Error adaptation in mental arithmetic

Charlotte Desmet\textsuperscript{a}, Ineke Imbo\textsuperscript{a}, Jolien De Brauwer\textsuperscript{b}, Marcel Brass\textsuperscript{a}, Wim Fias\textsuperscript{a}, Wim Notebaert\textsuperscript{a}

\textsuperscript{a}Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
\textsuperscript{b}Expertise centre CODE, Lessius University College, Jozef De Bomstraat 11, 2018 Antwerp, Belgium

Corresponding author:
Charlotte Desmet
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2
9000 Ghent
Belgium
E-mail: charlotte.desmet@Ugent.be
Phone: +32-9-2649418
Fax: +32-9-2646496

Acknowledgments: This work was supported by a grant from FWO (G.0630.08) and by the Belgian Science Policy, Interuniversity Attraction Poles program (P6/29). We acknowledge the support of Ghent University (Multidisciplinary Research Partnership “The integrative neuroscience of behavioral control”).
Abstract

Until now, error and conflict adaptation have been studied extensively using simple laboratory tasks. A common finding is that responses slow down after errors. According to the conflict monitoring theory, performance should also improve after an error. However, this is usually not observed. In this study, we investigated if the characteristics of the experimental paradigms normally used could explain this absence. More precisely, these paradigms have in common that behavioral adaptation has little room to be expressed. We therefore studied error and conflict adaptation effects in a task that encounters the richness of everyday life’s behavioral adaptation, namely mental arithmetic, where multiple solution strategies are available. In accordance with our hypothesis, we observed post error accuracy increases after errors in mental arithmetic. No support for conflict adaptation in mental arithmetic was found. Implications for current theories of conflict and error monitoring are discussed.
Introduction

Over the past decades, a lot of research has been conducted on how we adapt our behavior following errors. A well replicated finding is that response times slow down after we encounter an error. This has served as evidence for adaptive control mechanisms taking place after an error. More specifically, an error is assumed to alter the point on the speed accuracy trade-off curve to a more conservative level, such that behavior will be slower but more accurate (e.g., Brewer & Smith, 1984; Brewer & Smith, 1989, Rabbitt, 1979). These ideas have been integrated in current accounts of error monitoring. For example the conflict monitoring account, explains post error slowing in terms of an increase in response thresholds after an error or conflict (Botvinick, Braver, Barch, Carter, & Cohen, 2001). According to these traditional accounts, response times will thus slow down and accuracy will increase after an error. Although post error accuracy increases have been reported in some studies (e.g., Laming, 1968; Marco-Pallares, Camara, Munte, & Rodriguez-Fornells, 2008; Danielmeier, Eichele, Forstmann, Tittgemeyer, & Ullsperger, 2011; Seifert, von Cramon, Imperati, Tittgemeyer, & Ullsperger, 2011), other studies failed to find post error accuracy increases (e.g., Hajcak, McDonald, & Simons, 2003, Hajcak & Simons, 2008; King, Korb, von Cramon, & Ullsperger, 2010; Notebaert & Verguts, 2011) or even observed post error accuracy decreases (e.g., Fiehler, Ullsperger, & von Cramon, 2005; Rabbitt & Rodgers, 1977). This contradiction was addressed by Notebaert et al. (2009). According to these authors, it is not the erroneous nature of the (incorrect) response but the fact that it occurs infrequently that explains post error slowing. More precisely, an infrequent event attracts attention away from the task and in this way slows down subsequent processing.
Notebaert and colleagues confirmed this hypothesis by showing post error slowing after infrequent errors, but post correct slowing after infrequent correct responses. In addition, they showed that irrelevant and infrequent sounds also slowed down performance on subsequent trials. In sum, according to these authors, post error slowing should be considered as an attentional effect (or an orienting response) rather than as an adaptation effect. In contrast to the conflict monitoring account, these authors do not predict post error accuracy increases but rather post error accuracy decreases. Indeed, because infrequent events such as errors capture attention, task processing will be impaired, resulting in post error accuracy decreases.

In the laboratory, error monitoring has been studied extensively with a wide range of tasks. Typically, in these tasks, different stimuli are mapped in an arbitrary way onto different responses. For example, a red square has to be responded with a left button and a green square with a right button. To fulfill such a task it is important to notice the difference between the stimuli, to remember the stimulus-response mappings and to push the according button appropriately. Thus, correct task performance in these tasks, comprises a correct identification of the stimulus and the selection and execution of the corresponding button. In the remaining part of the manuscript we will refer to these tasks as direct mapping tasks.

However, in daily life, appropriate behavior is not only determined by a simple stimulus. Often, more complex cognitive processes are required, e.g., when multiple solution strategies are possible. For example when a traffic light suddenly turns orange, you can choose between stopping or driving through. However, the behavior that will eventually be chosen will not only depend on the stimulus (the orange light) but also on
other factors; the police car next to you, the fact that you are already late for an appointment, your driving speed,... It is clear that, depending on the situation, some strategies are more efficient than others. So in contrast to direct mapping tasks, the stimulus identification in itself is not enough to elicit the most efficient behavior. In parallel, error adaptation in direct mapping tasks will be restricted to specific strategies, namely paying more attention to the relevant stimulus and refreshing the appropriate stimulus-response mappings. For example, consider a task where participants should respond with a left button to the digit 3 and with a right button to the digit 8. When the digit 3 is categorized incorrectly with the right button, the only possible strategy to prevent this error in the future, is to pay more attention to the presented digit and to recall the rule: 3-left, 8-right. In a way, there is thus not much to do after an error has been made, except for looking more attentively to the screen and to remember the response rules. Therefore, post error effects, measured in direct mapping tasks might indeed only reflect an orienting response (as suggested by Notebaert et al, 2009).

However, error adaptation in more complex situations should not be restricted to an attention increase or rule refreshment and might therefore elicit, besides an orientation response, also an improvement in behavior (post error accuracy increase). Unfortunately, such more complex tasks have largely been neglected in classical error monitoring research. In this paper we will fill this gap and investigate error adaptation effects in a more complex and daily used task that involves more cognitive processes and permits a selection between different strategies, namely mental arithmetic. Participants were asked to verify simple multiplication problems (e.g., 4 x 6 = 24 correct / false?). Different strategies have been documented in the verification of multiplications.
Besides retrieving the answer from memory and comparing the retrieved answer to the presented one, participants can also use different rules to verify the presented solutions. For example, the ‘five rule’ comprises that when one of the operands is 5, the product should have 0 or 5 as a final digit (Campbell & Graham, 1985; Siegler, 1988). Another example is the ‘parity rule’, which states that if at least one of the operands is even, the outcome must also be even (Krueger, 1986; Lemaire & Fayol, 1995; Lemaire & Reder, 1999; Masse & Lemaire, 2001). (For other examples of strategies in arithmetic verification see Ashcraft & Stazyk, 1981; Winkelman & Schmidt, 1974; Zbrodoff & Logan, 1986).

The advantage of using multiplication verification (and not production or another operation, like addition) is that we can manipulate the table relatedness of the presented distracters. As we know from previous studies, it is more difficult to reject a closely related distracter (e.g., 4 x 6 = 28) compared to an unrelated distracter (e.g., 4 x 6 = 14) (Campbell, 1987; Stazyk, Ashcraft, & Hamann, 1982). The idea of a multiplication ‘network’ in which activation spreading is at work, is now generally accepted (Ashcraft, 1987; Campbell, 1995; McCloskey & Lindemann, 1992; Verguts & Fias, 2005) and explains the influence of table relatedness in a verification task. Hence, in the present study we defined ‘conflict’ in terms of table relatedness: 4 x 6 = 14 is a low conflict trial since 14 is not related to the table of 4 nor to the table of 6. On the other hand, 4 x 6 = 28 is a high conflict trial since 28 belongs to the table of 4 (4 x 7). Further, our high conflict trials were always one step away from the correct outcome (4 x (6+1) = 28) to maximize the amount of conflict (further called ‘distance 1 distracters’). In this study, conflict is thus defined as the distance between the correct answer and the presented
answer. Consequently, conflict is not defined at a response level or at a stimulus level, as is the case in most laboratory tasks, but at a higher cognitive level.

A robust finding in the cognitive control literature is the fact that interference effects (the difference between high and low conflict trials) are smaller after conflicting stimuli. This effect was initially demonstrated by Gratton, Coles and Donchin (1992) in the flanker task and has now been observed in a wide range of tasks (Simon tasks: Sturmer, Leuthold, Soetens, Schroter, & Sommer, 2002, Stroop Tasks: Kerns et al., 2004, and prime-target congruency effects: Kunde, 2003). The logic behind this observation is that conflict trials call for more control and therefore cause benefits on subsequent trials. A recent model for conflict adaptation by Verguts and Notebaert (2009) explains conflict adaptation by a strengthening of S-R associations (e.g., target arrow pointing leftwards means left response) at the moment conflict is detected. The crucial aspect of this Hebbian-like model is that only associations that are active at the time conflict is detected, will be strengthened. In mental arithmetic, there is support that “4 x 6” will activate certain table related solutions (Campbell, 1987; Stazyk, Ashcraft, & Hamann, 1982) but given the large amount of arithmetic problems and solutions, it is unlikely that performance on the immediately subsequent trial will benefit from this strengthened activation. In other words, what is the advantage of strengthening “4 x 6 = 24”, if on the next trial “5 x 3 = 18” is presented? Consequently, no conflict adaptation effects are expected.

In sum, conflict and error monitoring theories are based on studies using direct mapping tasks, in which paying more attention to the crucial stimulus and refreshing the stimulus
response rules are the only way in which performance can be increased. Although post error effects have been depicted as markers of adaptive behavior, not all data are in line with this hypothesis (i.e., the absence of post error accuracy increases). Recent accounts even question the adaptive nature attributed to post error effects (Notebaert et al., 2009). However, we believe that the use of more complex tasks (i.e., in which multiple solution strategies are possible) might provide additional information to this debate. More precisely, because behavioral adaptation is very limited in direct mapping tasks, post error effects in these tasks might predominantly reflect an orienting response. However, in more complex everyday tasks, behavior can be adapted in a countless number of ways. In other words, there is more room to improve subsequent behavior. One of the tasks where we expect post-error adjustments to be more than a generic slowing down is mental arithmetic. After an error in this task, participants have the opportunity to change strategies in order to improve performance. Subsequently, we expect post error accuracy increases in addition to post error slowing in mental arithmetic. Finally, due to the large amount of different stimuli and responses in our task, we predict no conflict adaptation in mental arithmetic.

Method

Participants

35 students at Ghent University (16 females) participated in this study (mean age = 19.3 years, SD = 1.2 years). The majority of the participants earned course credits in exchange for participation. The other participants were paid 8 euro.
Material

Stimuli were presented on a 17-inch computer screen. The viewing distance was about 50 cm. The multiplication problems were centered on the screen in the traditional format (e.g., 3 × 7 = 21) and presented in white on a black background (total outline: 4.2 cm × 0.6 cm). Responses were recorded by response boxes. The experiment was conducted using TsScope software (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006).

Stimuli

Three different types of multiplication problems were presented. Half the trials comprised problems presented with a correct solution (CORRECT: 4 × 6 = 24). The other half of the trials comprised problems presented with an incorrect solution (distracters). For the incorrect solutions, we manipulated the distance from the correct solution in the multiplication network. More precisely, in one fourth of the trials the distracter was one step away from the correct solution (DISTANCE 1: 4 × 6 = 28), whereas in the other fourth of the trials the distracter was unrelated to the correct solution (UNRELATED: 4 × 6 = 21). Transitions between these different trial types were pseudo-randomized over the experiment, in such a way that every possible transition occurred equally often. We selected problems ranging from 2 × 3 until 8 × 9. Tie problems were not included. This resulted in 28 problems. For unrelated distracters one unrelated outcome was chosen for each of the 28 problems. Every problem occurred in both the ‘larger x smaller’ and the ‘smaller x larger’ order. This resulted in 56 unique problems for correct and unrelated problem types. For correct problems, these 56
problems were repeated 8 times over the experiment, for unrelated problems they were repeated 4 times. For distance 1 distracters we included four different outcomes for each of the 28 problems: \((a+1) \times b\); \((a-1) \times b\); \(a \times (b+1)\); \(a \times (b-1)\). For distance 1 distracters, there were thus four lists of 28 problems\(^1\). Including the order of larger operand first /smaller operand first, there were 224 problems for distance 1 distracters. Every problem was repeated once during the experiment. In practice the four different distance 1 lists sometimes contained the same distracters. This was the case for problems with 2 or 9 as one of the operands (e.g., \(2 \times 7\) or \(9 \times 3\)) because problems with 1 (e.g., \((2-1) \times 7\)) or 10 as one of the operands (e.g., \((9+1) \times 3\)) were excluded from the stimulus set. In total, there were 896 experimental trials.

A number of restrictions were imposed on the stimuli. First of all, we ensured that the ‘split’ (i.e., the magnitude difference between the presented distracter and the correct product, Ashcraft & Stazyk, 1981; Koshmider & Ashcraft, 1991) did not differ significantly between distance 1 distracters (mean split = 5.5) and unrelated distracters \((M = 4.5)\), \(t(27) = 1.42, p = .17\). Second, the direction of the split was controlled: half of the distracters was larger than the correct product, the other half was smaller than the correct product, for both distracter types. Third, the magnitude of the presented distracters did not differ significantly between distance 1 \((M = 29.63)\) and unrelated distracters \((M = 28.39)\), \(t(27) = 1.23, p = .23\).

Procedure
Participants had to classify multiplication problems as correct or incorrect by pressing a button with their left or right index finger. The response mappings were counterbalanced between subjects.

In total there were four blocks of 226 trials resulting in 904 experimental trials. The experiment started with 8 practice trials. There was a short break after every block. During the break the mean reaction time of the participant appeared on the screen. The experiment lasted about 40 minutes. Participants were instructed to respond both fast and accurately.

Each trial started with the presentation of the following fixation mark ‘!’ for 500 ms. Then the verification problem appeared on the screen until participants responded or until the response deadline of 1500 ms had passed. After a correct response, a green circle was presented for 500 ms while after an erroneous response a red circle appeared. If participants did not answer within the response interval the words ‘TE TRAAG’ (‘too slow’ in Dutch) appeared on the screen for 500 ms. After a blank screen of 300 ms the following trial started, resulting in a response stimulus interval of 1300 ms.

Results

Three participants were removed from the analyses. The error rates of two participants were larger than 2 SD’s from the overall mean. The data of the third outlier indicated guessing behavior, shown by a lot of (23%) very fast (< 200 ms) responses. The mean response time of the remaining 32 participants was 832 ms (SD = 93 ms). The mean error rate was 14% (SD = 5%). Correct trials (in which a problem was presented with its correct product) were not included in the analyses since they only
served as control trials. In the final dataset we thus only included unrelated and distance 1 trials. Furthermore, responses exceeding the response deadline were discarded (3%, SD = 4%). We also excluded responses following these trials (3%, SD = 4%). In addition, errors on the current trial (17%, SD = 8%) were removed for response time analyses. For each dependent variable (RTs and accuracy) we first compared post correct performance with post error performance by means of a paired samples $t$-test. Second, we looked at conflict adaptation effects by means of a 2 x 2 repeated measures ANOVA on post correct trials including the factors INTERFERENCE N-1 (trial n-1 unrelated versus trial n-1 distance 1) and INTERFERENCE (unrelated versus distance 1). In the first section, we report the results for response times. In the second section, the results for accuracy rates are described.

**Response times**

Response times for error trials (882 ms, SD = 114 ms) were significantly slower than for correct trials (855 ms, SD = 98 ms), $t(31) = 2.52$, $p < 0.05$, standardized effect size (Cohen’s $d$) = 0.44. Participants responded slower after an error (951 ms, SD = 93 ms) than after a correct response (874 ms, SD = 103 ms), $t(31) = -6.37$, $p < 0.001$, $\eta_p^2 = -1.13$. See Figure 1. Further, the expected interference effect emerged. Participants were slower on distance 1 trials (897 ms, SD = 107 ms) than on unrelated trials (851 ms, SD = 105 ms), $F(1,31) = 81.44$, $p < 0.001$, $\eta_p^2 = 0.72$. However, the main effect of INTERFERENCE N-1 and the interaction between INTERFERENCE N-1 and INTERFERENCE did not reach significance, $Fs < 1$. Mean response times are shown in Table 1.
Accuracy

Participants were more accurate after an error (84%, SD = 8%) than after a correct response (81%, SD = 8%), $t(31) = -2.86, p = 0.01, \eta^2_p = -0.51$. See Figure 1.

Further, participants were less accurate on distance 1 trials (75%, SD = 10%) than on unrelated trials (87%, SD = 8%), $F(1,31) = 125.74, p < 0.001, \eta^2_p = 0.80$. There was no main effect of INTERFERENCE N-1, $F < 1$. The interaction between INTERFERENCE N-1 and INTERFERENCE was significant, $F(1,31) = 8.71, p = 0.01, \eta^2_p = 0.22$. However, the results do not support reduction of interference after high conflict. Rather they point into the opposite direction: there is more interference after distance 1 trials than after unrelated trials. Mean accuracy rates are shown in Table 1.

Discussion

The simplified nature of laboratory tasks might have narrowed the perspective on error monitoring. Moreover, post error accuracy increases are not always observed in
direct mapping tasks. As a result, post error effects in these tasks (i.e., post error slowing) have been explained by attentional effects (i.e., orienting to the error, attracting attention away from the task) rather than by adaptation effects. Our data suggest that in a more complex task, i.e., tasks where multiple solution strategies are possible and where people can adjust their behavior by flexibly switching between these strategies, attentional effects are not sufficient to explain the data pattern. In sum, post error adaptation effects (such as post error accuracy increases) might predominantly be observed in more complex tasks, as shown in the present study.

The difference in complexity between our task and direct mapping tasks is also expressed in the response times on errors. Namely, we found larger response times on errors compared to correct responses in mental arithmetic. In direct mapping tasks, where one strategy (i.e., paying more attention to the stimulus and refreshing the stimulus response rules) is the only way to improve performance, an error will primarily result from insufficient processing time. Errors in direct mapping tasks thus typically emerge on fast trials. However, in tasks that do not only rely on stimulus identification, errors will primarily occur on difficult trials (i.e., multiplication problems where the correct solution is not that straightforward) compared to trials where participants can immediately recollect the correct solution from memory. Consequently, over the whole experiment errors will be slower than correct responses.

Further, to investigate conflict adaptation, we manipulated the relatedness of the distracters: distance 1 distracters were responded slower and less accurate than unrelated distracters. However, reduction of interference after high conflict was not
observed. This indicates that conflict adaptation, a process at work in direct mapping
tasks, might not be at work in more complex tasks. In the introduction we argued that on
the basis of associative control models (Verguts & Notebaert, 2009), one would not
predict conflict adaptation in mental arithmetic, or any other task consisting of multiple
stimuli and responses (e.g., Braem, Verguts & Notebaert, 2011). This is in line with our
findings. On accuracy rates, the results even pointed into the opposite direction.
Interference was reduced after unrelated trials compared to distance 1 trials. At first
sight these results might seem odd. However, in a recent ERP-study, Tzur and Berger
(2007) showed that theta activity, a measure expressing ACC activity in error and
conflict detection (Luu, Tucker, & Makeig, 2004), is related to the salience of the rule
violation in mental arithmetic. More precisely, a larger deviation from the correct
response was related to more theta activity. In our experiment, unrelated distracters
were more salient violations from the correct response than were distance 1 distracters.
That is, distance 1 distracters are still related to the multiplication table of one of the
operands, whereas unrelated distracters are not related to the operands. In other words,
salience and conflict are not confounded in our design. In contrast, in direct mapping
tasks high conflict trials are often also the most salient trials. Our results thus might
suggest that interference effects are reduced after more salient events. This implicates
that not the level of conflict but rather the salience of the event is important to reduce
interference effects. Of course, future research is necessary to investigate this
possibility.

Taken together, the present study shows that it is crucial to investigate error
processing and conflict adaptation in tasks that resemble daily situations of flexible
behavior. Not all effects found in direct mapping tasks can be generalized to more complex tasks. We are convinced that broadening the domain by extending the sort of tasks being used, will gain new and interesting insights in the human ability of cognitive control, decision making, and flexible behavior.

Besides the broader view on cognitive control processes our study provides some important implications for research in mental arithmetic. Traditionally, researchers in this domain focus on RTs of correct responses and on percentages of errors (e.g., Campbell & Xue, 2001; Imbo & Vandierendonck, 2007a,b; LeFevre et al., 1996; Seitz & Schumann-Hengsteler, 2000; Siegler & Lemaire, 1997; Smith-Chant & LeFevre, 2003; etc.). Mostly, response times after errors are not discarded from the analyses. Nonetheless, errors are not that infrequent in mental arithmetic. For example, in a multiplication production task under time pressure, adults make between 1% and 35% errors (De Brauwer, Verguts & Fias, 2006; Imbo & Vandierendonck, 2010; Smith-Chant & LeFevre, 2003; Verguts & Fias, 2005). Future studies in the field of mental arithmetic need to be aware of post error effects. More specifically, we would suggest removing not only error trials but also trials that follow errors.
Foot Note

1. Because there were four possible distance-1 solutions per problem and only one possible unrelated solution per problem, we repeated all analyses restricted to the data gathered in the first block, thus only including the first presentation of both solution types. All results were replicated, indicating that the different presentation frequency of high and low conflict was not responsible for the pattern of results.
References


Table 1

*Mean response times in ms and accuracy rates (between brackets) in percentages, for previous trial type (unrelated and distance 1) and current trial type (unrelated and distance 1).*

<table>
<thead>
<tr>
<th>Current trial</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unrelated</td>
<td>Distance 1</td>
</tr>
<tr>
<td>Previous trial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrelated</td>
<td>851 (0.86)</td>
<td>894 (0.76)</td>
</tr>
<tr>
<td>Distance 1</td>
<td>850 (0.89)</td>
<td>900 (0.74)</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1. Observed mean response times (in ms) and mean accuracy rates (in percentages) after correct and error trials.