Effects of Subclinical Depression and Aging on Generative Reasoning About Linear Orders: Same or Different Processing Limitations?

Grzegorz Sedek
Warsaw School of Social Psychology and
Polish Academy of Sciences

Ulrich von Hecker
Cardiff University

The performance of older adults and depressed people on linear order reasoning is hypothesized to be best explained by different theoretical models. Whereas depressed younger adults are found to be impaired in generative inference making, older adults are well capable of making such inferences but exhibit problems with working memory (Experiments 1 and 2). Restriction of the available study time impairs reasoning by nondepressed control participants and, as such, proves to be a good model of older adults' but not depressed participants' limitations (Experiment 3). These results are replicated comparing depressed and older participants with a control group in the same study, providing increased power and linking the results to additional control measures of processing speed and working memory (Experiment 4).

Age-related declines have been documented in a variety of cognitive tasks (e.g., Craik & Salthouse, 2000; Kliegl, Mayr, & Oberauer, 2000; Perfect & Maylor, 2000). Similarly, depression (including subclinical depression) is associated with poor performance on numerous measures of cognitive functioning (e.g., Davidson, 2000; Williams, Watts, MacLeod, & Mathews, 1997). The current research attempts to compare declines for nondepressed older and depressed younger participants on generative reasoning tasks that require them to integrate piecemeal information into a coherent representation. Prior research has demonstrated that both depression and aging are associated with poor performance on similar tasks (Verhaeghen & Salthouse, 1997; von Hecker & Sedek, 1999). The central question of the current investigation was whether the cognitive deficits in generative reasoning in these two populations reflect the same or different underlying processing limitations.

The explanations for deficient information processing in depressed and older individuals are often similar, drawing on cognitive resources or memory limitations (Burt, Zembar, & Niederehe, 1995; Hasher & Zacks, 1979; Light, 1996), impaired inhibition (Hasher, Zacks, & May, 1999; Joormann, 2004), lowered efficiency of cognitive strategies (Oberauer & Kliegl, 2001; von Hecker & Sedek, 1999), or lack of cognitive initiative (Hertel, 1997; Hess, 2000) as explanatory concepts. However, some theoretical approaches to impaired cognitive performance are quite specific to the given area, such as the notion of cognitive slowing as a general mechanism to explain diminished cognitive effectiveness in older people (Salthouse, 1996; Verhaeghen & Salthouse, 1997) or preoccupation with negative thoughts as an interfering mechanism in depressed individuals (Ellis & Ashbrook, 1988; Lyubomirsky, Tucker, Caldwell, & Berg, 1999).

Cognitive deficits in older age and depression become especially obvious in complex tasks, particularly in tasks that require the integration of piecemeal information into a more coherent mental representation. Such integration is often required in problem solving or social perception. For example, performance by older adults on complex impression formation tasks is more varied.
able and motivation dependent compared with younger adults (Hess, Follett, & McGee, 1998). Performance by depressed individuals is impaired when they are asked to form a coherent impression of a social group, on the basis of pairwise sympathy relations (von Hecker & Sedek, 1999). Is integration of piecemeal information impaired, and if so, how is it impaired in the two populations, respectively? To study integrative deficits in more detail, we have deliberately focused our research on comparing the potential impairments of depressed and older adults in a linear order construction task that benefits from the online generation of a mental model (Johnson-Laird, 1983, 1996). We report four experiments in which we test contrasting predictions concerning the nature of possible dysfunctions in generative forms of reasoning among older adults versus depressed participants. In particular, we examine the hypothesis that both depression and aging can interfere with performance on linear order reasoning but that the mechanisms of those impairments are qualitatively different for younger depressed participants versus older adults. We demonstrate that depressed persons have problems with generative reasoning but their memory retrieval is intact. However, older participants, although having reasoning ability preserved, suffer from problems with memory retrieval.

Generative Reasoning on Linear Orders

The study of linear order reasoning has long been a topic of considerable interest in cognitive and developmental psychology (Piaget & Inhelder, 1974; Potts, 1972; Rabinowitz, Grant, Howe, & Walsh, 1994; Sternberg, 1980). One of the major findings in this area refers to the generative processes of rearranging incoming information into a comprehensive mental model (mental array). For example, on learning a series of statements such as “Paul is taller than George,” “John is taller than Paul,” and “George is taller than Ringo,” participants spontaneously rearrange these pieces of information into a coherent mental array (taller than): John > Paul > George > Ringo. Robust distance effects have been pointed out as evidence for generative processes happening during encoding (Mynatt & Smith, 1977; Potts, 1972, 1974; Pohl & Schumacher, 1991; K. H. Smith & Foos, 1975; for a recent review, see Leth-Steenens & Marley, 2000). According to these effects, response times are faster and accuracy is higher in a subsequent test stage, as the distance between queried elements on the hypothesized subjective model increases. Note that if the queried pair spans a distance wider than immediate neighbors on the array, it refers to a pair of elements that had not been presented in the first stage of the experiment. This effect reflects quicker and more precise verification responses for the John–Ringo type of query than, for example, the Paul–Ringo type of query. Likewise, the distance effect is especially pronounced for the end elements in the array, the so-called end-point effect. The response to the “end points” John and Ringo should therefore be especially precise, and latencies for correct responses should be fastest, despite that no statement pairing “John taller than Ringo” had been presented. It appears, then, that participants do not simply store the adjacent pairs during training and then use retrieval of them to draw transitive inferences at the time of testing. Rather, they seem to integrate all pairwise information into a unified episodic model from which they can easily “read off” the answer to a queried pair. In summary, distance effects are most commonly taken to reflect the outcome of a spontaneous generative, constructive mental activity, which leads to a unified, highly integrated memory representation. Such a representation appears similar to a mental model (Brewer, 1987) in that it gives ready access to implications that go beyond the information initially presented.

The bulk of research in the linear order area has used the so-called serial orders paradigm (for a recent review including a connectionist model, see Leth-Steenens & Marley, 2000). In this procedure, the participants typically study ordered pairs, and a repeated application of simple matching is sufficient to construct the mental array. For example, for the two relations “John is taller than Paul” and “Paul is taller than George,” “Paul” is a matching element. However, presenting pairs in this way, that is, with matching opportunities always provided by the sequence, tends to underestimate the demand on processing generally imposed by this type of reasoning problem (Halford, Wilson, & Phillips, 1998; Waltz et al., 1999). Therefore, in the present research, we use a procedure adapted from K. H. Smith and Foos (1975), in which difficulty levels of the load during the rearrangement of pairs into ordered sequences are experimentally manipulated (see Figure 1).

Overview of the Experimental Procedure

Figure 1 illustrates the two central within-participant variables of our research: difficulty of presented order (three levels) and pair distance (adjacent, two step, and end point). Participants study three pairs of relations at their own pace. Subsequently, they are queried about some relations between the presented persons. As participants study the three consecutive relations, each previous relation disappears when the next one is presented. In terms of encoding processes during the study phase, participants might simply try to rehearse and remember the presented pairs. Alternatively, they might try to construct a mental array (see the third column of Figure 1) from which they can later read off the queried information. We modified the original experimental procedure by K. H. Smith and Foos (1975) in three ways. First, participants were not explicitly asked to construct a mental array from the presented relations. Consequently, the encoding strategy was left to the participants. Immediately after studying, participants were asked about the presented and to-be-inferred relations. Thus, we were able to assess spontaneous tendencies to generate a mental model online. Second, Experiments 1, 2, and 4 used self-paced presentation, as opposed to fixed study times (K. H. Smith & Foos, 1975, and Experiment 3 of this article). Third, because our procedure was fully computerized, we were able to measure study times to address issues of general task motivation.

The Difficulty Variable

As in the article by K. H. Smith and Foos (1975), the difficulty level was operationalized by means of different types of pair orderings (see Figure 1). As illustrated in Figure 1, demands increased in terms of the number of constructive operations that the participant had to carry out to appropriately rearrange the sequence of pairs. The more such operations necessary, the more information had to be kept in working memory during processing, assuming that a participant applied a constructive strategy to build a mental array. Of course, if a participant, during learning, did not attempt to integrate the information but only stored the three pairs,
then the level of difficulty should be the same for all presented orders. Orders 1 and 2 represent the easiest difficulty level because each consecutive pair contains a matching element that facilitates the construction of a mental array. Orders 5 and 6 are most difficult because in these orders the second pair introduces a relation between two new persons, none of whom matches with one of the persons mentioned in the first sentence. Thus, participants had to hold both pairs in memory until the third pair was presented to complete integration of the four persons into a linear order. In Orders 3 and 4 (intermediary level of difficulty), participants who apply a constructive strategy should begin generating the mental array with the middle pair, expanding it later toward both end points.

Regarding study times, a participant who makes an active attempt at linear order construction should spend the most time studying the third pair because it is only after knowing the third pair that it becomes possible to construct the whole four-array mental model, which is especially obvious in the most difficult orders (5 and 6). Testing for this type of pattern was also crucial in view of potential differences in motivational levels of cognitive involvement among participants from different groups.

### Potential Models of Information Processing Limitations in Reasoning About Linear Orders

Using this methodology, we examined two basic descriptive models\(^2\) of processing limitations in linear order reasoning. Both of these models were based on the assumption that normally motivated students will show almost perfect accuracy in solving

\(^2\)This term is borrowed from Hambrick and Engle, 2002—see their conceptually similar methodological approach with descriptive models of cognitive performance in the domain of cognitive aging.

---

<table>
<thead>
<tr>
<th>Order difficulty</th>
<th>Abstract notation of presented orders</th>
<th>The verbal format of presented pair relations (exemplary names – in actual experiments names were never ordered alphabetically and they were different for each order)</th>
<th>The potentially generated mental array ((&gt; = ) relation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1. AB, BC, CD&lt;br&gt;2. CD, BC, AB</td>
<td>Alice smarter than Brenda, Brenda smarter than Carol, Carol smarter than Doris&lt;br&gt;Carol older than Doris, Brenda older than Carol, Alice older than Brenda</td>
<td>Alice &gt; Brenda &gt; Carol &gt; Doris&lt;br&gt;Alice &gt; Brenda &gt; Carol &gt; Doris</td>
</tr>
<tr>
<td>II</td>
<td>3. BC, CD, AB&lt;br&gt;4. BC, AB, CD</td>
<td>Brenda richer than Carol, Carol richer than Doris, Alice richer than Brenda&lt;br&gt;Brenda taller than Carol, Alice taller than Brenda, Carol taller than Doris</td>
<td>Alice &gt; Brenda &gt; Carol &gt; Doris&lt;br&gt;Alice &gt; Brenda &gt; Carol &gt; Doris</td>
</tr>
<tr>
<td>III</td>
<td>5. AB, CD, BC&lt;br&gt;6. CD, AB, BC</td>
<td>Alice stronger than Brenda, Carol stronger than Doris, Brenda stronger than Carol&lt;br&gt;Carol faster than Doris, Alice faster than Brenda, Brenda faster than Carol</td>
<td>Alice &gt; Brenda &gt; Carol &gt; Doris&lt;br&gt;Alice &gt; Brenda &gt; Carol &gt; Doris</td>
</tr>
</tbody>
</table>

**Figure 1.** Three categories of linear order difficulty.

---

**The Pair Distance Variable**

During the test phase, participants were queried about three types of pairs:

1. Pairs AB, BC, and CD (one-step relations or adjacent pairs), as a measure of memory retrieval.
2. Pairs AC and BD, which are not presented during studying. The queries about these pairs demand generative integrative reasoning across two steps on the hypothetical mental array (\(A > B > C > D\)).
3. Pair AD. The query for this pair is about the maximum array distance of three steps between the end points of the hypothetical array (see Figure 1).

The adjacent, two-step, and end-point questions define the independent variable pair distance. As a dependent measure during the test phase, accuracy in responding to the three types of queries was measured for each order difficulty.
linear problems, given the self-paced study procedure.\(^3\) Thus, the student control group should spend more time studying difficult linear orders, but assuming that they generate the proper mental model, their accuracy, taken as unimpaired baseline performance, is expected to be at least equally high across all difficulty and pair distance levels. If anything, accuracy may even increase from adjacent pairs to end points.

**Memory Retrieval Limitations (MR) Model**

This first deficit model to be considered assumes that the primary cognitive limitation lies in a limitation of memory retrieval. That is, the underlying problem may be a fundamental working memory limitation, which manifests itself in the difficulty in maintaining the necessary input information in an active state during processing of further input data (Miyake & Shah, 1999). Because of this limitation, differences between control and experimental groups should be visible for all types of queries across the board (a main effect of group), including adjacent queries. This is because the latter queries constitute a test of a participant’s ability to correctly retrieve the pairwise information. According to the MR model, any performance decrements for two-step and end-point queries are consequences of failures in memory retrieval. That is, the experimental group’s generative reasoning is assumed to be impaired, given that they work on correctly preserved input information. It follows from this that if levels of memory retrieval (performance on adjacent pairs) are controlled for statistically, the original effect of group end-point accuracy should be substantially attenuated.

**Generative Processing Limitations (GP) Model**

The second model to be considered assumes that the cognitive limitation lies in a genuine thinking limitation. Because generative thinking is impaired, the more premises that are to be integrated into a mental array, the more should the reasoning deficit become apparent. Accuracy should therefore decrease in a stepwise way from adjacent to end-point queries. The deficit assumed here concerns the integration of piecemeal information into a coherent mental model; this prediction holds irrespective of the level of order difficulty. It also follows from this perspective that even when we control for memory retrieval (performance on adjacent relations), the effect of group on end-point accuracy should remain significant.

**Evidence for Memory Retrieval Limitations in Reasoning Among Older Adults**

Salthouse and his collaborators (Salthouse, 1992; Salthouse, Legg, Palmon, & Mitchell, 1990; Salthouse, Mitchell, Skovronek, & Babcock, 1989) have examined the effects of adult age on integrative reasoning. The paradigm used by this group is close to ours but different in some important aspects, which we discuss below in some detail.

Salthouse and colleagues (Salthouse, 1992; Salthouse et al., 1990, 1989) used a reasoning task involving abstract verbal materials. One to three premises are presented individually on the screen in a self-paced procedure. Subsequently, test queries are presented. For example, one of the three-premises problems might look as follows: (a) “R and S do the opposite,” (b) “T and U do the same,” (c) “S and T do the opposite.” Subsequently, one of three different types of queries is asked:

1. **One-relevant** questions: “If T increases, what will happen to U?” Participants have to select between “decreases” and “increases”;
2. **All-relevant** questions: “If R decreases, what will happen to U?” Participants have to select between “decreases” or “increases”; and
3. **Recognition probes**: “T and U do the same?” Participants have to indicate whether the probe was either identical or different from one of the previously presented premises.

It is important to note the similarities to our procedure. First, exactly as in our end-point queries, giving a correct answer to an all-relevant query requires integration of information from all premises. Second, similar to our adjacent queries, a recognition probe does not require any reasoning steps but just the retrieval of a presented premise from memory. Third, similar to our linear order procedure, Salthouse et al.’s (Salthouse, 1992; Salthouse et al., 1990, 1989) integrative reasoning task places high demands on working memory. This is because on presentation of a premise, the previous premise, if there was one, is always removed. The principal difference between Salthouse et al.’s (Salthouse, 1992; Salthouse et al., 1990, 1989) reasoning task and the linear order task is that the linear order task encourages much more spontaneous generative reasoning during encoding of the presented premises. That is, the transitive character of presented relations may stimulate a participant to create a mental model (mental array) that represents, for example, who is fastest and who is slowest among the presented persons. This model may then be available as a retrieval device. In contrast, the more abstract nature of Salthouse et al.’s premises does not encourage the construction of any concrete mental model of comparable nature. Instead, in their procedure, participants have to wait for the test query before beginning a more or less complex integrative reasoning. This is so because only the test query provides a necessary initializing parameter (e.g., “T increases”). Thus, spontaneous generative thinking is not encouraged by the Salthouse et al. task because constraints are lacking during the studying of the premises. The task was deliberately designed to have this property, as is clear from the following citation from Salthouse et al. (1990):

> We assumed that cases such as this, in which neither a categorization nor an ordinal relation among the variables could be established until after the presentation of the final premise, would discourage efforts to organize or integrate the information after each successive premise. (p. 10)

\(^3\) For this reason, another well-known type of reasoning task, that is, syllogistic reasoning, was not applied here because most participants were unable to solve this task correctly (e.g., only 1 depressed and 1 nondepressed participant used consistently the correct, logical strategy in solving syllogisms, according to a study by Channon & Baker, 1994).
Conversely, spontaneous generative thinking may be observed in the linear order task in terms of distance effects and end-point related effects, as discussed above.

The replicated findings from Salthouse et al.'s (Salthouse, 1992; Salthouse et al., 1989, 1990) research are that age differences in the ability to integrate information from premises are nearly nonexistent, provided that correct information about the premises is available in memory. There were substantial effects of age on one-relevant questions; however, increasing the number of premises did not increase the effect of age on reasoning accuracy. As summarized by Salthouse (1992),

That is, the principal determinant of the variations in reasoning accuracy associated with additional premises or with increased age appears to be the ability to maintain previously presented information rather than the ability to integrate information that is available in memory. (p. 415)

Older adults’ deficiency in premise retrieval is attributed by Salthouse (1992) to some basic dysfunctions of working memory, which is conceived as a device for simultaneous maintenance and processing of incoming data (Baddeley, 1986; Daneman & Carpenter, 1980). According to this view, the limitations of older adults in integrative reasoning have to do with their problems concerning maintenance of relevant information, rather than with limitations concerning integrative reasoning per se. A similar interpretation has been made regarding more recent findings (Salthouse, 2000, 2001), demonstrating that the earlier results were not restricted to specific features of the particular integrative reasoning task. Nearly constant age–performance relations were found across more difficult problems in a range of divergent reasoning tests.

To summarize the implications from Salthouse et al.’s (Salthouse, 1992; Salthouse et al., 1990, 1989) research, one may expect that (a) older people in comparison with younger adults have problems with the active maintenance of presented premises during processing; (b) older adults have no problems with complex reasoning (i.e., they have no problems with the integration of premises), provided the necessary information can be remembered; and (c) the effect of age on reasoning accuracy is markedly attenuated when one controls for preservation of the premises in memory.

**Generative Reasoning Limitations Among Depressed Participants**

The most obvious evidence for generative reasoning problems among depressed people comes from studies on the construction of mental models. In line with the classic idea of Johnson-Laird (1983), a mental model is defined here as a construction based on incoming data (such as premises in reasoning tasks or text sentences in comprehension tasks). Mental models are generated online during task performance. This constructive activity might be contrasted with a mere preservation of presented information. Mental models are episodic representations, resembling structural and functional relations of real entities (Brewer, 1987). They may function in support of analogue simulations of the events in the world, either real or imaginary (Greeno, 1989).

In previous research (von Hecker & Sedek, 1999), we studied the cognitive consequences of an uncontrollability treatment as well as the consequences of subclinical depression in terms of the cognitive exhaustion model (Kofta & Sedek, 1998; Sedek & Kofta, 1990; Sedek, Kofta, & Tyszka, 1993). According to this model, cognitive dysfunctions in depressed individuals, such as difficulties in systematic testing (J. D. Smith, Tracy, & Murray, 1993), failures to initiate more complex strategies (Hertel & Rude, 1991), or encoding impairments (Channon, Baker, & Robertson, 1993a), are due to a deficit in the online generation of coherent episodic or situation models from incoming pieces of information.

In a social perception task, participants had to use detail information of varying diagnostic value to form an overall model of a social environment (von Hecker, 1997; von Hecker, Crockett, Hummert, & Kemper, 1996). It was found that depressed participants, like nondepressed participants, allocated more study times to diagnostic relations, thereby recognizing their diagnostic value as such. However, unlike nondepressed participants, depressed participants did not use those relations adequately for the purpose of model construction. In subsequent tests, the quality of the constructed social mental models was found to be impaired under both control loss experience and depression (von Hecker & Sedek, 1999, Experiments 2 and 3). These findings are in line with the cognitive exhaustion model. This model assumes that both the experience of control loss and depression should primarily interfere with the solving of tasks that require generative forms of thinking or tasks that are complex and cognitively demanding (cf. Kofta & Sedek, 1998; von Hecker, Sedek, & McIntosh, 2000). Our approach clearly concedes that processing effort as well as processing quality in less generative, more top-down or schema-guided tasks might be unimpaired or even enhanced under depression or control loss experience. Thus, participants in those states have been shown to meticulously process behavioral information to derive trait inferences from behavioral information (Gannon, Skowronski, & Betz, 1994; Weary, Marsh, Gleicher, & Edwards, 1993) and to be even more accurate than nondepressed control participants in processing situational versus dispositional information when making attributional judgments (Pittman & D’Agostino, 1989). Our point is, however, that despite their exerted cognitive effort, because of their cognitive exhaustion state, depressed participants’ performance will suffer particularly in those tasks that demand the deployment of integrative strategies for the participants to construct novel memory representations in a bottom-up fashion.

**Overview of Experiments**

The existing evidence suggests that in comparison to relevant control groups, depressed persons might have generative processing limitations when dealing with tasks that benefit from constructing a mental model (GP model, which we examine in Experiment 1). In contrast, older people should have problems primarily with memory retrieval (MR model, which we examine in Experiment 2). Experiment 3 investigates whether the pattern of results seen in a control group that works under fixed study-time conditions might be seen as a good model of thinking limitations in depressed participants or, alternatively, in older adults. Finally, Experiment 4 compares both older adult and depressed participants with nondepressed control participants within one and the same study and links the findings to measures of processing speed and working memory.
Experiment 1

Subclinically depressed students are expected to show an impairment in the construction of a linear mental model from pairwise information, thereby providing a conceptual replication of the study by von Hecker and Sedek (1999) as well as an extension into the domain of transitive order construction. The analysis of study times during the learning stage allowed us to assess general task motivation as well. If sufficiently motivated, participants in both groups should concentrate their study times on more difficult orders and should need more time to study third pairs than first or second pairs because at presentation of the third pair, all information can be integrated. However, if depressed participants are not sufficiently motivated or if they show a tendency to not try hard enough and thereby protect their self-esteem from threats by fear of cognitive failure (Kofka & Sedek, 1989), they should devote less study time to all presented relations compared with participants in the control group. Furthermore, depressed participants should also not be very sensitive to the difficulty of orders. They should also not show the same preference as control participants to focus on analyzing third pairs.

Depressed participants’ reasoning behavior should fit the GP model. This means we expected the following pattern of results: (a) decreasing accuracy of depressed participants as a function of pair distance, which would reflect a failure in online construction of a mental model, with this failure causing problems with reasoning about the nonpresented pairs, and (b) a statistically reliable relation between depression and reasoning accuracy concerning end points, even after we controlled for memory retrieval, that is, the accuracy for presented adjacent relations. According to the GP model, we should observe the most substantial differences between depressed and nondepressed groups in the accuracy for end-point relations but only minimal or no differences for adjacent relations because those accuracy values reflect memory retrieval.

If the performance of depressed participants was best described by the MR model, one should find only a main effect of group for reasoning accuracy. Furthermore, the statistical relation between depression and accuracy for end-point relations should be substantially attenuated when accounting for accuracy for adjacent relations.

Method

Participants

Students enrolled in psychological, historical, and educational sciences at the University of Potsdam participated in the experiment. The German version of the Beck Depression Inventory (BDI; Beck, 1967) was used as screening procedure. From this initial pool, a random sample of students who scored 5 and below or 10 and above on the BDI were selected for participation in the main session (7–10 days later). We readministered the BDI at the beginning of the main experimental session to ensure the stability of the depression state. Only those participants whose BDI scores remained within the prior category were assigned to the appropriate groups. In accordance with other procedures (Bargh & Tota, 1988; Edwards & Weary, 1993; von Hecker & Sedek, 1999), students with stable BDI scores ranging from 0 to 5 were assigned to the nondepressed group, and those with stable BDI scores of 10 and above were assigned to the depressed group. Participants who changed category (n = 31) were dismissed from the study and debriefed. The final sample comprised 38 students (mean age = 22.09 years, SD = 2.48 years), 19 in the nondepressed group (13 women and 6 men) and 19 in the depressed group (13 women and 6 men). The mean BDI scores at the second administration for the nondepressed and depressed participants were, respectively, 2.11 (SD = 1.41) and 15.00 (SD = 6.06). Participants could choose between course credits as partial fulfillment of study requirements or, alternatively, a monetary payment.

Materials

Six order types used by K. H. Smith and Foos (1975) were used as stimulus materials (see Figure 1). Capital letters in the second column of Figure 1 symbolize German first names that were actually used in this study. Orders 1–6 reflected an underlying array A–B–C–D. There were six different name sets (three female and three male), which were for each participant randomly assigned to the orders. Each name set comprised four names. Six German adjectives were used to denote six semantically different transitive relations (older, richer, taller, smarter, stronger, and faster). The assignment of adjectives to order types was counterbalanced across participants such that each participant was presented with each adjective and each order type exactly once. Each adjective appeared in each order type condition exactly once across every cycle of 6 participants.

Procedure

The main experimental session was initiated on the second appointment, directly after the participants completed the questionnaire, provided their assignment to either the nondepressed or depressed group had proven reliable. First, participants were given a general instructions sheet. They were instructed that the study was about learning and recalling rank relations within small groups of fictitious people. Participants were told that they were to study several sequences of person pairs. Each sequence would comprise three pairs among four fictitious persons. In each pair, a certain relation would hold, for example, “Christine is older than Maria” or “Thomas is taller than Markus.” Participants were asked to memorize the two names and the relation between them. They could study each relation at their own pace, initiating the first relation and moving from one relation to the next by pressing the space bar. Going through the three pairs in the described way constituted a learning stage.

During each learning stage, participants were presented with a sequence of three order relations, corresponding to one of the orders from Figure 1. All relations were phrased as sentences, for example, “Christine is older than Katharina” in normal ASCII letters in the center of the screen. Each sentence was preceded by a 1,500-ms presentation of an “X” in the center as a preparing signal.

The learning stage was immediately followed by a test stage. During each test stage, participants were asked about all six possible relations between the four persons, both presented and inferred. These relations were phrased as sentences as shown above and were presented in a random order. Sentences were either in correct format, conforming to the learned order (e.g., “Christine is older than Katharina”), or in false format, that is, contradicting the order (e.g., “Katharina is older than Christine”). Among the three adjacent test pairs, there was always exactly one randomly determined query in false format. Among the 2 two-step test pairs, one (by random) was in false and one was in correct format. The end-point test pair was always presented in correct format. Participants had to use the marked left and right arrow keys to indicate whether the query sentence reflected a correct statement. Half of the participants had to press left to indicate “correct” and right to indicate “false.” The instruction was reversed for the other half of participants. All participants were asked to respond as quickly and as accurately as possible.

The word LEARNING or, alternatively, TEST was always displayed appropriately in the upper left corner of the screen so that participants’ attention was focused on each particular task component. After each learning–test sequence, except for the last sequence, a four-item set of easy arithmetic was interspersed to clear participants’ memory from previous
materials. The arithmetic problems were in multiple-choice format and could be solved using marked keys on the keyboard. After two practice trials, participants began the main part of the experiment. This involved six blocks of learning–test–arithmetic trials (except the sixth block, after which no arithmetic task was given). The six order types (see Figure 1) were in a new random presentation order for each participant and then assigned to Blocks 1–6 in that presentation order.

To ensure that participants understood the instructions, we had them perform one full block of trials (learning phase, test phase, and arithmetic task) as practice before administration of the main task. After completing all experimental tasks, participants were debriefed. Sessions were conducted in small groups of 1 to 3 persons in a laboratory of the Social Psychology Unit of the Department of Psychology at the University of Potsdam, Potsdam, Germany, and lasted about 50–60 min.

Results

Study Times

To address motivational issues, we examined depressed versus nondepressed participants in terms of study times during the learning stage. Nondepressed and depressed participants were compared at the three levels of order difficulty and for three subsequently presented pairs. We calculated the size effects for all significant results using partial eta squared ($\eta^2$) measures. A 2 (depression: depressed vs. nondepressed students; between subjects) $\times$ 3 (order difficulty: Orders 1 and 2 vs. Orders 3 and 4 vs. Orders 5 and 6; within subjects) $\times$ 3 (sequential pair: first vs. second vs. third; within subjects) analysis of variance (ANOVA) on this measure showed a main effect for order difficulty, $F(2, 72) = 7.19$, $MSE = 50.77$, $p < .002$, $\eta^2 = .17$, and a main effect for sequential pair, $F(2, 72) = 15.13$, $MSE = 55.13$, $p < .001$, $\eta^2 = .30$. These effects were qualified by an Order Difficulty $\times$ Sequential Pair interaction, $F(4, 144) = 6.77$, $MSE = 29.53$, $p < .001$, $\eta^2 = .16$. The mean study times are shown in Figure 2.

To interpret the main effects, we applied a posteriori Newman–Keuls tests with pairwise comparisons. For order difficulty, this procedure revealed that the combined Orders 5 and 6 were studied longer than Orders 1 and 2 and Orders 3 and 4 ($p < .005$). The latter two order levels did not differ between each other. The comparisons for sequential pairs indicated that third pairs were generally studied longer than both first pairs and second pairs ($p < .001$). Study times for the latter two pairs did not differ among each other (see Figure 2).

To interpret the interaction effects, we applied planned comparisons. These analyses confirmed the predicted increase in processing demands across levels of order difficulty, as concerns study times for the second and especially for the third presented pair (see Figure 2). More specifically, inspection of the linear trend component across levels of order difficulty revealed no differences in study time for first pairs, $t(36) = 1.03$, $ns$. However, for second sequential pairs and especially for third pairs, the linear trends were significant, $t(36) = 2.65$, $p < .02$, and $t(36) = 3.28$, $p < .003$, respectively.

It is important to note that the main effect and the interaction effects of depression with other independent variables were not significant ($F$s $< 1$). Given the very similar patterns for depressed and nondepressed participants in Figure 2, we have no reason to assume differences in general task motivation between the two groups. Study behavior of both control and depressed participants was at comparable levels concerning the more extensive processing of more difficult orders and, likewise, the longer studying of second and especially third pairs.

Reasoning Accuracy

For each participant and each order, we determined the proportion of correct answers to queries for all three levels of pair distance. That is, means were calculated across the three adjacent pairs, AB, BC, and CD, and the two two-step pairs, AC and BD. A single value (0,1) was entered for the end-point pair, AD. Using these scores, we determined the average accuracy for each participant and for each pair distance level across pairs of Orders 1 and 2, Orders 3 and 4, and Orders 5 and 6. These average accuracies

![Figure 2](image-url)
created the three levels of the independent variable order difficulty. Retrieval accuracy data were analyzed by ANOVA as a function of depression, order difficulty, and pair distance.

The mean accuracy data are shown in Figure 3. A 2 (depression: depressed vs. nondepressed students; between subjects) × 3 (order difficulty: 1 and 2 vs. 3 and 4 vs. 5 and 6; within subjects) × 3 (pair distance: adjacent vs. two step vs. end point; within subjects) ANOVA on the retrieval accuracy measure demonstrated a main effect of depression, $F(1, 36) = 5.97, \text{MSE} = 2.917, p < .02, \eta^2 = .14$, with a higher proportion of accurate answers in the nondepressed ($M = .94$) than in the depressed group ($M = .84$). The only further significant effect in this analysis was the Depression × Pair Distance interaction, $F(2, 72) = 3.73, \text{MSE} = 661.65, p < .03, \eta^2 = .09$, indicating that both groups performed strikingly different as a function of pair distance (see Figure 3). In the nondepressed group, the proportion of correct answers was at a uniformly high level across all pair distances, which is consistent with model-based retrieval. In contrast, a planned comparison showed a significant linear trend component for the depressed group, $t(36) = 2.59, p < .02$, indicating that the proportion of correct answers decreased from adjacent pairs to end-point relations. Additionally, questions about the two types of inferred relations, that is, two-step and end-point pairs, were more accurately answered by nondepressed than by depressed participants, $t(36) = 2.06, p < .05,$ and $t(36) = 2.58, p < .02,$ respectively. There was no reliable difference between groups at the level of adjacent pairs (retrieval accuracy). This pattern contradicts an explanation in terms of memory retrieval limitations (MR model). Thus, in strong support for the hypothesis of generative processing limitations (GP model), we found a decreasing distance effect for depressed participants. The more inference steps were needed to answer a query, the more their reasoning appeared to suffer. At the same time, inspection of the two-way Group × Order Difficulty interaction ($F < 1$) and the three-way Group × Order Difficulty × Pair Distance interaction ($F < 1$) leads to the conclusion that depressed participants’ differential performance between adjacent pairs and end-point pairs is not systematically affected by the difficulty of orders.

Hierarchical Regression Analyses for End-Point Reasoning Accuracy

The data concerning end-point accuracy, which is the type of inference that demands the highest amount of generative mental activity, were subjected to a hierarchical regression analysis (see, Salthouse, 1992; Salthouse et al., 1989, 1990, for an identical approach in the case of abstract integrative reasoning across age groups). The aim of this analysis was to assess the effect of depression (categorical group variable) on end-point accuracy after partialing out the variance associated with memory retrieval (i.e., accuracy for adjacent relations that were presented in the learning phase). The results are summarized in Table 1. This analysis showed that the initial $R^2$ for the relation between depression (depressed vs. nondepressed) and end-point accuracy of .156 (significant at the .05 level) was reduced to .111 (still significant at the .05 level) after we controlled for memory retrieval, which corresponds to a rather modest attenuation of 28.8%. That is, the effect of depression on end-point accuracy did remain reliable, even after we controlled for memory retrieval.

Discussion

In the light of this experiment, mildly depressed individuals are relatively unsuccessful in constructing integrated linear arrays as memory representations of sets of learned transitive relations. Unlike nondepressed control participants, depressed participants did show a decrease in accuracy for inferred pairs, compared with the pairs they had seen during learning. This evidence speaks against retrieval based on an integrated mental model that was spontaneously constructed during learning. The evidence is however supportive of propositional reasoning at the time of the query. The notion here is that if asked for a specific relation, an individual collects all adjacent pairs needed for an answer as single propositions and then combines these propositions using the rules of deductive thinking (Gilhooly, 1998). Because this process is error prone, depressed participants are less accurate when responding to
inferred as compared with adjacent pairs. As such, the reported data support the assumptions outlined in the GP model: namely, that depressed mood is associated with a cognitive deficit, affecting the execution of flexible integrative steps of thinking (von Hecker & Sedek, 1999).

Nondepressed participants, conversely, showed high accuracy levels at all three pair distances, which is indicative of their use of mental models as an effective retrieval device (Johnson-Laird, 1983; Johnson-Laird & Oakes, 1990). The GP model was further supported by the regression analysis, in which the group factor remained significant after memory for adjacent pairs was taken into account. Also, there were no group differences in the accuracy of recalling adjacent pairs. Such a group difference would have been expected under the assumption that depressed participants had difficulties maintaining the presented materials in working memory.

In terms of study behavior, both groups showed comparable patterns. They devoted more time to third pairs, which allows an overall integration of the materials into a mental model, particularly in Orders 5 and 6. Earlier research has found those two orders most difficult (K. H. Smith & Foos, 1975), presumably because the third pair must be fit in between the previously unrelated first two pairs. The fact that even the depressed group exerted a considerable amount of effort to understand and integrate the information is in line with our earlier reasoning (von Hecker & Sedek, 1999) that depressed participants, despite engaging in the type of mental activity that is necessary to arrive at mental models, are actually not successful in doing so (see also Sedek & Kofta, 1990).

**Experiment 2**

This experiment was designed to examine older adults’ performance in linear order construction. In particular, because older adults are assumed to have problems with active maintenance of the premises (Salthouse, 1992), we expected a pattern of deficits in line with the MR model. Conversely, older adults are not expected to exhibit a generative processing deficit. They should be as good as younger students in generating mental arrays from individual relations, once they have successfully maintained the necessary input information in short-term memory. We kept the self-paced presentation method from Experiment 1 to address general task motivation in both groups.

Under the assumption of an underlying memory retrieval deficit in the older adult group, the expectations are as follows: (a) a main effect of group for reasoning accuracy, that is, age-related decrements for adjacent relations (memory test) as well as for the two types of to-be-inferred relations (two-step and end-point relations) and (b) a relation between age and reasoning accuracy concerning end points that is substantially attenuated when we account for level of memory retrieval (i.e., the accuracy for presented adjacent relations). The second prediction is especially crucial for a test of the MR model.

If, however, the processing limitations of older adults were better described by the GP model, one should predict a similar pattern of results as was found for depressed participants. Consequently, one should observe a significant interaction between age group and pair distance for accuracy reasoning and latencies for correct responses.

**Method**

**Participants**

The participants were 17 students from final high school classes in Warsaw, Poland (12 women and 5 men; mean age = 17.76 years, SD = 0.56 years; 11 classes of formal education) and 17 older adults (12 women and 5 men; mean age = 70.58 years, SD = 4.09 years). The older persons in this study were recruited through advertisements at the University of Third Age. This university has several locations in Warsaw and offers different kinds of lectures and seminars for interested older adults but is not equivalent to any formal university education. The criteria for the older adults’ participation were at least 11 classes of formal education, Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) scores of at least 28, and BDI scores below 10. Volunteers were paid for their participation and were thoroughly debriefed at the end of the study.

**Materials and Procedure**

All materials and instructions were translated without modifications from German to Polish. All procedural specifications were the same as in Experiment 1. Participants completed the experiment individually at the experimental laboratory of the Warsaw School of Social Psychology. The same computer program as in Experiment 1 controlled all consecutive procedures.

**Results**

**Study Times**

Study times in milliseconds were compared for older adults and high school students at the three levels of order difficulty and for
the three subsequently presented pairs. A 2 (age: older adults vs. students; between subjects) × 3 (order difficulty: Orders 1 and 2 vs. Orders 3 and 4 vs. Orders 5 and 6; within subjects) × 3 (presented pair: first vs. second vs. third) ANOVA on this measure showed effects similar to the corresponding analysis from Experiment 1. Specifically, there was a main effect for order difficulty, $F(2, 64) = 5.55, MSE = 140.73, p < .01, \eta^2 = .15$, and a main effect for presented pair, $F(2, 64) = 4.36, MSE = 191.65, p < .05, \eta^2 = .12$. These effects were qualified by an Order Difficulty × Presented Pair interaction, $F(4, 128) = 3.21, MSE = 104.85, p < .05, \eta^2 = .09$. According to Newman–Keuls a posteriori tests ($p < .05$), study times for Orders 5 and 6 ($M = 13.91$ s) were significantly longer than study times for Orders 1 and 2 ($M = 10.83$ s) and Orders 3 and 4 ($M = 10.28$ s). The latter two order levels did not differ with each other. Comparisons showed that third pairs ($M = 13.97$ s) were studied longer that both first pairs ($M = 10.15$ s) and second pairs ($M = 10.91$ s). Study times for the latter two pairs did not differ with each other. The interaction effect (see Figure 4) can be attributed to an increasing linear trend across order difficulty levels, but only for third pairs, $\eta(32) = 2.58, p < .02$.

Additionally, the analysis revealed an Age × Order Difficulty × Presented Pair interaction, $F(4, 128) = 2.69, MSE = 104.85, p < .05, \eta^2 = .08$. The decomposition of this interaction in terms of planned comparisons ($p < .05$) showed that there were no differences in study time between groups and presented pairs for the simpler orders (1 and 2 and 3 and 4). However, for the most difficult orders, 5 and 6, older adults needed an exceptionally long time to study third pairs ($M = 20.37$ s), that is, significantly more time than they needed to study any other type of relation. For the students, both second pairs ($M = 15.69$ s) and third pairs ($M = 17.58$ s) within the most difficult orders, 5 and 6, were studied significantly longer than first pairs ($M = 8.45$ s).

These analyses revealed a significant variation in difficulty between the orders, as reflected in the amount of processing time spent on Orders 5 and 6 in particular. The results also showed that both groups, older adults and high school students, spent most of their studying time on analyzing the third presented pair, which potentially enabled them to construct a mental model of the completely ordered array.

### Reasoning Accuracy

A preliminary ANOVA tested whether there were any significant differences in reasoning accuracy between control groups in Experiments 1 and 2. The design of this ANOVA was 2 (type of control group: university students from Potsdam vs. high school students from Warsaw) × 3 (order difficulty: Orders 1 and 2 vs. Orders 3 and 4 vs. Orders 5 and 6; within subjects) × 3 (pair distance: adjacent vs. two step vs. end point; within subjects). The main effect of type of control group was nonsignificant ($F < 1$), as were all interactions between type of control group and any other within-subjects variables. Therefore, we might assume that both control groups were equivalent in terms of reasoning accuracy.

Accuracy means (proportion of correct answers) are shown in Figure 5. A 2 (age: older adults vs. students; between subjects) × 3 (order difficulty: Orders 1 and 2 vs. Orders 3 and 4 vs. Orders 5 and 6; within subjects) × 3 (pair distance: adjacent vs. two step vs. end point; within subjects) ANOVA revealed a strong main effect of age, $F(1, 32) = 31.56, MSE = 2,684, p < .001, \eta^2 = .50$. The percentage of correct answers was lower for older adults ($M = 70\%$) than for high school students ($M = 93\%$). There was also a significant Age × Order Difficulty interaction, $F(2, 64) = 3.33, MSE = 1,471, p < .05, \eta^2 = .09$. As is clearly visible from Figure 5, the more difficult the order type, the higher was the proportion of correct answers in the group of older adults. This was confirmed by a significant linear trend in the older adult group across order difficulty levels, $\eta(32) = 2.74, p < .01$. There were no significant changes across order difficulty levels in the control group. To better understand the unexpected pattern of Age × Order Difficulty interaction, we carried out correlational analyses between accuracy for adjacent relations and study times for each presented pair. In the group of older adults, these correlations were very high: A zero-order Pearson correlation, $r(15)$, was equal to

---

![Figure 4. Experiment 2: The mean studying times as a function of group (age), order difficulty, and pair distance. Error bars represent standard errors of the mean.](image-url)
.55 for the first pair, .58 for the second pair, and .60 for the third presented pair (all significant at $p < .05$). In other words, the more older adults studied the presented relations, the better they preserved them during retrieval in the test phase. The most difficult relations were studied most intensively, and the retrieval of presented pairs (accuracy for adjacent relations) was also best for the most difficult orders. It is interesting to note that the above correlations were significant only in the group of older adults, and they were not significant for the depressed or control groups from Experiments 1 and 2.

In contrast to the previous study on depression (Experiment 1), there was no tendency for an interaction between age group and pair distance. As is visible from Figure 5, for both age groups, accuracy for inferred relations (two-step pairs and end-point pairs) is similar to accuracy for presented adjacent pairs.

Hierarchical Regression Analyses for End-Point Reasoning Accuracy

The data concerning end-point accuracy were again subjected to a hierarchical regression analysis. The aim of this analysis was to assess the effect of age (categorical group variable) on end-point accuracy after we had partialed out the variance associated with memory retrieval (accuracy for adjacent relations). The results are summarized in Table 1 (middle section), showing a pattern completely different from the findings in Experiment 1. The initial $R^2$ for the relation between age (older adults vs. students) and accuracy was .245 (significant at the .01 level), whereas the $R^2$ for age after memory retrieval was controlled for was dramatically reduced to .003 ($ns$), which corresponds to attenuation of 98.8%. That is, memory retrieval almost completely attenuated the effect of age on reasoning accuracy concerning end-point relations. This means that if the statistical procedure equalizes the memory for presented pair relations, age differences in reasoning concerning end-point relations are virtually nonexistent.

Discussion

Following the work by Salthouse and colleagues (Salthouse, 1992; Salthouse et al., 1989, 1990), we had predicted deficits among older adults in terms of maintenance functions in working memory. Older adults should not, however, show marked decrements in terms of integrative processing. In line with these predictions, this experiment revealed that differences in memory for adjacent relations almost completely explained the observed group difference in reasoning accuracy concerning end-point relations. Older participants performed worse than younger participants at all levels of pair distance, which disconfirms generative processing as a model of their deficit pattern. The GP model would have predicted an increase in deficit as a function of increasing pair distance. In terms of order difficulty, there was actually better performance in the older adult participants for the most difficult orders, 5 and 6, as compared with their performance on easier orders. During learning, older participants exposed themselves more extensively to the materials than did the younger participants. In conclusion, the MR model has received substantial support from the present data.

The finding that older participants actually performed better on more difficult orders was unexpected and may reflect greater thinking effort exerted by older adults vis-à-vis the more difficult learning materials, which would appear plausible from an adaptive processing viewpoint. It has, in fact, been shown that older participants allocate processing resources in a highly selective way, engaging in more resource-demanding operations only if cues during learning point to the necessity to engage in such operations (Hess, Bolstad, Woodburn, & Auman, 1999). In a similar vein, the presentation sequence of Orders 5 and 6, in which two unrelated pairs precede the third that finally allows integration, could have acted as a cue indicating increased difficulty. This, in turn, could have triggered the allocation of more resources to the memorization process. Also, the observation that older, unlike younger, adults spent disproportionately more time at studying the third pair in Orders 5 and 6 as compared with simpler orders is in line with this reasoning. This account is post hoc and must remain speculative, but it shows that the findings themselves can be accommodated within existing theories about adaptive cognition in older age.
Experiment 3

Experiments 1 and 2 showed that the linear order paradigm can reveal qualitatively different patterns of processing limitations in depressed students (GP model) and older adults (MR model), in comparison with control groups. In Experiment 3, we aimed (a) to test the differences between control participants and depressed participants in reasoning accuracy when both groups are exposed to restricted study times and (b) to examine whether a restriction of study times in the control group would eventually be a model of the specific processing limitations seen in older adults or, alternatively, in the depressed group.

Effects of Restricted Study Time on Reasoning Accuracy for Depressed and Nondepressed Participants

According to our explanation of the group differences in Experiment 1, nondepressed participants relied on a generative strategy of constructing a mental array during the learning stage from the incoming relations. Meanwhile, depressed participants spent their learning time on an attempt to store the information itself. Therefore, a straightforward implication is that a drastic restriction of the available study time should impair nondepressed participants’ reasoning, whereas it would leave depressed participants’ performance unaffected. In particular, nondepressed participants should not be able to effectively construct the mental model during a curtailed presentation interval. Conversely, depressed participants’ reasoning should be less, if at all, affected, under the assumption that the restriction still leaves them sufficient time to just memorize the presented relations. To ensure sufficient time to yield above-chance memory performance overall, we applied a fixed presentation interval of 5 s for all presented relations. This presentation interval has been used in previous research (e.g., Moeser & Tarrant, 1977; K. H. Smith & Foos, 1975) and has yielded satisfactory performance levels.

Simulation of Processing Limitations

Concerning a second goal of this experiment, we were influenced by Salthouse (1992), who showed that a cognitive load imposed on younger adults during the study phase induced a pattern similar to that previously obtained for older adults when integrative reasoning was subsequently tested. Furthermore, Kliegl, Krampe, and Mayr (2003) reviewed evidence supportive of the idea that the conceptually simpler procedure of restricted study time during encoding might be regarded as a basic standard procedure for modeling complex information processing deficits in control participants. Therefore, we were interested whether a restriction of study times would produce processing limitations in linear order reasoning similar to that revealed in older adults in Experiment 2 (retrieval processing limitations) or, alternatively, similar to the thinking problems of depressed students in Experiment 1 (generative processing limitations).

On the basis of Experiment 2, if a restriction of study time in the control group did produce a good fit to the MR model, then one should expect a main effect of group on reasoning accuracy (comparison between the self-paced control group from Experiment 1 vs. the restricted-study-time control group from Experiment 3). In particular, one should expect similar group differences (i.e., impairments in the control group with restricted study times) for adjacent relations as well as for to-be-inferred two-step and end-point relations. Also, the effect of group on reasoning accuracy concerning end-point relations should be substantially attenuated by memory retrieval (i.e., the accuracy for presented adjacent relations).

However, if the GP model was more adequate, then one should expect the control group with fixed study time to exhibit a pattern similar to that observed for depressed participants, hence, an interaction between group (control group without vs. control group with study time restriction) and pair distance for reasoning accuracy.

Method

Participants

Students enrolled in psychological, historical, and educational sciences at the University of Potsdam participated in the experiment. The assignment, credit, and payment procedures were the same as described in Experiment 1. The final sample comprised 36 female and 3 male students (mean age = 21.25 years, SD = 3.17 years, 19 in the nondepressed group (17 women and 2 men) and 20 in the depressed group (19 women and 1 man). The mean BDI scores at the second administration for the nondepressed and depressed participants were, respectively, 2.05 (SD = 1.61) and 15.45 (SD = 4.59).

Materials

The same materials as in Experiment 1 were used. All material specifications and all permutation and randomization measures were the same as in Experiment 1.

Procedure

The experimental sessions were initiated on the second appointment after participants completed the BDI questionnaire, provided the participant’s initial group assignment (control or depressed) was replicated. All procedural specifications were the same as in Experiment 1, with the exception of the learning stage. During each learning stage, participants were presented with three relations in the form of stimulus sentences, for example, “Christine is older than Katharina” in normal ASCII letters in the center of the screen. Each sentence was preceded by a 1,500-ms presentation of an "X" in the center as a preparing signal. The stimulus sentence was then presented and remained on the screen for 5,000 ms. All procedures in the test stage were identical to those in Experiment 1.

After completing all experimental tasks, participants were debriefed. Sessions were conducted in small groups of 1 to 3 persons in a laboratory of the Social Psychology Unit of the Department of Psychology at the University of Potsdam and lasted about 40–50 min.

Results

Results for Comparisons Between Depressed and Nondepressed Participants

As in Experiment 1, the accuracy index (proportion of correct answers) was examined by a 2 × 3 × 3 ANOVA as a function of group, order difficulty, and pair distance. The mean accuracy data are depicted in Figure 6. This analysis yielded only marginal significance for the main effect of pair distance, $F(2, 74) = 2.49, MSE = 888.00, p < .08, \eta^2 = .06$. Planned post hoc comparisons
showed that the accuracy for adjacent pairs was higher than for end-point pairs. When testing planned contrasts for the pair distance variable, we saw, as in Experiment 1, a clear linear trend in the depressed group, \( t(37) = 2.11, p < .05 \). This indicated that in this group accuracy levels decreased from adjacent to end-point pairs, as predicted by the GP model. Another planned contrast also revealed a marginally significant quadratic trend for the nondepressed group, \( t(37) = 1.84, p < .08 \).

Results for Simulation of Processing Limitations

**Retrieval accuracy.** A 2 (type of control group: self-paced from Experiment 1 vs. restricted study time from Experiment 3; between subjects) \( \times \) 3 (order difficulty: Orders 1 and 2 vs. Orders 3 and 4 vs. Orders 5 and 6; within subjects) \( \times \) 3 (pair distance: adjacent vs. two-step vs. end-point relations) ANOVA revealed only a main effect of type of control group, \( F(1, 36) = 6.13, MSE = 1.432, p < .02, \eta^2 = .15 \). The proportion of correct responses was lower in the control group with study time restriction. There was no indication of a Type of Control Group \( \times \) Pair Distance interaction (\( F < 1 \)). This supports the conclusion that the restriction of study times in fact produced memory retrieval limitations in the control group with study time restriction. There was no statistical relation between type of control group (free pace vs. restricted study time) and accuracy for end-point relations was .109 (significant at the .05 level), whereas the \( R^2 \) for type of control group after we controlled for memory retrieval was seriously reduced to .043, which corresponds to attenuation of 60.6%. That is, memory retrieval substantially attenuated (to a nonsignificant level) the effect of control group differences on reasoning accuracy concerning end-point relations. This offers further and the most compelling evidence that the MR model provides a valid explanation for the data in linear order reasoning obtained under restricted study times, similar to the type of limitation observed among older adults.

**Discussion**

This experiment was undertaken on the basis of our interpretation of the findings from Experiment 1. Nondepressed control participants, as we reasoned, spent their freely allocated study time on generating a mental array, whereas depressed participants simply rehearsed the individual relations. Given that nondepressed control participants presumably used a more complex strategy (maintenance and integration) than depressed participants (only maintenance), a time restriction would probably affect those more who used a more complex strategy (see Engle, 1996). The empirical question was in which way this would happen. As the comparison between control groups from Experiments 1 and 3 show, there is reason to assume that the limitation that control participants incurred as a result of time restriction is best described by the MR model, which also makes the time-restricted control condition a suitable model of age-related deficits in linear reasoning. As it seems from these findings, integrative reasoning in older adults, or in nondepressed younger people who work under time restrictions, is spontaneously engaged and mostly preserved, whereas the maintenance function of working memory suffers. It is important to note that the control participants, who were assumed to engage in a more complex processing strategy than the depressed participants, were the only ones to be affected by the time restriction in the present experiment.
Depressed participants’ performance, in contrast, was not at all impaired by the study time restriction. Their performance was strikingly similar to that in Experiment 1 and was again best described by the GP model. We conclude that the relatively simple rehearsal and maintenance strategy used by the depressed participants was not sufficiently curtailed by the restricted-time allocation to produce noticeable impairments. Depressed participants do not seem to have constructed mental array models from the presented pairs. Rather, and as a replication of Experiment 1, the linear trend in accuracy levels across pair distances suggests that they initiated propositional reasoning at the time of the query.

**Experiment 4**

The fourth experiment was undertaken for three reasons. First, we wanted to replicate the main findings from Experiments 1 and 2 in one and the same study to avoid the sole reliance on cross-experiment comparisons when assessing the relative merits of the MR versus the GP model in depression versus aging.

Second, the argument in refuting the MR model for depressed participants in Experiment 1 partly relies on a null effect. That is, no reliable difference between depressed and nondepressed participants in recall accuracy was observed for adjacent pairs. We used the above finding to argue that such a group difference would have been expected under the assumption that depressed participants had difficulties maintaining the presented pairs in working memory. Although other types of evidence support the general argument in favor of the GP model for depressed participants, we wanted to put this particular finding to a more rigorous test by increasing the sample size. If the memory retrieval deficit in depression is reliable but its effect size is relatively small, increasing the statistical power by more than double-sized samples should yield significant results. In a similar vein, this study also addresses the null effect problem concerning the absence of generative processing deficits among older adults in Experiment 2.

Third, we wanted to extend our findings by including measures of processing speed and working memory to see whether performance in one or both of our experimental populations was mediated by deficits in more elementary functions related to executive functioning. The results of Salthouse’s (1992) study on processing speed and working memory mediations of age differences in integrative reasoning demonstrated that statistical control of either processing speed or working memory resulted in big attenuation of age-related variance, with the role of processing speed being the more substantial.

The Digit Symbol Substitution test (DSST; a subtest of the Wechsler Adult Intelligence Scale—Revised [WAIS–R]; Wechsler, 1981) was used as a measure of processing speed. Our test of working memory was the Operation Span test (OSPAN; Turner & Engle, 1989), a measure that captures executive functions in terms of simultaneous maintenance and processing.

The processing speed theory (with the DSST as a principal measure) developed by Salthouse (1996) accounts for age-related differences in many cognitive tasks. According to this theory, the slower speed of executing cognitive operations in older age impairs performance in more complex tasks because of two mechanisms. The key defining feature of the first mechanism (time limited mechanism) is that because of time constraints, the processing of initial information is incomplete and defective. This is in contrast to the second mechanism (simultaneity mechanism) in which processing of initial information is complete but becomes unavailable in the later stages of processing because of decay. This second mechanism seems especially relevant here because it precisely captures the memory retrieval limitations observed among older adults in Experiment 2.

Older adults’ deficient performance in reasoning tasks might also be attributed to basic dysfunctions of working memory, especially to a dysfunction of the coordination between maintenance and processing (Miyake & Shah, 1999). In particular, OSPAN (Turner & Engle, 1989) is a measure that captures simultaneous maintenance and processing and is assumed to share variance with executive control functions. Engle and coworkers argued that the reason why a working memory measure such as OSPAN correlates with higher order cognitive functions is not individual differences in the storage component but individual differences in the central executive component—what they call a domain-unspecific executive attention component (Conway & Engle, 1996; Engle, Tuholski, Laughlin, & Conway, 1999; Kane & Engle, 2003).

Research on geriatric depression shows that older clinically depressed patients (65 years of age or older) are impaired on multiple cognitive and psychomotoric tests. Furthermore, these deteriorations may often be attributed to frontal lobe dysfunction in older depressed patients (Beats, Sahakian, & Levy, 1996; Boon et al., 1995; Elderkin-Thompson et al., 2003).

However, research evidence about the role of processing speed and working memory on cognitive functioning among younger and middle-aged depressed adults is inconsistent. In a study on executive functioning and verbal memory, Fossati, Amar, Raoux, Ergis, and Allilàire (1999) found no verbal memory deficits among depressed patients. However, these patients did demonstrate impairments on short-term memory tasks (verbal span tasks from the WAIS–R) and, much in line with our findings, showed pronounced deficits in complex integration, spontaneous cognitive flexibility, and initiation abilities. In a study on working memory dysfunction in depression (Pelosi, Slade, Blumhardt, & Sharma, 2000), depressed patients exhibited slower response times and more errors in a Sternberg working memory task. Channon, Baker, and Robertson (1993b) did not find differences among clinically depressed and control participants on articulatory loop and visuospatial sketch pad components of working memory. Among the range of different tasks measuring central executive, only backward digit span differed significantly between groups. Purcell, Maruff, Kyrios, and Pantelis (1997) demonstrated that middle-aged depressed patients were not impaired in cognitive speed but were worse in executive tasks like set shifting or planning. Very recent research (Rokke, Arnell, Koch, & Andrews, 2002) has shown that subclinically depressed students demonstrate executive attention deficits but only in complex dual-task conditions.

On the basis of this review, our predictions for Experiment 4 can now be more succinctly presented. If older adults’ performance was mediated by processing speed and operation span variables, then this would indicate a more general executive deficit in this population. In line with our previous findings, we predict that both DSST and OSPAN tasks should substantially attenuate the influence of aging on end-point accuracy. Conversely, for depressed individuals, only a highly specific and circumscribed impairment is predicted, that is, an impairment in the generation of complex mental representations (Sedek & Kočta, 1990; von Hecker &
Sedek, 1999). Therefore, even after the DSST and OSPAN tasks are taken into account, depression should remain highly predictive of end-point accuracy.

**Method**

**Participants**

The demographic characteristics of our samples are summarized in Table 2. High school students (nondepressed and depressed) had approximately one year less of formal education than older adults, $F(2, 114) = 17.16, p < .001, \eta^2 = .23$. The difference in gender proportion across groups was not significant, $\chi^2(2, N = 117) = 2.20, p < .33$.

The BDI selection criteria in both control and depressed groups were the same as in Experiments 1 and 3. Students filled out BDI questionnaires during morning classes, and those with a BDI score of 6 or above were tested experimentally during the same day in the afternoon. Older persons were recruited through advertisements at the University of Third Age. The criteria for older adults’ participation were again at least 11 classes of formal education and MMSE scores of at least 28. To make this sample more similar to samples reported in the research literature on old age, we tested experimentally during the same day in the afternoon. Older persons were tested first on a processing speed measure (DSST), then on a working memory test (OSPA), and finally on solving the linear order tasks.

**Processing speed.** We used the DSST (a subtest of WAIS–R; Wechsler, 1981). The number of correct responses during 60 s was used as the performance score.

**Working memory test.** The computerized version of the OSPAN (Turner & Engle, 1989) was used to assess working memory. On a computer screen, the participant is presented with an arithmetic problem in terms of the differences between groups for the most important dependent measures (post hoc analyses after a significant one-way ANOVA between groups).

<table>
<thead>
<tr>
<th>Demographic characteristic and cognitive task</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)***</td>
<td>17.92</td>
<td>0.47</td>
<td>17.88</td>
<td>0.40</td>
<td>69.65</td>
<td>2.96</td>
</tr>
<tr>
<td>Education (years)***</td>
<td>12.00</td>
<td>0.00</td>
<td>12.00</td>
<td>0.00</td>
<td>13.22</td>
<td>1.86</td>
</tr>
<tr>
<td>BDI***</td>
<td>2.97</td>
<td>1.35</td>
<td>17.78</td>
<td>4.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>44.17</td>
<td>7.64</td>
<td>42.72</td>
<td>5.88</td>
<td>21.70</td>
<td>6.38</td>
</tr>
<tr>
<td>DSST***</td>
<td>27.27</td>
<td>7.15</td>
<td>25.18</td>
<td>10.03</td>
<td>9.38</td>
<td>6.82</td>
</tr>
<tr>
<td>OSPAN***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear orders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Studying time (s)</td>
<td>10.92</td>
<td>4.61</td>
<td>9.37</td>
<td>3.95</td>
<td>12.80</td>
<td>11.60</td>
</tr>
<tr>
<td>Accuracy (% correct)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjacent relations***</td>
<td>88.89</td>
<td>9.66</td>
<td>85.69</td>
<td>14.45</td>
<td>64.86</td>
<td>11.86</td>
</tr>
<tr>
<td>Two-step relations***</td>
<td>91.04</td>
<td>9.69</td>
<td>84.79</td>
<td>15.89</td>
<td>61.71</td>
<td>16.95</td>
</tr>
<tr>
<td>End-point relations***</td>
<td>93.75</td>
<td>8.17</td>
<td>77.92</td>
<td>24.57</td>
<td>72.52</td>
<td>18.93</td>
</tr>
</tbody>
</table>

Note. For the nondepressed students, $n = 40$ (29 women, 11 men); for the depressed students, $n = 40$ (28 women, 12 men); for the older adults, $n = 37$ (31 women, 6 men). Means with different subscripts differ significantly at $p < .01$ by Newman–Keuls post hoc tests. BDI = Beck Depression Inventory; MMSE = Mini Mental State Examination; DSST = Digit Symbol Substitution test; OSPAN = Operation Span test. ***$p < .001$, one-way analysis of variance between groups significant.
Processing Speed and Working Memory

Processing speed (DSS scores; see Table 2) were analyzed in a simple ANOVA (group: depressed students vs. nondepressed students vs. older adults), resulting in a significant effect, $F(2, 114) = 134.47$, $MSE = 44.64$, $p < .001$, $\eta^2 = .70$. Newman–Keuls post hoc tests again revealed no difference between nondepressed and depressed students but revealed that these two groups scored higher than older participants ($ps < .001$). OSPAN scores (see Table 2) were analyzed in a similar way, yielding a significant effect, $F(2, 109) = 53.49$, $MSE = 66.40$, $p < .001$, $\eta^2 = .50$. Newman–Keuls post hoc tests revealed that nondepressed and depressed students were not reliably different in OSPAN but that these two groups had higher OSPAN scores than did older participants ($ps < .001$).

Study Times

Study times were analyzed in a 3 (group: depressed students vs. nondepressed students vs. older adults; between subjects) x 3 (order difficulty: Orders 1 and 2 vs. Orders 3 and 4 vs. Orders 5 and 6; within subjects) x 3 (sequential pair: first vs. second vs. third; within subjects) ANOVA. There was no group main effect (see Table 2). However, there were significant main effects of order difficulty, $F(2, 228) = 3.99$, $MSE = 90.24$, $p < .02$, $\eta^2 = .03$, and sequential pair, $F(2, 228) = 8.11$, $MSE = 105.38$, $p < .001$, $\eta^2 = .07$. Newman–Keuls tests revealed that Orders 5 and 6 needed more time to be studied than Orders 1 and 2 ($p < .01$) and that third pairs were studied longer than second ($p < .001$) and first pairs ($p < .02$). The two main effects were qualified by an interaction, $F(4, 456) = 6.79$, $MSE = 70.42$, $p < .001$, $\eta^2 = .06$, which was due to the third pair being studied at exceptionally long times in Orders 5 and 6; all pertinent Newman–Keuls comparisons for the three groups were at $p < .001$. These results are clear and replicate our earlier findings. Third pairs in the most difficult orders need longer than any other pair, and the three groups show similar patterns of study time allocation.

Reasoning Accuracy

Accuracy means (proportion of correct answers) are shown in Table 2. The scores were analyzed by a simple ANOVA (group: depressed students vs. nondepressed students vs. older adults) and 6 (within subjects) ANOVA. There was no group main effect ($p < .02$). Furthermore, for end-point queries, nondepressed participants performed better than both depressed, $t(114) = 3.83$, $p < .001$, and older participants, $t(114) = 5.03$, $p < .001$, but there was no reliable difference between depressed students and older adults ($p < .20$). In the nondepressed control group and for older adults, significant linear trends indicated increasing means from adjacent to end-point queries, $t(114) = 2.09$, $p < .05$, and $t(114) = 3.61$, $p < .001$, respectively. However, in the depressed group, the linear trend was also significant, but indicating an opposite effect, that is, a substantial decrease in performance across distance levels, $t(114) = 3.33$, $p < .001$. Different from Experiment 2, in this experiment, we did not observe any particular performance improvement of older adults in the most difficult linear order type. We analyzed study times across studies to understand this pattern. It appears that in Experiment 2, participants studied the difficult problems (5 and 6) significantly longer than the remaining problems: planned comparisons, $t(32) = 2.10$, $p < .05$. That is, they exerted exceptional efforts to analyze the most difficult part of the information. In Experiment 4, although older adults again studied the most difficult type of order longest, the associated increase in study time was not significant: results of planned comparisons, $t(114) = 1.55$, $p < .13$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>$\Delta F$</th>
<th>$df$</th>
<th>% attenuated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group: Depressed vs. student control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 1</td>
<td>.161</td>
<td>14.95***</td>
<td>1.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 2</td>
<td>.472</td>
<td>69.78***</td>
<td>1.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory retrieval</td>
<td>.570</td>
<td>.099</td>
<td>17.70***</td>
<td>1.77</td>
<td>38.5</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Group: Older adults vs. student control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 1</td>
<td>.358</td>
<td>41.90***</td>
<td>1.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 2</td>
<td>.425</td>
<td>55.52***</td>
<td>1.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory retrieval</td>
<td>.453</td>
<td>.028</td>
<td>3.76</td>
<td>1.74</td>
<td>92.2</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The dependent variable was accuracy for end-point relations; the memory retrieval variable was operationalized as the accuracy for adjacent relations. *** $p < .001$. 

Hierarchical Regression Analyses for End-Point Reasoning Accuracy

Attenuation analyses were carried out for the two experimental samples separately, each being compared with the nondepressed control group. The results are summarized in Tables 3 and 4. Table 3 replicates earlier results (Experiment 1), showing that the initial $R^2$ for the relation between depression (depressed vs. nondepressed students) and end-point accuracy was not significantly mediated.
by taking memory for adjacent pairs into account (39%), whereas the $R^2$ for age after the same control step was very substantially reduced by 92%. Again, memory retrieval almost completely attenuated the effect of age on reasoning accuracy concerning end-point relations but not the effect of depression on reasoning accuracy.

Table 4 shows an extension of the findings so far by looking at measures of working memory and mental speed. Both depression and OSPAN explained substantial proportions of variance concerning the generative reasoning (end-point accuracy) among non-depressed and depressed students, but the influence of processing speed was nonsignificant. It is apparent that the relation between depression and accuracy is not much attenuated by either OSPAN or DSST, whereas these two measures in conjunction reduce the impact of age differences on accuracy by 88%. Table 4 also shows that the OSPAN and DSST individually have a sizeable attenuation effect on age. The OSPAN and DSST were correlated significantly only among older adults, $r(35) = .36, p < .05$. This has implications for the relative role of DSST and OSPAN, depending on the order of entering (when OSPAN was entered first and DSST second, both predictors were significant at $p < .01$). However, independent of the order of entering, both variables combined attenuated the age effect to the same degree (87.7%).

Summary of Size Effects and Power Analyses

Table 5 presents the summary size effects and statistical power for Experiments 1, 2, and 4 on the two principal dependent measures. First included were the effects of groups (depression and age) on accuracy for adjacent orders (testing the memory retrieval limitations). Second, we included the effects of groups (depression and age) on an index of integration, calculated as the difference in accuracy between adjacent relations and end points. The more positive this difference, the more substantial are the generative

---

**Table 4**

Results of Hierarchical Regression Analyses on Generative Reasoning (Experiment 4, Including OSPAN and DSST)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>$\Delta F$</th>
<th>df</th>
<th>% attenuated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group: Depressed vs. student control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 1</td>
<td>Group</td>
<td>.161</td>
<td>14.95***</td>
<td>1, 78</td>
<td></td>
</tr>
<tr>
<td>Analysis 2</td>
<td>DSST</td>
<td>.005</td>
<td>0.39</td>
<td>1, 78</td>
<td></td>
</tr>
<tr>
<td>Analysis 3</td>
<td>Group</td>
<td>.174</td>
<td>15.75***</td>
<td>1, 77</td>
<td>-5.0</td>
</tr>
<tr>
<td>Analysis 4</td>
<td>OSPAN</td>
<td>.174</td>
<td>15.40***</td>
<td>1, 73</td>
<td>25.5</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>.294</td>
<td>12.23***</td>
<td>1, 72</td>
<td>23.6</td>
</tr>
<tr>
<td><strong>Group: Older adults vs. student control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis 1</td>
<td>Group</td>
<td>.358</td>
<td>41.90***</td>
<td>1, 75</td>
<td></td>
</tr>
<tr>
<td>Analysis 2</td>
<td>DSST</td>
<td>.302</td>
<td>32.50***</td>
<td>1, 75</td>
<td>82.7</td>
</tr>
<tr>
<td>Analysis 3</td>
<td>Group</td>
<td>.365</td>
<td>7.25**</td>
<td>1, 74</td>
<td></td>
</tr>
<tr>
<td>Analysis 4</td>
<td>OSPAN</td>
<td>.243</td>
<td>23.12***</td>
<td>1, 72</td>
<td>63.7</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>.373</td>
<td>14.74***</td>
<td>1, 71</td>
<td>87.7</td>
</tr>
</tbody>
</table>

Note. The dependent variable was accuracy for end-point relations. OSPAN = Operation Span test; DSST = Digit Symbol Substitution test. * $p < .05$. ** $p < .01$. *** $p < .001$.

**Table 5**

Effect Sizes and Statistical Power of the Main Results

<table>
<thead>
<tr>
<th>Statistical effect</th>
<th>Mean difference (%)</th>
<th>Effect size (partial $\eta^2$)</th>
<th>Statistical power ($1 - \beta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect of depression on adjacent accuracy</td>
<td>3.5</td>
<td>.037</td>
<td>.210</td>
</tr>
<tr>
<td>Effect of depression on adjacent end-point accuracy</td>
<td>13.2</td>
<td>.110*</td>
<td>.539</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect of age on adjacent accuracy</td>
<td>21.6</td>
<td>.512***</td>
<td>1.000</td>
</tr>
<tr>
<td>Effect of age on adjacent end-point accuracy</td>
<td>-1.0</td>
<td>.001</td>
<td>.053</td>
</tr>
<tr>
<td><strong>Experiment 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect of depression on adjacent accuracy</td>
<td>3.1</td>
<td>.017</td>
<td>.209</td>
</tr>
<tr>
<td>Effect of depression on adjacent end-point accuracy</td>
<td>12.7</td>
<td>.193***</td>
<td>.989</td>
</tr>
<tr>
<td>Effect of age on adjacent accuracy</td>
<td>24.0</td>
<td>.560***</td>
<td>1.000</td>
</tr>
<tr>
<td>Effect of age on adjacent end-point accuracy</td>
<td>-2.8</td>
<td>.010</td>
<td>.136</td>
</tr>
</tbody>
</table>

* $p < .05$. *** $p < .001$. 253 DEPRESSION, AGING, AND LINEAR ORDERS
processing limitations. A negative difference value indicates a distance effect (end points being better retrieved from the mental array than adjacent relations).

The results from Experiments 1 and 2 did replicate very well in Experiment 4. In Experiments 2 and 4, older adults were about 22%–24% less accurate than nondepressed students in preserving the premises for further reasoning (accuracy for adjacent relations), demonstrating clear memory retrieval limitations, and the results were very substantial in terms of size effects and statistical power. For both experiments (in Experiment 4, there were more than twice as many participants than in Experiment 1), the influence of depression on accuracy for adjacent relations was negligible.

Just the opposite results emerged for the composite measure of generative reasoning. Depressed participants in both Experiments 1 and 4 were around 13% worse in accuracy than nondepressed students on the generative reasoning measure (difference in accuracy between adjacent and end-point relations). The results for both experiments were very similar in mean differences, but effect size and power were higher in Experiment 4. Comparing nondepressed students and older adults, conversely, we found negligible differences for this composite score and only a very weak effect size.

**Discussion**

The present experiment has replicated the main pattern of findings from Experiments 1 and 2 in a single study, thereby increasing the internal validity of our comparisons between depressed and older adult populations. With substantially increased test power, the pattern now emerges even clearer. The linear increase in accuracy across pair distances in the nondepressed student group reflects a distance effect (Leth-Steensen & Marley, 2000; Potts, 1972, 1974; K. H. Smith & Foos, 1975). Moreover, it is also an indication of spontaneous mental model construction during the learning stage. Pairs are integrated online into a linear array (see Figure 1), which then serves as a retrieval device during the test stage.

The linear decrease in accuracy across pair distances in the depressed group demonstrates, once again, a lack of spontaneous integration, as hypothesized by the GP model. Participants in this group most likely use a simple rehearsal strategy during learning and engage in reasoning only at the time of the query. It is important to note that this interpretation is compatible with assumptions from control motivation theory (Pittman & D’Agostino, 1989; Weary et al., 1993) about unimpaired cognitive effort in depression. As the cognitive exhaustion model explains (Sedek & Kofta, 1990; von Hecker et al., 2000), exhaustion states are characterized by a selective failure to engage in complex, more demanding cognitive strategies, whereas simpler strategies are preserved. In line with this view, depressed students’ study behavior in this experiment again showed the same discrimination between easy and difficult pairs during learning as was shown by the nondepressed group. Depressed as well as nondepressed students do recognize the relative diagnosticity of a stimulus. In their attempts to further process these stimuli during the learning stage, as we hypothesize, nondepressed individuals spontaneously initiate an integrative strategy, whereas depressed individuals, because of their deficit, remain at the level of simpler rehearsal strategies.

The MR model received no support for the depressed group for two reasons. First, contrary to the prediction of this model, there is no difference in memory for adjacent pairs between depressed and nondepressed students, and we were able to test this comparison with increased power as compared with Experiment 1. Second, the decreasing linear trend in the depressed group is not compatible with the MR model.

The group of older adult participants again performed at a clearly lower level of accuracy overall, which again gives support to the MR model. Given that participants in this group were able to successfully recollect the pair information from the learning stage, their linear reasoning seems unimpaired, thus failing to support the GP model. Consistent with this interpretation, as regression analyses showed, the age effect on end-point accuracy was largely attenuated by taking memory for adjacent pairs into account. However, the present experiment goes beyond the earlier experiments in this series. This is so because measures of working memory and mental speed were included in further regression analyses, demonstrating that more fundamental executive functions, as measured by OSPAN, can account for the older adults’ but not the depressed participants’ performance. The fact that older adults’ performance was mediated by their OSPAN points to maintenance problems in the sense of a deficit in executive processes.

In Experiment 4, the older adults did not perform especially well in the most difficult problems, such as observed in Experiment 2. Additional analyses showed that in Experiment 2, older adults studied the most difficult orders exceptionally long, but this motivational tendency was not seen in Experiment 4. We attribute this difference to a negative impact of the included measure of working memory capacity (OSPA). Our argument is that when working on the OSP, older adults experienced a much larger proportion of failures than the younger participants and therefore were especially discouraged to invest more mental efforts in the subsequent, more complex linear order tasks. The series of failures in the memory task (clearly visible and largest among the older adults) might have had negative consequences for solving the target tasks (linear order problems) because of possible carryover effects such as self-stereotype activation (Wheeler & Petty, 2001). We focus more on the important problem of potential emotional and motivational carryover effects to be expected in repeated measures designs in the General Discussion.

Conversely, individuals in depressed states, as we argue (Sedek & Kofta, 1990; von Hecker & Sedek, 1999), are not impaired in executive processing per se but, more specifically, in generative thinking. In line with this view, depressed participants’ performance deficits in constructing linear orders, although clearly visible and replicated in this experiment, were not mediated by OSPAN or the DSST.

The results for effect sizes and the power analyses also support the assumption of specific and different cognitive limitations in linear order reasoning among older adults versus depressed students. These analyses yielded a replication across studies and ruled out the possibility that the obtained findings were due to insufficient statistical power. The replicated findings show that the effect sizes for memory retrieval limitation in depression and for general processing limitation among older adults were negligible even in larger samples.
General Discussion

In summary of our findings, the GP model was most successful in describing the performance of depressed students. Both in Experiment 1 (self-paced procedure) and Experiment 3 (restricted study time), a decreasing linear trend appeared across the levels of the pair distance variable. The larger the query distance, the more depressed students’ performance was impaired compared with the nondepressed students’ performance. Presumably, this was because greater distances demand more advanced integrative thinking. For adjacent pair queries, which constituted a memory test for just-presented materials, no differences between depressed and nondepressed students were observed in Experiments 1, 3, or 4. In the latter experiment, we replicated the earlier results with more test power. Additionally, regression analyses in Experiments 1 and 4 showed that even after we accounted for memory retrieval, the depressed–nondepressed group effect on end-point accuracy remained impressively unchanged. End-point queries capture the most generative aspect of reasoning in our paradigm. Both depressed and nondepressed students spent more time analyzing more difficult orders, and they devoted longer study times to third pairs, that is, pairs that allowed the integration of all premises into a coherent mental array model. Thus, the pattern of study time helps to refute an explanation of the generative deficit in terms of a generally lowered motivation in depressed people.

Alternatively, the MR model received support in older adults in three main findings: (a) In Experiment 2, older adults in comparison to younger students had serious limitations in remembering the premises; (b) older adults were clearly able to construct linear orders, in particular responding to end-point queries, provided they had correctly preserved the premises; and (c) Experiments 2 and 4 showed that the effect of age on generative reasoning was nearly completely attenuated in hierarchical regression analyses when we controlled for maintenance of adjacent relations in memory during processing. In terms of similar study time patterns between older adult participants and students, lack of motivation could not serve as an explanation for the older adults’ impaired reasoning performance.

A clear pattern emerges from this set of findings. There are qualitative rather than quantitative differences in the processing limitations among depressed students versus older adults when they construct transitive linear orders. Depressed students had limitations in generative processing but no problems with memorizing the presented premises. Older participants, on the contrary, suffered from serious limitations in remembering the premises, whereas their generative reasoning ability was impressively intact. We discuss these two findings in light of the literature on reasoning and mental model construction in depression and aging.

Generative Processing Limitations in Depression: Interpretation and Relations to Other Research

Our results from Experiments 1, 3, and 4 are incompatible with the capacity limitation view of depression (Ellis & Ashbrook, 1988; Ellis, Ottaway, Varner, Becker, & Moore, 1997). This explanation suggests that cognitive capacity is reduced in depression because depressed people allocate a portion of their attentional resources to ruminative thoughts or task-irrelevant features of the task. Therefore, for Experiment 1, the capacity limitation model would have predicted a different pattern than the one observed: The more cognitively complex the task, the more striking the reasoning differences should be between depressed and control students (for similar predictions concerning the effort required by a task, see also Hartlage, Alloy, Vazquez, & Dykman, 1993). This theory suggests we should have observed a disproportional decrease in performance as a function of order difficulty in the depressed group, as compared with the nondepressed group. In fact, however, neither Experiment 1 nor 4 showed such an interaction, despite considerable test power in Experiment 4. Furthermore, for Experiment 3, the capacity limitation model would also predict a larger overall difference between the two groups than was observed in Experiment 1. Restricting study time should render the reasoning task more cognitively demanding. Both predictions from the resource allocation model received no support in our studies.

Our data also speak to the lack-of-initiative model by Hertel and her associates (Hertel, 1997; Hertel & Rude, 1991), which was primarily developed to organize memory data. The notion here is that in depression, attentional strategies that are beneficial to task performance are not spontaneously engaged. This model could potentially explain our data, if one assumes that the strategies that support successful transitive inferences, hence the construction of a mental array, are not spontaneously engaged during the learning stage. Our paradigm allows us to distinguish evidence for simple memorization during learning from evidence for a more complex constructive strategy. Therefore, our findings suggest that it might be first and foremost those more complex, integrative functions and operations in working memory that suffer from a lack of spontaneous engagement. The more basic functions such as rote rehearsal or vigilance do not seem to be as much affected. Note that the diagnostic value of the third pairs in Orders 5 and 6, the ones that enable the final constructive step, are well noticed and attended to by depressed and nondepressed students alike. Independent evidence for the claim that depression affects complex strategies more than simple strategies comes from direct comparisons between single- and dual-task situations, in which depressed students consistently perform disproportionally worse in the latter, compared with their nondepressed counterparts (Rokke et al., 2002). However still, we would make a caveat as to the full applicability of the lack-of-initiative model to our present data. The model claims (for evidence on this point, see Hertel & Rude, 1991) that in a task that provides a tighter structure and guidance in maintaining the focus of attention, depressed individuals’ deficits should disappear, and they should be able to apply complex strategies just as well as nondepressed individuals. It is clearly inconsistent with this notion that in Experiment 3, which does provide considerably more structure, maintaining participants’ attentional focus by means of fixed presentation intervals, depressed participants’ performance (a) did not improve as compared with their overall accuracy level in Experiment 1 and, more importantly, (b) continued to show the same pattern of decreasing accuracy with pair distance—a finding which speaks against the spontaneous use of constructive strategies and in support of propositional reasoning at the time of the query.

We interpret the present findings concerning depression in terms of the cognitive exhaustion model (Kolfa & Sedek, 1998; von Hecker & Sedek, 1999). As outlined above, this model assumes that depressed individuals’ cognition is characterized by less sophisticated, less complex strategies as a result of ongoing ineffec-
tive mental effort, induced by circular thoughts, counterfactual thinking, or ongoing rumination (Nolen-Hoeksema, 2000). Specifically, it seems likely that constructive, integrative functions within working memory are primarily affected by ongoing sad mood. Recent attempts to provide more theory-based measurement models for working memory functions come from studies using latent variable analysis (e.g., Miyake et al., 2000; Oberauer, Süß, Wilhelm, & Wittmann, 2003). Within those approaches, executive functions in working memory are distinguished, such as “shifting,” “inhibition,” “updating” (Miyake et al., 2000), or “co-ordination” (Oberauer et al., 2003). The coordination function is defined as going beyond the mere maintenance of task-relevant information and dynamically manipulating the contents of working memory. As a speculation, we submit here that a functional framework of this kind may benefit from the consideration of an additional “integration” function, which may be partly overlapping with, but not identical to, coordination. We suggest that integrating piecemeal information into a more coherent mental representation, as it is crucial in solving transitivity problems, is also fundamental as an executive function when it comes to explain problem solving, social cognition (von Hecker & Sedek, 1999), and higher order thinking in general (for similar arguments, see also Halford, 1993; Waltz, Lau, Grewal, & Holyoak, 2000). Our data on depressed individuals constructing linear orders could be explained in the sense that the spontaneous deployment of the executive function integration is impaired in depressed mood.

What further evidence from research on depressed individuals’ reasoning might be accommodated by this approach? Weingartner and his associates (Silberman, Weingartner, & Post, 1983; Weingartner, 1986) found “thinking deficits” in abstract reasoning tasks and discrimination learning among depressed participants that could not be explained in terms of basic memory, logic, or attentional limitations. Also, J. D. Smith et al. (1993) found a lack of generative reasoning under depression, using category learning problems. In line with our own viewpoint, these authors also found that a version of their task in which participants could rely on simple fall-back strategies was unimpaired in the depressed sample. Notably, the depressed participants’ specific deficit in the more generative version could not be attributed to the general level of task difficulty.

Although it is the main theoretical framework of this article, addressing depressed participants’ performance exclusively from the perspective of a cognitive deficit may be overly limited. As a case in point, a comparison between Experiments 1 and 3 shows performance stability in depressed but not nondepressed students. Although this is clearly a comparison across studies, we still find it worthwhile to speculate on the possibility that depressed states are not always detrimental to task execution. Reliance on more basic cognitive strategies such as memory rehearsal may sometimes be functional. Research on social perception in depressed and helplessness-trained individuals has shown higher levels of detail accuracy in comparison with nondepressed individuals (Edwards & Weary, 1993), as well as depressed individuals’ particular diligence in processing impressions and schematic information (Gannon et al., 1994; Pittman & D’Agostino, 1989). Defocused, noncomplex strategies might be particularly useful to turn away from earlier, unsuccessful action plans or to abandon mental models that are no longer useful, while widening the scope of perception and screening the environment for alternative and potentially useful stimuli. Such situations might arise particularly subsequent to severe loss or failure. A more open, unfocused, unselective, low-effort mode of cognition in depressed mood (cf., Sullivan & Conway, 1989) would then possibly be beneficial. Similar arguments may be made concerning a possible adaptive value of intrusive thoughts and ruminations. Those thoughts may help an individual to better cope with a distressing event (Horowitz, 1975, 1983; Taylor & Schneider, 1989) or may allow the individual to approach alternative routes of action, given previous action plans that have turned out unsuccessful (Martin & Tesser, 1989). This is not to deny the well-documented cognitive deficits that are associated with depressed mood, but in the light of our data, it seems useful to speculate on broader implications of such moods in the context of behavioral and cognitive adaptation.

Memory Retrieval Limitations in Cognitive Aging: Interpretations and Relations to Other Research

On the basis of previous research by Salthouse and his associates (Salthouse, 1992; Salthouse et al., 1989, 1990), we had proposed the MR model as an explanatory framework for our data. The findings from Experiment 3 showed that for the control group, a restriction of study time might be seen as a conceptual model of reasoning accuracy among older adults (for the same argument, see Kliegl et al., 2003). All critical effects observed for older adults in Experiment 2 were observed when comparing the two control groups from Experiment 1 with self-paced procedure and from Experiment 3 with restricted study time. Regression analyses in Experiment 4 revealed that the deficits observed in the older adult participants were mediated by measures of working memory and mental speed. The results confirmed the predictions from Salt- house’s (1996) theory that age differences in more complex cognitive tasks might be substantially explained by older adults’ lower processing speed. More generally, the results for processing speed and OSPAN tasks are in line with recent research on aging and executive functions (Engle, 1996; Mayr, Spieler, & Kliegl, 2001; Verhaeghen & Cerella, 2002). According to this view, aging deficits in higher order cognitive tasks can be explained by a deterioration in basic executive control processes.

These results, indicating an insufficient memory encoding in older adults, were further corroborated in additional correlational analyses on the data of Experiment 2, with a characteristic pattern: The more older participants studied the premises, the better their performance in the reasoning task, resulting in an unexpected interaction effect between age and order difficulty. The more difficult the linear order problems, the more time the older adults spent on studying and, finally, the better the quality of their integrative reasoning.

Although working memory and mental speed seem to drive the effects of age, our data suggest that generative thinking is mainly preserved in older age. This view converges with recent research showing the absence of higher order reasoning problems in older participants. Salthouse (2000, 2001) found stable age–performance relationships across more or less complex problems, using different reasoning tasks. Radvansky and collaborators (Radvansky, Zacks, & Hasher, 1996; Radvansky, Zwaan, Curiel, & Copeland, 2001) reported that younger and older adults showed the same pattern of information integration, suggesting that they were able to generate similar types of situation models. Radvansky et al.
(2001) studied age differences in text comprehension and found that aging effects were significant at the surface and textbase levels but not at the level of the generated mental model on what the described situation was about. However, we hasten to admit that this does not mean that all deficits of older adults in cognitive processing of complex tasks should be explained by relatively basic working memory deficits. For example, recent findings in studies using novel methodologies such as time–accuracy functions, dual-task performance, or task switching demonstrated age differences in task complexity effects that cannot be explained in terms of basic memory retrieval limitations or cognitive slowing (Oberauer & Kliegl, 2001; Verhaeghen & Cerella, 2002; Verhaeghen, Kliegl, & Mayr, 1997).

Because we argue that older individuals do not primarily have problems with higher order cognition, it is useful to once again spell out the characteristics of our task and exactly what type of cognitive processes this task may shed light on. Whereas other research has also demonstrated that age-related differences in integrative reasoning are minor, provided that correct information is indeed preserved in memory (Salthouse, 1992; Salthouse et al., 1989, 1990), the focus of our task is on integrative reasoning as a spontaneous cognitive activity. Although Salthouse’s (Salthouse et al., 1989) task does demand integrative reasoning, it does not offer an opportunity for the generation of a mental model spontaneously during the encoding of incoming information. The linear order task can detect an individual’s failure to spontaneously engage in the construction of the mental array A–B–C–D (see Figure 1). This spontaneous engagement of generative thinking, according to our view, is the key distinction between the types of impairment seen in depressed versus older individuals. It seems to be substantially affected in depression but preserved in older age.

Unexplored Carryover Effects of Repeated Measures Designs in the “Cold Cognition” Area

As we observed in Experiment 4, subtle preexposure with a single working memory task changed the study strategy among older adults in comparison with Experiment 2. The experience with a task that indicated memory failures discouraged these participants from making extra efforts in studying the most difficult order types. In the area of “hot cognition,” much research centers around specific carryover effects of failure manipulations, uncontrollability, stereotype threat, or stress (Klein & Boals, 2001; Schmader & Johns, 2003; von Hecker & Sedek, 1999; Wheeler & Petty, 2001). Repeated exposure to failure and inefficient cognitive functioning during repeated testing with cognitive tasks might have had impairing effects because of evoked autostereotypes, cognitive exhaustion, or subjective distress. We underscore this problem because it is completely ignored in recent extensive repeated measures research on cognitive functioning based on structural equation models (e.g., Engle et al., 1999; Oberauer et al., 2003; Salthouse, Atkinson, & Berish, 2003). The lack of continuous monitoring of emotional and/or motivational changes (e.g., mood, subjective stress, self-evaluation, subjective interest or intrinsic motivation, and fatigue) during prolonged and extensive testing with a plethora of cognitive tasks might be a strong disadvantage because control of potentially important carryover effects is lacking. In conclusion, there might be systematic and unexplored interaction effects between individual differences variables (e.g., age and working memory capacity) and testing order produced by those evoked emotional and motivational changes (cf. Keppel, 1973, pp. 395–400).

Conclusions and Further Research

Older age and depression are factors that can induce substantial impairments to complex measures of higher cognitive functioning. We have argued in this article that, as far as measures of generative reasoning are concerned, low performance in both populations, relative to control groups, reflects clearly different impairments of more basic functions in working memory: maintenance in older age and integration in depression. An analysis of those different profiles between impaired groups is valuable for theoretical reasons because such distinctions can help further investigate how different components of executive functioning in working memory might be affected by various types of individual differences. For example, applying the situational model paradigm (Radvansky et al., 2001) to a depressed population might potentially demonstrate a qualitatively different pattern from what has been observed for older adults. As we would predict, depressed participants in this paradigm would not differ from control participants at the verbal-timetextbase levels of information processing but might show reliable deficits in creating situational mental models from input information.

Another interesting issue for further research is that the cognitive correlates of depression seem to be qualitatively different in younger and older persons. As mentioned before, research in the area of geriatric depression indicates massive impairments in basic cognitive and psychomotor functions among depressed older adults. However, both our research and findings from other labs indicate that depression in younger and middle age people is highly specific and impairs mostly higher order mental activities. Making these distinctions is also valuable for applied considerations, such as the development of training programs that may focus on the development of metacognitive skills (Hegarty & Steinhoff, 1997). For example, external memory stores may support older people’s complex thinking in the context of daily routines. A recent study with older couples showed that those who had developed social skills to alleviate each others’ memory load by externalizing part of the to-be-remembered contents about daily routines performed better in subsequent memory tasks (Johansson, Andersson, & Roenneberg, 2000). Using external memory devices of various sorts, if part of developed metacognitive skills, may help to overcome older people’s problems with maintenance in the course of more complex cognitive activities. Conversely, depressed individuals may benefit most from metacognitive skills and developing new habits that would help them focus their attention (Hertel, 2004) and initiate appropriate integrative steps of reasoning, which they are perfectly able to perform in principle but do not engage spontaneously.

Surprisingly, the literatures on the effects of depression among younger adults and the effects of healthy cognitive aging on more complex thinking have tended to evolve and progress largely in separation of each other, although, as it was mentioned in the beginning of this article, they share many common theoretical explanations (cognitive resources and memory limitations; impaired inhibition; lowered efficiency of cognitive strategies; lack of cognitive initiative). We hope that our research, demonstrating
so clearly the integrative benefits of comparing well-defined models of processing limitations in reasoning tasks across different populations, might stimulate a research agenda for more collaborative projects between “cognitive agers” and “cognitive psychopathologists.”

References

diagnosticity versus behavioral consistency as determinants of impression change in adulthood. Psychology and Aging, 1, 77–89.


Johansson, O., Andersson, J., & Roenneberg, J. (2000). Do elderly couples have a better prospective memory than other elderly people when they collaborate? Applied Cognitive Psychology, 14, 121–133.


Revision received January 22, 2004
Accepted January 26, 2004

Wanted: Old APA Journals!

APA is continuing its efforts to digitize older journal issues for the PsycARTICLES database. Thanks to many generous donors, we have made great strides, but we still need many issues, particularly those published in the 1950s and earlier.

If you have a collection of older journals and are interested in making a donation, please e-mail journals@apa.org or visit http://www.apa.org/journals/donations.html for an up-to-date list of the issues we are seeking.