Choking and excelling under pressure in experienced classifiers

DARRELL A. WORTHY, ARTHUR B. MARKMAN, AND W. TODD MADDOX
University of Texas, Austin, Texas

We extend previous work examining the effects of pressure on category learning to the effects of pressure on categorization performance in highly trained individuals. After extensive training on either a rule-based or an information-integration classification task, half of the participants performed the same task on a fifth day while under pressure to earn a monetary bonus ($50) for themselves and a partner. Performance of this group was compared with that of a low-pressure control group who performed without the pressure manipulation. Pressure caused performance decrements both for experienced classifiers performing rule-based tasks and for those performing information-integration tasks, as compared with control groups. These results contrast with those of previous research, where inexperienced classifiers choked on rule-based tasks but excelled on information-integration tasks. An additional "superpressure" block of trials was given at the end of the fifth session. Under this type of pressure, participants performing an information-integration task outperformed those performing rule-based tasks. Implications for theories of choking under pressure are discussed.

The anecdotal phenomenon of choking under pressure has been the subject of much laboratory research (e.g., Beilock & Carr, 2005; Beilock, Kulp, Holt, & Carr, 2004; Markman, Maddox, & Worthy, 2006; Masters, 1992). Choking under pressure occurs when someone underperforms on a task, relative to their normal performance, because of an acute stressor. Research has focused on two different explanations for choking. The *distraction theory*, which has some similarity to *processing efficiency theory* (e.g., Eysenck & Calvo, 1992), proposes that pressure leads to a decrease in available working memory (WM) resources, which in turn has a negative impact on the performance of cognitively demanding tasks (Wine, 1971). Alternatively, the *explicit monitoring theory*, which has some similarity to *reinvestment theory* (see, e.g., Masters & Maxwell, 2008), proposes that pressure causes increased attention to skill-focused processes, which disrupts the performance of proceduralized tasks (Baumeister, 1984).

We test these accounts using highly trained participants. To place this work in context, we first discuss previous research exploring the distraction and explicit monitoring theories. Then we present data from a study in which a total of 69 participants first received over 2,500 trials of classification learning across four separate training sessions, and were then put under pressure in a fifth experimental session.

Previous Research on Distraction and Monitoring

There is evidence supporting both the distraction and monitoring theories of choking under pressure. Studies supporting the distraction theory often come from WM-intensive tasks. For example, Markman et al. (2006) studied category-learning tasks. Participants performed either a *rule-based* task, which has been shown to recruit WM resources (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Maddox & Ashby, 2004; Maddox, Filoteo, Hejl, & Ing, 2004; Zeithamova & Maddox, 2006), or an *information-integration* task, which has been shown to recruit a procedural-based learning system that is not WM demanding (Ashby et al., 1998; DeCaro, Thomas, & Beilock, 2008; Maddox & Ashby, 2004; Maddox, Ashby, & Bohil, 2003). In accordance with the distraction theory, novice participants choked while performing the rule-based task under pressure, but they excelled while performing the information-integration task under pressure.

Further support for the distraction theory comes from Beilock and Carr (2005), who showed that performance of participants with high WM capacity declined more under pressure than did performance of those with low WM capacity. Additional studies using cognitively demanding tasks, such as math problems, also support the distraction theory (e.g., Ashcraft & Kirk, 2001; Beilock et al., 2004), and there is some evidence that skill-focused training can reduce choking under pressure in motor tasks, such as golf putting (Beilock & Carr, 2001, Experiment 3; Lewis & Linder, 1997).

Support for the monitoring theory has been demonstrated primarily in motor tasks, such as golf putting (Beilock & Carr, 2001, Experiment 4; Hardy, Mullen, & Jones, 1996; Masters, 1992), simulated baseball batting (Gray, 2004), and free-throw shooting (Liao & Masters, 2002). A common finding in many of these studies is

D. A. Worthy, worthyda@mail.utexas.edu

that attention to skill-focused processes during training disrupts participants' performance while under pressure, whereas training without their being instructed to focus on the processes at hand (Liao & Masters, 2002) or training while they perform a WM-demanding task (Hardy et al., 1996; Masters, 1992) leads to improved performance under pressure. The monitoring view assumes that skill-focused training gives participants explicit knowledge of the motor processes used to complete the task. Under pressure, participants then monitor their performance and call to mind this explicit information, which has the paradoxical effect of hurting performance.

Of greater relevance to the present study are experiments conducted using expert participants. Gray (2004) had Division 1-A collegiate baseball players with 14–18 years of experience perform a simulated batting task while under pressure. He found direct evidence that performance pressure caused increased attention to skill-focused processes, as confirmed by improved performance on a skill-focused task, and that this increased attention caused decrements in performance for expert participants. These results showed that, for expert baseball players, pressure did not cause a decrease in WM capacity, as is proposed by the distraction theory (Wine, 1971), but rather caused an inward shift in attentional focus, as is proposed by the monitoring theory (Baumeister, 1984).

Overview of the Present Study

In the present study, we examined the effects of pressure on the performance of participants given extensive practice with rule-based or information-integration classification tasks. Sixty-nine participants learned the categories in four 1-h sessions given on consecutive days, performing over 2,500 learning trials each. Each participant then completed a fifth experimental session.

A critical difference between rule-based and information-integration category structures is that the optimal strategy for rule-based tasks is explicit (i.e., verbalizable), whereas the optimal strategy for information-integration tasks is implicit (i.e., not verbalizable). For example, for the category structure shown in Figure 1A, the optimal rule is to classify stimuli with relatively low spatial frequency into one category and stimuli with relatively high spatial frequency into the other category. In contrast, there is no verbalizable rule that can be employed to optimally distinguish the categories in the information-integration structure shown in Figure 1B. Here the optimal rule may be verbalized as follows: "If the spatial orientation is greater than the spatial frequency, classify the stimuli into one category, but if the spatial orientation is less than the spatial frequency, classify the stimuli into the other category." However, this type of rule is difficult or impossible to implement to solve the task, because the relative magnitude of each spatial dimension cannot be easily compared.

Dual-systems theory of classification learning. There is abundant evidence that separate neural systems are recruited when one learns to classify stimuli from rule-based and information-integration category structures (Ashby et al., 1998; Maddox & Ashby, 2004; No-
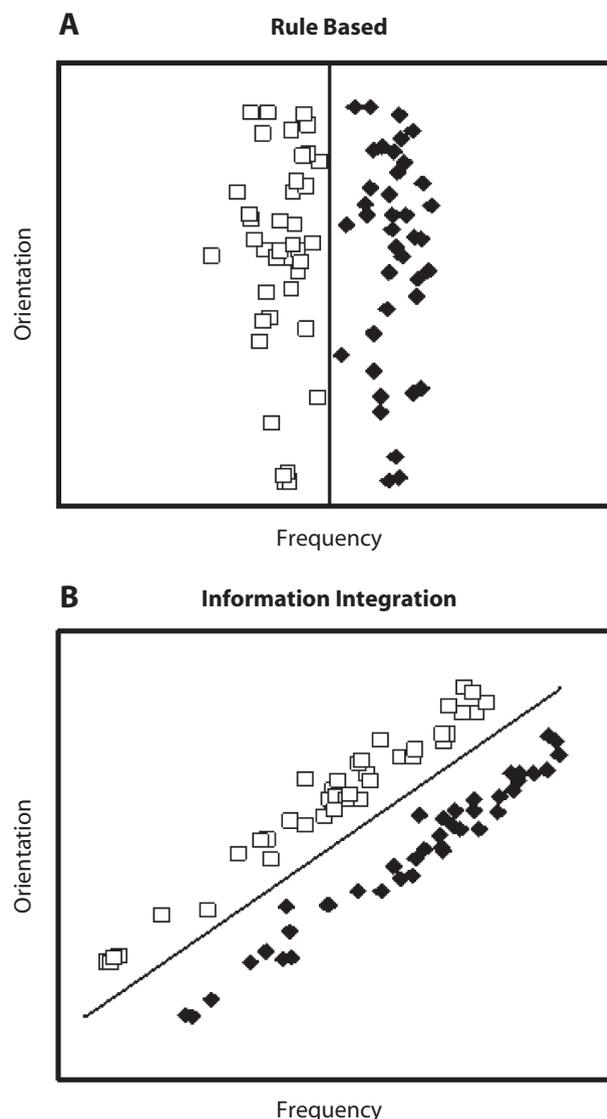

Figure 1. (A) Plot of stimuli from the rule-based category structure. (B) Plot of stimuli from the information-integration category structure.

mura et al., 2007). The competition between verbal and implicit systems (COVIS) model details these separate neural systems (Ashby et al., 1998). COVIS assumes that a frontally mediated hypothesis-testing system that depends heavily on WM and executive attention dominates rule-based learning, whereas a basal-ganglia-mediated procedural learning system that is sensitive to the timing of the feedback (but is not demanding of WM) dominates information-integration learning (Ashby et al., 1998; for a review, see Maddox & Ashby, 2004).

An important component of this paradigm is that rule-based category learning is more analogous to performing WM-demanding tasks, such as solving math problems, that have previously been used to study choking under pressure (e.g., Beilock & Carr, 2005). Information-integration category learning is more analogous to learning motor tasks, such as golf putting or baseball batting, that have previously been

used to study choking under pressure (e.g., Beilock & Carr, 2001; Gray, 2004). Information-integration category learning has been closely related to procedural learning (Ashby et al., 1998). For example, Ashby and colleagues found that switching the response keys between training and transfer phases in rule-based and information-integration classification tasks impaired information-integration transfer performance, but it did not impair rule-based transfer performance (Ashby, Ell, & Waldron, 2003). This suggests that the explicit rule-based system learns abstract category labels, whereas the procedural-based information-integration system learns response positions. Whereas links between information-integration classification learning and motor tasks should be made with caution, there is some evidence that information-integration learning requires a more procedural system.

Distraction and monitoring theories make contrasting predictions regarding learning under pressure for rule-based and information-integration tasks. The top rows of Table 1 summarize the learning predictions that are based on a dual-systems theory of classification. Distraction theory states that pressure will reduce the amount of available WM resources. Because rule-based category learning requires WM, it will suffer under pressure. However, information-integration category learning does not require WM; in fact, low WM has been shown to improve information-integration learning (DeCaro et al., 2008; however, see Tharp & Pickering, in press). Thus, the distraction theory predicts excelling under pressure in information-integration learning. The explicit monitoring theory makes the opposite predictions with respect to learning. The explicit monitoring theory states that pressure causes one to explicitly focus on the processes needed to carry out the task. Because rule-based learning relies on explicit hypothesis testing, the explicit monitoring theory predicts that pressure should improve rule-based learning. In contrast, monitoring of explicit processes is counterproductive to information-integration. Markman et al. (2006) examined the effects of pressure on rule-based and information-integration learning and found choking under pressure for rule-based learning, but they also found excelling under pressure for information-integration learning, in support of the distraction theory.

Table 1
Predicted Performance for Participants
Performing Under Pressure

	Rule Based	Information Integration
Learning (Dual-System)		
Distraction	Choking	Excelling
Monitoring	Excelling	Choking
Experienced (Cortical)		
Distraction	Excelling	Excelling
Monitoring	Choking*	Choking*
Summary of Results		
Pressure	Choking	Choking
Superpressure	Choking	Excelling

*As was explained in the introduction, monitoring theory may also predict a greater degree of choking for rule-based participants than for information-integration participants, due to a greater amount of explicit knowledge about the category structure available to rule-based participants.

An alternative theory of classification for experienced classifiers. Markman et al. (2006) examined classification *learning*, whereas the present study examines classification *performance* after extensive training. The neural systems used for learning the two category structures may be different from the systems used for classifying stimuli from each structure once the categories are well learned; thus, the effects of pressure may be different. Ashby and colleagues (1998; see also Ashby, Ennis, & Spiering, 2007) proposed a "subcortical pathways enable expertise development" (SPEED) theory of expert classification that assumes that the basal-ganglia-mediated procedural learning system used for learning information-integration category structures is gradually replaced by a much faster and more direct cortico-cortical system that involves projections from extrastriate visual cortices directly to motor and premotor areas. It is also reasonable to propose that a similar cortically mediated response system gradually replaces the frontal-based explicit system responsible for learning rule-based categories, as expertise develops and knowledge of the categories becomes automated. This view suggests that performance of experienced classifiers will be cortically mediated, regardless of which structure they are learning. The predictions for this view are summarized in the middle rows of Table 1.

If we assume that experienced classifiers use the same systems that have been shown to mediate responding during learning, then distraction theory would predict choking under pressure for participants performing rule-based tasks and excelling under pressure for participants performing information-integration tasks; explicit monitoring theory, then, would predict excelling for participants performing rule-based tasks and choking for participants performing information-integration tasks. In contrast, if we assume that the responses of highly experienced classifiers are mediated by a direct automated system, then distraction theory would predict excelling under pressure for participants performing both rule-based and information-integration tasks, whereas monitoring theory would predict choking under pressure, due to an increase in skill-focused attention that disrupts the automated system. Previous research suggests that distraction theory is more appropriate for performance under pressure during learning (e.g., Markman et al., 2006), whereas monitoring theory is more appropriate for performance under pressure for highly experienced participants (e.g., Gray, 2004; Masters, 1992).

Even though Gray (2004) and Masters (1992) assumed that experienced classifiers ultimately use the same system to make their classification responses regardless of the category structure, this view predicts that pressure will harm rule-based classification performance more than it will harm information-integration performance. This difference arises because of differences in the amount of explicit knowledge about the categories that is available to experts classifying stimuli from rule-based or information-integration category structures. Because rule-based tasks are learned via an explicit hypothesis-testing process, the explicit knowledge about the category structure available to participants performing rule-based tasks is much

greater than that for participants performing information-integration tasks. If it is assumed that pressure causes an increase in skill-focused attention that can disrupt the automated system (i.e., monitoring theory), the difference in explicit knowledge about the category structures may influence the extent of the observed performance decrements for highly trained participants performing each task while under pressure. Whereas monitoring theory predicts performance decrements for participants performing both tasks while under pressure (due to disruption of the automated system caused by an increase in skill-focused attention), the decrements may be greater for participants performing rule-based tasks because they have explicit knowledge of the category structure, which they can use to monitor their performance. In contrast, participants performing information-integration tasks may experience less severe decrements in performance because they have less explicit knowledge of the category structure to use to monitor their performance. This would corroborate the findings of several studies mentioned above, for which more explicit knowledge about the process of performing the task during training led to worse performance while the participants were under pressure (e.g., Gray, 2004; Liao & Masters, 2002; Masters, 1992).

To test these competing theories, we conducted an experiment in which participants were first given extensive training on either rule-based or information-integration classification tasks. After four 1-h sessions of training on the task, participants completed a fifth session. During the fifth session, half of the participants were given a pressure manipulation that has been used in many previous pressure experiments, and the other half were simply asked to do their best. We present this as Experiment 1.

After the fifth session, which consisted of eight blocks of 80 trials, we gave all participants a final 80-trial block and asked them to try to perform *perfectly* on this block. Participants who had performed Experiment 1 under both high and low pressure were told that if they did not make a single incorrect classification on the final block of trials, they would receive $100 cash. There were two motivations for this novel manipulation: We aimed to examine the effects of pressure across a shorter time span (one block vs. eight blocks), and we also wanted to explore the effects of increasing the incentives and bonus criterion on performance.

EXPERIMENT 1

In this experiment, we examined the effects of pressure on rule-based and information-integration classification performance in highly trained participants. Each participant received extensive initial training with either a rule-based or an information-integration task. By the end of the fourth training session, all participants were classifying with at least 85% accuracy.

At the beginning of the fifth session, half of the rule-based and half of the information-integration participants were put under high pressure by giving each the opportunity to earn a monetary bonus ($50) for excep-

tional performance (95% accuracy). Each participant in the high-pressure condition was informed that earning the bonus was a team effort. Specifically, they were told that they were being paired with a partner who had already reached the bonus criterion; but in order for either of them to receive the bonus money, both of them had to reach the bonus criterion on a randomly determined block during the final session. The other half of the rule-based and information-integration participants were put in the low-pressure condition. These participants were told just to continue classifying as they had done in the previous four sessions. This pressure manipulation was identical to that used in Markman et al.'s (2006) study of the effects of pressure on classification learning. Thus, we had good reason to believe that this manipulation would successfully induce pressure in our participants.

Method

Participants

Participants were 69 members of the University of Texas at Austin community who were paid for their participation. The experiment consisted of five sessions on five consecutive weekdays, with the constraint that no participant could start on a Tuesday, so that the pressure session would never occur immediately following a weekend. Participants who completed all five sessions received $30. Participants were randomly assigned to one of four between-participants conditions: rule-based low pressure ($n = 17$), rule-based high pressure ($n = 18$), information-integration low pressure ($n = 17$), or information-integration high pressure ($n = 17$). To ensure extensive initial learning, participants who did not reach a minimum of 85% accuracy across all blocks of Session 4 were excluded. After completing the fourth session of the experiment, participants who did not reach 85% accuracy level were paid and told that they did not need to complete the fifth session because they did not meet the minimum accuracy criterion. Nine participants failed to meet this performance criterion. Data from a total of 15 participants in each condition were used in the final analysis.

Materials and Stimuli

Participants performed the experiment on a personal computer using MATLAB software. The rule-based and information-integration structures are shown in Figures 1A and 1B. Stimuli were Gabor patches varying in the frequency (spacing of bars) and orientation of the grating relative to the x-axis of the computer screen. The experiment used the randomization technique introduced by Ashby and Gott (1988). Forty Category A and 40 Category B stimuli from the rule-based categories were generated by randomly sampling from two bivariate normal distributions. Each random sample (x_1, x_2) was converted to a stimulus by deriving the frequency [$f = .25 + (x_1/50)$] and the orientation [$o = x_2(\pi/500)$]. The stimuli for the information-integration categories were generated by rotating the 80 rule-based stimuli clockwise by 45º around the center of the spatial frequency/spatial orientation space and then shifting the spatial frequency and spatial orientation by an amount that resulted in the appropriate d'. The d' was 4.5 for rule-based categories and 8.0 for information-integration categories. The d' was increased for the information-integration category structure to equate performance accuracy between the two category structures. The category distribution parameters for both category structures are listed in Table 2. Each stimulus was created using the MATLAB Psychophysics Toolbox. The size of each stimulus was 200 × 200 pixels. Each stimulus was centered in a 650 × 650 pixel stimulus presentation box in the upper left-hand corner of the computer screen. On the right-hand side of the screen, a point meter kept track of the number of points earned during the current block.

Table 2
Category Distribution Parameters for the Rule-Based and Information-Integration Category Structures Used in Experiments 1 and 2

Category Structure	μ_x	μ_y	σ_x^2	σ_y^2	cov_{xy}
Rule Based					
Category A	280	125	75	9,000	0
Category B	320	125	75	9,000	0
Information Integration					
Category A	275	150	4,538	4,538	4,351
Category B	324	101	4,538	4,538	4,351

Procedure

Initial training. Every participant completed four sessions of initial learning. Each session consisted of eight 80-trial blocks. These sessions typically lasted just under 1 h. On each trial, a stimulus was presented; the participant generated a categorization response, and corrective feedback was provided. There was no response deadline. Participants were free to take as long as they wished when categorizing each stimulus. Feedback was given at the bottom left-hand corner of the screen, centered below the stimulus presentation box. If the participant's response was correct, the word "correct" was shown; if the participant's response was incorrect, the phrase "No, that was an (A or B)" was shown. Two points were added to a point meter on the screen when participants' responses were correct, and 0 points were added when they were incorrect. The counter was reset to 0 at the start of each block.

Pressure manipulation. At the beginning of the fifth session, half of the participants were introduced to the high-pressure manipulation, and half were told just to continue categorizing (low-pressure condition). Participants in the high-pressure condition were told that they would receive a monetary bonus of $50 if both they and a fictional partner exceeded a performance criterion (95% accuracy over a randomly determined block in the final session) and that neither would receive a bonus if one failed to reach the criterion. Participants were then informed that their partners had already exceeded the criterion, so the bonus for the participant and the partner rested on the participant's performance. Throughout the course of the fifth session, a line on the bonus meter indicated how many points were required to reach the bonus, and a box in the lower left-hand corner of the screen reminded the participants in the high-pressure conditions that his or her partner had reached the bonus criterion. Participants in the high-pressure conditions also heard a "ching" cash register sound when they made a correct classification; they heard a "buzz" sound when they made an incorrect classification. This type of manipulation has been used in numerous studies and has been shown to induce pressure (e.g., Beilock &

Carr, 2001, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004; Gray, 2004).

Results

We analyzed the data in two ways. First, we examined performance accuracy for the final training and pressure sessions. Next, we applied decision-bound models (Ashby & Maddox, 1993; Maddox & Ashby, 1993) to determine the specific strategies used for classifying the stimuli.

Training Performance

It was important to us to ensure that performance was equal across all conditions at the end of training. We conducted a 2 (category structure) × 2 (pressure level) ANOVA on the final session's accuracy. Neither the interaction nor the main effects of pressure level or category structure were significant on the final training session's accuracy. There were no significant differences between the low-pressure and high-pressure conditions on final training session accuracy for participants performing a rule-based task ($F < 1$) ($M = .890$ for the low-pressure condition, $M = .903$ for the high-pressure condition) or for participants performing an information-integration task ($F < 1$) ($M = .906$ for the low-pressure condition, $M = .898$ for the high-pressure condition). Thus, any pressure-driven performance differences cannot be attributed to differences in initial-task difficulty.

Performance Under Pressure

Accuracy-based analyses. To determine the effect of pressure on performance, we examined the change in performance from the final training session to the pressure session, as a function of category structure (rule based vs. information integration) and pressure level (high pressure vs. low pressure). Figure 2 displays these data. Positive values denote improved performance in the pressure session, whereas negative values denote a performance decrement. A 2 (pressure level) × 2 (category structure) ANOVA revealed a main effect of pressure [$F(1,58) = 9.81$, $p < .01$, $\eta^2 = .15$], with low-pressure participants ($M = .017$) showing significantly greater improvement than did high-pressure participants ($M = -.005$).
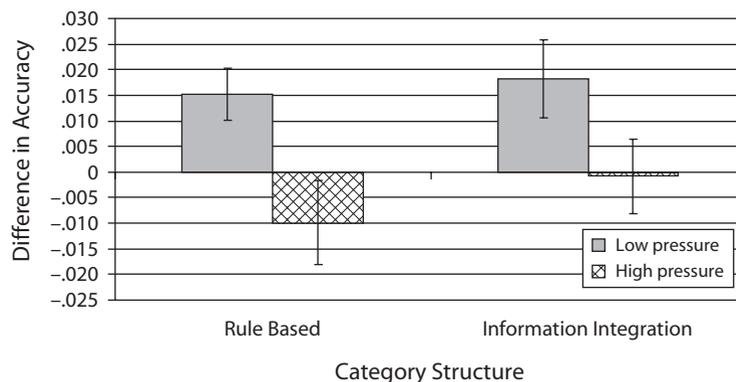

Figure 2. Difference in accuracy between the pressure and final prepressure sessions, collapsed across all blocks.

Thus, unlike in our previous category-learning results (Markman et al., 2006), experienced participants exhibited similar performance decrements under pressure for both rule-based and information-integration category structures; these decrements were more pronounced for participants given the rule-based task. This finding suggests that classification for experienced classifiers might be mediated by a system that is different from the dual-learning systems that mediate rule-based and information-integration category learning. Perhaps an automated cortical system, like that proposed in SPEED, mediates responding for highly experienced participants performing both rule-based and information-integration tasks and is disrupted when the participants are under pressure, due to an increase in skill-focused attention to the task. Therefore, the uniform decrement in performance across category structures appears to support the explicit monitoring theory. This finding is consistent with Gray's (2004) work examining the performance of expert participants under pressure, where the performance of highly experienced collegiate-level baseball players declined under pressure due to an increase in skill-focused attention on the task.

Model-based analyses. The accuracy-based analyses supported the proposal that pressure causes performance decrements for participants performing rule-based and information-integration classification tasks. Accuracy-based analyses are an informative measure of performance, but they provide no information regarding the specific strategies that participants use to classify stimuli. Knowledge of the effects of pressure on strategy use is of fundamental importance. To address this issue, we applied decision-bound models (Maddox, 1999; Maddox & Ashby, 1993) to the data from each individual participant on a block-by-block basis.

Decision-bound models assume that participants use a decision bound to separate stimuli into categories with stimuli on one side of the bound being classified into one category, and stimuli on the other side of the bound being classified into the other category. The optimal decision bound in the rule-based condition is depicted by the vertical line in Figure 1A. Examples of suboptimal strategies would be depicted by shifts of the decision bound along the frequency dimension, or by placement of a decision bound along the orientation dimension. The optimal decision bound in the information-integration condition is depicted by the diagonal line in Figure 1B. Depictions of suboptimal strategies might include altering the slope or *y*-intercept of the optimal decision bound or setting a decision bound on either the frequency or the orientation dimension.

With highly experienced participants, we expect a high proportion of optimal decision-bound use. Fits of the models will be useful in determining whether pressure causes a shift in classification strategies to a suboptimal—or even to a qualitatively different—strategy (e.g., from an optimal frequency-dimension decision bound to an orientation-dimension decision bound in the rule-based condition).

We fit seven decision-bound models separately to the data from each participant on a block-by-block basis. All of the analyses were performed at the individual-participant

level because of concerns about modeling aggregate data (e.g., Ashby, Maddox, & Lee, 1994; Estes, 1956; Maddox, 1999; Maddox & Estes, 2004; Smith & Minda, 1998).

Rule-based models. The *optimal unidimensional-frequency model* assumes that the participant uses the optimal criterion along the spatial-frequency dimension and applies the following rule: "Respond 'A' if the spatial frequency is low and 'B' if it is high." This model has one free parameter that represents the variance of internal (perceptual and criterial) noise; this model was fit only to data from participants performing rule-based tasks. The *generalized unidimensional-frequency rule model* assumes that the participant uses a criterion along the spatial-frequency dimension, but it allows the criterion value to be estimated from the data (two free parameters total). The *generalized spatial-orientation model* assumes that the participant uses a criterion along the spatial-orientation dimension, and allows the criterion value to be estimated from the data (two free parameters total). These two models were fit to data from all participants.

The *conjunction models* assume that the participant uses a conjunctive rule, in which he or she makes separate decisions about the levels of the two dimensions and then selects a response that is based on the outcome of these two decisions. Two conjunctive rules were examined:

1. Respond "A" if spatial frequency is low and orientation is large; otherwise respond "B."
2. Respond "B" if spatial frequency is high and orientation is small; otherwise respond "A."

Both rules partition the perceptual space into four regions. The first assigns one region to Category A and three regions to Category B, and the second assigns three regions to Category A and one region to Category B. The conjunction models have three parameters (an internal noise parameter and a criterion on each dimension). The conjunction models were fit to data from all of the participants.

Information-integration models. The *general linear-classifier model* (GLC) assumes that the decision bound between each pair of categories is linear. This produces an information-integration decision strategy because it requires linear integration of perceived frequency and orientation. The GLC has three parameters: an internal noise parameter and the slope and intercept of the linear bound. The *optimal GLC model* assumes that the participant will use the linear bound that maximizes accuracy. This model has only one free parameter, representing the internal noise. The GLC was fit to all data, and the optimal GLC was fit only to data from information-integration participants.

Random-responder model. Each participant's data were also fit by a one-parameter random-responder model that assumed a fixed probability (estimated by the model) of responding "A" for all the stimuli. This model never provided the best account of any participant's data and is not discussed further.

Model Fitting

Each model was fit separately to the data from each of the eight blocks of trials in Sessions 4 and 5, for each

participant. The model parameters were estimated using maximum log-likelihood (Wickens, 1982), and the goodness-of-fit statistic used was AIC $= -2\ln L + 2k$, where k is the number of free parameters (Akaike, 1974). The AIC statistic penalizes models with extra free parameters. The best-fitting model is the model with the smallest AIC value. For each block, for each participant in each session, we determined which model provided the best fit to the data.

Final training session analysis. We found that the optimal or generalized unidimensional-frequency model fit best in every block of trials for rule-based low-pressure- and high-pressure-condition participants. In addition, we found that the optimal or GLC fit best in every block of trials for information-integration low-pressure- and high-pressure-condition participants. Thus, participants used a strategy of the same form as the optimal model, but in some cases the strategy was suboptimal. We compared the proportion of participants best fit by the optimal model in each condition. As expected, there were no significant differences between the low-pressure and high-pressure conditions for either the rule-based or the information-integration category structures ($F < 1$).

Pressure session analysis. As with the final learning session, we found that rule-based and information-integration participants in the low-pressure and high-pressure conditions used a strategy of the same form as the optimal model, but in some cases the strategy was suboptimal.

Figure 3A displays the proportion of information-integration participants in each condition whose data were best fit by the optimal model in each block. A larger proportion of information-integration low-pressure participants' data was best fit by the optimal model in seven of eight blocks ($p < .05$, by sign test). Binomial tests revealed that a significantly higher proportion of data sets from low-pressure participants was best fit by the optimal model in Block 8 ($p < .05$), and a marginally higher proportion was best fit by the optimal model in Block 2 ($p < .10$).

Figure 3B displays the proportion of rule-based participants in each condition whose data were best fit by the optimal model in each block. A larger proportion of rule-based low-pressure participants' data was best fit by the optimal model in three of eight blocks (n.s., based on a sign test). However, significantly more low-pressure-condition participants were best fit by the optimal model in Blocks 4 and 7 ($p < .05$, by binomial test).
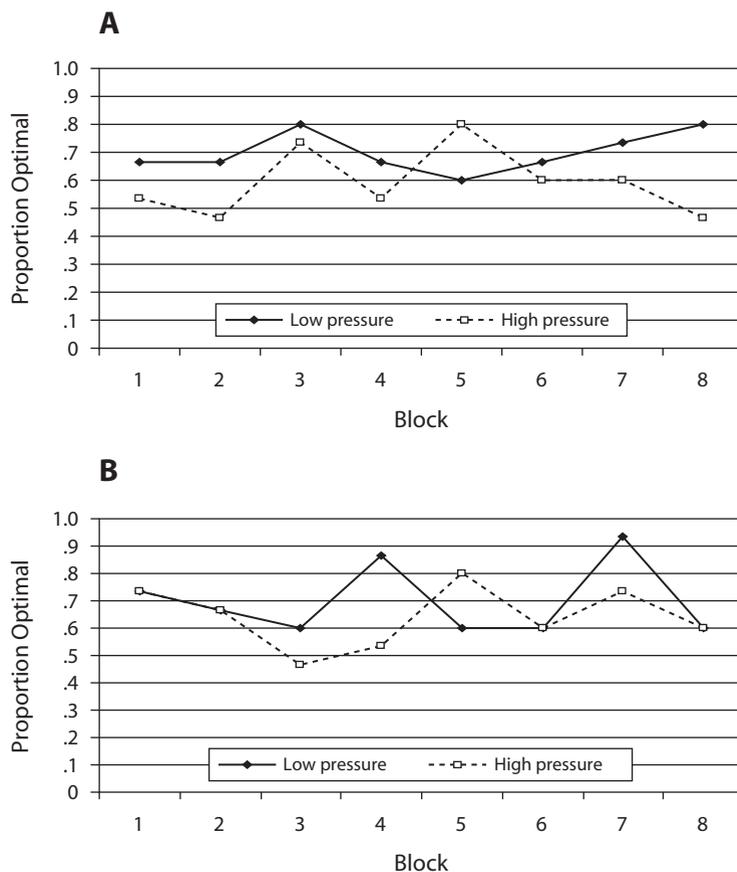

Figure 3. (A) Proportion of information-integration participants best fit by the optimal general linear classifier model for each block of the pressure session. (B) Proportion of rule-based participants best fit by the optimal unidimensional frequency model for each block of the pressure session.

Taken together, these findings suggest that pressure does not lead to a qualitative shift in strategy (e.g., from an information-integration strategy to a rule-based strategy), but can lead to an increase in the use of a suboptimal strategy (albeit a strategy of the same form as the optimal model).

Noise From Optimal Models

Although pressure did not lead to large qualitative shifts in strategy use, it is likely that pressure leads to more subtle changes. As a first pass, we examined the estimated noise parameter values from fits of the optimal model to each participant's data. Larger values of the noise parameter reflect greater noise in the perception of each stimulus and/or greater variance in the memory for, and application of, the decision bound.

We subtracted the optimal model noise estimate averaged across all blocks for the pressure session from that averaged across all blocks for the final training session. Positive values indicate less perceptual noise and/or less variance in the memory for, and application of, the decision bound in the pressure session, whereas negative values indicate more perceptual noise and/or greater variance in the memory for, and application of, the decision bound in the pressure session.

Figure 4 displays the noise difference measure for each of the four conditions. A 2 (pressure level) $\times$ 2 (category structure) ANOVA revealed a main effect of pressure on this measure, with low-pressure participants yielding a larger (and positive) noise difference (i.e., less noise in the pressure session, relative to the final training session) than did high-pressure participants, who yielded a negative noise difference [$F(1,58) = 4.21, p < .05, \eta^2 = .07$]. The noise differences were larger for participants performing information-integration tasks than for those performing rule-based tasks, but this was likely due to the greater distance from the decision bounds for the information-integration stimuli than for the rule-based stimuli.

The combined modeling results indicate that pressure caused a decrement in optimal strategy use for both rule-based and information-integration high-pressure participants. High-pressure participants were more likely to be fit by a suboptimal model that had the same form as the optimal model. High pressure did not cause a complete shift in classification strategy or a complete collapse in performance. This pattern suggests that the decrements in accuracy for participants under high pressure resulted from a disruption of the execution of the optimal classification strategy that was reflected by increased perceptual and criterial noise.

EXPERIMENT 2

In many domains of expert performance, there is little room for error. Performance by athletes, doctors, soldiers, pilots, students, and other high-stakes performers can be disrupted even by small decrements in accuracy or execution. Very often, choking involves only the smallest mistake or blunder. A game can be won or lost by a missed free throw, and admission into a prestigious academic institution can be gained or lost by errors on a few test problems. Our highly experienced classifiers did not show major qualitative drops in performance, but they did significantly underperform relative to equally experienced peers who were under low pressure. Thus, the amount of pressure placed on students, athletes, and others striving to perform optimally can influence how well they do.

We used a pressure manipulation similar to many manipulations that have been shown to induce pressure in previous laboratory studies (e.g., Beilock & Carr, 2001, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004; Gray, 2004). This type of pressure manipulation involves a combination of social and monetary incentives. Participants in our experiment were asked to achieve at least 95% accuracy on each block. This amounted to making 76 out of 80 correct classifications. We used this manipulation because it followed directly from previous work studying choking under pressure (e.g., Beilock & Carr, 2001, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004; Gray; 2004; Markman et al., 2006). However, we also wanted to examine performance in an extremely high-stakes situation. We added a final "superpressure" block of trials to the end of the final experimental session, in which all participants were given a high monetary incentive ($100 in cash) to perform perfectly. This manipulation was added to examine the effects of different levels of pressure on performance.

Method

Participants performed Experiment 2 as an addition to the fifth session of Experiment 1. After completing eight blocks of 80 trials during the fifth and final session, participants were asked to perform one final 80-trial block under "superpressure." Participants performed the same categorization task they had been performing throughout the entire 5-day experiment. We informed all participants, including those in the low-pressure condition, that they had the opportunity to earn $100 in cash if they did not make a single incorrect classification on the final block of trials. After completing the superpressure block, participants were debriefed regarding the first pressure manipulation and given any bonus money they had earned for their performance.
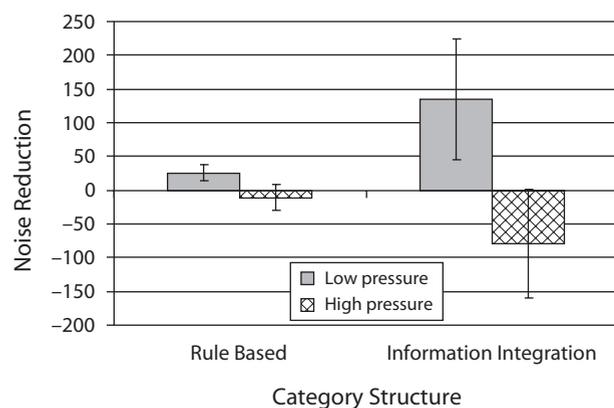

Figure 4. Difference in criterial noise from the optimal decision-bound model between the final training session and the pressure session.
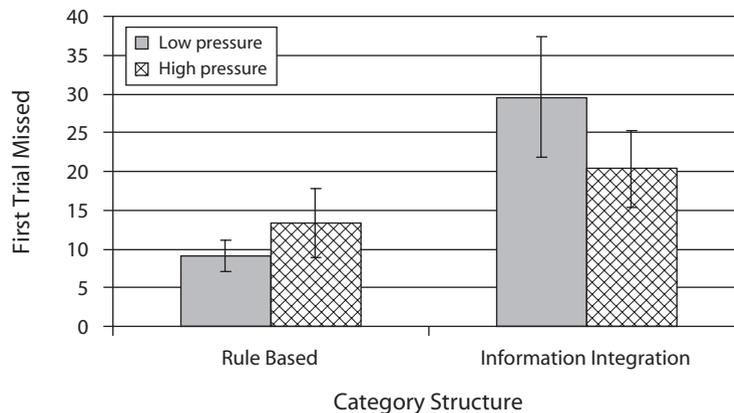
Figure 5. Average first trial missed on the superpressure block.

Results

Because performance for many participants quickly dropped off after they made their first incorrect classification during the superpressure block (i.e., they gave up), we compared this with the trial on which they made their first incorrect classification in the other condition. Figure 5 shows the average first trial missed for each condition. A 2 (category structure) × 2 (pressure level) ANOVA showed a main effect of category structure [$F(1,50) = 4.99$, $p < .05$, $\eta^2 = .091$].[1] Participants performing an information-integration task ($M = 21.30$) went significantly longer before making an incorrect response than did participants performing a rule-based task ($M = 11.44$). This finding is consistent with the idea that pressure causes an increase in skill-focused attention that disrupts people's ability to respond automatically. This increase in monitoring hurt rule-based participants more than it did information-integration participants because they had explicit knowledge of the category structure during training. The additional knowledge of the category structure eventually disrupted their performance under pressure (e.g., Gray, 2004; Liao & Masters, 2002; Masters, 1992).

DISCUSSION

This study examined the influence of pressure on experienced participants' rule-based and information-integration classification performance. We employed two types of pressure manipulations that we call "pressure" and "superpressure." For the pressure manipulation, we created a situation where participants in the high-pressure conditions were told that they must reach a very high accuracy criterion (95%) on a randomly chosen block to earn $50 for themselves and their partners, who had already done their part in earning the bonus. Thus, the participants were faced with the uncertainty of which block would be chosen to determine whether they would receive the bonus, and with the knowledge that failure on their own part to reach the bonus criterion would cause a worthy partner not to receive the bonus. A manipulation very similar to this has been used in numerous previous studies

and has been shown to induce pressure (e.g., Beilock & Carr, 2001, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004; Gray, 2004; Markman et al., 2006).

This type of pressure generally harmed experienced participants performing both a rule-based classification task and an information-integration task. These results are in direct contrast to those of studies of classification performance under pressure in novices (e.g., Markman et al., 2006). Previous research has demonstrated that learning the rule-based task is WM intensive, whereas learning the information-integration task actually benefits from low WM capacity (DeCaro et al., 2008; however, see Tharp & Pickering, in press). Thus, the observation that pressure hurts both tasks suggests that performance of experienced participants involves different processes than does performance of participants learning the task. This finding is consistent with the view that expertise involves training of cortico-cortical connections that can swiftly and efficiently handle task performance (see, e.g., Ashby et al., 2007).

Model-based analyses indicate that the pressure manipulation caused an increase in the use of suboptimal strategies, albeit strategies of the same form as the optimal strategy. Significantly more blocks of data from participants in the information-integration high-pressure condition were best fit by a suboptimal model, relative to model fits for data from the final training session. Noise parameter estimates from fits of the optimal models to the data indicate that pressure caused an increase in perceptual and criterial noise for participants performing both rule-based and information-integration tasks under pressure. These results suggest that the pressure manipulation disrupted the processing in the categorization system that was most optimal for participants performing both types of tasks under pressure.

After all participants had completed five sessions of 640 trials each, we employed a superpressure manipulation that was different in many ways from the pressure manipulation. All participants were given a final 80-trial block in which they were told that perfect performance would yield a bonus of $100 in cash. Unlike the pressure

manipulation, this manipulation had no social component. Participants were trying to earn the large monetary bonus only for themselves. The superpressure manipulation also differed from the pressure manipulation in that there was less uncertainty about which block would be used as the criterion for earning the bonus. For the pressure manipulation, participants were told that a randomly selected block would be used to determine whether they should receive the bonus. If they made over 95% correct classifications on this randomly selected block, they would receive $50 each for themselves and a partner. Thus, there was some uncertainty about which block would be used for the determination of whether participants would earn the bonus. There was no such uncertainty with the superpressure manipulation. Participants knew the specific block that would determine whether they met the bonus criterion. Finally, the two manipulations differed in that the superpressure manipulation offered absolutely no chance for failure. The 95% accuracy criterion given in the pressure manipulation offered some small room for mistakes, but any inaccuracy in the superpressure manipulation would lead to failure.

We found a different pattern of results between the two manipulations. For the pressure manipulation, participants choked under pressure, relative to participants given no pressure, regardless of whether they had learned the explicit rule-based task or the implicit information-integration task. In contrast, for the superpressure manipulation, participants performing the rule-based task performed much worse than participants performing the information-integration task.

The results of the superpressure manipulation are consistent with the possibility that explicit knowledge obtained during rule learning disrupted the procedural skill developed over the extensive training period. In this view, the superpressure manipulation caused participants to explicitly monitor their performance, but only rule-based participants had explicit knowledge of the category structure that could influence their performance. In this view, information-integration participants were more immune to the deleterious effects of monitoring because they did not have explicit knowledge about the category structure from their early learning. It should be acknowledged, however, that we can only assume that the superpressure manipulation further raised anxiety levels. Future efforts should be made to include manipulation checks (see, e.g., Houtman & Bakker, 1989).

The acute nature of the superpressure manipulation may have brought out more differences between participants performing rule-based and information-integration tasks than did the nature of the regular pressure manipulation. The uncertainty about which block's accuracy score would determine the bonus for the pressure manipulation, as well as the longer time span of the pressure manipulation, relative to the time span of the superpressure manipulation (i.e., eight blocks for the pressure manipulation, as compared with one block for the superpressure manipulation), may have reduced some advantages for participants performing information-integration tasks. The superpressure manipulation required only one flaw-

less string of trials to achieve the bonus, whereas the pressure manipulation required consistently high performance throughout eight 80-trial blocks. It may have been more difficult to maintain high performance over such a long time span.

Future work should investigate different types of pressure manipulations. Many real-world events contain pressure situations that are more similar to either our pressure or our superpressure manipulation. Exams and sporting events, for example, may have more in common with our pressure manipulation. These events occur over long periods of time. Participants do not necessarily have to perform perfectly, but they must maintain very high performance throughout. Team sporting events involve a social component similar to the one in our pressure manipulation. Teammates depend on each other to reach their common goals. Other real-world events, such as surgeries, military missions, and chess matches, more closely resemble our superpressure manipulation. These events offer essentially no room for failure. The differences that we found in performance between the pressure and superpressure manipulations we used may have implications for differences in the types of pressure people face in real-world situations.

Implications for Theories of Choking

These results suggest that the effects of pressure on experienced classification are more in line with the predictions from the monitoring theory than with those from the distraction theory. The results are also consistent with a recently proposed theory of experienced information-integration classification (i.e., Ashby et al., 2007) and may extend the theory to account for experienced rule-based classification. The middle rows of Table 1 predict a uniform decrement in performance while participants are under pressure for both rule-based and information-integration category structures. This pattern was obtained for the pressure manipulation. During the pressure session, participants under pressure were less accurate, relative to the final prepressure session, than were participants under no pressure. This finding supports the prediction of monitoring theory, that pressure causes an increase in skill-focused attention leading to lower accuracy.

One puzzling aspect of these data, however, is the performance in the superpressure block. Participants given the rule-based task choked under superpressure, whereas participants performing the information-integration task showed much smaller performance decrements under superpressure. This pattern was observed during rule-based and information-integration classification learning (Markman et al., 2006) and was interpreted as being support for the distraction theory. It is possible to view these results in a way that is consistent with the monitoring theory. In this interpretation, pressure decreases performance for rule-based participants because they possess explicit task-related knowledge acquired during learning. This explicit knowledge disrupts the automated system when the participant performs under pressure. In further support of this view, the first pressure manipulation ap-

pears to have caused greater decrements in performance for participants performing rule-based tasks than for those performing information-integration tasks.

Masters and colleagues' reinvestment theory, a refinement of explicit monitoring theory (see, e.g., Masters & Maxwell, 2008), states that individuals engage in conscious self-regulation of their performance when there is a discrepancy between their current performance level and their desired performance level. This theory could explain why participants performing information-integration classification tasks did not show as much choking on the superpressure block as they had during the first pressure manipulation. The shorter time span to perform under pressure may have given participants less time to ruminate about the discrepancy between their current level of performance and their desired level of performance. Thus, over longer periods of time, participants may "break" under pressure and engage in explicit monitoring, which leads to choking under pressure.

The distraction theory is consistent with the pattern of data obtained in the superpressure block, assuming that the neural systems that mediate classification during learning are the same as those that mediate classification after the categories have been well learned. However, this view cannot account for the uniform decrements in performance found across both category structures after the first pressure manipulation. The combined pattern of data offers the most support for the monitoring theory of choking under pressure in highly experienced participants.

One view that is very similar to the distraction theory is processing efficiency theory (Eysenck & Calvo, 1992; Wilson, Smith, Chattington, Ford, & Marple-Horvat, 2006). Like distraction theory, processing efficiency theory states that pressure causes worry that reduces available WM resources. However, processing efficiency theory further predicts that pressure may also stimulate increases in on-task effort, which may partially or totally compensate for reduced performance effectiveness. This increase in on-task effort could be viewed as being similar to the proposed mechanism of choking in explicit monitoring theory. Thus, processing efficiency theory could combine both hypothesized mechanisms of choking into one theory. Such a "double whammy" effect of pressure has been proposed elsewhere (e.g., Beilock & Gray, 2007).

Finally, individual differences in performing under pressure should be addressed in future work. Several studies have already investigated dispositional traits, such as reinvestment (e.g., Masters, Polman, & Hammond, 1993; Shea, Wulf, Whitacre, & Park, 2001). Participants who score high on the reinvestment scale (Masters et al., 1993) tend to choke more when performing motor tasks under pressure. Other individual-differences variables, such as action orientation (Raab & Johnson, 2004), skill-focused attention (Baumeister, 1984), and internal and external attentional foci (Shea et al., 2001), may also affect performance under pressure.

Much work remains to be done to examine the effects of pressure on performance—particularly for the performance of experts. Many of the anecdotal phenomena of choking under pressure involve highly experienced individuals placed in high-stakes situations. In the present study, we demonstrated that pressure affects novice and expert participants differently. We also illuminated the need for further development of models of expertise. The conclusions to be drawn regarding the effects of pressure on expert participants should inform models of expertise in category learning. Neuroimaging studies are another avenue for further exploration to help determine the processes compromised by performance pressure.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

ASHBY, F. G., ALFONSO-REESE, L. A., TURKEN, A. U., & WALDRON, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.

ASHBY, F. G., ELL, S. W., & WALDRON, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, **31**, 1114-1125.

ASHBY, F. G., ENNIS, J. M., & SPIERING, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, **114**, 632-656.

ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.

ASHBY, F. G., & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.

ASHBY, F. G., MADDOX, W. T., & LEE, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity–choice model. *Psychological Science*, **5**, 144-151.

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

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

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. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, **16**, 101-105.

BEILOCK, S. L., & DECARO, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 983-998.

BEILOCK, S. L., & GRAY, R. (2007). Why do athletes choke under pressure? In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 425-444). Hoboken, NJ: Wiley.

BEILOCK, S. L., KULP, C. A., HOLT, L. E., & CARR, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, **133**, 584-600.

DECARO, M. S., THOMAS, R. D., & BEILOCK, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, **107**, 284-294.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, **53**, 134-140.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, **6**, 409-434.

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.

Houtman, I. L. D., & Bakker, F. C. (1989). The anxiety thermometer: A validation study. *Journal of Personality Assessment*, **53**, 575-582.

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

Liao, C.-M., & Masters, R. S. W. (2002). Self-focused attention and performance failure under psychological stress. *Journal of Sport & Exercise Psychology*, **24**, 289-305.

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics*, **61**, 354-374.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, **66**, 309-332.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 650-662.

Maddox, W. T., & Estes, W. K. (2004). Predicting true patterns of cognitive performance from noisy data. *Psychonomic Bulletin & Review*, **11**, 1129-1135.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 227-245.

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, **17**, 944-948.

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., & Maxwell, J. (2008). The theory of reinvestment. *International Review of Sport & Exercise Psychology*, **1**, 160-183.

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

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., et al. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, **17**, 37-43.

Raab, M., & Johnson, J. G. (2004). Individual differences of action orientation for risk-taking in sports. *Research Quarterly for Exercise & Sports*, **75**, 326-336.

Shea, C. H., Wulf, G., Whitacre, C. A., & Park, J.-H. (2001). Surfing the implicit wave. *Quarterly Journal of Experimental Psychology*, **54A**, 841-862.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.

Tharp, I. J., & Pickering, A. D. (in press). A note on DeCaro, Thomas, and Beilock (2008): Further data demonstrate complexities in the assessment of information-integration category learning. *Cognition*.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.

Wilson, M., Smith, N. C., Chattington, M., Ford, M., & Marple-Horvat, D. E. (2006). The role of effort in moderating the anxiety–performance relationship: Testing the prediction of processing efficiency theory in simulated rally driving. *Journal of Sports Sciences*, **24**, 1223-1233.

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

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, **34**, 387-398.

NOTE

1. Due to experimenter error, 3 participants in the rule-based low-pressure group and 3 participants in the information-integration low-pressure group did not perform the superpressure block.