Stopping Decisions: Information Order Effects on Nonfocal Evaluations

Michael Yu and **Cleotilde Gonzalez**, Carnegie Mellon University, Pittsburgh, Pennsylvania

Objective: We investigated how the order in which information is presented affects when a person decides to stop performing a task.

Background: A *stopping decision* is a decision to stop performing a task on the basis of a sequence of cues. Previous order-effects models do not account for how these contexts limit available working memory for making such decisions.

Method: Participants decided how long to perform a task known as the Work Hazard Game that began by rewarding points but later cost points if work continued after an unannounced "emergency." An additive sequence of cues indicated the probability of an emergency. Study 1 involved a three-group design with cue sequences that indicated the same risk at each decision point but whose final cue presented a high, medium, or low probability. Study 2 had a 2 × 2 design with high or low final cues and an easy or a challenging task.

Results: In Study 1, participants stopped sooner when the most recent cue presented a high rather than low probability (p = .09), despite the same emergency risk. In Study 2, participants stopped sooner when the most recent cue presented a high rather than low probability for the challenging task but not for the easy task (p = .08).

Conclusion: Stopping decisions appear sensitive to the most recent cue observed while experiencing task load. Participants responded to the same risks differently only on the basis of a change in presentation.

Application: Findings may be relevant for research and training for hazardous jobs, such as subsurface coal mining, fishing, and trucking.

Keywords: order effects, recency, attention, working memory, stopping decisions, coal mining

Address correspondence to Cleotilde Gonzalez, Dynamic Decision Making Laboratory, Carnegie Mellon University, 4609 Winthrop Street, Suite 102, Pittsburgh, PA 15213; e-mail: coty@cmu.edu.

HUMAN FACTORS
Vol. 55, No. 4, August 2013, pp. 732-746
DOI:10.1177/0018720812468152

INTRODUCTION

In many real-world tasks, decision makers must focus on performing their while at the same time evaluating safety conditions. For example, in subsurface coal mining, while a miner may be engaged in operating or maintaining heavy machinery (e.g., a continuous miner, shearer, or conveyor), he or she must also keep track of and evaluate cues of a potential mine emergency (e.g., shifts in air flow, increases in temperature, visual signs of smoke, or the weakening of supports). Researchers have studied how people evaluate a series of cues (e.g., Hogarth & Einhorn, 1992). However, their focus has been on cases in which the evaluation is the person's primary task. In subsurface coal mining, one would expect miners' primary task to be performing their jobs, whereas evaluating safety conditions is an important but peripheral concern.

In this article, we investigate how people evaluate a series of cues when this evaluation is not their primary task. We refer to this scenario as a *stopping decision*, because the result of this evaluation is often a decision to either continue or stop performing the primary task. The primary task is called the *focal task*, and the evaluation of whether to stop is called the *nonfocal evaluation*. For subsurface coal miners, the focal task includes normal mining operations, and the nonfocal evaluation is the evaluation of safety conditions. We suspect that the nonfocal nature of the evaluation can lead to inconsistent decisions. To evaluate this possibility, we ran two studies in which participants performed nonfocal evaluations, and we tested for information order effects. Specifically, we tested whether people would respond to the same information differently, depending on the order in which it was presented to them.

This potential for inconsistent decisions can be problematic in subsurface coal mining. First, it presents a challenge for safety trainers to communicate the appropriate safety environments in which specific actions can be taken. Second, as subsurface coal miners work in crews, differing evaluations between individual crew members can result in delayed reactions to dangerous conditions or lost productivity in safe conditions.

Stopping decisions are not unique to subsurface coal mining. Truckers and fishermen, for example, need to evaluate weather cues to determine how long to continue driving or fishing. Athletes may need to evaluate physical injury cues to determine whether to stay in the game. Stopping decisions are also prevalent in nonhazardous professions. A factory supervisor may need to determine when production needs to be stopped for machine maintenance. Office workers often need to determine when to interrupt their work, given signs of a potential computer or software failure.

For our purposes, we used a computer-based task called the Work Hazard Game, which includes the essential characteristics of stopping decisions in coal mining and other contexts (an important factor for making connections to real world tasks; see Gonzalez, Vanyukov, & Martin, 2005). In this game, participants (who belong to a general population) evaluated a series of cues to determine when to stop performing a money-making task. As with subsurface coal mining, continuing the focal task provides rewards and requires ongoing use of working memory; continuing the task past a certain point, indicated by the cues, results in a loss, though. A potential critique of this method would be that subsurface coal miners have expertise in evaluating the risk of a mine emergency, which our participants, facing a novel task, do not. However, prior research has shown that experts who are faced with rare events, changing cues, and limited decision aids can also make poor decisions (Shanteau, 1992). Another critique is that subsurface coal miners would be better motivated to pay attention to cues of mine emergency risk, given more severe consequences from a mistake. However, miners who fail to perform their jobs effectively also face more severe consequences, suggesting that there is also a stronger motivation to focus on continuing work.

In our experiments, we find that nonfocal evaluations are characterized by recency: People are more sensitive to later cues than earlier cues, although our findings are tempered by low significance levels. Our second study reproduces the recency effect with a modified version of the Work Hazard Game, strengthening our confidence in our initial finding. Additionally, we find evidence that low working memory availability can increase the recency effect. This finding supports the argument that the nonfocal nature of the evaluation task contributes to the recency effect.

Information Order Effects

Early research on information order effects showed that the order in which a series of cues was presented could affect how they were evaluated. However, these results were inconsistent, sometimes indicating greater sensitivity to earlier cues (the primacy effect; Nisbett & Ross, 1980), to later cues (the recency effect; Anderson, 1981; Davis, 1984), or to a mix of both (Murdock, 1962). The belief adjustment model, proposed by Hogarth and Einhorn (1992), reconciled these findings by connecting the differing results to different combinations of the task's characteristics: complexity, series length, consistency of information, response mode (evaluating at the end of the series or after each cue), and the type of judgment (estimating a value or evaluating a hypothesis). In this model, initial cues were used to develop an initial belief about the evaluation. Later cues were used to update this belief. However, the ways in which the initial and later cues contributed to this belief depended on the characteristics of the task. These task characteristics proved predictive in applied domains, including in auditing (Ashton & Ashton, 1988; Tubbs, Messier, & Knechel, 1990) and among Patriot air defense officers (Adelman & Bresnick, 1992; Adelman, Tolcott, & Bresnick, 1993). These studies suggested that important, practical decisions could be influenced by the order in which the information was presented.

In other research, however, it has been suggested that the original belief adjustment model did not capture all task characteristics relevant to real-world decision making. Authors of a follow-up study on Patriot air defense officers found evidence of primacy when the meaning of later cues could be reinterpreted rather than recency, as the belief adjustment model would have predicted (Adelman, Bresnick, Christian, Gualtieri, & Minionis, 1998). Stopping decisions are characterized by a split in focus between the focal task and the nonfocal evaluation. To date, order effects studies have not explicitly accounted for this focal and nonfocal split as a task characteristic. Authors of research in working memory (Baddeley, 1992, 2000; Baddeley & Hitch, 1974), however, suggest that a split in focus may affect how people process information and thus how cues are integrated into stopping-decision evaluations.

Working Memory and Stopping Decisions

Working memory describes the ability to maintain and process information. Greater working memory availability has been linked to better performance. For example, people with higher working memory capacity demonstrate greater controlled attention (Engle, 2002; Engle, Kane, & Tuholski, 1999), and people perform better when faced with low task loads (Anderson, Reder, & Lebiere, 1996; Gonzalez, 2005).

In a study in which U.S. Navy officers were asked to categorize unidentified planes, high task load was associated with lower information recall regarding planes viewed as nonthreatening (Perrin, Barnett, Walrath, & Grossman, 2001). This selective focus may have been adaptive, leading to devoting more attention to the threatening cues. For subsurface coal mining, however, it may lead to neglecting cues that only appear nonthreatening. Besides neglecting information, limited working memory may also lead an individual to be more selective in information processing. For example, research in hypothesis generation links constraints on working memory to decreases in the number of hypotheses that are generated and pursued (Dougherty & Hunter, 2003; Flin, Stewart, & Slaven, 1996; Thomas, Dougherty, Sprenger, & Harbison, 2008). For subsurface coal mining, this link means that miners may neglect evaluating safety conditions completely if normal mining operations are particularly demanding.

Stopping decisions impose constraints on working memory by splitting the focus between the focal task and the nonfocal evaluation. These constraints can reduce attention to the nonfocal evaluation. For subsurface coal mining, in which cues of a potential emergency evolve and accumulate over time, we expect that this effect would be stronger at the start of the evaluation. In the beginning, the accumulated cues are likely to appear less threatening, and normal mining operations are likely to demand more of the miners' working memory. Insufficient attention paid to initial cues has been shown to reduce primacy effects (Wilson, Houston, Etling, & Brekke, 1996). As we expect stopping decisions to reduce attention at the start of the evaluation, we expect stopping decisions to be characterized by recency.

In two studies, we test our prediction about recency and the role of working memory in stopping decisions using a simulation known as the Work Hazard Game. In the first study, we look for the overall effect of recency by varying the cue sequence of the nonfocal evaluation. In the second study, we consider the role of working memory by manipulating both the cue sequence of the nonfocal evaluation and the working memory demands of the focal task.

STUDY 1

Participants

Participants were recruited through an online participant pool from Carnegie Mellon University and the University of Pittsburgh to participate in the Work Hazard Game for a $4 base payment and a potential bonus of up to $8. Participants had to complete a quiz on the game's mechanics after training, and those who passed continued on to the actual study. A total of 18 participants who were recruited passed the quiz. Of those who passed, 44% were female and the median age was 21.

Materials

The Work Hazard Game is a simulation designed for the study of nonfocal evaluations in stopping decisions. In the game, participants initially earn points by performing a focal task. As the game progresses, there is a probability that the game changes and that performing the

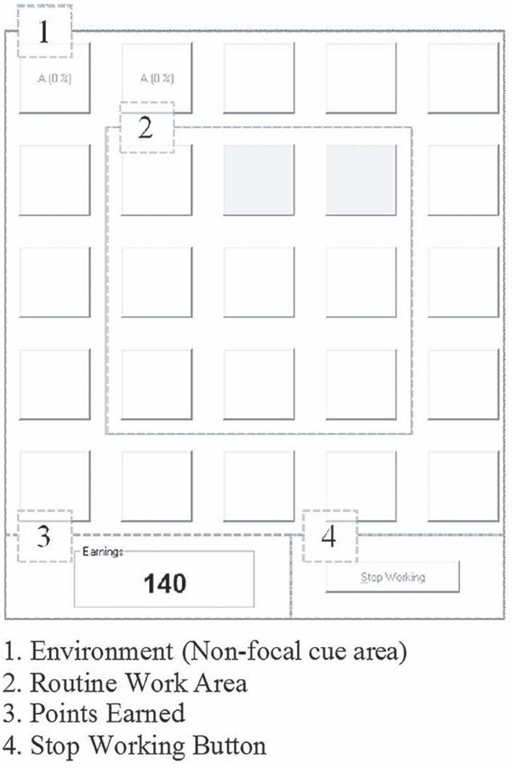

Figure 1. Diagram of Work Hazard Game, Study 1.

focal task will instead lose points. Participants are not informed of when the game changes but are provided a sequence of cues indicating the probability that the shift has happened. Participants must determine when to stop performing the focal task. The interpretation of the cue sequence and the decision of when to stop represent the nonfocal evaluation.

This simulation is set up on a 5 × 5 grid of squares, with the center 3 × 3 squares designated as the "routine work area" and the surrounding 16 squares as the "work environment," as presented in Figure 1. The focal task is performed in the routine work area. The cues appear in the work environment. The game proceeds in turns and rounds, with each round composed of five turns. Each turn, the participant is asked to perform the focal task. At the end of each round, participants have the option of continuing with the focal task or stopping work.

Focal task. The focal task is a pattern completion task. At the start of a turn, two adjoining squares in the work area will light up, one after the other. The participant must click on a third square in the routine work area that is to the left, right, above, or below the second square to earn points. Clicking on a square that satisfies these requirements rewards the participant with 10 points, whereas clicking on a different square results in 0 points. All squares are then turned off and the next turn starts.

Nonfocal evaluation. The nonfocal evaluation involves evaluating a sequence of cues to determine when to stop the focal task. At the start of each turn, a cue may appear in the environment. At most, one cue will appear each turn. Cues appear in the work environment starting from the upper-left corner. Subsequent cues fill the work environment clockwise. There are five types of cues, each labeled with a letter and a percentage: "A–0%", "B–2.5%", "C–5%", "D–7.5%", and "E–10%." After they appear, cues remain visible for the rest of the game.

At the end of the first turn of each round (every five turns), the game may change to an "emergency" state. In that state, participants lose points for continuing the focal task. The probability that an emergency state occurs is equal to the sum of the percentages displayed on all visible cues. For example, if two cues were in the environment, "B–2.5%" and "E–10%," the risk of transitioning to an emergency state is 12.5%. When started, an emergency state remains for the rest of the game. Participants lose 15 points for each turn completed, resulting in a net loss of 5 points if the focal task is completed correctly and a net loss of 15 if the focal task is completed incorrectly. Point losses are not shown to the participant until after work has stopped so as to prevent participants from using their point total to determine whether an emergency state has started.

To avoid working in an emergency state, participants may stop performing the focal task before completing the first turn in each round. If they stop, they are told whether an emergency had occurred and are provided information about points earned, points lost, and the final score. If the participants continue, they must complete the next round (5 turns) before they have an option of stopping again. If the participant reaches

20 rounds (100 turns) without stopping work, the game ends automatically.

Design

The study involved three within-subject treatments, each with a different cue sequence. Whenever participants had to decide whether to stop performing the focal task, all treatments indicated the same probability of an emergency state. In the "increasing" treatment, cues within each round indicated successively larger increases in probability. The "decreasing" treatment had the same cues from the increasing treatment, but the order was reversed within each round. The "flat" treatment repeated the same cue within each round, such that the total risk was the same as in the other treatments.

For example, in Round 12, the increasing treatment included a cue of "A–0%" in Turn 57 followed by a cue of "C–5%" in Turn 59; the decreasing treatment included a cue of "C–5%" in Turn 57 followed by a cue of "A–0%" in Turn 59; and the flat treatment included a cue of "B–2.5%" in Turn 57 followed by a cue of "B–2.5%" in Turn 59. Thus, at the start of Round 13, the total risk that an emergency would be triggered was increased by 5%, regardless of treatment. The full cue sequence across all treatments is included in Figure 2.

Procedures

Participants first reviewed a computer-based tutorial on the Work Hazard Game. This tutorial explained all of the game mechanics described earlier. Then participants completed two practice versions of the game, one of which included no cues indicating an emergency. Participants then completed a quiz on the game's mechanics. The quiz tested participants' understanding of the turn–round structure, when a shift to an emergency state would affect their score, how the score display on the game worked, and the consequences of stopping or continuing work on the basis of the emergency state. Those who passed then answered demographic questions and continued with the study. Those who failed were paid their show-up fee and did not continue. Participants continuing with the study performed all three treatments. The treatments were counterbalanced across participants. After each treatment, participants indicated their belief in how likely it was that they had stopped during an emergency state (1 = *very unlikely* to 5 = *very likely*). Participants were then paid a bonus based on one of the three treatments selected at random. The bonus ranged from \$1 to \$8 and was determined by the equation 1 + max(0, score/1000 × 7), rounded to the nearest dollar.

RESULTS

We performed a repeated-measures ANOVA with the number of turns before stopping the focal task as our dependent variable and the treatment as our independent variable, controlling for participant-level effects. The ANOVA suggested a potential difference in treatments, $F(2, 17) = 2.62$, $p = .09$, with the mean number of turns worked as greatest in the decreasing treatment ($M = 63.89$, $SD = 18.36$), followed by the flat treatment ($M = 61.94$, $SD = 17.51$), and then the increasing treatment ($M = 55.83$, $SD = 17.51$). We test for differences in the treatments while adjusting for multiple comparisons using Hothorn, Bretz, and Westfall's (2008) and Bretz, Hothorn, and Westfall's (2010) methodology, as implemented in the multcomp package in R. We find potential differences between increasing and decreasing treatment ($Z = 2.19$, $p = .07$) but not between the decreasing and flat ($Z = 0.53$, $p = .86$) or the increasing and flat ($Z = 1.66$, $p = .22$) sequences. Figure 3 shows the percentage of participants remaining in the game at a given turn. Lower curves imply that participants generally stopped work earlier, and higher curves indicate that they stopped work later. In general, the curve for the increasing sequence falls below that of the flat sequence and the flat sequence below that of the decreasing sequence. The flat sequence also appears to cross over both sequences, suggesting that it may be difficult to evaluate its effects relative to the other two. Note that the increasing sequence is consistently equal to or lower than the decreasing sequence.

Two participants stopped the focal task before a nonzero risk cue appeared in the environment. Participants can always increase their score by continuing work in these scenarios. As such, it is possible that these participants did not

Round	Turn	Cue Appearance			Risk of Emergency Starting
		Increasing	Flat	Decreasing	
1	1				0.0%
	2				
	3				
	4				
	5				
2	6				0.0%
	7				
	8				
	9				
	10				
3	11				0.0%
	12				
	13				
	14	A -- 0%	A -- 0%	A -- 0%	
	15	A -- 0%	A -- 0%	A -- 0%	
4	16				0.0%
	17				
	18				
	19				
	20				
5	21				0.0%
	22				
	23				
	24				
	25				
6	26				0.0%
	27				
	28				
	29				
	30				
7	31				0.0%
	32				
	33				
	34				
	35				
8	36				0.0%
	37	A -- 0%	B -- 2.5%	C -- 5%	
	38	B -- 2.5%	B -- 2.5%	B -- 2.5%	
	39	C -- 5%	B -- 2.5%	A -- 0%	
	40				
9	41				7.5%
	42				
	43				
	44				
	45				
10	46				7.5%
	47				
	48				
	49				
	50				

Round	Turn	Cue Appearance			Risk of Emergency Starting
		Increasing	Flat	Decreasing	
11	51				7.5%
	52	B -- 2.5%	C -- 5%	D -- 7.5%	
	53	D -- 7.5%	C -- 5%	B -- 2.5%	
	54				
	55				
12	56				17.5%
	57	A -- 0%	B -- 2.5%	C -- 5%	
	58				
	59	C -- 5%	B -- 2.5%	A -- 0%	
	60				
13	61				22.5%
	62	A -- 0%	C -- 5%	E -- 10%	
	63				
	64	E -- 10%	C -- 5%	A -- 0%	
	65				
14	66				32.5%
	67				
	68				
	69				
	70				
15	71				32.5%
	72	C -- 5%	D -- 7.5%	E -- 10%	
	73	E -- 10%	D -- 7.5%	C -- 5%	
	74				
	75				
16	76				47.5%
	77	C -- 5%	D -- 7.5%	E -- 10%	
	78	E -- 10%	D -- 7.5%	C -- 5%	
	79				
	80				
17	81				62.5%
	82				
	83				
	84				
	85				
18	86				62.5%
	87				
	88				
	89				
	90				
19	91				62.5%
	92				
	93				
	94				
	95				
20	96				62.5%
	97				
	98				
	99				
	100				

Figure 2. Cue Sequences for Study 1

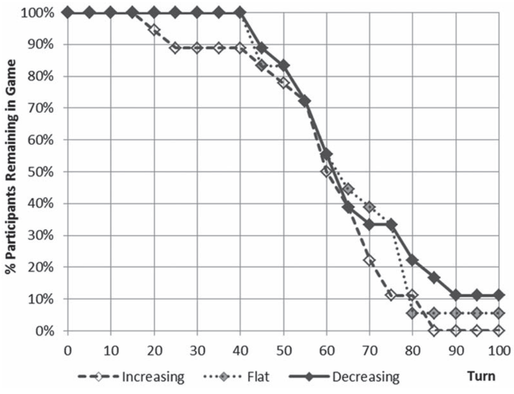

Figure 3. Participants remaining by turn, Study 1.

understand the game. Dropping all responses from these two participants results in a loss of significance from the ANOVA, $F(2, 15) = 1.22$, $p = .31$.

We also tested for a relationship between the participant's belief in having stopped in an emergency state and the treatments. In neither the full data set nor the reduced data set did the ANOVA suggest any relationship, $F(2, 17) = 0.56$, $p = .58$; $F(2, 15) = 0.38$, $p = .69$.

DISCUSSION

The results provide tentative support for recency. Participants stop earlier in the increasing treatment than in the decreasing treatment. As the increasing treatments have higher-probability cues appearing at the end, this behavior is consistent with participants' being more sensitive to later cues. That finding noted, we are cautious about our conclusions, given no significant difference between the flat treatment, a somewhat high alpha ($\alpha = .10$) required for significance between the increasing and decreasing treatments, and a loss in significance when we drop the 2 participants who may have misunderstood the game. Directional consistency is maintained in both these cases, providing encouragement for a follow-up study.

We discuss several alternative explanations for our findings. First, the different treatments may have caused participants to use different probability thresholds to determine when to stop working. In other words, participants may have deliberately chosen to stop at lower probabilities of an emergency state for the increasing sequence than for the decreasing sequence. However, our results indicate no significant difference in the perceived likelihood of an emergency state across treatments. This finding suggests that stopping at different probabilities may not have been deliberate.

Next, participants may have tried to "predict" future cues on the basis of the pattern of cues observed. The increasing treatment could be interpreted as suggesting that future cues will have high probabilities. However, the simulation was designed such that new cues would affect the risk of an emergency only in the following round. As participants could stop at the start of each round, participants would be affected only by cues they have observed and not by cues that they expect to see. As such, predictions of future cues should not affect behavior. Although some participants may have been confused, they were trained and quizzed on this specific aspect of the game. This training should have reduced this possible confusion.

Finally, as cues remained visible after they appeared and as participants could stop only at the start of each round, participants may have evaluated the cues as a set rather than one at a time. Research on the perceptions of sets suggests that people recall the mean and range of a set but not individual items (Ariely, 2001). In this case, we would expect no differences in stopping decisions. These findings suggest that our effects may have been weakened.

A second study was designed to address some of the outstanding questions from Study 1. Specifically, in the second study, we looked to replicate Study 1's results with a larger sample size and greater power, to test for the role of working memory directly, and to resemble realistic stopping decisions more closely.

STUDY 2

Participants

Participants were recruited through an online participant pool from Carnegie Mellon University and the University of Pittsburgh to participate in the Work Hazard Game for a \$4 base payment and a potential bonus of up to \$10. Participants had to complete a quiz on the game's mechanics after training, and those who passed continued on to the actual study. A total of 43 participants who were recruited passed the quiz. Of those who passed, 48% were female and the median age was 22.

Materials

Similar to Study 1, participants played the Work Hazard Game and were asked to perform a task that initially rewarded points but then later changed into a task that cost points on the basis of visible cues. However, important changes were made to both the focal task and the nonfocal evaluation.

Focal task. The focal task was changed from a pattern completion task to a pattern matching task. Within the 3 × 3 routine work area, several squares would light up and turn off, one after the other. Participants had to click on the squares in the same order they lit up. Participants earned 10 points for successfully replicating the pattern and 0 points if they made a mistake. By changing the length of the pattern, we could manipulate the working memory demands of the task.

The design of the focal task was similar to that of the Corsi Block Tapping task (Corsi, 1972), which has been used to measure working memory (e.g., Berch, Krikorian, & Huha, 1998; Cavallini, Fastame, Palladino, Rossi, & Vecchi, 2003; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Park et al., 2002). The original Corsi task involves nine blocks arranged in seemingly random locations. The experimenter begins by tapping blocks one after another, starting with two blocks. The participant must tap the same two blocks in the same order. The task repeats with another pattern of two blocks, then moves to two different patterns of three blocks, then two different patterns of four blocks, and so on. Participants continue with the experiment if they can successfully replicate at least one of the two patterns for a given pattern length. Participants then received a score based on the number of patterns successfully replicated multiplied by the length of the longest pattern replicated.

Nonfocal evaluation. As before, participants had to evaluate a series of cues in the environment to determine when to stop the focal task. The cues were changed from text and percentage labels to a random pattern of red squares that

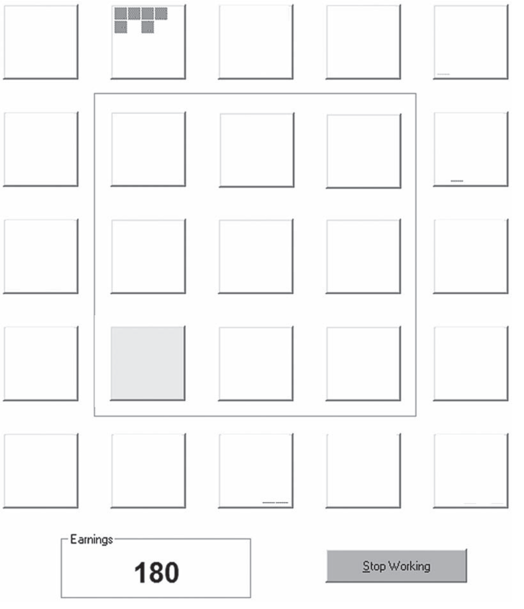

Figure 4. Diagram of Work Hazard Game, Study 2.

appeared in each cue. Each square was explained as a hazard and represented a 0.5% risk of triggering an emergency. As before, cues appeared starting from the upper-left corner and continued clockwise. When a cue appeared, it increased the risk of triggering an emergency for all subsequent rounds. However, cues were visible only for a single turn and would disappear after the turn was completed. An example of a revised cue appears in the context of the game board in Figure 4. The location of hazards within each cue was randomized before every game, such that the visual representation of the hazards differed even when playing the same game multiple times. These modifications were designed to provide a less abstract representation of hazards and to encourage evaluation of the risk cues in sequence.

The game was similarly divided into turns and rounds. As before, at the start of each round, there was a probability of transitioning into an emergency state, and participants could decide whether to continue the focal task. Because of the increased time required to perform the focal task, each round was to three turns, compared with the five turns in the original Work Hazard Game in Study 1.

Round	Turn	Number of Hazards		Risk of Emergency
		Increasing	Decreasing	
1	1			0.0%
	2			
	3			
2	4			0.0%
	5			
	6			
3	7			0.0%
	8	1	5	
	9	3	3	
4	10	5	1	4.5%
	11			
	12			
5	13			4.5%
	14			
	15	2	6	
6	16	6	2	8.5%
	17			
	18			
7	19			8.5%
	20			
	21			
8	22			8.5%
	23	1	5	
	24			
9	25	5	1	11.5%
	26			
	27	2	4	
10	28	4	2	14.5%
	29			
	30			

Round	Turn	Number of Hazards		Risk of Emergency
		Increasing	Decreasing	
11	31			14.5%
	32			
	33			
12	34			14.5%
	35	1	3	
	36	2	2	
13	37	3	1	17.5%
	38			
	39			
14	40			17.5%
	41			
	42			
15	43			17.5%
	44	1	3	
	45			
16	46	3	1	19.5%
	47			
	48			
17	49			19.5%
	50			
	51	2	6	
18	52	6	2	23.5%
	53			
	54			
19	55			23.5%
	56			
	57			
20	58			23.5%
	59			
	60			

Figure 5. Cue Sequences for Study 2

Design

We created four treatments in a 2 × 2 design, manipulating the sequence order (increasing, decreasing) and the difficulty of the focal task (repeating a three-square pattern, "easy"; repeating a five-square pattern, "challenging"). As before, the probability of changing to an emergency state was the same at the start of each round across all four treatments. The full cue sequence across all treatments is included in Figure 5.

Procedure

Participants first reviewed a computer-based tutorial on the Work Hazard Game and ran through one practice version of the game. Participants then completed a quiz on the game's mechanics. Those who passed answered some demographic questions and continued with the study. Those who failed were paid their show-up fee and did not continue. Participants continuing with the study performed all four treatments. The order in which the participants took the

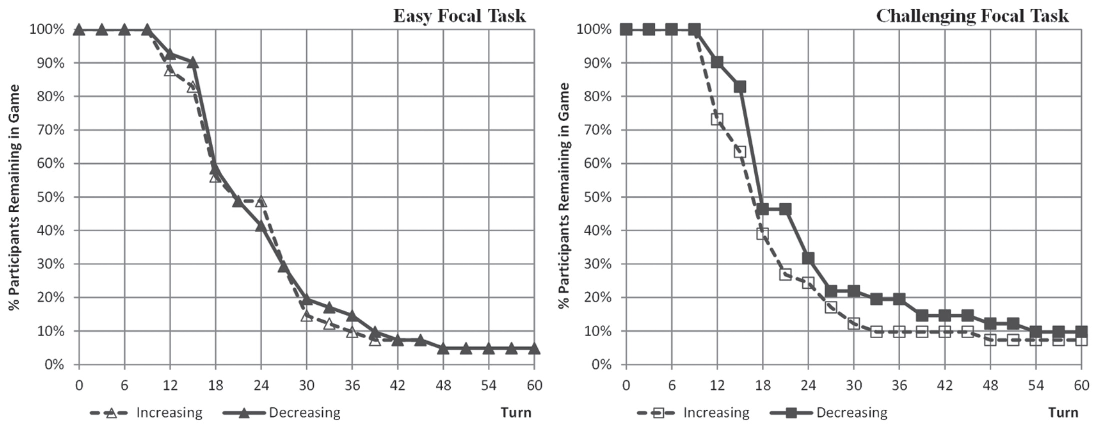

Figure 6. Participants remaining by turn, Study 2.

treatments was counterbalanced. After completing all four treatments, participants were asked to perform the Corsi block-tapping task, as implemented in PEBL (Psychology Experiment Building Language; Kessels, van Zandvoort, Postma, Kappelle, & de Haan, 2000; Mueller, 2010). Participants were then paid a bonus on the basis of their performance on one of the four treatments selected randomly. Bonuses ranged from \$1 to \$9 and were determined by the following equation: $1 + \max(0, \text{score}/600 \times 9)$.

Results

Number of turns. We performed a repeated-measures ANOVA with the number of turns before stopping as our dependent variable and the treatments as our independent variable, while controlling for participant-level effects. The ANOVA suggested potential main effects for the treatments, $F(3, 42) = 2.44$, $p = .07$. For the challenging focal task, the mean number of turns worked was higher in the decreasing treatment ($M = 23.86$, $SD = 16.04$) than in the increasing treatment ($M = 20.02$, $SD = 14.79$). For the easy focal task, the mean number of turns worked was also higher in the decreasing treatment ($M = 24.56$, $SD = 14.08$) than in the increasing treatment ($M = 23.86$, $SD = 14.10$). Adjusting for multiple comparisons, we tested for differences between the increasing and decreasing treatments for each task difficulty separately and found a potential difference for the challenging focal task ($Z = 2.06$, $p = .08$) but no difference for the easy focal task ($Z = 0.37$, $p = .92$). Figure 6 shows the percentage of participants remaining in the game for a given turn. Lower curves imply that participants generally stopped work earlier and higher curves indicate that they stopped later. The easy focal task shows that the difference between the increasing and decreasing sequences is small, although the curve for the increasing sequence is generally lower than the decreasing sequence. The challenging focal task shows greater difference between the increasing and decreasing sequences, with the increasing sequence more consistently equal to or lower than the decreasing sequence.

We tested whether working memory mediates the recency effect using two measures. First, we considered the effects of task load. We performed a two-way repeated-measures ANOVA using sequence order and task difficulty as our two factors. The interaction term in the ANOVA was not significant, $F(1, 42) = 2.62$, $p = .11$. The two-way ANOVA run without the interaction term indicated a potential main effect for the sequence order, $F(1, 42) = 3.77$, $p = .06$, but it did not reach significance for the changes in focal task difficulty, $F(1, 42) = 1.76$, $p = .19$. Next, we considered the effects of individual differences in working memory capacity as measured by the Corsi task. We ran a repeated-measures ANCOVA using sequence order and task difficulty as factors and working memory capacity as a covariate. The three-way interaction of sequence order,

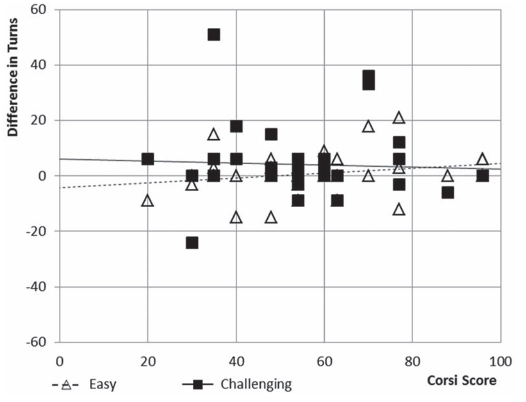

Figure 7. Difference in turns, decreasing to increasing, by working memory capacity.

task difficulty, and working memory was not significant, $F(1, 41) = 1.16$, $p = .29$. A plot of the difference in number of turns worked between the decreasing and increasing sequences, shown in Figure 7, suggests that although increased working memory capacity may lead to more consistent behavior across decreasing and increasing sequences (differences closer to 0), there is also substantial subject-level variability.

Likelihood of emergency. We also tested for any relationship between the perceived likelihood of an emergency state and the treatments. The ANOVA suggested main effects for the treatments, $F(3, 42) = 5.55$, $p < .01$. For the challenging focal task, the mean rating for likelihood was higher in the decreasing treatment ($M = 3.63$, $SD = 0.87$) than in the increasing treatment ($M = 3.00$, $SD = 1.18$). For the easy focal task, the mean rating was also higher in the decreasing treatment ($M = 3.49$, $SD = 1.05$) than in the increasing treatment ($M = 3.16$, $SD = 1.04$). Differences in the perceived likelihood of an emergency state across different sequence orders were significant for the challenging focal task ($Z = 3.62$, $p < .01$) but not for the easy focal task ($Z = 1.88$, $p = .12$).

Discussion

In Study 2, we found further support for recency, although still requiring a somewhat high alpha for significance ($\alpha = .10$). Along with the results from Study 1, however, this strengthens our belief in the effect. Study 2 also provides tentative evidence that the complexity of the focal task, and thus working memory, may play a role in moderating the effects.

However, Study 2 also provides evidence that participants may have deliberately chosen to stop the focal task at different perceived probabilities of an emergency state. This explanation was an alternative considered but dismissed in Study 1. Potential mechanisms for this effect are discussed in the general discussion.

GENERAL DISCUSSION

In the two studies discussed, stopping decisions seem to be characterized by recency—when people are determining whether to continue performing an ongoing task, their decisions appear to be more strongly influenced by the most recent cue that they observed. This effect may be moderated by working memory availability. With heavy task load, when working memory is limited, this effect appears stronger. With light task load, when working memory is more freely available, this effect may disappear. In our two studies, we found that participants responded differently to the same information about risk when the information was reordered and task load was increased.

In our research, we advance the belief adjustment model (Hogarth & Einhorn, 1992) by adapting the model to a more dynamic environment for decision making and by evaluating the influence of working memory on its predictions. In addition, our results enhance and relate to other concepts in the human factors literature. Research in situation awareness (Endsley, 1995) focuses on problems in which knowledge of the environmental state is important. Research in mode awareness (Sarter & Woods, 1995) focuses on how people use this knowledge to change how they interact with a given set of equipment, such as a cockpit, in different environmental states. Stopping decisions are similar to mode awareness problems in that environmental cues are used to drive changes in behavior, although stopping decisions may not be directly tied to any equipment. Limited working memory has been linked to greater errors in situation awareness tasks (Endsley, 1995). To our knowledge, however, there is no work in situation awareness that focuses on evaluating the sequences of cues and the evolution and dynamics of cues over time. Our study may provide a starting point for work in this particular area.

Alternatively, our results may also be considered in the framework of risks as feelings. In this article, we treated nonfocal evaluations as a cognitive task in which one interprets a series of cues. However, our studies used terms such as *risks, hazards*, and *emergencies*, which are affectively rich. Risks have been shown to be multidimensional and include more than just numeric probabilities (Fischhoff, Watson, & Hope, 1984), and people's response to risks can be based on how they experience them as feelings (Loewenstein, Weber, Hsee, & Welch, 2001). By limiting working memory, our studies may have caused participants to rely more heavily on their feelings rather than their risk assessments. Risks as feelings can explain why participants may have chosen to leave given different perceived probabilities of an emergency state. For example, our increasing treatment may have induced a greater discomfort in participants, leading them to stop the focal task earlier. If our participants were evaluating the risk of an emergency state affectively, the peak-end effect, which suggests that affective experiences are evaluated on one's most extreme and final feelings (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993), would also produce predictions similar to our recency hypothesis. Nonetheless, working memory would still play a similar role in these affect-based models, and the solutions previously suggested should still reduce the information order effects.

Implications to Coal Mining

Our results have important implications for many fields in which humans must share attention between performing their focal task while continuously evaluating cues of safety conditions. In the case of subsurface coal mining, evaluating safety conditions plays an important part in protecting the well-being of miners and in developing awareness of potential emergencies. Similar to the task in our studies, miners must perform this evaluation while performing other tasks associated with normal mining operations. If miners are subject to the same recency effects, then they would evaluate the same set of cues differently depending on the order in which they were presented. This inconsistency poses a challenge both for miners trying to accurately evaluate safety conditions and for mine safety researchers who would want to provide concise, consistent guidance on how to appropriately respond to emergency cues.

To promote more consistent evaluations, mine safety researchers may want to pursue policies that reduce the demands on the miners' working memory. Authors of studies in other environments have considered reducing these demands, for example, by providing electronic information displays (Adelman, Bresnick, Black, Marvin, & Sak, 1996). These displays reduce the demands of the nonfocal evaluation by helping to maintain the cues as well as to present the cues in a way that promotes a more consistent evaluation. Such technological solutions seem ill suited to a mining environment, though, where damage may be likely. Even pencil-and-paper equivalents may be problematic, given low-light conditions and the ease at which these tools may be misplaced.

Another solution may be to train miners to evaluate safety conditions right before they start work and each time they take a break. This solution may not be a substantial change from what the miners naturally do. However, if done explicitly, it may strengthen the miners' impressions of earlier and intermediate emergency cues. Nonetheless, miners may still neglect cues that are presented while they are in the middle of some other task.

Safety evaluations may also be improved by sharing the responsibility among those with fewer working memory demands. Currently, the crew foreman is the main person responsible for mine safety. However, the foreman is likely to be engaged in a large number of concurrent tasks. Thus, he or she is likely to be working with high task load and may pay insufficient attention to initial mine emergency cues. Crew members with more limited responsibilities may be better equipped to recall early and later cues more consistently. Sharing safety responsibilities with these crew members may provide a more consistent check on the foreman's evaluations.

Finally, it may be possible to reduce the working memory demands of evaluation by improving a miner's ability to perform the task. Research has shown that participants training in a slower-paced, controlled environment can perform better than those trained in environments designed to emulate the real scenario (Gonzalez & Brunstein, 2009). It seems plausible that practiced evaluations can help alleviate working memory demands and reduce the recency effect.

Limitations and Future Work

The aforementioned implications must be considered according to the constraints of the simplicity of the task used in our studies. Our research involved abstracted laboratory experiments with university undergraduate and graduate students. Experienced miners may make better decisions than our laboratory participants when considering the naturalistic conditions and the context experience they bring to the job. However, studies suggest that the characteristics of mine safety evaluation—rare events, changing cues, and limited decision aids—can lead to poor performance even among experts (Shanteau, 1992), and given the relatively uncommon nature of mining emergencies, even experienced miners have little actual exposure to them. Additionally, reliance on recency in mine safety evaluation may be an advantage rather than a problem. For example, if mine emergencies are characterized by the presence of more important cues over time, focusing on the last cues may be sufficient, as long as response to those later cues implicitly account for the earlier ones. Such behavior is consistent with the fast-and-frugal model proposed by Gigerenzer and Goldstein (1996). Although such strategies may be generally effective, policy makers must decide whether the number of times they do not work is worth addressing.

We propose several areas for further work on stopping decisions. From a theoretical perspective, additional work to clarify the mechanisms underlying the recency effect is warranted. Although we found evidence that working memory may play a role, the high alpha level required for significance ($\alpha = .10$) suggests that more rigorous testing may be required. Furthermore, other mediators, such as affective processing, should also be considered. A clear understanding of these factors will help in better design and targeting of interventions.

Further efforts to simplify the Work Hazard Game and make it more intuitive may lead to reduced noise and more accurate measures of its effect. The changes to the focal task in Study 2 not only allowed us to manipulate the focal task's working memory requirements but provided a more familiar task for participants. However, a more contextualized game could make the nature of cue integration and the unannounced emergency more intuitive. For example, one contextualization of the cues is to tell participants that they are performing a stylized mining task. As they work, they receive messages, such as "There is a small increase in methane levels" or "The air flow stopped for about 30 seconds." Training prior to the game will allow participants to interpret these cues, for example, telling them that repeated malfunctions in the ventilation system indicate an increasing risk of that the system will fail. This method replicates our cue accumulation process more intuitively. Note that in this case, the stylized context makes the nature of the cues easier to understand but is unlikely to accurately reflect actual mining if designed for the general public. Nonmining contexts may be used, as well. For example, participants may be told that they are operating a printing press, receiving messages about possible issues with the printing process. In this version, the unannounced emergency may also be more intuitively expressed, such as if participants are told that they cannot check the results of the press until the press is stopped.

Recent innovations in subsurface coal mine training also offers new ways to adapt the paradigm of the Work Hazard Game into a more realistic simulation. Virtual reality simulators can facilitate use of real-world cues and experienced participants to help better demonstrate the degree to which our findings apply to these environments and to identify other potential task characteristics that play an important role in stopping decisions.

ACKNOWLEDGMENTS

The authors are thankful to Hau-yu Wong and Tomas Lejarraga of the Dynamic Decision Making

Lab for providing editorial comments. This research was partially supported by a National Institute for Occupational Safety and Health (Centers for Disease Control and Prevention, Award No. 200-2009-31403) award to professor Cleotilde Gonzalez.

KEY POINTS

- Stopping decisions, which are decisions to stop performing a task on the basis of a sequence of cues, were predicted to correlate with the most recent cue observed.
- Participants played the Work Hazard Game, in which they earn money for performing a task in its initial state but lose money for the task in an "emergency" state. Cues indicating the same risk of an emergency state were presented to participants but developed in different orders.
- In Study 1, participants stopped earlier when cue sequences ended with a high-risk rather than a low-risk cue (p = .09). Study 2 reproduced this finding (p = .08) and showed directional evidence of working memory as a mediator.
- People are likely to make more consistent stopping decisions when they are not distracted by other tasks requiring working memory.

REFERENCES

Adelman, L., & Bresnick, T. (1992). Examining the effect of information sequence on Patriot air defense officers' judgments. *Organizational Behavior and Human Decision Processes*, *53*, 204–228. doi:10.1016/0749-5978(92)90062-C

Adelman, L., Bresnick, T. A., Black, P. K., Marvin, F. F., & Sak, S. G. (1996). Research with Patriot air defense officers: Examining information order effects. *Human Factors*, *38*, 250–261. doi:10.1518/001872096779048002

Adelman, L., Bresnick, T. A., Christian, M., Gualtieri, J., & Minionis, D. (1998). Demonstrating the effect of context on order effects for an Army air defense task using the Patriot simulator. *Journal of Behavioral Decision Making*, *10*, 327–342. doi:10.1002/(SICI)1099-0771(199712)10:4<327::AID-BDM259>3.0.CO;2-8

Adelman, L., Tolcott, M. A., & Bresnick, T. A. (1993). Examining the effect of information order on expert judgment. *Organizational Behavior and Human Decision Processes*, *56*, 348–369. doi:10.1006/obhd.1993.1059

Anderson, N. H. (1981). *Foundations of information integration theory*. New York, NY: Academic Press.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*, 221–256. doi:10.1006/cogp.1996.0007

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162. doi:10.1111/1467-9280.00327

Ashton, A. H., & Ashton, R. H. (1988). Sequential belief revision in auditing. *Accounting Review*, *63*, 623–641.

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559. doi:10.1126/science.1736359

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 42–89). New York, NY: Academic Press.

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi Block-Tapping Task: Methodological and theoretical considerations. *Brain and Cognition*, *38*, 317–338. doi:10.1006/brcg.1998.1039

Bretz, F., Hothorn, T., & Westfall, P. (2010). *Multiple comparisons using R*. Boca Raton, FL: CRC Press.

Cavallini, E., Fastame, M. C., Palladino, P., Rossi, S., & Vecchi, T. (2003). Visuo-spatial span and cognitive functions: A theoretical analysis of the "Corsi" task. *Imagination, Cognition, and Personality*, *23*, 217–224. doi:10.2190/HWQ6-35GU-C9VA-WVHA

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Unpublished doctoral dissertation). McGill University, Montreal, Canada.

Davis, J. H. (1984). Order in the courtroom. In D. J. Miller, D. C. Blackman, & A. J. Chapman (Eds.), *Perspectives in psychology and law* (pp. 251–265). London, UK: Wiley.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neruopsychologia*, *37*, 1189–1199. doi:10.1016/S0028-3932(98)00159-6

Dougherty, M. R., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica*, *113*, 263–282. doi:10.1016/S0001-6918(03)00033-7

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32–64. doi:10.1518/001872095779049543

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23. doi:10.1111/1467-8721.00160

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York, NY: Cambridge University Press.

Fischhoff, B., Watson, S. R., & Hope, C. (1984). Defining risk. *Policy Sciences*, *17*, 123–139. doi:10.1007/BF00146924

Flin, R., Stewart, K., & Slaven, G. (1996). Emergency decision making in the offshore oil and gas industry. *Human Factors*, *38*, 262–277. doi:10.1518/001872096779048110

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669. doi:10.1037/0033-295X.103.4.650

Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors*, *47*, 92–101. doi:10.1518/0018720053653767

Gonzalez, C., & Brunstein, A. (2009). Training for emergencies. *Journal of Trauma*, *67*, S100–S105. doi:10.1097/TA.0b013e3181adc0d5

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*, 273–286. doi:10.1016/j.chb.2004.02.014

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55. doi:10.1016/0010-0285(92)90002-J

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*, 346–363. doi:10.1002/bimj.200810425

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*, 401–405. doi:10.1111/j.1467-9280.1993.tb00589.x

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology*, *7*, 252–258. doi:10.1207/S15324826AN0704_8

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*, 267–286. doi:10.1037/0033-2909.127.2.267

Mueller, S. T. (2010). *PEBL: The Psychology Experiment Building Language* (Version 0.11) [Software]. Available from http://pebl.sourceforge.net/

Murdock, B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488. doi:10.1037/h0045106

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of human judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*, 299–320. doi:10.1037/0882-7974.17.2.299

Perrin, B. M., Barnett, B. J., Walrath, L., & Grossman, J. D. (2001). Information order and outcome framing: An assessment of judgment bias in a naturalistic decision-making context. *Human Factors*, *43*, 227–238. doi:10.1518/001872001775900968

Sarter, N., & Woods, D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, *37*, 5–19. doi:10.1518/001872095779049516

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*, 252–266. doi:10.1016/0749-5978(92)90064-E

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, *115*, 155–185. doi:10.1037/0033-295X.115.1.155

Tubbs, R., Messier, W. F., & Knechel, W. R. (1990). Recency effects in the auditor's belief-revision process. *Accounting Review*, *65*, 452–460.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*, 387–402. doi:10.1037/0096-3445.125.4.387

Michael Yu received his master's degree in social and decision sciences from Carnegie Mellon University in 2010, where he is currently a PhD student.

Cleotilde Gonzalez is an associate research professor in the Department of Social and Decision Sciences and director of the Dynamic Decision Making Laboratory at Carnegie Mellon University. She received her PhD in management information systems in 1996 from Texas Tech University.

Date received: October 22, 2010
Date accepted: October 21, 2012