memory demands. As we noted earlier, traditional tests of ToM used with adult populations typically require participants to compare two or more perspectives; these tasks are likely to place high demands on central processing resources. In the current experiment, as with older adults in Experiment 1, we compared two different types of ToM tasks. The first required participants to consider one perspective only (i.e., first-order ToM task); the second required participants to compare two perspectives simultaneously (second-order ToM task). We expected the interfering effects of the *n*-back task to be greater for the two perspective taking task than for the one perspective taking task due to increased working memory load and competition for resources. Alternatively, if additional working memory resources are not required for tasks requiring second-order ToM reasoning, no additional interference would be expected.

The participants in our study were neurologically-intact young adults. Our goal was to examine directly the role of working memory in performance of ToM tasks that have been used previously with adult populations. Hence, we compared performance across two different groups of participants. Participants in the *dual-task condition* were required to complete simultaneously the ToM and working memory tasks. By contrast, participants in the *control condition* completed the ToM task only.

3.1.1. Method

3.1.1.1. Participants. The participants in this study were 20 members ($M = 22.4$ years of age; mean years of education: 15.35) of the University of Toronto community. Half of the participants were in the *dual-task* condition; the remaining participants formed the control group. Participants received partial course credit or nominal payment for their participation in the experiment, which took approximately 45 min. All participants had normal or corrected-to-normal vision and hearing.

3.1.1.2. Stimulus materials

3.1.1.2.1. ToM task. The ToM task was identical to that in Experiment 1, with the exception that two stories were eliminated in order to reduce the overall time to completion of the experiment.

3.1.1.2.2. Two-back task. Participants were presented with a continuous sequence of high (2250 Hz), medium (750 Hz), and low (250 Hz) tones. Tone frequencies were chosen so as to be double in psychological intensity across the three tone types. Each tone had a duration of 3000 milliseconds (ms); the silent interval between successive tones was approximately 35 ms. Participants were required to determine if the tone currently heard was the 'same' as or 'different' from the tone heard two tones previously (e.g., H M L M – 'same' versus H M L H – 'different'). Tones were presented in a random order with the constraint that no two tones of the same frequency were heard consecutively.

3.1.1.3. Apparatus

3.1.1.3.1. ToM task. The stimuli were presented using the presentation software Microsoft Power Point installed on an IBM computer. Participants used the space bar of the keyboard to initiate trials. Responses were recorded verbatim by the experimenter.

3.1.1.3.2. Two-back task. The stimuli were digital sound files created using the program Sound Edit installed on a Macintosh computer (iMac). Stimulus presentation and response recording were controlled by a customized software program connected to a Dell (*Dimension XPS T550*) computer. Stimuli were presented at a comfortable listening level via computer speaker in a quiet room.

3.1.1.4. Procedure

3.1.1.4.1. Dual-task condition. Participants were tested individually and received instructions both verbally and on the computer screen. Prior to the test session, participants completed two practice sessions. Each practice session consisted of a 1-minute session of the 2-back task followed by a longer session that involved presentation of a test story and questions along with the 2-back task. Participants were encouraged to ask questions during the practice sessions.

Following administration of the practice session, participants were required to perform the 2-back task for a period of 3 min. This provided a *pre-test* baseline measure of participants' performance on this measure that could be subsequently compared to performance when both tasks (2-back task and ToM task) were completed simultaneously.

The actual test session consisted of six ToM stories presented along with the 2-back task. Participants were required to complete both tasks simultaneously. They were instructed not to focus their attention on one task only, but rather to attempt to complete each task as successfully as possible.

Participants read each story silently. They were encouraged to read the story carefully and to take as much time as necessary to understand its contents fully. The story remained on screen during presentation of the test questions. In order to reduce potential demands on working memory, participants were encouraged to refer back to these stories while forming their responses. Questions were presented individually on the computer screen. Responding was self-paced; participants pressed the 'space bar' of the keyboard to initiate the visual display. Participants read each question silently and then provided an oral response. Responses to the questions were recorded verbatim by the experimenter. No feedback was given regarding correct or incorrect responses.

The 2-back task was presented auditorily On each trial, participants heard one of three different tones and judged whether the tone was the 'same' as or 'different' from the tone heard two tones previously. Responding was self-paced; participants indicated their response by pressing a key ('1' for 'same'; '2' for 'different') on the computer keyboard. Participants could respond at any time during presentation of the test tone; responding terminated the presentation of this tone. No feedback was given regarding correct or incorrect responses.

Following completion of the test session, participants completed a *post-test* measure of 2-back performance; they performed this task alone for 3 min.

3.1.1.4.2. Control condition. The procedure was *identical* for the control condition, with the exception that during the actual test session, participants were required to complete the ToM task only.

3.1.2. Results

3.1.2.1. ToM task. As expected, simultaneous performance of a working memory task interfered with performance of the ToM task. Moreover, the effect of this interference was greater for second-order questions (relative to first-order questions) that required participants to compare simultaneously multiple perspectives (see Fig. 4).

The results were analyzed using a 2×2 mixed-design ANOVA. Group (dual-task versus control condition) was treated as a between-subjects factor. Question type (first-order versus second-order questions) was treated as a within-subjects factor. There was a main effect of group [$F(1,18) = 25.60$, $p < .001$]; performance was better in the control as compared to dual-task condition. There was also a main effect of question type [$F(1,18) = 45.97$, $p < .001$]; participants exhibited superior performance for first-order as compared to second-order ToM questions. The interfering effects of the working memory task appeared mediated by question type; there was an interaction between group and question type [$F(1,18) = 26.97$, $p < .001$]. Although the effect was greater for the second-order questions, simultaneous performance of a working memory task interfered with performance of both first-order [$t(9) = -2.28$, $p < .05$]. and second-order [$t(9) = -5.76$, $p < .001$] ToM questions. Moreover, dual-task conditions resulted in greater interference for second-order as compared to first-order ToM questions [$t(9) = -6.51$, $p < .001$]; performance on

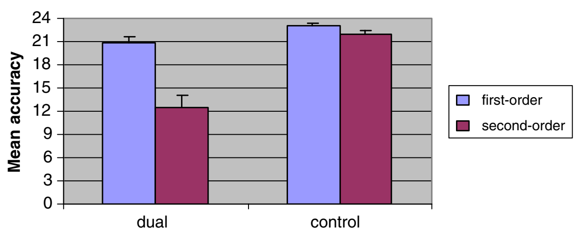

Fig. 4. Mean accuracy across task conditions and ToM question types.

first-order as compared to second-order ToM questions did not differ in the control condition ($p > .05$).

Errors were divided further into descriptive categories in order to identify the locus of response errors on ToM questions in the dual-task and control conditions. For second-order ToM questions, five categories of errors were identified: (i) those that involved adopting the wrong perspective (that of a single person only, rather than considering both perspectives simultaneously) – 15.32% of errors in dual-task condition (17 errors total); 9.5% of errors in control condition (2 errors total); (ii) errors that involved a failure to consider the different types of information held in mind by the two characters described in the story (e.g., known information that differed across the parties) – 16.21% of errors in dual-task condition (18 errors total); 0% of errors in control condition; (iii) incorrect attributions of mental states – 18.9% of errors in dual-task condition (21 errors total);; 43.5% of errors in control condition (10 errors total); (iv) non-mental explanations – 9% of errors in dual-task condition (10 errors total); 14.2% of errors in control condition (3 errors total); (v) failure to respond (i.e., 'I don't know') – 41% of errors in the dual-task condition (45 errors total); 28.5% of errors (6 errors total) in control condition. Hence, more than 30% of the total errors committed in the dual-task condition involved a failure to consider multiple pieces of information simultaneously either by adopting a single perspective only or by failing to consider the differing information held in mind by the two parties described in the social scenarios; another 30% of the errors involved an incorrect attribution or a failure to evoke a mental state attribution for characters' responses. Finally, a larger percentage of participants in the dual-task condition as compared to control condition (16% more participants) were unable to provide an explanation of the thoughts or feelings of characters described in the stories. Interestingly, nearly 50% of the 'I don't know' category of errors were committed by two subjects (48.8% of these errors or 22/45 errors), indicating that the dual-task condition did not simply overwhelm the majority of participants' cognitive resources such that responding was not possible under these strenuous conditions.

Three categories of errors were identified for the single-perspective ToM questions: (i) incorrect mental state attributions – 66.67% of errors in the dual-task condition (20 errors in total); 100% of errors in the control condition (10 errors) (ii) failure to respond (i.e., 'I don't know') – 33.33% of errors in the dual-task condition (10 errors total); 0% of errors in control condition; (iii) non-mental explanations – 0% of errors in the dual-task and control conditions.

3.1.2.2. Two-back task. Simultaneous performance of the ToM task also interfered with performance of the 2-back task. Specifically, simultaneous performance of the ToM task had the effect of lowering accuracy and increasing RT in the dual-task condition when performance was compared to our baseline measures.

3.1.2.2.1. Accuracy. We analyzed these results using a one-way repeated-measures ANOVA treating condition (pre-test versus experimental session versus post-test) as the factor of interest (see Fig. 5). There was a main effect of condition [$F(2,8) = 21.35$, $p < .001$]. Post hoc testing confirmed that participants were less accurate in the dual-task experimental condition as compared to both the baseline

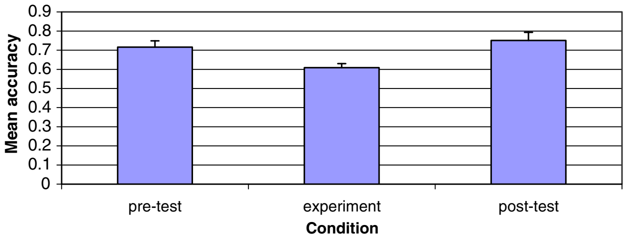

Fig. 5. Mean accuracy on the 2-back task across exposure conditions.

[$t(9) = 3.11$, $p < .02$] and post-test [$t(9) = 3.91$, $p < .01$] conditions. Accuracy did not differ between the pre- and post-test conditions (ps < .05).

3.1.2.2.2. Reaction time. We also analyzed these results using a one-way repeated-measures ANOVA treating condition (pre-test versus experimental session versus post-test) as the factor of interest (see Fig. 6). There was a main effect of condition [$F(2,8) = 11.35$, $p < .01$]. Whereas response time did not differ between the pre-test and experimental condition ($p > .05$), response times were faster in the post-test as compared to experimental condition [$t(9) = 2.62$, $p < .03$]. Response times were also faster in the post- as compared to pre-test condition [$t(9) = 4.92$, $p = .001$].

3.1.3. Summary – Experiment 4

There were two main findings in this experiment. The first concerns the role of domain-general processing resources in ToM reasoning. Consistent with our indirect findings concerning the poor performance of older adults on resource demanding ToM tasks, simultaneous performance of a 2-back task and a ToM task resulted in performance decrements across on both tasks tested in this dual-task paradigm. These results provide a direct demonstration that central processing resources (i.e., working memory) contribute to performance of the type of ToM task tested in our studies.

A secondary goal of this experiment was to determine if the demands placed on working memory are greater for ToM tasks that require participants to integrate multiple perspectives (i.e., second-order ToM questions; What does X think Y

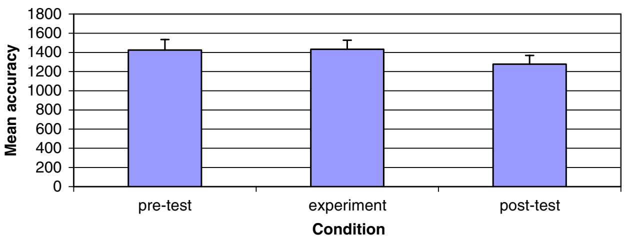

Fig. 6. Mean reaction times on the 2-back task across exposure conditions.

thinks/feels?) as compared to those that require participants to consider one perspective only (i.e., first-order ToM questions; What does X think/feel?). In Experiment 1, older adults showed performance decrements for second- (but not first-) order ToM tasks. Consistent with this finding, in the present experiment, the interference observed in the dual-task condition was greater for those second-order ToM questions that required participants to integrate multiple perspectives and, hence, placed greater demands on working memory, than for the first-order condition. This result is consistent with performance accounts of social reasoning (e.g., Leslie et al., 2004) that suggest the contribution of central processing resources is greater for those ToM tasks involving high EF demands.

Although the magnitude of interference, as predicted, was far greater for the second-order ToM questions, interference was also observed on presumably less difficult first-order questions. This deficit is not attributable to a failure to hold information in mind over time; the ToM stories were available for inspection while questions were being answered and the ToM questions themselves were displayed on screen while participants formed their responses. Instead, the results suggest that operating on this information to make ToM judgments requires general cognitive resources, and more so for the two-perspective task than for the one-perspective task.

Although we observed a high number of non-responses for our second-order dual-task ToM questions, it is unlikely that this pattern of responding stemmed simply from participants being so overwhelmed by the competing cognitive demands that they abandoned the ToM task entirely; nearly 50 percent of these "I don't know" errors were committed by two subjects while the remainder of subjects committed relatively few of these errors in the same condition. Indeed, these same two subjects made many of the same type of errors for the first-order ToM tasks. The remaining second-order ToM errors were attributable primarily to responses that relied on an understanding of one person's perspective only (misinformation or wrong perspective) or an incorrect or non-mental attribution.

In a second experiment, we turn to an alternate type of social reasoning task, deontic reasoning, to determine if the same pattern of findings will extend to a different task which shares structural similarities with typical ToM tasks.

3.2. Experiments 5

In this experiment, we asked our participants to complete simultaneously two different tasks: the same social contract or precaution versions of the deontic selection task tested in older adults and the *n*-back task. If as our experiments with older adults suggested, deontic reasoning depends directly on the same domain-general processing resources that mediate performance on the *n*-back task, then performance should suffer on both tasks under dual-task conditions.

As in our dual-task ToM experiment, we reasoned that, if the working memory demands of social versions of the deontic tasks are substantial, then competition for working memory resources between the *n*-back and a deontic task would result in performance decrements on both tasks when performed simultaneously. Moreover, competition for central processing resources (i.e., working memory) would presumably eliminate any performance benefit for deontic tasks requiring the detection of cheaters or the avoidance of hazardous conditions; performance should be similar to that for those deontic tasks that are descriptive or abstract in nature. Alternatively, if "hazardous" and "cheater" versions of the task rely strictly (or preferentially) upon pre-existing constructs (or "modules"), they should, by definition, place few demands on general working memory resources. Indeed, performance should vary little between the full and divided attention conditions and maintain the benefits for reasoning about the detection of cheaters or the avoidance of hazardous conditions as compared to the descriptive versions of the task.

The participants in our study were neurologically-intact young adults. Our goal was to examine directly the role of working memory in performance of deontic selection tasks that have been used previously with adult populations. Hence, we compared performance across two different groups of participants. Participants in the *dual-task condition* were required to complete simultaneously the deontic and working memory tasks. By contrast, participants in the *control condition* completed the deontic task only; this condition was included to replicate earlier findings indicating that college-aged students exhibit performance benefits for deontic tasks requiring the detection of cheaters or the avoidance of hazardous conditions.

3.2.1. Method

3.2.1.1. Participants. The participants in this study were 32 members ($M = 20.52$ years of age; mean years of education: 14.00) of the University of Toronto community. Half of the participants were in the dual-task condition; the remaining participants formed the control group. Participants received partial course credit or nominal payment for their participation in the experiment, which took approximately 45 min. All participants had normal or corrected-to-normal vision and hearing.

3.2.1.2. Stimulus materials

3.2.1.2.1. Deontic selection task. Social contract version. The materials in this experiment were the same as those used in an earlier experiment by Cosmides (1989) and in Experiment 2.

Precaution version. The materials in this experiment were the same as those used in an earlier experiment by Fiddick et al. (2000) and in Experiment 3.

3.2.1.2.2. Two-back task. The 2-back task was the same as that used in the ToM dual-task experiment (Experiment 4).

3.2.1.3. Apparatus

3.2.1.3.1. Deontic selection task. Dual-task condition. The stimuli were presented using the presentation software Microsoft Power Point installed on an IBM computer. The stimuli were presented on two computer screens. The first screen comprised the unfamiliar scenario; the second screen displayed the four selection cards. Participants used the space bar of the keyboard to initiate trials. Responses were recorded verbatim by the experimenter.

Control condition. The stimuli were presented on two sheets of paper. The first sheet comprised the unfamiliar scenario; the second sheet displayed the four selection cards. Responses were recorded verbatim by the experimenter.

Two-back task. The stimuli were digital sound files created using the program Sound Edit installed on a Macintosh computer (iMac). Stimulus presentation and response recording were controlled by a customized software program connected to a Dell *(Dimension XPS T550)* computer. Stimuli were presented at a comfortable listening level via computer speaker in a quiet room.

3.2.1.4. Procedure

3.2.1.4.1. Dual-task condition. Participants were tested individually and received instructions both verbally and on the computer screen. Prior to the test session, participants completed two practice sessions. Each practice session consisted of a 1-min session of the 2-back task followed by a longer session that involved presentation of a short story describing complex social situations and a series of related questions along with the 2-back task. Participants were encouraged to ask questions about the procedure during the practice sessions. Otherwise, the procedure mirrored that of Experiment 4.

The practice session was followed by two test sessions. The actual test sessions consisted of presentation of one of the deontic selection tasks along with the 2-back task. Half of the participants completed the social contract and precaution versions of the deontic selection task; the remaining half of participants completed the descriptive versions of these tasks. The order of administration of the social contract/descriptive task and precaution/descriptive tasks was counterbalanced across participants.

Participants were required to complete both the deontic and 2-back tasks simultaneously. They were instructed not to focus their attention on one task only, but rather to attempt to complete each task as successfully as possible.

Participants read each of the deontic scenarios silently. They were encouraged to read the story carefully and to take as much time as necessary to understand its contents fully. Responding was self-paced. Following presentation of the scenario, participants pressed the 'space bar' of the keyboard to initiate a visual display of the 4 hypothetical selection cards. Participants viewed these screens silently and then provided an oral response. Responses selections were recorded verbatim by the experimenter. No feedback was given regarding correct or incorrect responses.

Following the post-test, we again presented dual-task participants with the deontic selection tasks used in the test session. Participants were required to reread the scenarios tested and to choose new (or the same) stimulus cards based upon their rereading of the text.

3.2.1.4.2. Control condition. Participants in the control condition completed the deontic selection tasks only. Administration of this task was identical to the dual-task condition with the exception that the tasks were presented on sheets of paper, rather than on a computer screen. In order to avoid order of administration effects, presentation of the deontic tasks was counter-balanced.

3.2.2. Results

3.2.2.1. Deontic selection tasks

3.2.2.1.1. Social contract version. As predicted, simultaneous performance of a working memory task eliminated performance differences between the descriptive versions of the deontic task and the social contract and precaution versions. Specifically, participants in the dual-task condition exhibited equal levels of performance on the social contract and precaution versions as compared to descriptive versions of this task. By contrast, as in previous studies (e.g., Cosmides, 1989), participants in the control condition exhibited superior levels of performance for the social contract and hazardous versions as opposed to descriptive version of this task.

The scoring procedure was identical to that used in Experiment 3. As Fig. 7 indicates, only a small percentage of participants in the dual-task condition were able to select correctly the two cards required to answer the problem posed in the stories. Our analysis failed to reveal any advantage for the social contract/precaution versions of the deontic task in the dual-task conditions; performance, however, was better for the social contract and precaution versions of the task in the control conditions.

These results were confirmed using 2×2 ANOVAs treating both condition (dual-task versus control) and task version (social contract versus descriptive and precaution versus descriptive) as between-subjects factors. This analysis revealed main effects of condition type: social contract $[F(1,28) = 8.91,\ p < .01]$ and precaution $[F(1,28) = 7.63,\ p < .02]$; performance was better in the control as compared to dual-task conditions. There was also a main effect of stimulus type for the social contract $[F(1,28) = 6.55, p < .02]$ session; performance was better for social contract as compared to descriptive version of the deontic task. The main effect of stimulus type was not significant ($p > .05$) in the precaution session. This analysis, however, confirmed that the effect of stimulus type was mediated by type of condition tested (significant condition by stimulus type interaction) for both the social contract $[F(1,28) = 8.91,\ p < .01]$ and precaution $[F(1,28) = 5.47,\ p < .05]$ sessions. Post hoc analyses failed to reveal an effect of stimulus type for participants in the social contract and precaution dual-task sessions ($p > .05$). By contrast, the effect of stim-

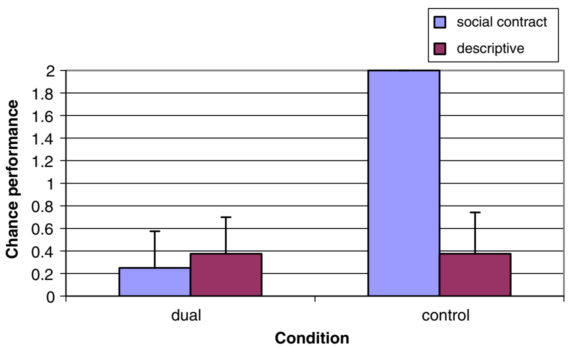

Fig. 7. Social contract version: Mean performance across task conditions and stimulus types.

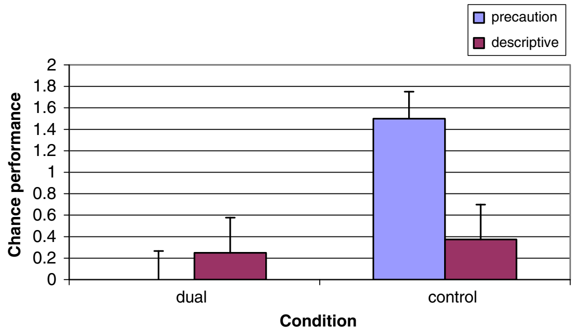

Fig. 8. Precaution version: Mean performance across task conditions and stimulus types.

ulus type was highly significant for participants in the control condition of the social contract [$F(1,28) = 15.81, p < .001$] and precaution [$F(1,28) = 7.31, p < .02$] sessions (see Fig. 8).

Interestingly, 50 percent of participants in the social contract version of the dual-task condition succeeded in identifying correctly the appropriate stimulus cards when the scenarios were presented to them individually (without the 2-back task) following the test session; only 25 percent of participants in the descriptive condition succeeded in successfully identifying these cards upon second individual presentation. Indeed, 62.5 percent of participants in the social contract version performed above chance in this selection phase as compared to only 25 percent of participants in the descriptive version of the task. These results were identical for the precaution versus descriptive conditions of the experiment.

3.2.2.2. Two-back task

3.2.2.2.1. Social contract accuracy. There was some evidence that simultaneous performance of a social contract/descriptive version of the deontic selection task resulted in decreased performance on the 2-back task relative to our post-test measure of baseline performance. Specifically, 2-back accuracy was lower in the dual-task condition (social contract and descriptive versions) as compared to accuracy for the post-test measure of 2-back performance (see Fig. 9).

We confirmed these results using a 2×3 mixed-design ANOVA treating deontic condition (social contract versus descriptive) as a between-subjects variable and exposure condition (pre-test versus dual-task versus post-test) as a within-subjects factor. There was no effect of condition type ($p > .05$); performance on the 2-back task was equivalent across the social contract and descriptive versions of the task. There was, however, a main effect of effect of exposure condition, $F(2,28) = 4.77$, $p < .02$. Post hoc tests indicated that 2-back accuracy was higher in the post-test as compared to dual-task condition, $t(15) = 2.85$, $p < .02$; no significant difference in response accuracy was found between the pre-test and dual-task conditions $p > .05$. Application of the Bonferonni correction [experiment-wise error rate divided by a priori number of tests of interest ($.05/2 = .025$)] did not alter this pattern of find-

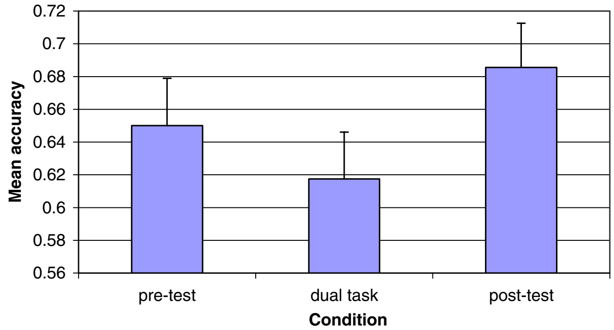

Fig. 9. Social contract task: Mean accuracy on the 2-back task across exposure conditions.

ings. The interaction between deontic condition and exposure condition failed to achieve significance ($p > .05$).

3.2.2.2.2. Social contract reaction times. Simultaneous performance of a social reasoning task resulted in longer reaction times (correct responses only) in the dual-task condition as compared to the post-test measure of 2-back performance. Specifically, 2-back reaction times were longer in the dual-task condition (social contract and descriptive versions) as compared to reaction times in the post-test measure of 2-back performance (see Fig. 10).

We confirmed these results using a 2×3 mixed-design ANOVA treating deontic condition (social contract versus descriptive) as a between-subjects variable and exposure condition (pre-test versus dual-task versus post-test) as a within-subjects factor. There was no effect of condition type ($p > .05$); performance on the 2-back task was equivalent across the social contract and descriptive versions of the task. There was, however, a main effect of effect of exposure condition, $F(1.52, 21.27) = 4.53$, $p < .04$, with Huyn-Feldt correction. Post hoc tests indicated that reaction times were shorter in the post-test as compared to dual-task condition, $t(15) = 2.52$, $p < .03$; no significant difference in reaction time was found between the pre-test and dual-task conditions $p > .05$. Application of the Bonferonni correction [experiment-wise error rate divided by a priori number of tests of interest (.05/

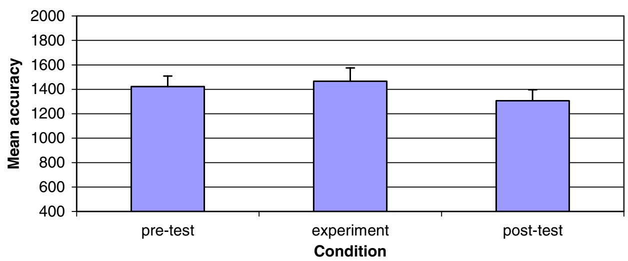

Fig. 10. Social contract task: Mean reaction time on the 2-back task across exposure conditions.

2 = .025)] did not alter this pattern of findings. The interaction between deontic condition and exposure condition failed to achieve significance ($p > .05$).

3.2.2.2.3. Precaution accuracy. There was some evidence that simultaneous performance of a precaution/descriptive version of the deontic selection task resulted in decreased performance on the 2-back task relative to our pre- and post-test measures of baseline performance. Specifically, 2-back accuracy was lower in the dual-task condition (social contract and descriptive versions) as compared to accuracy for the pre- and post-test measures of 2-back performance (see Fig. 11).

We confirmed these results using a 2×3 mixed-design ANOVA treating deontic condition (precaution versus descriptive) as a between-subjects variable and exposure condition (pre-test versus dual-task versus post-test) as a within-subjects factor. There was no effect of condition type ($p > .05$); performance on the 2-back task was equivalent across the precaution and descriptive versions of the task. There was, however, a main effect of effect of exposure condition, $F(2, 28) = 9.73$, $p < .01$. Post hoc tests indicated that 2-back accuracy was higher in the pre- [$t(15) = 2.53$, $p < .03$] and post-test [$t(15) = 4.41$, $p < .001$] conditions as compared to dual-task condition. Application of the Bonferonni correction [experiment-wise error rate divided by a priori number of tests of interest (.05/2 = .025)] did not alter this pattern of findings. The interaction between deontic condition and exposure condition failed to achieve significance ($p > .05$).

3.2.2.2.4. Precaution reaction times. Reaction times on the 2-back task (correct responses only) did not differ between the dual-task condition and pre- and post-test measures of baseline performance (see Fig. 12).

We confirmed these results using a 2×3 mixed-design ANOVA treating deontic condition (precaution versus descriptive) as a between-subjects variable and exposure condition (pre-test versus dual-task versus post-test) as a within-subjects factor. There was no effect of condition type ($p > .05$); performance on the 2-back task was equivalent across the social contract and descriptive versions of the task. Reaction times did not differ between the different exposure conditions ($p > .05$). The interaction between deontic condition and exposure condition also failed to achieve significance ($p > .05$).

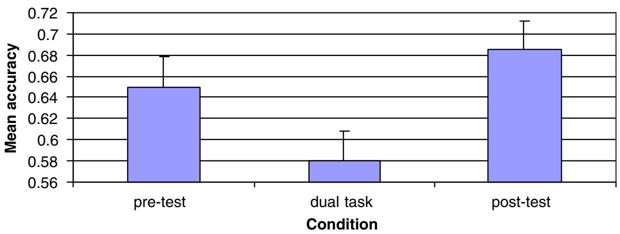

Fig. 11. Precaution version: Mean accuracy on the 2-back task across exposure conditions.

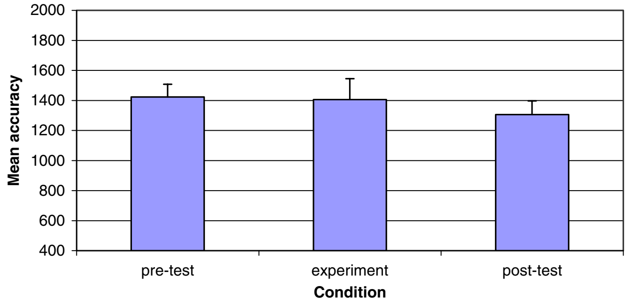

Fig. 12. Precaution version: Mean reaction times on the 2-back task across exposure conditions.

3.2.3. Experiment 5 – Summary

These results extend our initial finding that performance on social reasoning tasks (i.e., a ToM task; Experiment 4) is impaired under conditions involving competition for shared working memory resources. The main finding in Experiment 5 was that dual-task conditions resulted in decrements on both versions of the deontic tasks tested. Moreover, no differences in performance were observed between versions of the task that required the detection of cheaters (Cosmides, 1989) or the avoidance of hazardous conditions (Fiddick et al., 2000) and versions of the task that were abstract or descriptive in nature under dual-task conditions. We also found evidence that simultaneous performance of deontic selection tasks interfered with performance of our working memory task. These result is consistent with our earlier finding that older adults performed worse than younger adults on social versions of the deontic task and provides a direct demonstration of the contribution of central processing resources to task performance. Indeed, it appears that a domain-general mechanism, working memory, contributes to performance on the deontic selection task, no matter what the type. Finally, these results confirm our view that structural similarities among social reasoning tasks result in a pattern of impaired performance under conditions that tax the working memory resources upon which these tasks rely.

4. General discussion

We wish to highlight a number of findings in our experiments. Using a variety of different methods, our results provide novel evidence that domain-general resources, such as working memory, contribute to performance on two different social reasoning tasks, theory of mind (ToM) and social versions of the deontic selection task. Specifically, both older adults and younger adults under conditions of divided attention showed performance decrements on these tasks. Our results are consistent with previous demonstrations illustrating the penetrability of social reasoning tasks to central processing resources that, when in operation, preclude obligatorily output of potentially modular components of task performance. These central processing resources appear requisite to performance on social reasoning tasks in older adults and in younger adults without disorder, suggesting that EFs contribute to social reasoning beyond a period of initial belief construction in early childhood (e.g., Moses, 2001; Perner & Lang, 1999). Moreover, consistent with performance accounts of social reasoning (see, for example, Leslie et al., 2004), performance varied with the demands these tasks placed on EFs (e.g., first- versus second-order ToM tasks). Finally, our observation that central processing resources impact on performance of both ToM and social versions of the deontic task highlights the shared nature of responding on these two different social reasoning tasks that, until now, have been described as largely dissimilar.

4.1. Performance versus emergence accounts of social reasoning

Our results are broadly consistent with performance accounts (Leslie et al., 2004) of social reasoning that emphasize the on-going contribution of EFs to social reasoning tasks, provided that task demands are such that central processing resources are required (e.g., ToM tasks requiring multiple levels of inhibition). This view contrasts with emergence accounts (e.g., Moses, 2001; Perner & Lang, 1999) that emphasize the role of executive functions in the initial construction of social beliefs; once constructed, EFs are thought unnecessary for the expression of these beliefs. If performance on social reasoning tasks no longer requires EFs, and, with age, operates in what some authors have described as a preferentially modular fashion (e.g., Baron-Cohen, 1995; Cosmides, 1989), these tasks should not be affected by deterioration or damage to central processing resources that presumably have little or no involvement in task performance.

In Experiment 1, however, we presented evidence showing that older adults perform poorly on a ToM task. Specifically, relative to younger adults, normally aging older adults were impaired on second-order ToM tasks that required them to integrate competing perspectives by answering questions about the thoughts and feelings of two characters described in complex social scenarios; no such deficits were observed for first-order ToM questions where participants had to consider the thoughts and feelings of one person only. These results are in line with earlier findings that older adults are impaired on ToM tasks with high central processing demands (e.g., Maylor et al., 2002).

When older adults were tested on a second social reasoning task, deontic reasoning, which appears to involve similar EF demands as ToM tasks, a similar pattern of deficit emerged. Specifically, in two experiments (Experiments 2 and 3), we demonstrated selective impairment in older adults on two versions of the deontic task (social contract and precaution) on which younger adults have previously been shown to demonstrate benefits for social reasoning. These findings are in line with earlier demonstrations that older adults perform poorly on ostensibly unfamiliar (and presumably more difficult cf. Pollack et al., 1995) versions of deontic selection tasks (Overton et al., 1986) where domain-general processing requirements also appeared to affect task performance. If performance on these tasks relies preferentially on modular components that no longer require

EFs for their expression, one would expect a preservation of performance with cognitive aging. The opposite proved to be true in our experiments, however, with declines in central processing resources appearing to contribute to an overall pattern of impairment on deontic tasks, even when social reasoning versions of the task were tested.

Our results are best explained by performance accounts (Leslie et al., 2004) of social reasoning that predict performance deficits in older adults with executive dysfunction, particularly when these tasks involve a high level of executive processing demands. Indeed, older adults' performance on the ToM task was impaired only for the second-order task requiring the greatest amount of central processing resources. Here, limits placed on social reasoning by executive functioning appear to precede the formulation of other people's desires and beliefs, restricting their expression when performance demands overwhelm the processing capacity of the respondent.

Although these results are consistent with the notion that well-documented declines in EFs among older adults contribute to the pattern of poor performance observed on resource-demanding versions of these social reasoning tasks, the evidence for this contribution remains incomplete. Equally plausible is the suggestion that general deterioration accompanying aging affects modules specialized for social reasoning. Although we consider this unlikely, given that those social reasoning tasks requiring the greatest amount of EFs (e.g., second-order ToM) were most impacted by aging, we nonetheless set out to provide a direct demonstration of the role of central processing resources in social reasoning.

Here, we turn to our results from the dual-task interference experiments. In a pattern broadly consistent with that which we observed in older adults, young adults under divided attention showed impairment on both ToM and social versions of the deontic selection task (Experiments 4 and 5). Moreover, in each case, interference was also observed on the working memory task performed simultaneously. As for older adults, the magnitude of this interference was greatest when the task was presumably more complex and drew more heavily on EFs, as is the case for first- versus second-order ToM. We believe these interference effects stem from competition between the working memory and social reasoning tasks for the shared central processing resources upon which they rely. These results provide a *direct* demonstration of the contribution of domain-general resources to social reasoning that is complementary to the indirect evidence we present from older adults. Furthermore, the results of these dual-task experiments provide additional evidence that EFs continue to contribute to social reasoning in younger adults well past a period of early belief formation in childhood (Leslie et al., 2004, 2005).

4.2. Shared resources for social reasoning tasks

Interference by central processing resources on both the ToM task and deontic tasks, and our finding of impairment in older adults and young adults under DA, is strongly indicative of the reliance of both types of social reasoning tasks on shared resources. These studies represent the first demonstration of this association. Given this evidence, we cannot support the view advanced by some proponents of performance accounts of ToM reasoning (Leslie et al., 2005) that suggests "theory of mind" executive functions can be wholly or partially dissociated from domain-general executive functions, where executive functions required for social reasoning are thought specific to the ToM domain. We believe our work speaks to this account, where the executive functioning demands we placed on younger adults under divided attention were designed to be domain-general, rather than specific to ToM. Here, domain-general processing demands interfered with performance on both the deontic reasoning and ToM task.

These indirect (older adults) and direct (younger adults) demonstrations of the role of working memory in ToM performance lend strong support to an emerging body of work suggesting that domain-general resources, including, but certainly not limited to, working memory, contribute to performance on ToM tasks. For example, developmental work suggests that the appearance of ToM skills in young children is correlated with the emergence of numerous executive functions, including working memory (Gordon & Olson, 1998) and children's ability to switch their perspective from one focus to another (Frye, Zelazo, & Palfai, 1995). Moreover, recent work (Channon & Crawford, 2000) demonstrating that selective ToM deficits in patients with frontal lobe damage are correlated with their performance on other tasks with working memory demands, including the Raven's progressive matrices task, is also in line with our findings. Our observation in Experiment 5 of interference on both the social contract and precaution version of the deontic reasoning tasks, where previously observed benefits for social reasoning disappeared under dual-task conditions, is also consistent with growing claims in the research literature that domain-general requirements may influence performance of deontic reasoning tasks (e.g., Almor & Sloman, 2000; Oaksford, Morris, Grainger, & Williams, 1996).

Evidence for additional EF contributions to social reasoning is particularly strong in development, where numerous studies demonstrate that ToM development is correlated with a host of other domain-general functions including, but certainly not limited to, inhibitory control (Flynn, O'Malley, & Wood, 2004; Ozonoff et al., 1991) and mental set shifting (Frye et al., 1995). Similarly, deontic reasoning has been shown to rely on multiple domain-general resources (e.g., Almor & Sloman, 1996) including attention (Oaksford et al., 1996) and the ability to hold in mind an information set (Almor & Sloman, 2000). These resources may act in tandem with modular components of ToM (e.g., Leslie et al., 2004) and social reasoning, perhaps overwhelming the operating capacity of such "modules" when sufficient executive functioning demands exist. More rigorous tests, involving careful identification of the cognitive demands of the wide variety of social tasks reported in the literature, however, will be required to identify accurately and to dissociate domain-general and domain-specific contributions to social reasoning tasks.

4.3. Limits placed on modular components of social reasoning

Our results speak to the non-modular components of performance on social reasoning tasks, and, in no way, negate the possibility that some components of social reasoning tasks are, in fact, modular. Indeed, work is underway in numerous laboratories to identify potentially modular and non-modular aspects of social reasoning. In the meantime, our results, contribute to the existing literature setting boundary conditions on what it means for social reasoning to be modular by suggesting that domain-general resources play a crucial role in task performance. Specifically, our experiments suggest that, when central processing demands are sufficient, their operation may overwhelm at least two attributes that have been previously ascribed to modules (Fodor, 1983; Moscovitch & Umilta, 1990): automaticity and cognitive impenetrability. In our experiments, social reasoning performance was affected by deterioration of or competition for central processing resources that presumably would have little or no involvement in task performance if its execution were mediated only by the modular components of social reasoning. This view is complementary to the elegant set of constraints already posited in the developmental literature (e.g., Leslie et al., 2004) where both modular and non-modular components are thought to underlie task performance. We have no reason to believe that these modular components are not preferentially active when central processing demands are low. Indeed, our results cannot exclude the possibility that select modular components of social reasoning (e.g., emotion comprehension; see Winston, Strange, O'Doherty, & Dolan, 2002 for evidence of the automaticity of this function) do not operate in tandem with non-modular components even under high levels of executive functioning.

Interestingly, one caveat to our findings is that although older adults were impaired relative to younger adults on the precaution version of the deontic task, their overall performance on both the descriptive and precaution versions of the task combined was only marginally impaired relative to younger adults, a finding stemming in part from the overwhelming difficulty of the descriptive task for both older and younger participants. Future tests are awaited to determine if, for example, differences in central processing demands between the social contract version of the task (on which older adults were significantly impaired relative to younger adults) and precaution version of the task contribute to the pattern of somewhat differential performance observed in our aging population.

We also observed some evidence of a selective sparing of performance on the social contract as opposed to descriptive version of the deontic task in older adults. Specifically, older adults exhibited a small performance benefit for the social contract as opposed to descriptive version of the task. Here, we speculate that the enhanced familiarity (detection of cheaters versus anthropological investigation) of the social contract relative to descriptive version of the task may have allowed participants to 'schematize' more readily the information contained in the social contract story. A resulting reduction in the amount of information to be held in mind and operated on in the social contract version of the task would, in essence, free up working memory resources for use in the social contract version of the selection task and contribute to the slightly enhanced performance observed in older adults on this version of the task. Future studies may address the potential contribution of such schematization to task performance by manipulating the demand characteristics of different versions (e.g., increasing the familiarity of descriptive versions) of the task to allow for more ready schematization and by providing participants with explicit instructions to try to create schemas from the information presented in the stories. This performance benefit may also be related to pre-existing cognitive bias, formed in early childhood, or innately specified, that give a slight boost to performance, even under cognitively taxing conditions.

On balance, we believe that additional studies involving other social reasoning paradigms and different subject populations may reveal further whether potentially modular aspects of social reasoning act in tandem (or in isolation) with domain-general resources, such as working memory, inhibition and attention, to satisfy the multiple processing requirements of typical social reasoning tasks. Moreover, it will be important to determine whether those modular aspects of social reasoning performance exhibit other characteristics previously ascribed to modules such as innateness and rapidity.

Appendix A. Experiment 1

A.1. Social contract version

You are a Kaluame, a member of a Polynesian culture found only on Maku Island in the Pacific. The Kaluame have many strict laws that must be enforced, and the elders have entrusted you with enforcing them. To fail would disgrace you and your family.

Among the Kaluame, when a man marries, he gets a tattoo on his face; only married men have tattoos on their faces. A facial tattoo means that a man is married, an unmarked face means that a man is a bachelor.

Cassava root is a powerful aphrodisiac – it makes the man who eats it irresistible to women. Moreover, it is delicious and nutritious – and very scarce.

Unlike cassava root, molo nuts are very common, but they are poor eating – molo nuts taste bad, they are not very nutritious, and they have no other interesting 'medicinal' properties.

Although everyone craves cassava root, eating it is a privilege that your people closely ration. You are a very sensual people, even without the aphrodisiacal properties of cassava root, but you have very strict sexual rules. The elders strongly disapprove of sexual relations between unmarried people and particularly distrust the motives and intentions of bachelors.

Therefore, the elders have made laws governing rationing privileges. The one you have been entrusted to enforce is as follows:

"If a man eats cassava root, then he must have a tattoo on his face".

Cassava root is so powerful an aphrodisiac that many men are tempted to cheat on this law when the elders are not looking. The cards below have information about four young Kaluame men sitting in a temporary camp; there are no elders around. A tray filled with cassava root and molo nuts has just been left for them. Each card represents one man. One side of the card tells which food a man is eating and the other side tells whether or not the man has a tattoo on his face.

Your job is to catch the men whose sexual desires might tempt them to break the law – if any get past you, you and your family will be disgraced. Indicate only those card(s) **you definitely need to turn over to see if any of these Kaluame men are breaking the law.**

A.2. Descriptive version

You are an anthropologist studying the Kaluame people, a Polynesian culture found only on Maku Island in the Pacific. Before leaving for Maku Island you read a report that says some Kaluame men have tattoos on their faces, and that they eat either cassava root or molo nuts, but not both. The author of the report, who did not speak the language, said the following relation seemed to hold:

"If a man eats cassava root, then he must have a tattoo on his face".

You decide to investigate your colleague's particular claim. When you arrive on Maku Island, you learn that cassava root is a starchy staple food found on the south end of the island. Molo nuts are very high in protein, and grow on Molo trees, which are primarily found on the island's north shore.

You also learn that bachelors live primarily on the north shore, but when men marry, they usually move to the south end of the island. When a Kaluame man marries, he gets a tattoo on his face; only married men have tattoos on their faces. A facial tattoo means that a man is married, an unmarked face means that a man is a bachelor. Perhaps men are simply eating foods which are most available to them.

The cards below have information about four Kaluame men sitting in a temporary camp at the center of the island. Each man is eating either cassava root or molo nuts which he has brought with him from home. Each card represents one man. One side of a card tells what food a man is eating and the other side tells whether or not a man has a tattoo on his face.

The rule laid out by your colleague may not be true; you want to see for yourself. Indicate only those card(s) **you definitely need to turn over to see if any of the Kaluame men are breaking the rule.**

Selection cards

Tatoo
No Tatoo
Eats Cassava Root
Eats Molo Nuts

Appendix B. Experiment 2

B.1. Hazard version

You are a Kalama tribesman. While you are away on a hunting trip, some anthropologists visited your village. The anthropologists often bring gifts for your tribe and this time you notice they brought and left several bright orange jackets. You can't quite figure out why the anthropologists brought so many of the jackets so you ask one of your fellow tribesman "What are these for?" He tells you "If you go hunting, then you wear these jackets to avoid being shot".

You think the jackets are a great idea, but you are concerned that some of the other tribesman might not know about the jackets and are needlessly endangering themselves. You decide to watch what some of them do. The cards below represent four tribe members that you watched. Each card represents one person. One side of the card tells whether or not the person went hunting, and the other side of the card tells whether or not that person wore an orange jacket. Indicate only those card(s) **you definitely need to turn over to see if any of these people are needlessly endangering themselves.**

B.2. Descriptive version

You are a Kalama tribesman. While you are away on a hunting trip, some anthropologists visited your village. The anthropologists often bring gifts for your tribe and this time you notice they brought and left several bright orange jackets. You can't quite figure out why the anthropologists brought so many of the jackets so you ask one of your fellow tribesman "What are these for?" He tells you "If you go hunting, then you wear these jackets to avoid being shot".

You are not sure if what he said is true so you decide to watch what some of them do. The cards below represent four tribe members that you watched. Each card represents one person. One side of the card tells whether or not the person went hunting, and the other side of the card tells whether or not that person wore an orange jacket. Indicate only those card(s) **you definitely need to turn over to see if what your tribesman said ("If you go hunting, then you wear these jackets to avoid being shot") is true.**

Selection cards

Went hunting
Did not go hunting
Wore orange jacket
Did not wear orange jacket

References

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: a disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*(2), 225–235.

Almor, A., & Sloman, S. A. (1996). Is deontic reasoning special?. *Psychological Review 103*, 374–380.

Almor, A., & Sloman, S. A. (2000). Reasoning versus text processing in the Wason selection task: A nondeontic perspective on perspective effects. *Memory and Cognition, 28*, 1060–1070.

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press/Bradford Books.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Case, R. (1992). *The mind's staircase: Exploring the conceptual underpinnings of children's thoughts and knowledge*. Hillsdale, NJ: Erlbaum.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology, 37*(5), 715–730.

Channon, S., & Crawford, S. (2000). The effects of anterior lesions on performance on a story comprehension test: Left anterior impairment on a theory of mind-type task. *Neuropsychologia, 38*(7), 1006–1017.

Charman, T., & Baron-Cohen, S. (1992). Understanding drawings and beliefs: a further test of the metarepresentation theory of autism: A research note. *Journal of Child Psychology and Psychiatry, 33*(6), 1105–1112.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187–276.

Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 384–420). Princeton, NJ: Van Nostrand Reinhold.

Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science, 3*(5), 155–158.

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), 1–79.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science, 7*(1), 103–115.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: Bradford.

Fodor, J. A. (2000). Why we are so good at detecting cheaters. *Cognition, 75*, 29–32.

Frye, D., Zelazo, P. D., & Burack, J. A. (1998). Cognitive complexity and control: I. Theory of mind and rule-based reasoning. *Current Directions in Psychological Science, 7*, 116–121.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*(2), 127–171.

Girotto, V., Kemmelmeier, M., Sperber, D., & van der Henst, J. B. (2001). Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition, 81*(2), B69–B76.

Gordon, A. C., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology, 68*(1), 70–83.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., & Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain, 125*, 752–764.

Griggs, R. A., & Jackson, S. L. (1990). Instructional effects on the responses in Wason's selection task. *British Journal of Psychology, 18*, 197–204.

Happe, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*(2), 129–154.

Happe, F. G., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology, 34*(2), 358–362.

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34*(6), 1326–1339.

Kirk, R. E. (1995). *Experimental Design: Procedures for the Behavioral Sciences* (4th ed.). Pacific Grove, CA: Brooks/Cole.

Klein, D. (1976). Attentional mechanisms and perceptual asymmetries in tachistoscopic recognition of words and faces. *Neuropsychologia, 14*(1), 55–66.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind. *Trends in Cognitive Sciences, 8*(12), 528–533.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*(1), 45–85.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251.

Maylor, E. A., Moulson, J. M., Muncer, A.-M., & Taylor, L. A. (2002). Does performance on theory of mind tasks decline with age?. *British Journal of Psychology 93*, 465–485.

Meguro, Y., Fujii, T., Yamadori, A., Tsukiura, T., Suzuki, K., Okuda, J., et al. (2000). The nature of age-related decline on the reading span task. *Journal of Clinical and Experimental Neuropsychology, 22*(3), 391–398.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*(1), 3–65.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104*, 749–791.

Moscovitch, M. (1976). On the representation of language in the right hemisphere of right-handed people. *Brain and Language, 3*(1), 47–71.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience, 4*, 257–267.

Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and the medial temporal cortex. *Neuropsychology, 8*, 524–534.

Moscovitch, M., Fernandes, M. A., & Troyer, A. K. (2002). Working-with-memory and cognitive resources: A component process account of divided attention and memory. In M. Naveh-Benjamin, M. Moscovitch, & H. L. I. Roediger (Eds.), *Perspectives on human memory and cognitive aging: Essays in honor of Fergus Craik* (pp. 171–207). Philadelphia, PA: Psychology Press.

Moscovitch, M., & Klein, D. (1980). Material-specific perceptual interference for visual words and faces: implications for models of capacity limitations, attention, and laterality. *Journal of Experimental Psychology: Human Perception and Performance, 6*(3), 590–604.

Moscovitch, M., & Umilta, C. (1990). Modularity and neuropsychology: implications for the organization of attention and memory in normal and brain-damaged people. In M. F. Schwartz (Ed.), *Modular deficits in Alzheimer-type dementia*. Cambridge, MA: MIT Press.

Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development, 72*(3), 688–690.

Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 476–492.

Overton, W. F., Yaure, R., & Ward, S. L. (1986). Deductive reasoning in young and elderly adults. (Paper presented to the Conference on Human Development, Nashville, TN).

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology & Psychiatry, 32*, 1081–1105.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Science, 3*(9), 337–344.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that...": Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437–471.

Pollack, R. D., Overton, W. F., Rosenfeld, A., & Rosenfeld, R. (1995). Formal reasoning in late adulthood: The role of semantic content and metacognitive strategy. *Journal of Adult Development, 2*, 1–14.

Pratt, M. W., Diessner, R., Pratt, A., Hunsberger, B., & Pancer, S. M. (1996). Moral and social reasoning and perspective taking in later life: A longitudinal study. *Psychology and Aging, 11*, 66–73.

Ragland, J. D., Turetsky, B. I., Gur, R. C., Gunning-Dixon, F., Turner, T., Schroeder, L., et al. (2002). Working memory for complex figures: An fMRI comparison of letter and fractal n-back tasks. *Neuropsychology, 16*(3), 370–379.

Ruthruff, E., Pashler, H., & Hazeltine, E. (2003). Dual-task interference with equal task emphasis: Graded capacity sharing or central postponement? *Perception & Psychophysics, 65*, 801–816.

Salthouse, T. A. (1992). Working-memory mediation of adult age differences in integrative reasoning. *Memory and Cognition, 20*(4), 413–423.

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology, 8*, 535–543.

Saltzman, J., Strauss, E., Hunter, M., & Archibald, S. (2000). Theory of mind and executive functions in normal human aging and Parkinson's disease. *Journal of the International Neuropsychological Society, 6*(7), 781–788.

Tisserand, D. J., van Boxtel, M. P., Gronenschild, E., & Jolles, J. (2001). Age-related volume reductions of prefrontal regions in healthy individuals are differential. *Brain and Cognition, 47*, 182–185.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*(3), 277–283.

Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science, 7*(4), 121–126.

Zelazo, P. D., Jacques, S., Burack, J. A., & Frye, D. (2002). The relation between theory of mind and rule use: Evidence from persons with autism-spectrum disorders. *Infant & Child Development. Special issue: Executive functions and development: Studies of typical and atypical children, 11*, 171–195.