Journal of Experimental Psychology: General
2004, Vol. 133, No. 4, 584–600

0096-3445/04/$12.00 DOI: 10.1037/0096-3445.133.4.584

More on the Fragility of Performance: Choking Under Pressure in Mathematical Problem Solving

Sian L. Beilock
Miami University

Catherine A. Kulp
University of Michigan

Lauren E. Holt
Miami University

Thomas H. Carr
Michigan State University

In 3 experiments, the authors examined mathematical problem solving performance under pressure. In Experiment 1, pressure harmed performance on only unpracticed problems with heavy working memory demands. In Experiment 2, such high-demand problems were practiced until their answers were directly retrieved from memory. This eliminated choking under pressure. Experiment 3 dissociated practice on particular problems from practice on the solution algorithm by imposing a high-pressure test on problems practiced 1, 2, or 50 times each. Infrequently practiced high-demand problems were still performed poorly under pressure, whereas problems practiced 50 times each were not. These findings support *distraction theories* of choking in math, which contrasts with considerable evidence for *explicit monitoring* theories of choking in sensorimotor skills. This contrast suggests a skill taxonomy based on real-time control structures.

The desire to perform as well as possible in situations with a high degree of personally felt importance is thought to create *performance pressure* (Baumeister, 1984; Hardy, Mullen, & Jones, 1996). Paradoxically, despite the fact that performance pressure results from aspirations to do one's best, pressure-packed situations are often where suboptimal skill execution is most visible. The term *choking under pressure* has been used to describe this phenomenon. Choking, defined as performing more poorly than expected given one's skill level, is thought to occur across diverse task domains where incentives for optimal performance are at a maximum (Beilock & Carr, 2001; Lewis & Linder, 1997; Masters, 1992). In everyday life, people talk about "bricks" in basketball when the game-winning shot is on the line, "yips" in golf when an easy 3-foot putt to win the tournament stops short, or "cracking" in important test-taking situations where a course grade or college admission is at stake as unmistakable instances of such incentive- or pressure-induced performance decrements.

Surprisingly, although research concerning the cognitive mechanisms governing superior task performance is abundant across both cognitive and sensorimotor skill domains (Anderson, 1982, 1993; Ericsson & Charness, 1994; Proctor & Dutta, 1995; Rosenbaum, Carlson, & Gilmore, 2001), substantially less attention has been devoted to suboptimal skill execution, especially in situations in which optimal task performance is not only desired but expected. Insight into the mechanisms governing execution failure is important, as it will further understanding of not only the variables responsible for skill decrements but those responsible for success as well. That is, a careful cognitive analysis of choking under pressure opens a new kind of window into the organization and operation of the information-processing mechanisms that underlie performance.

Theories of Choking Under Pressure

Why do skills fail in high-pressure situations? Two main types of explanations have been put forth. *Self-focus* or *explicit monitoring* theories propose that performance pressure increases anxiety and self-consciousness about performing correctly, which in turn enhances the attention paid to skill processes and their step-by-step control. People try harder to exert conscious control over the steps they need to accomplish, in the hope that being careful in this way will increase their chances of success. Attention to performance at such a component-specific level is thought to disrupt the proceduralized or automated processes of high-level skills that normally run outside the scope of working memory during performance (Baumeister, 1984; Beilock & Carr, 2001; Kimble & Perlmuter, 1970; Langer & Imber, 1979; Lewis & Linder, 1997; Masters, 1992).

In contrast, *distraction theories* suggest that pressure fills working memory with thoughts about the situation and its importance that compete with the attention normally allocated to execution. Pressure serves to create a dual-task environment in which controlling the execution of the task at hand and worries about performance vie for the attentional capacity once devoted solely to

Sian L. Beilock and Lauren E. Holt, Department of Psychology, Miami University; Catherine A. Kulp, Department of Mathematics Education, University of Michigan; Thomas H. Carr, Department of Psychology, Michigan State University.

Correspondence concerning this article should be addressed to Sian L. Beilock, Department of Psychology, Miami University, Room 202, Benton Hall, Oxford, OH 45056-1601. E-mail: beilocsl@muohio.edu

primary task performance (Beilock & Carr, in press; Lewis & Linder, 1997).

Thus, distraction and explicit monitoring theories make contrasting predictions concerning the mechanisms responsible for performance decrements under pressure. Although distraction theories suggest that pressure creates a distracting environment that draws attention away from primary skill execution, explicit monitoring theories suggest the opposite—that pressure prompts too much attention to performance processes and procedures.

To date, explicit monitoring theories have received the most support in accounting for choking under pressure. Some of this support comes from training studies. For example, Beilock and Carr (2001) examined performance under pressure in a golf putting task to determine whether practice at dealing with the causal mechanisms proposed by each theory (i.e., explicit attention vs. distraction) would reduce pressure-induced failure. Participants were trained to a high putting skill level under one of three different learning conditions and then exposed to a pressure situation. The first training condition involved ordinary single-task practice, which provided a baseline measure of choking. The second training condition involved practice in a dual-task environment (putting while monitoring an auditory word list for a target word). This condition was designed to distract attention from the primary putting task with execution-irrelevant activity in working memory—the aspect of a pressure situation that distraction theories propose causes skill failure. In the third *self-conscious* or *skill-focus* training condition, participants learned the putting task while being videotaped for subsequent public analysis by experts, a manipulation first used by Lewis and Linder (1997). This manipulation was designed to expose performers to having attention called to themselves and their performance in a way intended to induce explicit monitoring of skill execution—the aspect of pressure that explicit monitoring theories propose causes failure. After training, all groups were exposed to the same pressure situation created by a performance-contingent monetary award.

Choking occurred for those individuals who were trained on the putting task in the single-task condition used as a baseline and also for individuals trained in the dual-task environment that created distraction. However, choking did not occur for those trained in the self-conscious condition. Beilock and Carr (2001) concluded that training under conditions that prompted attention to the component processes of execution enabled performers to adapt to the type of attentional focus that often occurs under pressure. In this way, self-consciousness training served to inoculate individuals against the negative consequences of over-attending to well-learned performance processes—the mechanism that explicit monitoring theories suggest is responsible for performance decrements in high-pressure situations.

Lewis and Linder (1997), using the same technique of videotaping for subsequent analysis by experts, have also demonstrated that learning a golf putting skill in a self-awareness–heightened environment inoculates individuals against pressure-induced failure at high levels of practice. Like Beilock and Carr (2001), Lewis and Linder found that pressure caused choking in those individuals who had not been adapted to self-awareness. Furthermore, they found that the introduction of a secondary task (counting backward from 100) while performing under pressure helped to alleviate the performance decrements shown by the nonadapted putters. Lewis and Linder concluded that the secondary backward-counting task occupied working memory, preventing their participants from focusing attention on the proceduralized processes that controlled performance. As a consequence, choking under pressure was ameliorated—another finding that is consistent with explicit monitoring theories.

Thus, both Beilock and Carr (2001) and Lewis and Linder (1997) have shown that skill training that induces attention to performance may protect individuals from the negative effects of pressure. Although these types of training methods lend indirect insight into the cognitive mechanisms driving skill failure in high-stakes situations, it is also possible to more directly assess the impact of processes that might be responsible for pressure-induced performance decrements. Recently, Beilock, Carr, MacMahon, and Starkes (2002) directly manipulated the attentional focus of experienced soccer players performing a ball-dribbling task rather than imposing pressure and inferring attentional focus from performance outcomes. The goal was to directly test explicit monitoring and distraction theories' predictions regarding the causal mechanisms that underlie skill failure.

Experienced soccer players dribbled a soccer ball through a series of pylons while performing either a secondary auditory monitoring task (designed to distract attention away from execution, mimicking distraction theories' proposed choking mechanism) or a skill-focused task in which individuals monitored the side of the foot that most recently contacted the ball (designed to draw attention to a component process of performance, mimicking explicit monitoring theories' proposed choking mechanism). Performing in a dual-task environment did not harm experienced soccer players' dribbling skill in comparison to a single-task practice condition used as a baseline. However, when the soccer players were instructed to attend to performance (i.e., monitoring the side of the foot that most recently contacted the ball), their dribbling skill deteriorated in comparison to both the dual-task condition and a single-task baseline. Consistent with the evidence presented above in support of explicit monitoring theories of choking, step-by-step attention to skill processes and procedures harms well-learned performance. Gray (2004) reported analogous results in an investigation of baseball batting.

Supporting evidence regarding the differential impact of distraction versus skill-focused attention has also been obtained from a different kind of manipulation: speed versus accuracy performance instructions. Beilock, Bertenthal, McCoy, and Carr (2004) found that simply limiting the opportunity for skill-focused explicit monitoring through instructions to perform a putting task rapidly improved the performance of experienced golfers relative to a condition in which the same golfers were told to take as much time as they needed to be accurate. The impact of this manipulation was phenomenologically noticeable: Several golfers reported that the speed instructions aided their performance by keeping them from thinking too much about execution.

Is the Issue Settled? Differences Due to Task Control Structure

The work reviewed above suggests that maladaptive explicit monitoring is responsible for choking. Given the range and consistency of this evidence, it may seem unlikely that distraction theories could provide additional insight. However, the proceduralized sensorimotor skills used in the extant choking research may not possess the right control structures to be susceptible to pressure-induced failure via distraction. Proceduralized sensorimotor skills are thought to run outside of working memory, under the control of an integrated motor program (Fitts & Posner, 1967; Proctor & Dutta, 1995), and, as suggested by the results already reviewed, are largely robust to decrements resulting from distracting, dual-task situations (Beilock, Carr, et al., 2002; Gray, 2004). Thus, one reason why distraction theories of choking may not have received much support is because they have not been tested in the appropriate skill domains.

Academic Skills and the Impact of Test Anxiety

However, there is a literature to look to for clues concerning how pressure might influence skills that are more vulnerable to online demands that threaten the capacity of working memory. Within the test anxiety literature, it has been suggested that anxiety manifests itself in the form of intrusive thoughts or worries about the situation and its outcome (Ashcraft & Kirk, 2001; Eysenck, 1979, 1992; Wine, 1971). Because these thoughts are attended to, a portion of working memory normally devoted to primary problem-solving activity is consumed and therefore not available for the processing of task-relevant cues and the execution of task-relevant computations. For tasks that rely heavily on working memory for online execution, such as the kinds of problems encountered on academic tests like the Standardized Aptitude Test or the Graduate Record Examination, this decrease in capacity is thought to cause suboptimal performance outcomes (Ashcraft & Kirk, 2001; Sorg & Whitney, 1992; Tohill & Holyoak, 2000). Such problems include verbal comprehension, logical reasoning, and mathematical computation. The latter—mathematical computation—is our present focus.

Although a number of models depict working memory's organization, in tasks with heavy real-time processing demands, such as mental arithmetic and other forms of mathematical computation, Baddeley's (1986; Baddeley & Logie, 1999) multicomponent model has most often been used to explore the role of working memory in performance (Trbovich & LeFevre, 2003). Baddeley's original model had three major components: a limited-capacity central executive, a phonological loop for storing verbal information, and a visual-spatial sketchpad for storing visual images. Recently, a fourth component has been added: an episodic buffer for storing a situation model of the event currently being experienced (Baddeley, 2000).

In the context of this conception of working memory, it has been suggested that the impact of distracting thoughts and worries on performance in tasks such as mental arithmetic results from a disruption of the central executive component of working memory that controls and applies the sequence of arithmetic operations during problem solving (Ashcraft & Kirk, 2001). The central executive is thought to be especially important for mathematical procedures that are difficult because of their complexity, such as carrying during addition and borrowing during subtraction. Such procedures require an extended sequence of steps, as well as the maintenance of intermediate products, to be completed successfully (Fürst & Hitch, 2000).

Recent research has found support for the idea that anxiety-induced worries disrupt mathematical problem solving by consuming working memory capacity. Ashcraft and Kirk (2001) examined low- and high-math-anxious individuals' ability to simultaneously perform a mental addition task and a memory task involving the short-term maintenance of random letter strings for later recall. Difficulty levels of both the primary math task and the secondary memory task were manipulated. Performance was worst (mainly in the form of increased math task error rates) in instances in which individuals, regardless of math anxiety level, performed both a difficult math task and a difficult memory task simultaneously. Furthermore, in comparison to less anxious individuals, participants high in math anxiety showed an exaggerated increase in performance errors under the difficult math and memory task condition. The authors concluded that performance deficits under demanding dual-task conditions were most pronounced in high-math-anxiety individuals because anxiety, similar to a demanding secondary task, drains the attentional capacity that might otherwise be available for primary skill execution.

This work lends support to the notion that performance decrements may result from anxiety-induced worries that decrease task-relevant processing resources, at least in certain kinds of tasks. If, as proposed by distraction theories, pressure serves to create a distracting environment via worries about the situation, then situational pressure should impose constraints on working memory–intensive tasks similar to those of chronic math anxiety. The lack of support for distraction theories in the choking literature, then, may be a function of the types of skills under investigation. If tasks with control structures susceptible to dual-task decrements are tested, support for distraction theories of choking may appear.

Present Experiments

Our aim in this work was to examine performance under pressure in a task with working memory requirements that are likely to make it susceptible to choking according to distraction theories. We chose Gauss's *modular arithmetic* task (as described in Bogomolny, 1996) for this investigation. The object of modular arithmetic is to judge the truth value of problem statements such as $51 \equiv 19 \pmod{4}$. To do this, the problem's middle number is subtracted from the first number (i.e., $51 - 19$) and this difference is divided by the last number (i.e., $32 \div 4$). If the dividend is a whole number (as here, 8), the problem is true. Modular arithmetic is advantageous as a laboratory task because it is unusual and hence its learning history can be controlled. However, because it is based on common arithmetic operations carried out in a particular order, it is also similar to the kinds of math problems encountered in the real world. Thus, it should be learned in a fashion similar to the way other types of mental arithmetic are learned.

According to Logan's (1988) instance-based theory of how mental arithmetic is learned, a rule-based algorithm is initially used to solve unpracticed problems. Here, problem solutions are dependent on the explicit application of a capacity-demanding step-by-step process that must be maintained and controlled online by working memory during execution. With practice on particular problems, the reliance on this procedure decreases and past instances of problem solutions are retrieved directly or automatically from long-term memory into working memory (similar to how one's multiplication tables might be retrieved from memory), whereas new problems continue to engage the algorithm. Because the algorithm does not change with practice, new problems remain slow and capacity demanding to solve.

Logan's (1988) model proposed an alternative view of automaticity to traditional theories of proceduralization, which suggest that skill learning results in the compilation of the step-by-step processes once used by novices into encapsulated procedures that require little attention and operate largely outside of working memory (Anderson, 1987, 1993; Fitts & Posner, 1967; Keele & Summers, 1976; Squire & Knowlton, 1994). Logan's model has been quite successful in accounting for changes in speed and accuracy of performance with practice on cognitive tasks such as alphabet arithmetic, lexical decision, and semantic categorization. However, contrary to Logan's assumption that the algorithm does not change with practice, there is evidence that unfamiliar problems based on algorithmic computations are solved more efficiently with general algorithm practice (Rickard, 1997; Touron, Hoyer, & Cerella, 2001) Yet, even practiced algorithms may not be governed by the same type of control structures as, for example, proceduralized motor programs. An algorithmic solution procedure, regardless of each component's efficiency, is based on a hierarchical and sequentially dependent task representation in which initial steps must be held and acted on in working memory to generate subsequent steps and final solutions. Well-learned sensorimotor skills that operate largely outside of working memory (Beilock & Carr, 2001; Fitts & Posner, 1967; Gray, 2004) are most likely not governed by the same type of working memory–dependent representation. Hence, math problem solving may look different than sensorimotor skill performance in pressure situations that serve to alter what occupies working memory and how attention is paid to execution. We report three experiments in which we investigate these issues.

In Experiment 1, we examined unpracticed modular arithmetic performance under pressure. According to distraction theories, pressure-induced worries may compromise the resources participants rely on to solve such problems, especially unpracticed problems with heavy working memory demands. In contrast, explicit monitoring theories suggest that pressure should not harm participants' ability to solve unpracticed modular arithmetic problems, regardless of working memory demands. Pressure-induced attention to execution should not hurt information that is already explicitly attended to and maintained online. Indeed, Beilock and Carr (2001) found that pressure actually enhances the performance of novices in golf putting, despite harming the execution of more practiced individuals.

Experiment 2 extended the exploration of choking under pressure to highly practiced modular arithmetic. To the extent that the control structure of modular arithmetic changes to automatic answer retrieval with practice, distraction theories propose that the ability to solve heavily practiced problems should not be harmed when working memory is consumed. Explicit monitoring theories might make a different prediction—at least according to Masters, Polman, and Hammond (1993). At high levels of practice, pressure "may result in a return to an explicit, algorithmic-based control of behavior through disruption of automatic retrieval of skill-based information from memory" (Masters et al., 1993, p. 664). Such a regression would slow performance and increase the opportunity for error, which would create poorer performance. Furthermore, if, rather than a shift to automatic answer retrieval as Logan (1988) would propose, modular arithmetic automates via proceduralization of the algorithm, explicit monitoring theories would also predict performance decrements under pressure at high practice levels. Such failure may be due to pressure-induced attentional control that increases the time or error associated with maintaining, rehearsing, or acting on the well-learned algorithm.

To pursue these possibilities, in Experiment 3, we examined participants' susceptibility to choking after different amounts of exposure to specific modular arithmetic problems. Problems were presented either 1, 2, or 50 times each during practice, after which participants were given a high-pressure test. In the context of Logan's (1988) instance-based theory of how a task like modular arithmetic should change with practice, the two theories of choking again make very clear predictions. If choking is due to pressure-induced capacity limitations, as distraction theories propose, then regardless of how many different problems individuals have been exposed to, only those problems that have been practiced enough to produce instance-based answer retrieval (a minimum of 36 to 72 exposures, according to Klapp, Boches, Trabert, & Logan, 1991) should be inoculated against the detrimental capacity-limiting effects of pressure. In contrast, explicit monitoring theories make an opposite prediction, as long as automatic memory retrieval is disrupted in the manner proposed by Masters et al. (1993). Namely, pressure-induced attention may disrupt automatic answer retrieval by reverting it back to algorithmic control.

These predictions are modified somewhat if the algorithm does change with practice, such that unfamiliar problems still based on algorithmic computations are solved more efficiently because of practice with the algorithm on other problems. First, novel problems should be solved faster and more accurately after the algorithm has been practiced than when the performer had no experience with modular arithmetic. Second, if the algorithm changes so much that it becomes proceduralized, then—consistent with explicit monitoring theories—individuals might be susceptible to choking on novel problems via pressure-induced attention that disrupts or slows down proceduralized performance processes that normally run outside of conscious control. In contrast, if practicing algorithms makes them more efficient but the algorithms do not actually become proceduralized, then such problems should still impose attentional demands and thus should be performed at a suboptimal level under pressure according to distraction theories. This is due to the fact that the intermediate operations of the algorithmic solution procedure would still be held online in working memory, even though each operation is performed more efficiently once it is implemented. This possibility can be explored by examining pressure-induced failure as a function of problems' working memory demands. If novel problems based on practiced algorithms are performed poorly as a result of distraction, then the more capacity demanding the novel problem, the more vulnerable the problem should be, regardless of the amount of practice that has been devoted to the general algorithm by which the problem is solved.

Experiment 1

In Experiment 1, individuals were randomly assigned to either a low-pressure or a high-pressure group prior to performing three blocks of novel modular arithmetic problems. The first block of problems served as a pretest measure of performance and the second block served as a small amount of practice at the algorithm and was designed to stabilize performance. Immediately preceding the last block of problems, the low-pressure group was simply informed that they would be performing another set of problems, whereas the high-pressure group was given a scenario intended to create a high-pressure environment.

The working memory demands of the modular arithmetic problems being performed were manipulated within each block: Single-digit problems without a borrow operation created the least online capacity demands (low-demand problems), and double-digit problems with a borrow operation were the most working memory demanding (high-demand problems). Large numbers (single digit vs. double digit) combined with borrow operations were chosen as the means to establish these comparisons, as it is well established in the math-problem–solving literature that complex problems involving large numbers and borrow operations place greater demands on the capacity of the working memory system than do problems with small numbers and no borrow operation (Ashcraft, 1992; Ashcraft & Kirk, 2001).

From the standpoint of distraction theories, the performance of unpracticed modular arithmetic should fail under pressure—especially problems with the heaviest working memory demands. However, if explicit monitoring theories are correct in the domain of mathematical problem solving and performance pressure prompts explicit attention to skill execution, then performance of unpracticed modular problems in Experiment 1, regardless of working memory demands, should not be harmed by increased pressure. Here, pressure-induced attention to execution should not impact information already explicitly attended online. Indeed, such attention may even help performance by gathering and focusing the attentional resources devoted to the problem.

Method

Participants

Participants were students enrolled at Michigan State University who were not math majors and reported no previous exposure to modular arithmetic. Participants were randomly assigned to either a low-pressure group ($n = 40$) or a high-pressure group ($n = 40$), provided their accuracy in the pretest block was greater than 55% correct. This minimum accuracy criterion was implemented across all three experiments to ensure that individuals were performing above chance on the modular arithmetic task prior to the implementation of any experimental manipulations. No participants were excluded from Experiment 1 on the basis of this criterion.

Procedure

Participants filled out a consent form and a demographic sheet detailing previous math experience. They were informed that the purpose of the study was to examine how individuals learn a new math skill. Participants were seated in front of a standard computer and introduced to modular arithmetic through a series of written instructions presented on the computer screen. Individuals were informed that they would be judging the validity of modular arithmetic problems and were provided with several examples. Participants were instructed to judge the validity of the problems as quickly as possible without sacrificing accuracy and, when they had derived an answer to the problem presented on the screen, to press the corresponding *T* or *F* key on a standard keyboard set up in front of them. Participants were instructed to rest their right and left index fingers on the *T* and *F* keys, respectively, throughout the experiment.

The stimuli were digits; the word *mod*, designed to denote the modular arithmetic statement; and a congruence sign ($\equiv$). Each trial began with a 500-ms fixation point in the center of the screen, which was immediately replaced by a problem that was present until the participant responded. The problem was then removed and the word "Correct" or "Incorrect" was displayed on the screen for 1,000 ms, providing feedback. The screen then went blank for a 1,000-ms intertrial interval.

All participants performed three blocks of 24 modular arithmetic problems each; the blocks were separated by a short break. Each block consisted of 8 low-demand problems requiring a single-digit no-borrow subtraction operation, such as $7 \equiv 2 \pmod 5$, and 8 high-demand problems requiring a double-digit borrow subtraction operation, such as $51 \equiv 19 \pmod 4$. An additional 8 problems with intermediate attentional demands requiring a double-digit no-borrow subtraction operation, such as $19 \equiv 12 \pmod 7$, were also included in each of the problem blocks. These intermediate-level problems served as filler problems, intended to diminish the contrast between the low-demand and high-demand problems. Half of the problems within each demand level were true, half were false.

In addition, each true problem had a false correlate that only differed as a function of the number involved in the mod statement. For example, if the true problem $51 \equiv 19 \pmod 4$ was presented, then a false correlate problem $51 \equiv 19 \pmod 3$ was also presented at some point within the same problem block. This pairing was designed equate the true and false problems as much as possible in terms of the specific numbers used in each equation.

Problems within each block were presented in a different random order to each participant, and each problem was presented only once across the entire experiment. Finally, the problems presented in the last two blocks were counterbalanced across participants. This counterbalancing was done to assure that performance in the last block of problems was independent of the particular problems individuals were exposed to.

The first block of problems served as a pretest measure of modular arithmetic performance (pretest block) for both the low- and high-pressure groups. Individuals were asked simply to perform as best they could—to solve the problems as quickly as possible without sacrificing accuracy. Similar instructions were given to both the low- and the high-pressure groups prior to the second, practice block of problems. Immediately preceding the last block of problems (posttest block), individuals in the low-pressure group were simply informed that now they were going to be performing another set of problems, whereas individuals in the high-pressure group were given a scenario designed to create a high-pressure situation.

The high-pressure scenario was based on several sources of pressure that commonly exist across skill domains: monetary incentives, peer pressure, and social evaluation. Although exactly how these different sources of pressure exert their influence is an empirical question, our purpose in the present work was to capture the real-world phenomenon of choking. Thus, we created a pressure scenario that incorporated as many components of high-pressure performance as possible. In athletics, for example, performance is frequently scrutinized by others, there are often monetary consequences for winning and losing, and team success is dependent on the performance of individual athletes, which may generate peer pressure to perform at an optimal level. In more academic arenas, monetary consequences for test performance are manifested in terms of scholarships and future educational opportunities, and social evaluation of performance comes from mentors, teachers, and peers.

Specifically, participants in the high-pressure group were informed that the computer used a formula that equally takes into account reaction time and accuracy in computing a "modular arithmetic score." Participants were told that if they could improve their modular arithmetic score by 20% relative to the preceding practice trials, they would receive $5. Participants were also informed that receiving the monetary award was a "team effort." Specifically, participants were told that they had been randomly paired with another individual, and, to receive their $5, not only did the participant presently in the experiment have to improve in the next set of problems, but the individual he or she was paired with had to improve as well. Next, participants were informed that this individual, "their partner," had already completed the experiment and had improved by the required amount. If the participant presently in the experiment improved by 20%, both the participant and his or her partner would receive \$5. However, if the present participant did not improve by the required amount, neither the participant nor his or her partner would receive the money. Finally, participants were told that their performance would be videotaped during the test situation so that local math teachers and professors in the area could examine their performance on this new type of math task.

The experimenter set up the video camera on a tripod directly to the right of participants, approximately 0.61 m away. The field of view of the video camera included both the participant and the computer screen. Participants in the high-pressure group then completed the last block of problems. The experimenter then turned off the camera and faced it away from the participants.

After completion of the last block of modular arithmetic, participants in both the low- and the high-pressure groups filled out a number of questionnaires designed to assess their feelings of anxiety and performance pressure. Individuals first filled out the State Anxiety form of the State–Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970). The STAI is a well-known measure of state anxiety, consisting of 20 questions designed to assess participants' feelings at a particular moment in time. Individuals are instructed to assign a value to, for example, the degree to which they feel calm or at ease on a 4-point scale ranging from 1 (*not at all*) to 4 (*very much so*). The State Anxiety form of the STAI has been used in a number of studies investigating the impact of anxiety on complex task performance (e.g., in analogical reasoning ability; see Tohill & Holyoak, 2000).

After taking the STAI, all participants answered a number of questions related to their perceptions of performance in the posttest. Individuals were asked to indicate on a 7-point scale (a) how important they felt it was for them to perform at a high level in the posttest, with answers ranging from 1 (*not at all important to me*) to 7 (*extremely important to me*); (b) how much performance pressure they felt to perform at a high level in the posttest, with answers ranging from 1 (*very little performance pressure*) to 7 (*extreme performance pressure*); and (c) how well they thought they performed in the posttest, with answers ranging from 1 (*extremely poor*) to 7 (*extremely well*). Individuals were then debriefed, and participants from the high-pressure group were given the monetary award regardless of their performance.

Results

Questionnaires

Importance. Participants in the low-pressure group ($M = 4.63$, $SE = 0.21$) and the high-pressure group ($M = 5.03$, $SE = 0.19$) did not significantly differ in terms of their perceptions of the importance of performing at a high level in the posttest problem block, $t(78) = 1.40$, $p = .17$. On average, participants in both groups reported that it was at least moderately important to perform at a high level on these problems.

State anxiety. Participants in the high-pressure group ($M = 42.68$, $SE = 1.87$) showed significantly higher levels of state anxiety than did participants in the low-pressure group ($M = 32.08$, $SE = 1.20$), $t(78) = 4.79$, $p < .01$, $d = 1.07$.[1]

Performance pressure. Participants in the high-pressure group ($M = 5.08$, $SE = 0.21$) reported feeling significantly more pressure to perform at a high level in the posttest than did individuals in the low-pressure group ($M = 3.95$, $SE = 0.24$), $t(78) = 3.53$, $p < .01$, $d = 0.79$.

Performance success. Participants in the high-pressure group ($M = 4.03$, $SE = 0.20$) had significantly worse perceptions of their performance in the posttest problem block in comparison to participants in the low-pressure group ($M = 4.98$, $SE = 0.19$), $t(78) = 3.40$, $p < .01$, $d = 0.76$.

These questionnaire results demonstrate that our manipulation was successful in increasing participants' feelings of performance pressure and anxiety. Furthermore, because participants in the low- and high-pressure groups assigned equal importance to performing at a high level on the posttest block of equations, it would be difficult to explain any differences between groups that might be observed in modular arithmetic performance merely by general differences in motivation without more specific reference to pressure and anxiety. We now turn to the performance data to determine whether participants' self-reports of pressure and its consequences parallel its objectively measured impact.

Accuracy and Response Time

We began the analysis of accuracy and response time (RT) by removing outliers from the data. RTs were computed for each problem. RTs more than 3 standard deviations below or above an individual's mean RT for each block of problems were considered outliers, and the trials on which such RTs occurred were removed from the data set. This resulted in the dismissal of 11 RTs and their corresponding accuracy scores.

Next, accuracies and RTs for trials on which responses were correct were subjected to separate three-factor analyses of variance (ANOVAs). Pressure (low pressure, high pressure) varied between groups. Test (pretest, posttest) and problem demand (low demand, high demand) varied within subjects.

The central result of the experiment comes from the analysis of accuracy, which revealed a significant Pressure Group × Test × Problem Demand interaction, $F(1, 78) = 4.01$, $p < .05$, $MSE = 0.01$, $\eta_p^2 = .05$.[2] As can be seen in Figure 1, the impact of pressure was quite different depending on the working memory demand of the problems being performed. On low-demand problems, accuracy increased significantly from the pretest to posttest problem blocks for both the low-pressure group, $t(39) = 2.98$, $p < .01$, $d = 0.65$, and the high-pressure group, $t(39) = 2.40$, $p < .03$, $d = 0.5$. On high-demand problems, the low-pressure group directionally (but not significantly) increased in accuracy from the pretest to the posttest, $t(39) = 1.17$, $p = .25$. In contrast, the high-pressure group significantly declined in accuracy from the pretest to the posttest problem block, $t(39) = 2.77$, $p < .01$, $d = 0.37$.

To interpret this decline by the high-pressure group, it is important to establish that the low- and high-pressure groups did not differ in high-demand problem accuracy in the pretest, $t(78) = 0.51$, $p = .61$. This was not the case in the posttest, however, where the high-pressure group performed significantly less accurately on the high-demand problems in comparison to the low-pressure group, $t(78) = 3.41$, $p < .01$, $d = 0.77$. This pattern of

[1] Cohen's d was used as the measure of effect size for the simple effects tests (for equations, see Dunlap, Cortina, Vaslow, & Burke, 1996). Cohen (1988, 1992) has suggested that 0.20 is a small effect size, 0.50 is a medium effect size, and 0.80 is a large effect size.

[2] Partial eta squared (η_p^2) was used as the measure of effect size for the repeated measures ANOVAs (Tabachnick & Fidell, 1996). The omnibus ANOVAs for both accuracy and RT were also performed with modular arithmetic problem answers (i.e., *true* or *false*) included as an additional variable. Across all three experiments, no interaction with this variable was obtained. Thus, this factor was not included in the presentation of the results.

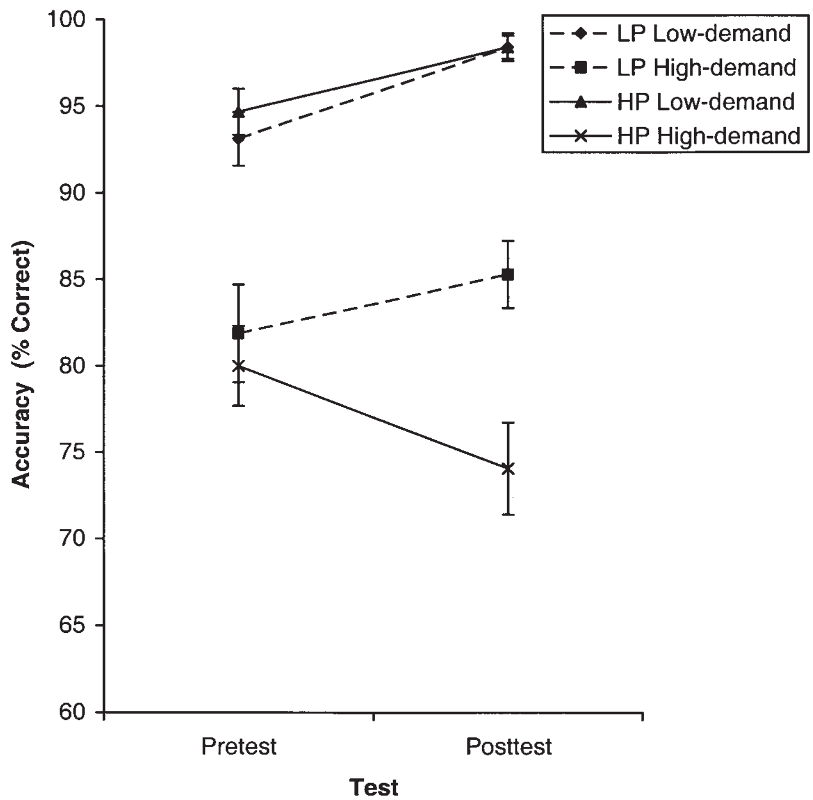

Figure 1. Mean accuracy (% correct) for the low-pressure group (LP) and high-pressure group (HP) for the low-demand and the high-demand problems in the pretest and posttest problem blocks in Experiment 1. Error bars represent standard errors.

data supports the notion that in mathematical problem solving, pressure harms performance on only those problems that rely most heavily on working memory for successful execution. Problems that do not heavily burden working memory are not harmed—in fact, they may even improve.

Analysis of RTs did not alter these conclusions. As described below, a number of significant effects occurred in the RTs, and the overall pattern was not the same as in accuracy. In particular, there was no significant three-factor interaction in the RTs. However, the results that occurred in RT neither contradicted the conclusion that pressure harms performance on only high-demand problems nor indicated that this effect could be attributed to speed–accuracy trade-off.

The three-factor ANOVA on RTs revealed main effects of test, $F(1, 78) = 48.77$, $p < .01$, $MSE = 15.03 \times 10^5$, $\eta_p^2 = .38$, and problem demand, $F(1, 78) = 615.54$, $p < .01$, $MSE = 40.68 \times 10^5$, $\eta_p^2 = .89$; a Test $\times$ Problem Demand interaction, $F(1, 78) = 9.97$, $p < .01$, $MSE = 11.86 \times 10^5$, $\eta_p^2 = .11$, in which there were greater differences in RT from the pretest to the posttest for the high-demand in comparison to the low-demand problems; and a Problem Demand $\times$ Pressure Group interaction, $F(1, 78) = 5.07$, $p < .03$, $MSE = 40.68 \times 10^5$, $\eta_p^2 = .06$, in which problem demand had a greater impact on the RTs of the low-pressure group in comparison to the high-pressure group. There was no Test $\times$ Pressure Group $\times$ Problem Demand interaction, $F < 1$. Across both the low- and the high-pressure groups, RTs were faster for the low-demand in comparison to the high-demand problems. Additionally, for the low-pressure group, RTs significantly decreased from the pretest to the posttest for the low-demand problems, $t(39) = 5.05$, $p < .01$, $d = 0.78$ (pretest, $M = 2{,}444$ ms, $SE = 95$ ms; posttest, $M = 1{,}983$ ms, $SE = 91$ ms), and the high-demand problems, $t(39) = 2.60$, $p < .02$, $d = 0.37$ (pretest, $M = 8{,}815$ ms, $SE = 450$ ms; posttest, $M = 7{,}817$ ms, $SE = 405$ ms). Similarly, for the high-pressure group, RTs significantly decreased from the pretest to the posttest for the low-demand problems, $t(39) = 7.71$, $p < .01$, $d = 1.08$ (pretest, $M = 2{,}531$ ms, $SE = 109$ ms; posttest, $M = 1{,}846$ ms, $SE = 85$ ms), and the high-demand problems, $t(39) = 5.20$, $p < .01$, $d = 0.67$ (pretest, $M = 8{,}118$ ms, $SE = 423$ ms; posttest, $M = 6{,}432$ ms, $SE = 358$ ms).

The absence of any interaction between pressure group, test, and problem demand level in RT suggests that the pressure-induced differences in high-demand problem accuracy reported above were not produced by a trade-off with RT. However, it should be noted that although the difference was not significant, the high-pressure group did decrease in high-demand-problem RT from the pretest to posttest somewhat more than did the low-pressure group. To ensure that the high-demand problem accuracy differences reported above were not the product of a trade-off with RT, we first looked at a correlation of the difference scores (from the pretest to posttest) for high-demand-problem accuracy (i.e., pretest high-demand accuracy − posttest high-demand accuracy) and RT (i.e., pretest high-demand RT − posttest high-demand RT). There was no significant correlation of accuracy and RT for either the low-pressure group ($r = -.09$, $p = .6$) or the high-pressure group ($r = -.15$, $p = .34$). These negative correlations are actually in the opposite direction of what one would expect if speed–accuracy trade-off were playing a role.

To further examine the possibility of speed–accuracy trade-off, we also performed a 2 (low-pressure group, high-pressure group) $\times$ 2 (pretest, posttest) ANOVA on high-demand problem accuracy with RT differences from the pretest to the posttest (i.e., pretest RT $-$ posttest RT) covaried out. If a significant interaction occurs, this would suggest that the significant difference in high-demand-problem accuracy from the pretest to the posttest as a function of pressure group cannot be accounted for by RT differences. A significant Pressure Group $\times$ Test interaction was obtained, $F(1, 77) = 7.28$, $p < .01$, $MSE = .01$, $\eta_p^2 = .09$. In fact, the F value for this interaction was slightly bigger than in the same Pressure Group $\times$ Test interaction for high-demand-problem accuracy in which RT was not covaried out. Thus, accounting for high-demand-problem RT differences did not diminish the interaction of pressure group and test for high-demand-problem accuracy.

Finally, to directly examine the relationship between declines in accuracy in the high-pressure condition and expressed perceptions of performance pressure, we computed the correlation between participants' self-report measures of performance pressure and their high-demand problem accuracy difference scores (i.e., high-demand pretest accuracy $-$ high-demand posttest accuracy). Across all participants, there was a significant positive correlation between self-report measures of performance pressure (with higher scores indicating increased feelings of pressure) and high-demand accuracy difference scores (with higher scores indicating a greater decrease from the pretest to posttest; $r = .36$, $p < .01$). That is, the more pressure individuals felt in the posttest, the bigger their decline in high-demand problem accuracy.

Discussion

We designed Experiment 1 to examine the impact of pressure on a task whose performance should be susceptible to choking according to distraction but not explicit monitoring. Individuals assigned to either a low-pressure or a high-pressure group performed unpracticed modular arithmetic equations that varied as a function of their working memory demands. The pressure scenario increased participants' feelings of performance pressure and anxiety and reduced accuracy on high-demand problems that draw heavily on working memory.

Experiment 1 lends support to distraction theories as an explanation for the choking phenomenon in the domain of mathematical problem solving. However, participants began the experiment as novices at modular arithmetic, and, when pressure was applied, they were still relatively unpracticed. Therefore, it is remains possible that performance decrements under pressure may occur at higher levels of practice on modular arithmetic via the mechanisms proposed by explicit monitoring theories. In Experiment 2, we tested this notion.

Experiment 2

In Experiment 2, individuals performed low-demand and high-demand modular arithmetic problems under pressure both prior to and after extended modular arithmetic practice. Participants had no previous exposure to the specific problems on which they were tested prior to the first high-pressure situation. By the time of the second high-pressure test, however, participants had been exposed to each problem appearing in the test 49 times.

As in Experiment 1, participants working on modular arithmetic problems should be susceptible to choking under pressure early in learning according to distraction theories. These failures should be most pronounced when participants are trying to solve high-demand problems that incur the highest working memory load. At higher levels of problem-specific practice, however, when answers to now well-practiced problems are being retrieved directly from long-term memory into working memory rather than being computed via a step-by-step algorithm that must be held and manipulated online in working memory, such capacity-related failures should diminish. This is because these practiced problems should rely less on working memory than do their unpracticed counterparts.

There is evidence that the central executive of working memory is involved in even simple mental arithmetic problems that are often based on direct answer retrieval (De Rammelaere, Stuyven, & Vandierendonck, 2001). Thus, one might suggest that performance on highly practiced problems, whose answers are retrieved into working memory, should still be harmed by pressure-induced disruption of working memory. However, although practiced problems do involve working memory (indeed, their answers are retrieved into working memory), they do not use these resources to the same extent as an unpracticed problem does. Thus, if distraction theories of choking are applicable to mathematical problem solving, unpracticed problems should show larger pressure-induced performance decrements than repeatedly practiced equations do.

Choking might also occur at high levels of practice via the mechanism proposed by explicit monitoring theories. Well-learned problems, based on the stimulus-driven retrieval of past problem instances from memory, may still be performed poorly under pressure because pressure-induced attention disrupts automatic answer retrieval (Masters et al., 1993). If so, then participants should choke on all highly practiced problems regardless of demand—at least to the extent that the problems are solved via automatic answer retrieval.

Furthermore, even if practice serves to proceduralize the algorithm rather than shift performance to automatic answer retrieval as Logan (1988) proposed, the performance of practiced modular arithmetic problems might still fail via pressure-induced attention that serves to disrupt well-learned algorithmic processes. Similar to the disruption of automatic retrieval from memory, disruption of a proceduralized algorithm should also harm performance on all problems, as practiced algorithms, regardless of working memory demands, should be harmed by the instantiation of explicit attentional control mechanisms that slow down or disrupt highly efficient computations (although perhaps the impact would be greater on high-demand problems whose algorithmic solutions include more steps). These differing mechanisms and domains of applicability of distraction versus explicit monitoring theories allow for contrasting predictions concerning when and with which problems choking should be observed in Experiment 2.

Method

Participants

Participants ($N = 22$) were students enrolled at Michigan State University who were not math majors and reported having no previous exposure to modular arithmetic. An additional participant's data were not included in the following analyses because accuracy in the first low-pressure test was less than the 55% correct minimum accuracy criterion.

Procedure

Participants filled out a consent form and a demographic sheet detailing previous math experience and were introduced to the modular arithmetic task used in Experiment 1. Individuals first performed 12 practice problems, presented in a different random order to each participant. Four low-demand problems required a single-digit subtraction operation, such as $7 \equiv 2 \pmod 5$, and 4 high-demand problems required a double-digit borrow subtraction operation, such as $51 \equiv 19 \pmod 4$. An additional 4 problems with intermediate attentional demands requiring a double-digit no-borrow subtraction procedure, such as $15 \equiv 10 \pmod 3$, were also included. As in Experiment 1, these intermediate-level problems served as filler problems, intended to diminish the contrast between the low-demand and the high-demand problems. Half of the problems within each demand level were true, half were false, and each problem was presented once.

In addition, each true problem had a false correlate that only differed as a function of the number involved in the mod statement. As in Experiment 1, this pairing was designed to equate the particular numbers used in the true and false problems. In Experiment 2, this pairing was also designed to help prevent individuals from merely using stimulus recognition of the first few numbers in repeated problems to retrieve the verification answer. That is, because true and false problems only varied as a function of the mod number, individuals were forced to encode the entire problem to retrieve the correct answer into working memory. The goal was to ensure that savings as a function of practice were mainly due to the shift from algorithmic processing to retrieval rather than due to incomplete encoding of the problem's terms.

After the practice problems, individuals completed a 12-problem low-pressure test (LP1) and a 12-problem high-pressure test (HP1), which were separated by a short break. The problems in LP1 and HP1 were presented in a different random order to each participant. Each problem appeared only once in either LP1 or HP1, and the problems in LP1 and HP1 were counterbalanced across participants. Within both LP1 and HP1, there were 4 low-demand problems, 4 high-demand problems, and 4 filler problems. Half of the problems within each demand level were true, half were false, and each true problem had a false correlate.

To the participant, LP1 appeared to be just another series of practice problems. That is, participants were simply told to perform the next block of practice. After LP1, participants were given the same pressure scenario used in Experiment 1, with the exception that individuals in Experiment 2 were informed that they were about to enter the first of two test situations in the experiment and that they had to improve their modular arithmetic performance by the required amount in both test situations to receive the monetary award for themselves and their partner.

After HP1, individuals were informed that they would return to practicing modular arithmetic problems (modular arithmetic training). Participants were presented with 12 new problems: 4 low-demand problems, 4 high-demand problems, and 4 intermediate-level filler problems. Half of the problems within each demand level were true, half were false, and each true problem had a false correlate. Each problem within the training session was repeated 48 times for a total of 576 trials, which were separated into three blocks of 192 problems each, with a short break after each block. Within each block, each problem was repeated 16 times.

Participants then took part in the second 12-problem low-pressure test (LP2) and the second 12-problem high-pressure test (HP2). The problems within LP2 and HP2 were the same 12 problems that were presented 48 times each during training and were presented in a different random order to each participant. There was no need to counterbalance these problems across the second low- and high-pressure tests, as the same 12 problems were used in both LP2 and HP2.

As in LP1, participants were not made aware of the LP2 test situation. They were not told that this block would be shorter, nor were they given any other cues. To the participant, LP2 appeared to be just another series of practice problems. The experimenter then informed participants that they were about to take part in the second test situation, repeated the high-pressure scenario—including the instruction that performance needed to improve by 20% in both tests—and turned on the video camera. Participants then completed HP2. Individuals were debriefed and given the monetary award regardless of their performance. In total, participants were exposed to 50 presentations of each of the 12 training problems (48 exposures during the training session, 1 exposure in LP2, and 1 exposure in HP2).

The self-report measures of performance pressure and anxiety administered in Experiment 1 were not used in Experiment 2. Experiment 2 was a completely within-participant design. Although one might imagine that the questionnaires could be administered multiple times to create a within-participant comparison (i.e., once prior to and once after each high-pressure scenario), pilot testing revealed that asking for such offline measures of performance pressure prior to the introduction of a high-pressure situation significantly increased participants' skepticism regarding the validity of the pressure manipulation. Therefore, to present the strongest pressure manipulation possible, the questionnaires were excluded. As is seen later, however, behavioral evidence of choking made it clear that the manipulation was again effective.

Results

As in Experiment 1, data were purged of outliers on the basis of RT. RTs were computed for each problem. RTs more than 3 standard deviations below or above an individual's mean RT for each block of problems were considered outliers, and the trials on which such RTs occurred were removed from the data set. This resulted in the dismissal of 3 RTs and their corresponding accuracy scores from the following analyses.

Accuracies can be seen in Figure 2. A 2 (low-pressure test, high-pressure test) $\times$ 2 (before training, after training) $\times$ 2 (low-demand problems, high-demand problems) ANOVA on accuracy revealed a significant Pressure Test $\times$ Training $\times$ Problem Demand interaction, $F(1, 21) = 16.03$, $p < .01$, $MSE = 0.01$, $\eta_p^2 = .43$, which was the central result of the experiment. In addition, there was a main effect of pressure test, $F(1, 21) = 8.05$, $p < .01$, $MSE = 0.01$; a main effect of training, $F(1, 21) = 28.71$, $p < .01$, $MSE = 0.02$; a main effect of problem demand, $F(1, 21) = 12.82$, $p < .01$, $MSE = 0.02$; a Training $\times$ Pressure Test interaction, $F(1, 21) = 7.15$, $p < .02$, $MSE = 0.02$; a Training $\times$ Problem Demand interaction, $F(1, 21) = 19.05$, $p < .01$, $MSE = 0.02$; and a Pressure Test $\times$ Problem Demand interaction, $F(1, 21) = 11.34$, $p < .01$, $MSE = 0.01$.

To pursue the three-factor interaction, we analyzed data from the low- and high-pressure tests before training and after training separately. A 2 (low-pressure test, high-pressure test) $\times$ 2 (low-demand problems, high-demand problems) ANOVA prior to training revealed a main effect of pressure test, $F(1, 21) = 8.66$, $p < .01$, $MSE = 0.02$, and a main effect of problem demand, $F(1, 21) = 18.77$, $p < .01$, $MSE = 0.01$, which were qualified by a significant Pressure Test $\times$ Problem Demand interaction, $F(1, 21) = 16.74$, $p < .01$, $MSE = 0.02$, $\eta_p^2 = .44$. In the tests before

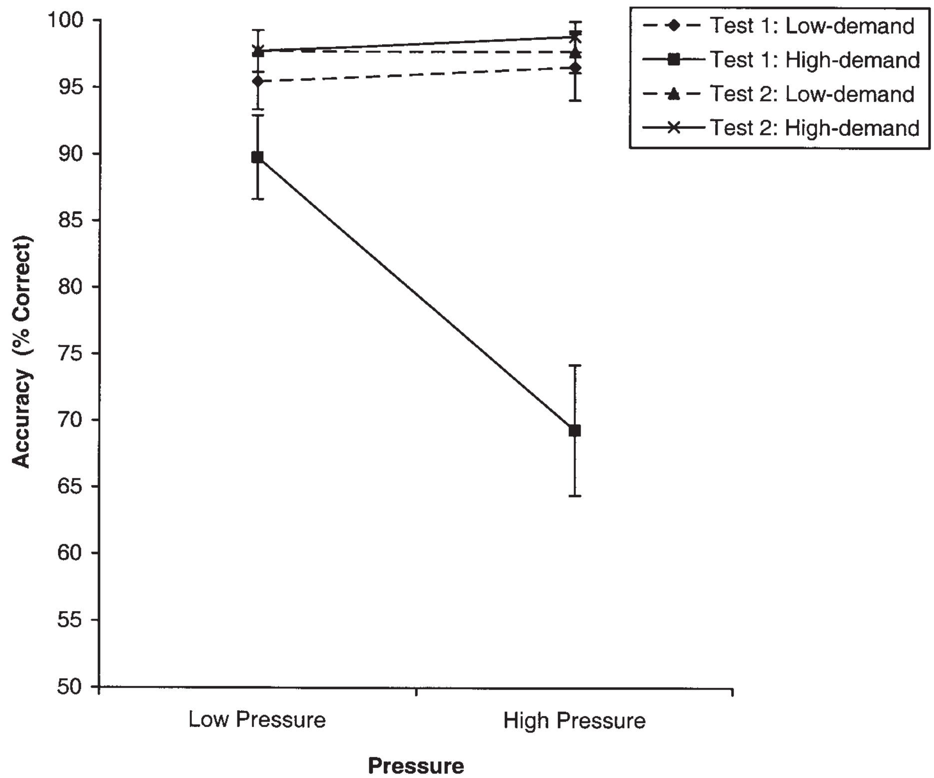

Figure 2. Mean accuracy (% correct) for the low- and high-pressure tests prior to modular arithmetic training (Test 1) and after modular arithmetic training (Test 2) for the low-demand and the high-demand problems in Experiment 2. Error bars represent standard errors.

training, whereas performance on the low-demand problems did not differ from the low- to high-pressure test, $t(21) = 0.44$, $p = .67$, performance on the high-demand problems got significantly worse, $t(21) = 3.81$, $p < .01$, $d = 1.05$. This finding replicates the outcome of Experiment 1 and suggests that in mathematical problem solving, pressure impacts performance on only those problems that rely most heavily on working memory. A similar 2 (low-pressure test, high-pressure test) $\times$ 2 (low-demand problems, high-demand problems) ANOVA after training revealed no Pressure Test $\times$ Problem Demand interaction, $F < 1$.

Thus, modular arithmetic performance decrements under pressure occurred prior to extended problem training but not after it. This pattern of data supports the predictions of distraction theories as an explanation for the choking phenomenon in that only novel modular arithmetic problems, requiring the execution of a capacity-demanding rule-based solution algorithm, were performed poorly under pressure. After participants extensively practiced the problems being tested, behavioral evidence of choking was no longer observed.

As in Experiment 1, analyses of RTs did not alter these conclusions. A 2 (low-pressure test, high-pressure test) $\times$ 2 (before training, after training) $\times$ 2 (low-demand problems, high-demand problems) ANOVA on RT indicated main effects of training, $F(1, 21) = 71.49$, $p < .01$, $MSE = 69.73 \times 10^5$, and problem demand, $F(1, 21) = 33.58$, $p < .01$, $MSE = 63.91 \times 10^5$, which were qualified by a Training $\times$ Problem Demand interaction, $F(1, 21) = 30.91$, $p < .01$, $MSE = 61.95 \times 10^5$, $\eta_p^2 = .60$. There was no Pressure Test $\times$ Training $\times$ Problem Demand interaction, $F < 1$.

The low-demand problem RTs were faster than the high-demand problem RTs. In addition, for both the low-demand and high-demand problems, RTs were slower prior to modular arithmetic training in LP1 and HP1 (low-demand problems in LP1, $M = 2{,}302$ ms, $SE = 170$ ms; low-demand problems in HP1, $M = 2{,}216$ ms, $SE = 154$ ms; high-demand problems in LP1, $M = 6{,}474$ ms, $SE = 813$ ms; high-demand problems in HP1, $M = 6{,}634$ ms, $SE = 810$ ms) than after training in LP2 and HP2 (low-demand problems in LP2, $M = 1{,}054$ ms, $SE = 83$ ms; low-demand problems in HP2, $M = 905$ ms, $SE = 51$ ms; high-demand problems in LP2, $M = 1{,}198$ ms, $SE = 111$ ms; high-demand problems in HP2, $M = 1{,}006$ ms, $SE = 65$ ms). Finally, problem demand level had an effect on RT prior to but not after modular arithmetic training. This indicates that practicing these problems eliminated the initial time differences in their solution. Such a finding is consistent with the kind of shift from algorithmic execution to direct answer retrieval from memory for all problems proposed by Logan's (1988) theory of retrieval-based automaticity. Thus, if Masters et al. (1993) were correct in asserting that pressure disrupts direct retrieval of solutions from memory, we should have observed choking in the high-pressure test after training as we did in the high-pressure test before training, but we did not.

Finally, it is necessary to check for evidence of a speed–accuracy trade-off. As in Experiment 1, the lack of a Pressure Test $\times$ Training $\times$ Problem Demand interaction in RTs suggests that the above reported accuracy results are not the product of a speed–accuracy trade-off. In fact, as accuracy for the high-demand problems declined from LP1 to HP1, RT increased, although not significantly, $t(21) = 0.38$, *ns*.

Discussion

In Experiment 2, performance on novel modular arithmetic problems, whose solutions require the maintenance of intermediate problem steps and their products in working memory, declined under pressure. However, as in Experiment 1, these decrements were limited to the most working memory–demanding problems. In contrast, well-learned modular arithmetic problems, thought to be supported by the one-step direct retrieval of past problem instances from long-term into working memory, showed no signs of performance failure under pressure, regardless of problem demand.

According to distraction theories, modular arithmetic performance should be most susceptible to pressure-induced decrements at low levels of practice when working memory demands are greatest and pressure-induced worries impinge on task-relevant processing resources. This prediction was clearly borne out. Not only was choking solely observed prior to modular arithmetic training, but, similar to Experiment 1, performance decrements under pressure were limited to problems that incurred the highest working memory load.

The first two experiments provide support for distraction theories and add weight to the argument against explicit monitoring theories as an explanation for choking under pressure as observed in the working memory–intensive task of modular arithmetic. In Experiment 3, we further explored performance under pressure in modular arithmetic by considering the role of general practice at the algorithm. This was achieved by comparing participants' ability to solve infrequently practiced problems under pressure with their ability to solve heavily practiced problems under pressure at similarly high overall levels of general algorithmic practice.

Experiment 3

Participants performed over 700 modular arithmetic practice problems (presented 1, 2, or 50 times each) prior to being exposed to low- and high-pressure tests. As previously discussed, the task control structures of modular arithmetic problems should change most dramatically as a function of specific problem exposure, not necessarily experience at performing many different problems (Logan, 1988). Although practice at many different problems may improve the efficiency of the algorithm somewhat (Rickard, 1997; Touron et al., 2001), application of the algorithm to an unpracticed problem ought still to engage the control and storage operations of working memory in the algorithm's execution. Thus, if choking is due to pressure-induced capacity limitations, as distraction theories would propose, then regardless of how many different problems individuals have been exposed to, performance on only high-demand problems that have not been repeatedly practiced to the extent that their answers are retrieved directly from long-term memory into working memory should be harmed by pressure.

If practice produces sufficiently substantial increases in efficiency and proceduralization of algorithmic computations, it is possible that performance failures under pressure on novel problems after extensive practice of the algorithm could be explained by explicit monitoring theories. By analogy to what happens with well-practiced sensorimotor skills (Beilock & Carr, 2001), pressure-induced attention may serve to disrupt highly efficient, proceduralized algorithmic computations. If so, this type of failure should be evident across all problem demand levels, as practiced algorithms, regardless of working memory demands, should be harmed by the instantiation of explicit attentional control mechanisms that slow down or disrupt highly efficient computations. Thus, in contrast to Experiment 1, even performance on novel low-demand problems should suffer under pressure, as well as that on novel high-demand problems, if the algorithm has become proceduralized enough to be disrupted by pressure-induced explicit monitoring. As we said before, it is possible that this impact would be greater for high-demand problems whose algorithmic solutions include more steps. Nonetheless, if pressure-induced explicit monitoring is responsible for the disruption of sufficiently proceduralized algorithms, at least some sign of performance decrement in low-demand problems should be observed.

Method

Participants

Participants ($N = 28$) were students enrolled at Miami University who were not math majors and reported no previous exposure to modular arithmetic. As in Experiments 1 and 2, a minimum 55% correct accuracy criterion was applied in Experiment 3. Four additional participants' data were not included in the following analyses because their low-pressure test accuracy was less than 55% correct for one or more of the problem repeat types (see definitions below).

Procedure

Participants filled out a consent form and a demographic sheet detailing previous math experience and were introduced to the modular arithmetic task. Participants first performed 720 training problems divided into three blocks of 244 problems, 232 problems, and 244 problems, respectively. Across all training problems, there were 12 problems (6 low demand and 6 high demand) repeated 50 times each (multiple-repeat problems), 48 problems (24 low demand and 24 high demand) repeated once (once-repeat problems), and 24 problems (12 low demand and 12 high demand) shown one time (no-repeat problems). Within each problem demand and repeat level, half the problems were true, half were false, and each true problem had a false correlate.

After practice, participants took part in a 36-problem low-pressure test and a 36-problem high-pressure test. The low-pressure test consisted of the 6 low-demand and 6 high-demand problems repeated 50 times each during training (multiple repeats), 6 low-demand and 6 high-demand problems repeated once (once repeats), and 6 low-demand and 6 high-demand problems not previously presented (no repeats). The high-pressure test consisted of the 6 low-demand and 6 high-demand multiple repeats, 6 new low-demand and 6 new high-demand once repeats, and 6 new low-demand and 6 new high-demand no repeats. Problems within these tests were presented in a random order to each participant, and the once- and no-repeat problems used in the low- and high-pressure tests were counterbalanced across participants. To the participant, the low-pressure test appeared to be just another series of practice problems. Individuals were then given the same high-pressure scenario used in the first two experiments.

Because there was only one high-pressure test and it came at the end of the problem-solving protocol, we were able to return to our strategy in Experiment 1 of questioning participants about their phenomenological reactions to the high-pressure situation. After completing the high-pressure test, participants were first asked to fill out a retrospective verbal report questionnaire intended to shed light on their thoughts during the high-pressure test (Verbal Thought Questionnaire). Specifically, individuals were given a questionnaire that stated, “We all have several thoughts that run through our mind at any given time. Please describe everything that you remember thinking about as you performed the last set of modular arithmetic problems.”

As in Experiment 1, participants then filled out the State Anxiety form of the STAI (Spielberger et al., 1970) and answered questions regarding their perceptions of the importance of performing at a high level in the high-pressure test, their perceptions of performance pressure, and how well they felt they performed in the high-pressure test. Participants were fully debriefed and given the monetary award regardless of their performance.

Results

Questionnaires

Verbal Thought Questionnaire. Responses were divided into five categories:

1. Thoughts about the high-pressure situation and its consequences. Some examples include “I am not gonna get the money,” “I hope I don’t look stupid,” “It can’t all depend on me,” and our favorite (although X-rated), “Oh f***. There is no way I am going to increase my score by 20%.”

2. Thoughts related to carrying out the steps involved in performing the math problems. For example, “19 does not equal 9,” “First I looked at the digit in the ones place of the number being subtracted.”

3. Thoughts related to increased focus on a specific aspect of performing the math problems. An example is “A desire to focus on the screen.”

4. Thoughts concerning a lack of perceived pressure. For example, “Since I did not know the person, the pressure to improve was not that heavy.”

5. Thoughts unrelated to the experimental situation. An example includes “I thought about what I was going to do today.”

The first category was designed to test distraction theories’ hypothesis that performance pressure results in thoughts or worries about the high-pressure situation and its outcome. The second and third categories were designed to test explicit monitoring theories’ hypothesis that pressure enhances the attention individuals allocate to specific aspects of performing the math problems. Two experimenters independently coded the Verbal Thought Questionnaire data. Interexperimenter reliability was high across all five categories (Category 1, $r = .86$; Category 2, $r = .91$; Category 3, $r = .89$; Category 4, $r = .92$; Category 5, $r = .94$).

On average, participants reported about 5.5 thoughts in total ($M = 5.46$, $SE = 0.60$). These thoughts broke down into the following: Category 1, $M = 52.75\%$, $SE = 7.24\%$; Category 2, $M = 11.98\%$, $SE = 4.13\%$; Category 3, $M = 2.17\%$, $SE = 1.12\%$; Category 4, $M = 3.78\%$, $SE = 2.54\%$; and Category 5, $M = 29.32\%$, $SE = 6.64\%$. Overall, thinking or worrying about the high-pressure situation and its consequences accounted for slightly more than half of participants’ reported thoughts during the high-pressure test, and other potentially distracting thoughts not directed at the steps of performance per se (Categories 4 and 5) accounted for another third. As is shown below, the Verbal Thought Questionnaire data, when combined with the self-report measures of perceived pressure and anxiety and actual modular arithmetic performance, provide converging evidence that pressure-induced distraction is responsible for choking under pressure in mathematical problem solving.

Importance. As in Experiment 1, participants reported that it was at least moderately important for them to perform at a high level in the last block of modular arithmetic problems, the high-pressure test ($M = 4.12$, $SE = 0.32$).

State anxiety. Participants reported high state anxiety levels ($M = 38.96$, $SE = 1.70$). Because Experiment 3 was entirely within subjects, we compared these state anxiety scores with those reported by the low-pressure group in Experiment 1 to demonstrate that the high-pressure manipulation in Experiment 3 significantly increased participants’ state anxiety beyond that of a control group that did not receive a pressure manipulation. The state anxiety scores reported by participants in Experiment 3 were significantly higher than those of the low-pressure group in Experiment 1 ($M = 32.08$, $SE = 1.20$), $t(66) = 3.41$, $p < .01$, $d = 0.83$.

Performance pressure. Participants’ perceptions of performance pressure ($M = 4.95$, $SE = 0.25$) in the last block of problems were also significantly higher than those reported by the low-pressure group in Experiment 1 ($M = 3.95$, $SE = 0.24$), $t(66) = 2.82$, $p < .01$, $d = 0.70$.

Performance success. Participants’ perceptions of their performance in the last block of problems ($M = 4.12$, $SE = 0.29$) was significantly lower than that reported by the low-pressure group in Experiment 1 ($M = 4.98$, $SE = 0.19$), $t(66) = 2.57$, $p < .02$, $d = 0.62$.

Thus, participants in Experiment 3 reported significantly higher levels of state anxiety, reported significantly heightened perceptions of performance pressure, and felt they performed significantly worse than did the low-pressure group in Experiment 1. This suggests that our pressure manipulation was again successful in achieving its goal. We now turn to actual task performance.

Accuracy and RT

RTs were computed for each problem and retained for only those problems answered correctly. As in Experiments 1 and 2, RTs more than 3 standard deviations below or above an individual’s mean RT for each block of problems were used to identify trials that were outliers, and these trials were removed. This resulted in the dismissal of 20 RTs and corresponding accuracy scores from the following analyses.

We first compared the accuracy of the no-repeat and once-repeat problems in a 2 (low-pressure test, high-pressure test) × 2 (once-repeat problems, no-repeat problems) × 2 (low-demand problems, high-demand problems) ANOVA, which revealed no Pressure Test × Problem Repetition × Problem Demand interaction, $F < 1$. Thus, only one encounter with a problem during training was not enough to change participants' performance on it relative to the completely novel no-repeat problems.

Data from the no-repeat problems were then used to compare the performance of the least practiced problems with the most practiced problems as a function of pressure and problem difficulty. As seen in Figure 3, the central result of the experiment was again a three-factor interaction for accuracy. A 2 (low-pressure test, high-pressure test) × 2 (no-repeat problems, multiple-repeat problems) × 2 (low-demand problems, high-demand problems) ANOVA on accuracy revealed a main effect of problem repetition, $F(1, 27) = 45.81$, $p < .01$, $MSE = 0.02$; a main effect of problem demand, $F(1, 27) = 40.31$, $p < .01$, $MSE = 0.02$; a Pressure Test × Problem Demand interaction, $F(1, 27) = 6.74$, $p < .02$, $MSE = 0.01$; and a Problem Repetition × Problem Demand interaction, $F(1, 27) = 12.35$, $p < .01$, $MSE = 0.02$, which were all qualified by the significant Pressure Test × Problem Repetition × Problem Demand interaction, $F(1, 27) = 5.28$, $p < .03$, $MSE = 0.01$, $\eta_p^2 = .16$.

We pursued this three-factor interaction by analyzing the heavily practiced multiple-repeat problems and novel no-repeat problems separately. A 2 (low-pressure test, high-pressure test) × 2 (low-demand problems, high-demand problems) ANOVA within the multiple-repeat problems revealed no Pressure Test × Problem Demand interaction, $F > 1$. In contrast, a 2 (low-pressure test, high-pressure test) × 2 (low-demand problems, high-demand problems) ANOVA within the no-repeat problems revealed a significant Pressure Test × Problem Demand interaction, $F(1, 27) = 9.12$, $p < .01$, $MSE = 0.01$, $\eta_p^2 = .25$. When we used 95% confidence intervals, accuracy for the no-repeat low-demand problems was significantly higher in the high-pressure test than in the low-pressure test, $d = 0.49$, whereas accuracy for the no-repeat high-demand problems was significantly lower in the high-pressure test than the low-pressure test, $d = 0.37$. Under pressure, performance on the low-demand problems improved at the same time that performance on the high-demand problems declined.

Again, the analysis of RT data did not alter the conclusions supported by the accuracy analysis. A 2 (low-pressure test, high-pressure test) × 2 (no-repeat problems, multiple-repeat problems) × 2 (low-demand problems, high-demand problems) ANOVA on RT revealed main effects of problem repetition, $F(1, 27) = 84.64$, $p < .01$, $MSE = 11.09 \times 10^5$, and problem demand, $F(1, 27) = 90.31$, $p < .01$, $MSE = 33.56 \times 10^5$, which were qualified by a significant Problem Repetition × Problem Demand interaction, $F(1, 27) = 57.33$, $p < .01$, $MSE = 13.35 \times 10^5$, $\eta_p^2 = .68$. No Pressure Test × Problem Repetition × Problem Demand interaction, $F(1, 27) = 1.65$, $p = .21$, $MSE = 46.45 \times 10^4$, was obtained.

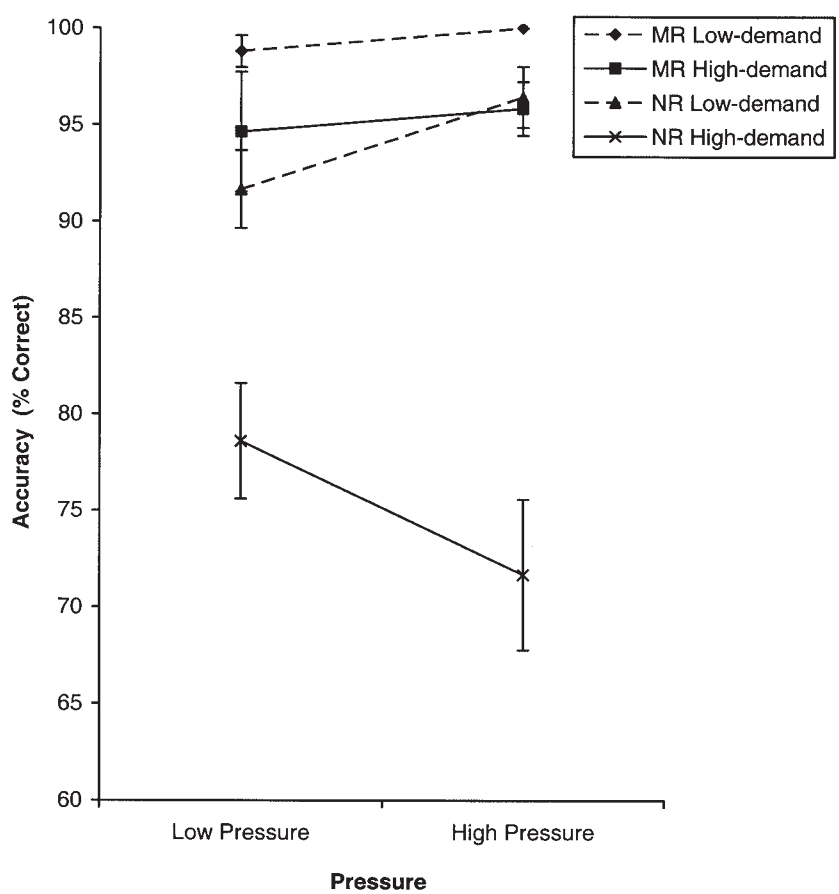

Figure 3. Mean accuracy (% correct) for the low- and high-pressure tests for the multiple-repeat (MR) and no-repeat (NR) low-demand and high-demand problems in Experiment 3. Error bars represent standard errors.

Multiple-repeat problem RTs were faster than no-repeat problem RTs for the high-demand problems (multiple repeat, M = 2,305 ms, SE = 257 ms; no repeat, M = 4,769 ms, SE = 392 ms), more so than for the low-demand problems (multiple repeat, M = 1,147 ms, SE = 87 ms; no repeat, M = 1,273 ms, SE = 87 ms).

As in Experiments 1 and 2, the lack of a Pressure Test $\times$ Problem Repetition $\times$ Problem Demand interaction in RTs suggests the accuracy results are not the product of a speed–accuracy trade-off. However, to further guard against the possibility that a pressure-induced reduction in accuracy (as seen in the no-repeat high-demand problems) was accompanied by the opposite effect in RT, we examined the correlation between (a) the accuracy difference score of the no-repeat high-demand problems from the low-pressure test (M = 78.6%, SE = 3.0%) to the high-pressure test (M = 71.7%, SE = 3.9%) and (b) the RT difference score of the no-repeat high-demand problems from the low-pressure test (M = 4,962 ms, SE = 493 ms) to the high-pressure test (M = 4,576 ms, SE = 417 ms). A nonsignificant correlation was found ($r = -.21$, $p = .29$). This negative correlation is in the opposite direction of what one would expect if a speed–accuracy trade-off were playing a role. Finally, we compared no-repeat high-demand problem accuracy in the low-pressure test and the high-pressure test while covarying out the no-repeat high-demand RT difference score (i.e., low-pressure test RT $-$ high-pressure test RT). The significant pressure-induced drop in accuracy reported above remained significant, $F(1, 26) = 4.75$, $p < .04$, $MSE = 0.02$, $\eta_p^2 = .16$.

Discussion

Experiment 3 further supports distraction theories of choking in the working memory–intensive task of modular arithmetic. Participants were asked to report their thoughts in the high-pressure test. Over half of these reports consisted of worries about the situation and its consequences. Furthermore, individuals reported high levels of state anxiety and perceptions of performance pressure in the high-pressure test. These self-reports were accompanied by behavioral evidence of pressure-induced working memory consumption. Although modular arithmetic problems practiced 50 times each and thus not heavily reliant on working memory were not performed poorly under pressure, problems presented only once were. Furthermore, these failures were limited to the no-repeat problems that placed the heaviest demands on working memory. In fact, the performance of the no-repeat low-demand problems significantly improved under pressure, similar to performance of such problems in Experiment 1. This outcome is not compatible with explicit monitoring of proceduralized algorithms playing a role in response to pressure over and above that played by distraction. That is, had the algorithm become sufficiently proceduralized that pressure-induced explicit monitoring could cause its disruption, no improvement under pressure should have been observed, even for the simpler version of the algorithm. The current data are completely consistent with the predictions of distraction theory.

Similar to the results from Experiments 1 and 2, these results parallel findings in the test anxiety literature demonstrating the most pronounced performance decrements in anxiety-provoking and hence capacity-limiting situations for those problems with large online working memory demands (Ashcraft & Kirk, 2001; Darke, 1988). The results of Experiment 3 also suggest that it is not merely an issue of general task practice that earmarks those skills most susceptible to choking but rather the structural demands of the specific problems being performed.

General Discussion

The present work was designed to explore the impact of pressure in a task with a control structure that might make performance susceptible to choking via distraction, at least at low levels of practice, with a possible shift of mechanisms to choking via explicit monitoring at high levels of practice. Explicit monitoring theories suggest that performance pressure prompts attention to skill processes and their step-by-step control. Attention to execution at this component level is thought to disrupt the proceduralized or automated processes of high-level skills that are normally run off without such explicit attention (Baumeister, 1984; Beilock & Carr, 2001; Lewis & Linder, 1997; Masters, 1992), and it has been speculated that attention at this level might also disrupt automated one-step memory retrieval (Masters et al., 1993). However, distraction theories propose that pressure serves to create a dual-task environment in which controlling execution of the task at hand and performance worries divide the attentional capacity once devoted solely to primary task performance (Beilock & Carr, in press; Lewis & Linder, 1997). Although explicit monitoring theories have received substantial support in accounting for the choking phenomenon, most of this evidence has come from well-learned sensorimotor tasks that automate via proceduralization (Beilock & Carr, 2001; Marchant & Wang, 2001). These skills may not be an adequate domain in which to test distraction theories' predictions.

In Experiment 1, individuals assigned to either a low- or a high-pressure group performed novel modular arithmetic, problems whose performance should be susceptible to decrements under pressure as a result of distraction but not explicit monitoring. Individuals in the high-pressure group had significantly increased levels of state anxiety and perceptions of performance pressure compared with the low-pressure group participants. Additionally, individuals in the high-pressure group performed at a significantly lower accuracy level on the modular arithmetic problems than their low-pressure counterparts did. However, this lower accuracy was limited to those problems with the heaviest working memory demands. In fact, problems that did not incur heavy working memory demands were performed significantly better under pressure.

Experiment 2 extended the examination of performance under pressure in modular arithmetic to include highly practiced problems. Again, only the most capacity-demanding modular arithmetic problems were performed poorly under pressure. Furthermore, these pressure-induced failures were limited to the performance of problems with low levels of practice (in which problem solutions were based on algorithms requiring the online maintenance of intermediate results in working memory). Once problems were repeatedly practiced so that their answers were retrieved directly from long-term memory into working memory, choking under pressure was no longer observed.

In Experiment 3, individuals performed modular arithmetic practice problems presented either 1, 2, or 50 times and were then exposed to a high-pressure test. Again, participants showed performance decrements under pressure only on working memory–intensive modular arithmetic problems that had not been highly practiced. And again, performance significantly improved on unpracticed problems with low working memory demands under pressure.

These findings are consistent with distraction theories of choking and suggest that pressure-induced capacity limitations may result in performance decrements in tasks with the right properties to be harmed by such constraints. In particular, choking via distraction may occur in tasks that require a sequence of mental operations with interdependent demands on storage and processing rather than direct retrieval of an answer from long-term memory (Logan, 1988; Zbrodoff & Logan, 1986) or the execution of a proceduralized motor program (Beilock & Carr, 2001; Brown & Carr, 1989; Keele, 1986; Keele & Summers, 1976).

It should be noted that in two of our three experiments, the low-pressure tests always preceded the high-pressure tests. This set order was designed to make the low-pressure test as innocuous and inconspicuous as possible. In fact, across all experiments, to the participant, the low-pressure test appeared to be just another series of practice problems. This set pressure order does allow for the possibility that the effects reported above are due to the order of the tests rather than pressure per se. However, having the low-pressure test always precede the high-pressure test should only make it harder to find pressure-induced performance decrements. Individuals always had more practice by the time they reached each high-pressure test than they had at the time of the comparison low-pressure test. Thus, it should be the most difficult to find performance decrements (relative to a less practiced low-pressure test) at this point. It is possible that the lack of pressure-induced performance decrements in the second high-pressure test in Experiment 2 was due to this additional practice. However, the fact that pressure-induced performance decrements were found after extended practice in Experiment 3 for problems that were not repeatedly practiced but not for problems that were highly practiced (similar to the problems in Experiment 2) suggests that the order of the second high-pressure test in Experiment 2 was not responsible for the observed results.

Additionally, one might be concerned about the generalizability of the above results based on a verification task (in which the goal is to produce an answer of either *true* or *false*) to other math tasks that involve production (in which the goal is to produce an answer of a specific number). For unpracticed problems, one must work through the problem to arrive at the solution regardless of whether the answer is *true* or *false* (as in modular arithmetic) or a number (as in a math production task). Similarly, in terms of performance on problems that have been practiced to the extent that their answers are being retrieved directly from long-term memory into working memory, a similar process should be used to retrieve either a numerical answer or a *true* or *false* answer. That is, both a practiced production task and a practiced verification task rely on perceptual cues to automatically retrieve the answer trace into working memory. The actual form of the answer (i.e., *true*, *false*, or a number) should not matter.

Although finding support for distraction theories in mathematical problem solving sheds new light on the phenomenon of choking under pressure, it also begs additional questions. Namely, given the extensive support for explicit monitoring theories outlined in the introduction, how can distraction and explicit monitoring both be viable explanations for choking?

Working Memory–Intensive Tasks Versus Automated Skills

It may be that a simple dichotomy based on controlled and working memory–intensive versus automatic processing is sufficient to solve this problem: Distraction creates performance decrements under pressure in tasks that engage attention and make heavy demands on the limited resources of working memory, whereas explicit monitoring creates choking in skills that have become automatic through practice. However, although this hypothesis is at first glance appealing, it is too simple. Two kinds of evidence work against it.

The first piece of evidence is that not all “automatic” performances show signs of pressure-induced failure. This applies to Logan’s (1988) alphabet arithmetic task (Beilock & Carr, 2001), as well as to the experiments on more complicated mathematical problem solving in the present work. Tasks such as these are thought to automate via a shift to direct retrieval of answers from memory, as opposed to tasks that automate via a shift to reliance on proceduralized motor programs. Thus, the mechanism that underlies the automated version of performance is not the same for all types of tasks. And this difference in how tasks automate appears to matter in terms of whether a task is susceptible to choking at high levels of practice.

The second piece of evidence is the converse of the first. Not all unpracticed and hence working memory–intensive tasks demonstrate performance failures under pressure: Sensorimotor skills do not, and this is a puzzle. Novice sensorimotor tasks are thought to be based on declaratively accessible performance rules (Proctor & Dutta, 1995) and have been shown to be harmed by dual-task manipulations (Allport, Antonis, & Reynolds, 1972; Beilock, Carr, et al., 2002; Gray, 2004). Hence, participants performing these tasks should show signs of choking under pressure due to distraction, just as participants in the present work showed for math problem solving. To test this idea, Beilock and Carr (2001) had participants practice a golf putting task. The participants were exposed to high-pressure situations both early and late in practice. Early in practice, pressure to do well actually facilitated execution. Only at the later stages of learning did performance decrements under pressure emerge.

Pressure’s Dual Impact: Cognitive Versus Sensorimotor Skills

Thus, a working memory versus automated distinction does not appear to adequately explain the choking-under-pressure results to date. It may be that rather than pressure having one kind of impact on attentional control in one type of task and another kind of impact on attentional control in a different task, the imposition of pressure creates two effects that alter how attention is allocated to execution: (a) Pressure induces worries about the situation and its consequences, thereby reducing working memory capacity available for performance, as distraction theories would propose, and (b) at the same time, pressure prompts individuals to attempt to control execution to ensure optimal performance, in line with explicit monitoring theories. It may be that these two effects are differentially relevant to performance depending on the specific composition of the control structures governing performance and that cognitive and sensorimotor skills often differ in this regard.

For example, cognitive tasks whose performance shows decrements under pressure seem to rely on working memory in a very different way than either novice or well-learned sensorimotor skills do. That is, skills such as modular arithmetic appear to be based on a hierarchical and sequentially dependent task representation in which initial steps and memory for their outcomes and products are used to generate subsequent processes and final solutions. In modular arithmetic problems such as $72 \equiv 39$ (mod 4), for example, the derivation of a final problem solution is dependent on the correct answer to the first subtraction operation, which is in turn reliant on the successful maintenance of the intermediate steps necessary to produce the borrow operation.

Well-learned sensorimotor skills that operate largely outside of working memory are almost certainly not based on such a representation (Fitts & Posner, 1967; Proctor & Dutta, 1995). Furthermore, novice sensorimotor skills do not appear to depend on this type of task representation either. Despite the fact that unpracticed motor skills may be based, in part, on explicitly accessible declarative knowledge (Beilock, Wierenga, & Carr, 2002), this knowledge is not organized in such a fashion that the execution of each element of performance is dependent on the maintenance of every prior step. Novice golfers may have explicit access to such skill rules as "keep knees bent." However, subsequent steps in performance such as "bring club back straight" are not dependent on this knowledge in the same way as a borrow operation in modular arithmetic is dependent on the maintenance of the specific numbers necessary to carry out this operation. Hence, it may be the sequentially dependent interweaving of processing and information-storage demands that makes a complex cognitive task susceptible to choking via distraction. If so, this would elucidate why most of the support for explicit monitoring theories has originated in sensorimotor skills, whereas support for performance decrements being a result of distracting environments has been found in working memory–intensive cognitive tasks such as mental arithmetic (Ashcraft & Kirk, 2001) and analogical reasoning (Tohill & Holyoak, 2000). Future research examining the idea that the manner in which skill performance fails varies as a function of the specific composition of the control structures supporting execution seems very likely to further knowledge of the choking-under-pressure phenomenon.

In conclusion, the findings of the three experiments in the present study lend support to the notion that both explicit monitoring and distraction are viable explanations for the choking phenomenon. More research in this area is needed to identify the precise mechanisms of the cognitive and sensorimotor skills that demonstrate performance patterns under pressure consistent with each type of theory. One product of such research will be a taxonomy of skills based on a clear understanding of the nature and representation of their control structures at different levels of expertise. Continued exploration of the choking phenomenon across diverse skill domains will speak to task type, skill level, and individual differences in susceptibility to performance failures and ultimately to means of engineering training regimens to diminish such susceptibility—knowledge that will benefit researchers, practitioners, and performers alike.

References

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

Anderson, J. R. (1982). Acquisition of a cognitive skill. *Psychological Review, 89,* 369–406.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192–210.

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

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44,* 75–106.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130,* 224–237.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), Models of working memory: Mechanisms of active maintenance and executive control (pp. 28–61). New York: Cambridge University Press.

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology, 46,* 610–620.

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review, 11,* 373–379.

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130,* 701–725.

Beilock, S. L., & Carr, T. H. (in press). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied, 8,* 6–16.

Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A,* 1211–1240.

Bogomolny, A. (1996). *Modular arithmetic*. Retrieved March 1, 2000, from http://www.cut-the-knot.com/blue/Modulo.shtml

Brown, T. L., & Carr, T. H. (1989). Automaticity in skill acquisition: Mechanisms for reducing interference in concurrent performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 686–700.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Darke, S. (1988). Effects of anxiety on inferential reasoning task performance. *Journal of Personality and Social Psychology, 55,* 499–505.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved? *Memory & Cognition, 29,* 267–273.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725–747.

Eysenck, M. W. (1979). Anxiety learning and memory: A reconceptualization. *Journal of Research in Personality, 13,* 363–385.

Eysenck, M. W. (1992). *Anxiety: The cognitive perspective*. Hillsdale, NJ: Erlbaum.

Fitts, P. M., & Posner, M. I. (1967). *Human performance.* Belmont, CA: Brooks/Cole.

Fürst, A. J., & Hitch, G. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition, 28,* 774–782.

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied, 10,* 42–54.

Hardy, L., Mullen, R., & Jones, G. (1996). Knowledge and conscious control of motor actions under stress. *British Journal of Psychology, 87,* 621–636.

Keele, S. W. (1986). Motor control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 1–60). New York: Wiley.

Keele, S. W., & Summers, J. J. (1976). The structure of motor programs. In G. E. Stelmach (Ed.), *Motor control: Issues and trends* (pp. 109–142). New York: Academic Press.

Kimble, G. A., & Perlmuter, L. C. (1970). The problem of volition. *Psychological Review, 77,* 361–384.

Klapp, S. T., Boches, C. A., Trabert, M. L., & Logan, G. D. (1991). Automatizing alphabet arithmetic: II. Are there practice effects after automaticity is achieved? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 196–209.

Langer, E., & Imber, G. (1979). When practice makes imperfect: Debilitating effects of overlearning. *Journal of Personality and Social Psychology, 37,* 2014–2024.

Lewis, B., & Linder, D. (1997). Thinking about choking? Attentional processes and paradoxical performance. *Personality and Social Psychology Bulletin, 23,* 937–944.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Marchant, D. B., & Wang, J. (2001, May). *Choking: Current issues in theory and practice*. Paper presented at the 10th World Congress of Sport Psychology, Skiathos Island, Greece.

Masters, R. S. W. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology, 83,* 343–358.

Masters, R. S. W., Polman, R. C. J., & Hammond, N. V. (1993). "Reinvestment": A dimension of personality implicated in skill breakdown under pressure. *Personality and Individual Differences, 14,* 655–666.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance.* Thousand Oaks, CA: Sage.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126,* 288–311.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology, 52,* 453–470.

Sorg, B. A., & Whitney, P. (1992). The effect of trait anxiety and situational stress on working memory capacity. *Journal of Research in Personality, 26,* 235–241.

Spielberger, C. C., Gorsuch, R. L., & Lushene, R. (1970). *State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychology Press.

Squire, L. R., & Knowlton, B. J. (1994). The organization of memory. In H. Morowitz & J. L. Singer (Eds.), *The mind, the brain, and complex adaptive systems* (Vol. 22, pp. 63–97). Reading, MA: Addison-Wesley.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics.* New York: HarperCollins College.

Tohill, J. M., & Holyoak, K. (2000). The impact of anxiety on analogical reasoning. *Thinking and Reasoning, 6,* 27–40.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2001). Cognitive skill acquisition and transfer in younger and older adults. *Psychology and Aging, 16,* 555–563.

Trbovich, P. L., & LeFevre, J. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31,* 738–745.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76,* 92–104.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case of study arithmetic. *Journal of Experimental Psychology: General, 115,* 118–130.

Received January 30, 2003
Revision received June 21, 2004
Accepted July 6, 2004 ■

New Editor Appointed for *History of Psychology*

The American Psychological Association announces the appointment of James H. Capshew, PhD, as editor of *History of Psychology* for a 4-year term (2006–2009).

As of January 1, 2005, manuscripts should be submitted electronically via the journal's Manuscript Submission Portal (www.apa.org/journals/hop.html). Authors who are unable to do so should correspond with the editor's office about alternatives:

James H. Capshew, PhD
Associate Professor and Director of Graduate Studies
Department of History and Philosophy of Science
Goodbody Hall 130
Indiana University, Bloomington, IN 47405

Manuscript submission patterns make the precise date of completion of the 2005 volume uncertain. The current editor, Michael M. Sokal, PhD, will receive and consider manuscripts through December 31, 2004. Should the 2005 volume be completed before that date, manuscripts will be redirected to the new editor for consideration in the 2006 volume.