Title

Individual Differences in Causal Judgment under Time Pressure: Sex and Prior Video Game Experience as Predictors

Permalink

https://escholarship.org/uc/item/5989z7d3

Journal

International Journal of Comparative Psychology, 24(1)

ISSN

0889-3675

Authors

Young, Michael E.
Sutherland, Steven C.
Cole, James J.

Publication Date

2011

DOI

10.46867/ijcp.2011.24.01.04

International Journal of Comparative Psychology, 2011, 24, 76-98.

Individual Differences in Causal Judgment under Time Pressure: Sex and Prior Video Game Experience as Predictors

Michael E. Young, Steven C. Sutherland, and James J. Cole
Southern Illinois University at Carbondale, U. S. A.

Individual differences in the effects of stress on causal attribution were studied in the context of a first-person-shooter video game. Participants were tasked with identifying the source of an explosion by repeatedly choosing among three possible enemy targets that were firing their weapons at random. In each trio of possible targets, the true enemy (the cause) produced these explosions at a delay of either 0.5, 1.0, or 2.0 seconds and with a probability of 100%, 75%, or 50%; condition varied across trios of targets. In Experiment 1, half of the participants made these choices while under stress (by being under fire by snipers in the hills surrounding the choice area) and half were not under fire. Men had higher accuracies and shorter latencies, and being under fire produced lower accuracy but had no effect on latency. In Experiment 2, a more explicit form of time pressure was used in which participants had a fixed amount of time in which to make their choice. This form of time pressure succeeded in dramatically reducing decision latency with an associated drop in accuracy. There was unreliable evidence of a higher accuracy for men. Neither experiment revealed a relationship between self-reported video game play and performance. The results suggest that causal decisions are negatively affected by time pressure, and the manipulations affected men and women similarly.

The study of the human attribution of causality has a long history in the field of psychology. Researchers have confirmed that Hume's (1739/1969) cues to causality (e.g., spatial and temporal contiguity, covariation, and temporal priority) are key determinants of the induction or perception of causality (for reviews, see Scholl & Tremoulet, 2000; Shanks, 1993; Young, 1995). These studies either involve situations in which the experimenter dictates the data available to the participant (analogous to classical conditioning) or the participant can explore and interact with the environment in order to obtain information about the extant causal relations (analogous to operant conditioning).

Despite the substantive body of evidence on human understanding of causality, there has been little exploration of either individual differences or the effect of time pressure and other environmental constraints on the process. Data is typically averaged across participants, and individual differences variables like sex and age are rarely collected or analyzed. Although there are studies reporting differential strategy use (e.g., Kao & Wasserman, 1993; Shaklee & Tucker, 1980; Young, Rogers, & Beckmann, 2005), these studies primarily document clusters of behaviors with little consideration of what gives rise to these different clusters. A handful of studies, however, have explored differences in causal judgment as a function of age, sex, and depression (note, the focus here is on adult performance; we are not considering the large number of studies of the development of causal judgment and reasoning).

This research was supported by AFOSR grant FA9550-07-1-0429 and NSF grant SES-0720588. We would like to thank Nam Nguyen, Tara Webb, Victoria Babb, Amber Blacharski, Garren King, James McCoy, and Andrew Vaz for their assistance in running the experiment.Correspondence concerning this article should be addressed to Michael E. Young, Department of Psychology, Mailcode 6502, Southern Illinois University, Carbondale, IL, 62901-6502, U.S.A. (meyoung@siu.edu).

We could identify only two lines of research explicitly attending to individual differences in causal judgment, one involving age and one involving depression. In a series of studies on the effect of aging on causal judgment (Mutter, DeCaro, & Plumlee, 2009; Mutter, Strain, & Plumlee, 2007; Mutter & Williams, 2004), Mutter and colleagues have consistently found that older adults have a more difficult time with negative contingencies (i.e., inhibitory relationships). They believe that this difficulty may be specific to problems with the identification of the importance of absent events. In studies involving mood, greater levels of depression have often been associated with higher accuracy on contingency judgment tasks (Msetfi, 2007; Msetfi, Murphy, Kornbrot, & Simpson, 2008; Msetfi, Murphy, Simpson, & Kornbrot, 2005), but there is mounting evidence that depressed individuals are instead showing an impairment in contextual processing that results in more accurate judgment under some conditions.

In one of the few studies reporting sex differences in causal judgment, Wasserman, Chatlosh, and Neunaber (1983) examined performance in a free-operant task across three experiments and only in their final experiment did they find that men showed greater sensitivity to the presence of negative contingencies than women did. The authors were at a loss as to how to explain the result given that the experiment was not substantively different from the other two experiments, but they did note that Shaklee and Hall (1983) also found a sex difference in a contingency judgment task. Wasserman, Elek, Chatlosh, and Baker (1993), however, found no sex differences in a free-operant task that involved one-minute exposure to each contingency.

Thus, most studies of contingency, covariation, and causality judgment have not examined sex differences, and the few that have only occasionally found differences. We became interested in sex and other individual differences because of their practical implications for training military personnel to make better battlefield decisions. In our laboratory, we have begun using a first-person-shooter video game design because these decisions are more similar to those faced in the field than those typically studied in causal judgment experiments (for a discussion of the advantages of microworld designs, see Brehmer, 1996; DiFonzo, Hantula, & Bordia, 1998; Gray, 2000). Game players engaged in a task that is analogous to a classical conditioning paradigm rather than the operant paradigms reported above. We have frequently found sex differences in our game-based causal decision making tasks. For example, Young and colleagues (Nguyen, Young, & Cole, 2010; Young & Nguyen, 2009) examined participants' abilities to identify the cause of explosions in the simulated environment; they reported higher accuracies and shorter latencies for men (about 10% higher and 10 s shorter) than for women. Although the sex difference in accuracies could be explained by differences in the amount of self-reported video game experience, latency differences remained even after partialling out variance due to game experience.

The current study revisited the issue of sex differences in causal decision making as a function of time pressure. Some experimental studies have noted that decision making under time pressure can reduce the accuracy of a decision, limit information use, and produce extreme judgments. In the limited study of the differential effects of time pressure on men and women, differences appear to vary as a function of the task. For example, Ibanez, Czermak, and Sutter (2009) reported that when not under time pressure, women spent more time than was optimal obtaining additional bids in a bidding game whereas men spent less time than was optimal. When subsequent mild time pressure was applied, women's performance was improved whereas men's performance was unchanged. In a study of anticipatory stress (not time pressure) on the Iowa Gambling Task (Preston, Buchanan, Stansfield, & Bechara, 2007), men's performance was superior to women's when not under stress, but women's performance was superior to men's when anticipatory stress (the expectation of giving a public speech) was introduced.

In our previous studies examining causal decision making in a video game environment, women's latencies were significantly longer than men's (Nguyen et al., 2010; Young & Nguyen, 2009). This outcome raises the possibility of two plausible outcomes for a differential effect of time pressure on causal decision making as a function of sex. First, women might actually perform better if the environment encourages faster decision making thus paralleling the behavior of men. Second, women might perform worse if their natural predisposition toward longer latencies is discouraged by the time pressure, thus producing hasty decisions.

Because some of the sex differences that we have observed in our video game (e.g., some differences in accuracy) might be accounted for by differential experience playing video games, we also collected data on self-reported amount and type of prior video game experience. Individual differences are often explained by a vague appeal to differences in pre-experimental history, so we made an attempt to evaluate this hypothesis through self-reports of an experience that might be directly related to performance in the task. Green and Bavelier (2006) have discovered that experience playing first-person-shooter video games improves both visual discrimination and attention, two skills that may affect decision making in our video game. Thus, the specific game-playing history of our participants may be predictive of decision accuracy or latency.

In our first experiment, time pressure was created in a naturalistic way by placing the participants' avatar under fire. A small number of enemies were hidden in the mountains and shot at the player at random. If the player remained stationary for significant periods of time, the likelihood of the avatar dying significantly increased (death resulted in the need to repeat the level). Participants who found the task especially difficult could try to locate "health packs" in the environment to help them finish without their avatar dying. Although the presence of health packs reduces the experienced time pressure, they helped to insure that we would not lose too many participants due to a failure to complete the game in a timely manner. In our second experiment, we created time pressure more explicitly by giving the participants a discrete time period in which they needed to make each decision.

Experiment 1

Method

Participants

A total of 82 introductory psychology students (41 men and 41 women) at Southern Illinois University at Carbondale received course credit for their voluntary participation.

Game Environment and Design

The Torque Game Engine (obtained from www.garagegames.com) was adapted as the platform for game development. Torque's first-person-shooter starter kit involves a rich world containing hills, mountains, buildings, lakes, a crossbow that shoots exploding projectiles, and orcs – monster-like characters (see Fig. 1). Using the terrain editor, a perimeter of mountains was created around the village to constrain participants from wandering outside of, while still allowing free movement within, the village. Buildings were constructed as the targets of enemy weapon fire. The critical variables modified the consequences of the orcs' weapon fire. The weapon projectiles were not visible as they traveled because this provides an all-too-obvious causal link between the cause and its effect.

The game environment contained seven separate regions, each populated by three visually identical orcs (an example of three orcs in a region is shown in Fig. 1A; the seven regions are shown in Fig. 1B from an aerial view of the terrain). For simplicity, the orcs were stationary and oriented toward a region (e.g., a building) that the player was directed to protect. Each group of three orcs was oriented toward a different building to maintain a distinctive trio. Every 4 s (on average), each orc fired its weapon (an orc's firing was noticeable from the recoil of the weapon and an audible click, although it may or may not have produced an explosion). In each 3-orc region, the firing of one of the orcs (the enemy orc) produced contingent explosions on the building. The player's task was to identify the enemy orc that was producing the explosions and destroy it. Destruction of a single orc required eight shots, as our pilot studies revealed that participants showed greater discernment under these conditions because poor choices would lengthen the game. For a video clip showing a participant observing a trio, destroying an orc (i.e., making a choice), and observing the consequences of their choice: http://www.psychology.siu.edu/bcs/facultypages/young/Research/Supplemental.html. This clip also shows a bird's eye view of the entire game region.

Once all of the enemies were destroyed, the game environment was reset with the same environment but with different programmed delays and outcome likelihoods (the probability of an explosion given that the enemy had fired) assigned to the enemies. This resetting was done three times thus producing four levels of the game. Level 1 (their initial experience in the game) contained no weapons' delays and was used as a method for orienting the participant to the game's requirements and to assess their understanding. Although Level 1 performance was noted, it was tangential to the main variables of interest. Levels 2 through 4 each used different levels of outcome likelihoods (50%, 75%, or 100%) for the enemy weapons, counterbalanced across levels. Within each level, the enemy orc in each trio either fired a weapon with a 0 s delay and 100% likelihood (the control trio) or a 0.5, 1.0, or 2.0 s delay and with the targeted outcome likelihood that was programmed for that level. The left to right location of the enemy in each trio was randomly determined for each trio for each participant. The assignment of outcome likelihood to level was counterbalanced across participants with each receiving one of the following orders for Levels 2 through 4: 50/75/100, 75/100/50, or 100/50/75.

Two versions of the game environment were used, one intended to induce time pressure and one that did not. For the time pressure condition, we added two orcs that were hidden in the mountains and shot at the participants throughout all four levels (including the orientation level). Explosions created by these snipers sounded differently from those normally produced by the crossbow (to minimize confusion with the explosions produced by enemies). Furthermore, the magnitude of the explosion was significantly decreased (to minimize visual obstruction due to smoke and debris), and the amount of damage produced was much lower and delayed (to avoid making the task too difficult by destroying the participants' avatars too early; the delay also helped them avoid damage by staying on the move rather than remaining stationary). Each orc in the mountains fired every 2 s on average (randomly chosen from the [0 s, 4 s] interval) and, if the participant was hit, caused a reduction in the participant's health by up to 3%. For the no-time-pressure condition, the orcs in the mountains did not fire. Health packs were distributed throughout each level to allow participants to increase their health if they were close to being destroyed.

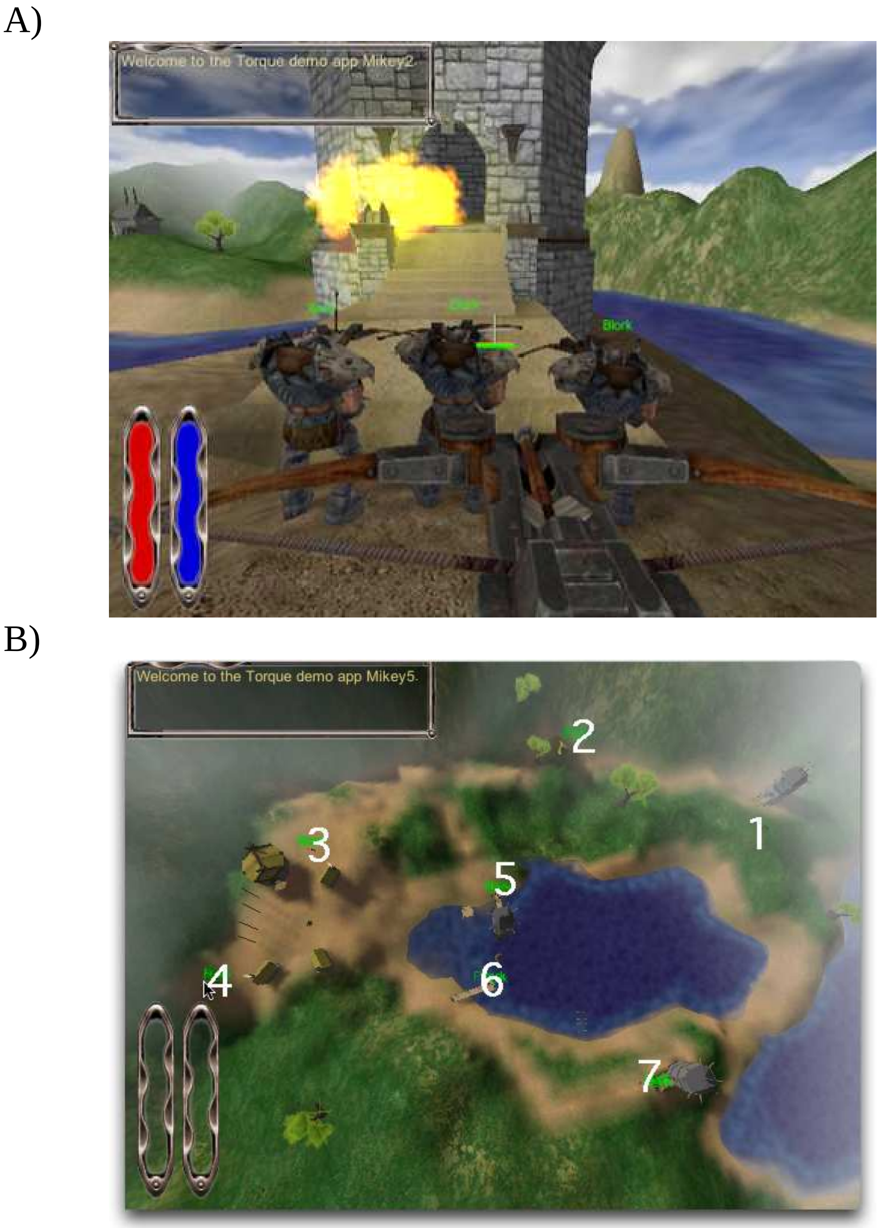


Figure 1. Screen shots from the video game used in Experiment 1. A) One trio (three orcs) is shown as well as the outcome (an explosion on the surface of a building). B) A top-down view of the relative positions of the seven trios in the game's landscape.

Procedure

The participants were seated at one of four identically configured 1.25 GHz Mac Mini computers. Each participant received verbal instructions, including a description of the task, using a reference screenshot of what they would see once the experiment was started. In addition to the task instructions, participants were advised on how to navigate within the environment. Once each participant indicated an understanding of the procedure, the experimenter started each of the programs.

Upon completion of the experiment, the participants completed a demographic questionnaire asking their sex, self-rated video game experience during elementary school, middle school, high school, and college (Likert scales ranging from “0” indicating none to “6” indicating daily), and types of video games that they play. Because a principal components analysis revealed no distinct roles for when the experience occurred, we summed across the four periods of experience to create scores that could range from 0 to 24. We performed a cluster analysis of types of video game experience to determine if participants’ behavior could be segregated categorically and whether these categories would predict performance on our tasks. Once these categories were identified, we examined whether participants in each category differed in terms of their sex distribution and amount of experience playing video games, accuracy, and latencies. We performed a K-means cluster analysis using the participants’ responses to the types of games that they played and identified three distinct clusters. To maximize our sample size, this analysis was actually performed across both experiments.

Results

All 82 participants completed the task with only four participants experiencing “death” at some point during the game. Of these four participants, three of them experienced a single death in the first level and the remaining participant experienced a single death in each of the first two levels. Given that initial level performance is not analyzed, the impact of deaths was thus negligible. In our presentation of the results, we first show the results of our cluster analysis and then examine the effects of our variables on initial choice accuracy and on the latency between the previous accurate decision and the first choice for the subsequent trio (we evaluated only the accuracies and latencies for the initial orc destroyed by the participant for each trio).

Self-reported video game responses fell primarily into three clusters (see Fig. 2). The action gamer cluster also self-reported as playing more often (M = 16 out of a maximum score of 24 on our experience scale) than the moderate and recreational gamers (Ms = 13 and 9, respectively). The action and moderate gamer clusters comprised more males than females (14 vs. 9 for action and 52 vs. 17 for moderate) whereas the recreational gamer cluster comprised more females than males (52 vs. 18).

The overall effects of our within-subject variables are shown in Figure 3A. As delay increased, initial choice accuracy decreased and latency increased suggesting that delayed causation was more difficult to discern. As likelihood decreased, initial choice accuracy at the 0.5 s and 1.0 s delays decreased, but latency increased only for the 1.0 s delays. For the longest delays, likelihood had no clear effect on accuracy or latency. Finally, placing the avatar under fire reduced accuracy but had little impact on latency.

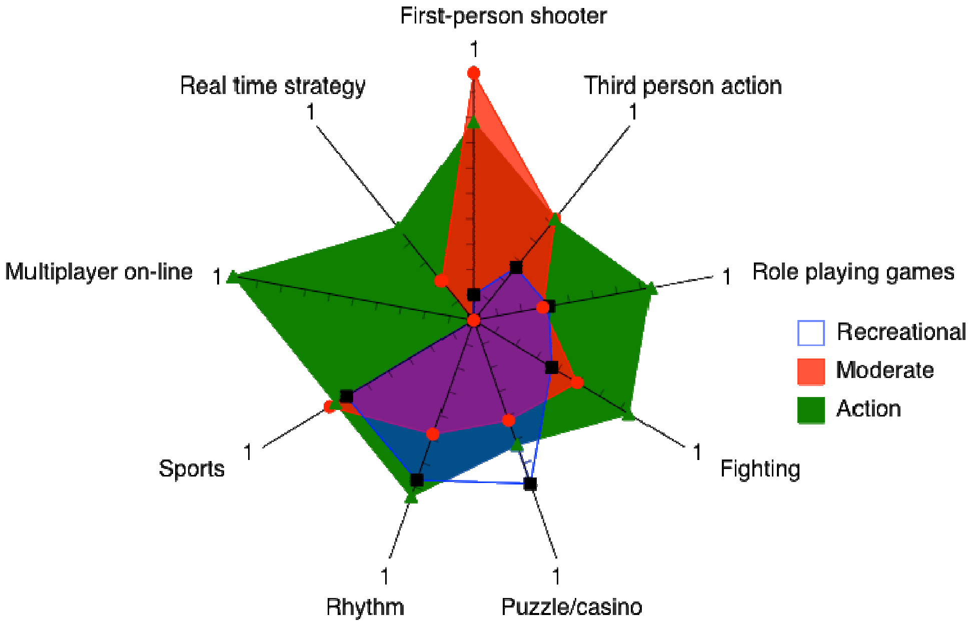
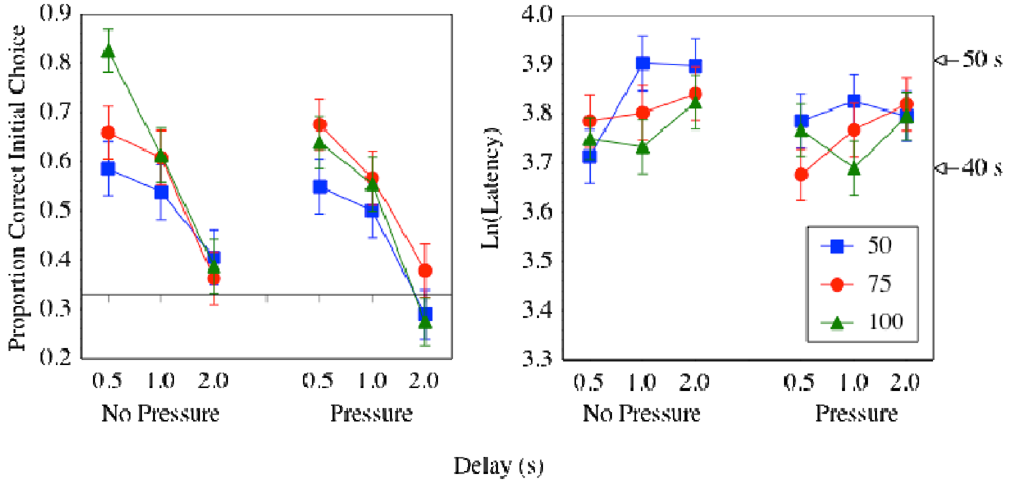


Figure 2. Cluster analysis results from Experiments 1 and 2 combined. For Experiment 1, the *N*s were 12, 34, and 33, for the action, moderate, and recreational gamers. For Experiment 2, the *N*s were 11, 35, and 37.

In our analyses, we used linear mixed effects modeling in the R statistical platform (available at http://www.r-project.org/). Mixed effects (or multilevel) modeling allows the estimation of parameters at both the group level (fixed effects) and at the individual participant level (random effects); for a more detailed discussion, see the Appendix. Individual differences in parameter estimates were allowed only when there was sufficient statistical evidence that the estimates varied significantly across participants (i.e., allowing them to vary produced a better fit). We fit a series of models that included and excluded each of our fixed effects (delay, likelihood, sex, time pressure, prior experience amount, and prior experience type) and their interactions until we discovered the model with the best fit (as assessed using the Akaike Information Criterion or AIC, Akaike, 1974; Myung, 2000; Pitt & Myung, 2002). Models could allow either the intercept to vary across participants (i.e., was the participant overall more accurate or slower) and/or the slope effects of delay and likelihood to vary across participants (i.e., did delay and/or likelihood have a larger effect on some participants than for others?). We only present the results of the best fitting model for each analysis in order to simplify presentation.

A) Experiment 1



B) Experiment 2

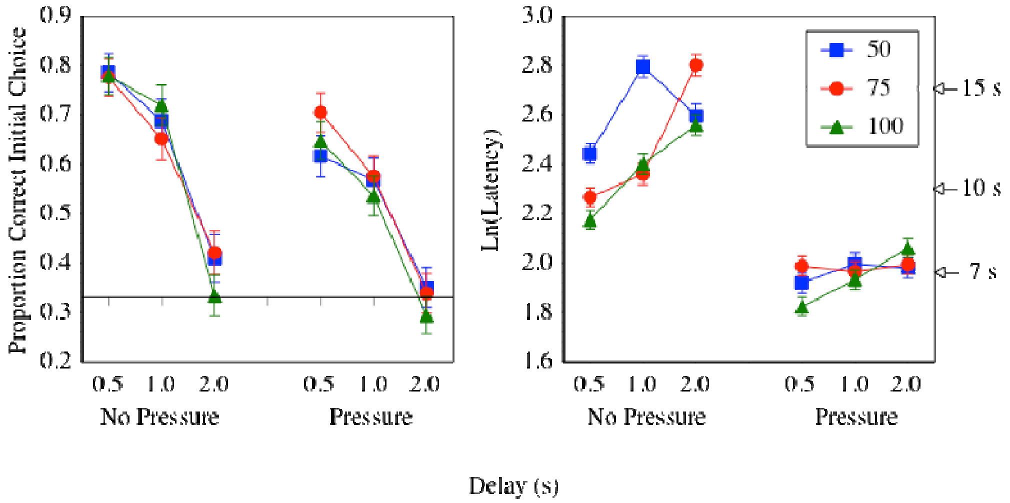


Figure 3. A) Experiment 1's mean (+/- 1 standard error) performance for the effects of delay and likelihood on initial choice accuracy (left graph) and log latency in seconds (right graph). B) Experiment 2's mean (+/- 1 standard error) performance for the effects of delay and likelihood on initial choice accuracy (left graph) and log latency in seconds (right graph). The horizontal lines in the left graphs indicate chance performance (33%).

Accuracy

For the accuracy analyses, we used a binomial error distribution (family = binomial in R's lmer) because accuracy was coded as a 1/0 variable. The best fitting model for accuracy was:

$$\text{Correct} \sim \text{Sex} + \text{TimePressure} + \text{Delay} \times \text{Likelihood}$$

Thus, the model included sex, time pressure, and the full factorial combination of likelihood and delay as fixed effects. This best model also allowed the intercepts to vary across participants (see the Appendix). Because of the binomial distribution of errors, the analysis returns z values rather than Fs. Across participants there were main effects of sex, $z = 2.84$, $p < 0.01$, with men having higher accuracy, and time pressure, $z = 2.53$, $p < 0.05$, with time pressure reducing accuracy; neither the amount nor types of prior video game experience contributed to predicting individual differences in accuracy. Within-participants, there was a main effect of likelihood, $z = 2.72$, $p < 0.01$, and delay, $z = 10.03$, $p > 0.50$, and a significant Delay × Likelihood interaction, $z = 2.24$, $p < 0.05$ (see the left graph of Fig. 3A for the nature of the interaction; delay and likelihood were centered for the interaction). Accuracy did not change across the game levels (2 through 4; the first level was not part of the analysis).

In linear mixed effects modeling, it is helpful to examine the best fitting parameter values and their standard errors to identify each predictor's effect on performance. The best fitting estimates are shown in Table 1. Because it is hard to interpret the effects of these coefficients on accuracy, it is often easier to describe variable effects when the other factors are held at their average value. Under these conditions, time pressure reduced predicted accuracy (from 59% to 52%) and men had a higher predicted accuracy than women (60% vs. 52%).

Table 1

The parameter estimates from the best-fitting mixed effects model of correct initial choices for Experiment 1.

Parameter	Estimate	Std. Error
Intercept	0.38	0.22
Sex (male)	0.34	0.12
Time Pressure	-0.31	0.12
Likelihood Slope	0.67	0.25
Delay Slope	-0.73	0.07
Likelihood × Delay Slope (each centered)	-0.80	0.36

Note: Delays varied from 0.5 to 2.0 s and likelihoods varied from 0.50 to 1.00).

Latency

We also examined the amount of time that lapsed between a correct choice (the last one) for one trio and the first choice for the next trio. This latency measure includes travel time, observation time, and firing time. It proved untenable to cleanly distinguish among these factors (e.g., some players would fire while in motion), but initial analyses revealed no systematic differences among firing rates or movement rates as a function of our independent variables. Thus, we treat these factors as contributors to error variance, not systematic variance.

For the latency analyses, latencies were first log-transformed to normalize their distribution. The best fitting model was:

$$\text{Ln(Latency)} \sim \text{Sex} + \text{Delay} + \text{Likelihood} + \text{Level}$$

Across participants there was a main effect of sex, *F*(1, 1466) = 6.52, $p < 0.01$, with men showing shorter latencies, but neither time pressure nor the video game variables were retained. Within-participants, there were main effects of likelihood, *F*(1, 1466) = 9.35, $p < 0.01$, delay, *F*(1, 1466) = 15.03, $p < 0.01$, and game level, *F*(1, 1466) = 34.04, $p < 0.01$. The intercepts (i.e., overall speed of choice) and the level effect (i.e., slope effect of level on latency) were allowed to randomly vary across participants. Latencies decreased across game levels (*Mdns* = 47, 42, and 40 s, for game levels 2 through 4, respectively) with the rate of decrease varying across participants.

The best fitting estimates are shown in Table 2. Because it is hard to interpret the effects of these coefficients on log latency, we translated these effects into their impact on latency on the original scale while holding the other factors constant. Men's latencies were shorter than those of the women (41 vs. 47 s); the effects of likelihood and delay are illustrated in the right graph of Fig. 3A with untransformed labels on the right side of the plot. We also examined the smoothed frequency distributions (density plots, Fig. 4A) for the latencies to determine if there were any differences in the nature of the latency distribution as a function of time pressure or sex. The distributions of the log latencies had similar variances and symmetry across these variables suggesting that the time pressure and sex conditions differed only in the central tendency of the distributions.

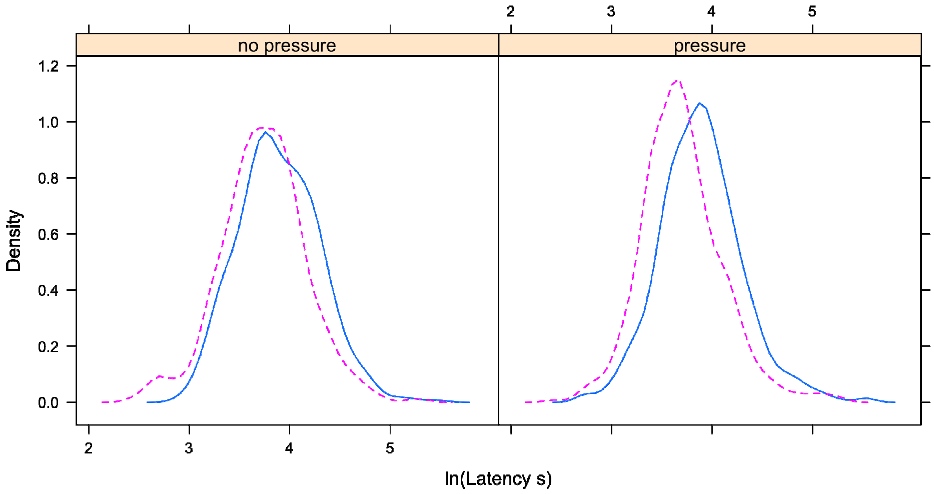
Table 2

The parameter estimates from the best-fitting mixed effects model of Ln(Latency) for Experiment 1.

Parameter	Estimate	Std. Error
Intercept	4.14	0.07
Sex (male)	-0.13	0.05
Likelihood Slope	-0.15	0.04
Delay Slope	0.05	0.01
Level Slope (2 through 4)	-0.08	0.01

Note: Delays varied from 0.5 to 2.0 s and likelihoods varied from 0.50 to 1.00).

A) Experiment 1



B) Experiment 2

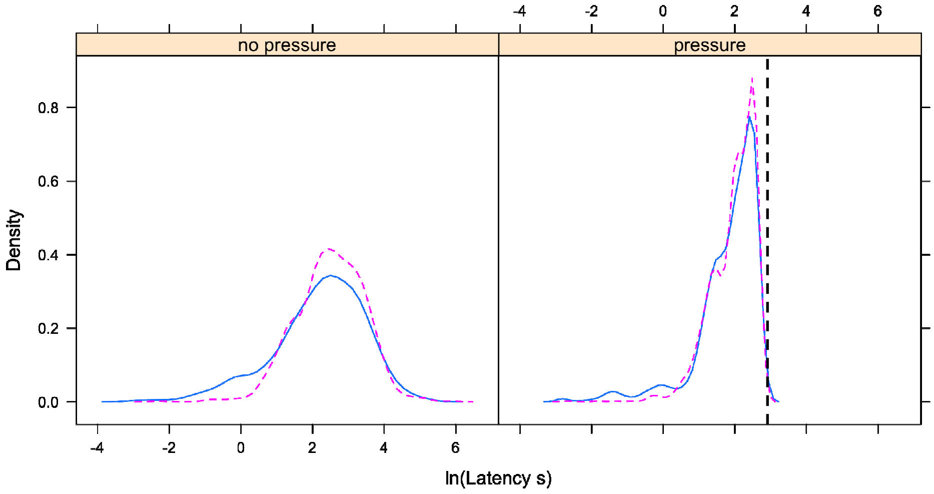


Figure 4. Density plots showing the distribution of latencies for men (dashed lines) and women (solid lines) when under pressure in each of the two experiments. Note the difference in scaling for the x-axes for Experiments 1 and 2. The vertical dashed line in the right figure indicates the maximum allowable latency for the time pressure condition; the curve in the pressure condition continues past this limit is due to the smoothing algorithm.

Discussion

Longer cause-effect delays produced much lower accuracies and longer latencies, and lower likelihoods had modest detrimental effects on each measure (see Fig. 3A). Although our attempt to create time pressure by placing the avatar under fire produced lower accuracies, it did not succeed in producing shorter latencies. Given that participants did not make faster decisions, the manipulation may have produced a different form of stress than the one intended. Indeed, an examination of the way that participants navigated the game environment (not shown here but available from the first author) reveals that participants often responded to the enemy sniper fire by constantly moving around in order to avoid being shot. This more erratic behavior likely resulted in deteriorated observations of the cause-effect relations that may have led to the lower accuracy while under enemy fire.

Our analyses revealed that women had lower accuracies and longer latencies than men, replicating the results of Young and Nguyen (2009). In contrast to this earlier report, however, none of the difference could be explained by an appeal to differences in the amount or type of self-reported video game experience. Additionally, Green and Bavelier (2003) documented an advantage for first-person-shooter experience on visual attention tasks whereas our participants with self-reported high experience with these types of video games (the action and moderate gamers, see Fig. 2) did not show an advantage in our causal judgment task. Finally, the participants showed some improvement in their performance as the game proceeded; latencies decreased across the final three game levels. Thus, practicing the identification of causes may have produced an overall benefit to the game players.

The finding that our attempt to create time pressure had effects other than those intended (e.g., erratic maneuvering) leaves unaddressed the question of how time pressure affects causal decision making in our paradigm. We believed that quicker decisions would result in fewer cause-effect instances being observed thereby creating a detrimental effect on performance. Furthermore, we were interested in whether men and women would react differently in the face of this time pressure given the natural predisposition for women to wait longer when not under pressure. Thus, Experiment 2 used a more direct manipulation of time pressure in that participants would have a discrete amount of time in which to make a decision before being placed in a “penalty box” for a fixed duration before being released to continue the task.

Furthermore, our measure of latency includes travel time between decision points, the amount of time that the candidate causes and the effect are observed, and the time to complete the eight shots necessary to destroy the chosen orc. Although unreported analyses indicate that our independent variables are not related to travel rate nor firing rate, in Experiment 2 we addressed the issue more directly by placing the decision points within buildings thus requiring the player to enter before being able to observe the cause-effect interactions. By indexing latency to building entry, travel time between the buildings no longer contributes to our measure. Furthermore, we only required one shot to destroy an orc thus eliminating firing time as a factor in measuring latency. This change necessitated a different form of disincentive for hasty decisions – participants who made incorrect decisions had their avatar placed in the same “penalty box” that was used when the time constraint was exceeded, and they received a longer penalty time for incorrect decisions than they did for not having made a decision.

Experiment 2

Method

Participants

A total of 83 introductory psychology students (45 men and 38 women) at Southern Illinois University at Carbondale received course credit for their voluntary participation.

Design

The task in Experiment 2 was similar to the task in Experiment 1 (i.e., selecting the true “enemy” orc from among three possible orcs); however, there were several changes made to the game environment and the contingencies. The game environment in Experiment 1 contained seven separate regions with each region populated by three visually identical orcs; however, for Experiment 2, the seven regions were replaced by nine individual round huts, each containing a stationary trio of visually identical orcs. The buildings were equally spaced and arranged in a circle (an example of the trio of orcs as viewed within a building is shown in Fig. 5A; Fig. 5B shows an aerial view of the circle of buildings). As in Experiment 1, participants were free to select the order in which they approached each trio. Rather than being oriented towards a building that they fired upon, the orcs in Experiment 2 were oriented towards an “energy crystal” at the rear of the building that they fired upon. A trio of orcs did not fire their weapons until the participant’s avatar entered the building in which the trio was located. In order not to obstruct the participant’s view of the explosions, no orc was placed directly between the crystal and the entryway. Instead, the orcs in each hut were randomly assigned to three of four possible positions, two on either side of the room. Additionally, to eliminate the possibility of firing into a hut from outside, the participant’s weapon was disabled when they were not inside of a building. Upon entering a building, a door closed behind the participant’s avatar, the trio of orcs began firing, and the participant’s weapon was enabled. Unlike the eight shots required to destroy an orc in Experiment 1, in Experiment 2 destroying an orc required only a single shot.

Participants in the no-time-pressure condition were permitted to observe the orcs’ weapons’ fire for as long as they wished. If the participant destroyed the correct orc, the door to the building opened and the participant was permitted to move on to the next building. If, however, the participant destroyed an incorrect orc, the two remaining orcs disappeared (participants were not required to return to the building to select between the two remaining orcs), the door to the building opened, and the participant was instantly teleported to a “penalty box” for 60 s. At the end of the 60 s penalty, participants were teleported back to the center of the circle of buildings and were permitted to continue. The penalty box was a hole in the terrain, designed to restrict the movement of the participant’s avatar and to provide limited visual stimulation. These penalties were meant to punish incorrect decisions by increasing the total task time for poorer performance and replaces the multiple-shot contingency used in Experiment 1.

In the time pressure condition, correct and incorrect decisions were met with the same contingencies as in the no time pressure condition; however, participants were given only 15 s to make a decision. Participants were given a visual cue for their time limit in the form of a time bar in the lower left side of the screen (the right bar in Fig. 5B) that decreased at a steady rate over the 15 s. Additionally, an auditory cue was provided in the form of three beeps when the participants had only 3 s remaining to make their decision. If the participant failed to make a decision after 15 s, the trio of orcs remained in the building (participants were required to return to the building to make a selection), the door to the building opened, and the participant was instantly teleported to a “penalty box” for 30 s. At this time, the identity of the enemy in the given hut was re-randomized to prevent participants from accumulating observation data over several visits to that building. At the end of the 30-second penalty, the participant was teleported back to the center of the circle of buildings and allowed to continue. This penalty was meant to punish not making a selection, but it was meant to be a lesser punishment than that for an incorrect decision as we did not want to reinforce faster incorrect decisions that might be made to hurry through the experiment.

Figure 5. Screen shots from the video game used in Experiment 2. A) One trio (three orcs) is shown inside of a building. B) A top-down view of the relative positions of the nine buildings in the game's landscape.

Once all of the orcs were destroyed, the game environment was reset with the same environment but with different programmed delays and outcome likelihoods assigned to the orcs. This resetting was done four times thus producing five levels of the game (Experiment 1 contained only four levels). Level 1 was again used as an orienting level and contained no weapons' delays and 100% outcome likelihoods for all nine orc trios. Experiment 2 contained the same likelihoods (50%, 75%, or 100%) and delays (0.5, 1.0, or 2.0 s) as Experiment 1. Whereas the first experiment programmed delays to vary within level and likelihoods to vary between levels, the current experiment varied delay and likelihood within each level (three delays × three likelihoods producing nine types of trios, with one assigned randomly to each building). The intent of this change was to allow for a cleaner test of behavioral change across levels that would not be partially confounded by one of the independent variables. The left to right location of the orc in each trio was again randomly determined for each trio for each participant.

Procedure

The procedure for Experiment 2 was identical to that in Experiment 1, except that a separate K-means cluster analysis was not conducted. Instead, participants were included in a cross-experiment analysis as discussed earlier. This action was taken to maintain consistency between the two experiments in the classification of prior video game play.

Results

All 83 participants completed the first three levels of the task. One male participant (in the no pressure condition) only completed three levels and seven females only completed the first four levels (three of these were in the time pressure condition). Given that all variables were manipulated within level, all participants were retained regardless of completion. The distribution of participants across the three clusters was very similar to that for Experiment 1 (see Fig. 2).

The overall effects of our within-subject variables are shown in Figure 3B. As delay increased, initial choice accuracy decreased and latency increased (for more strongly for the no pressure condition). As likelihood decreased, initial choice accuracy appears largely unaffected with some indication of longer latencies for some of the lower likelihoods, but only when there was no time pressure.

Accuracy

The best fitting model for accuracy was:

$$\text{Correct} \sim \text{TimePressure} + \text{Delay} + \text{Level}$$

Unlike in Experiment 1, there were no significant effects of sex or likelihood, and performance varied across the last four game levels. This best model also allowed the intercepts and the delay/level slopes to vary across participants. Across participants there was a main effect of time pressure, $z = 2.37$, $p < 0.05$, with time pressure reducing accuracy; neither prior video game experience (amount or type) nor sex contributed to predicting individual differences in accuracy (note that there was a difference that is consistent with that observed in Experiment 1, *M*s = 58% vs. 51%, for men and women, respectively). Within-participants, there were main effects of delay, $z = 9.56$, $p < 0.01$, and game level, $z = 3.32$, $p < 0.01$. Accuracy increased across game levels (47%, 52%, 53%, and 59%, for levels 2, 3, 4, and 5, respectively). The best fitting estimates for the model are shown in Table 3. When the other factors were held constant, time pressure reduced predicted accuracy (from 62% to 53%).

Table 3
The parameter estimates from the best-fitting mixed effects model of correct initial choices for Experiment 2.

Parameter	Estimate	Std. Error
Intercept	1.04	0.22
Time Pressure	-0.37	0.15
Delay Slope	-1.03	0.11
Level Slope (2 through 5)	0.14	0.04

Note: Delays varied from 0.5 to 2.0 s and likelihoods varied from 0.50 to 1.00).

Latency

We also examined the amount of time that lapsed between entry into each hut and the first choice. Unlike the latency measure used in Experiment 1, this latency measure includes little travel time (only movements that occurred within the hut), observation time (capped at 15 s in the pressure condition), and no firing time (because only one shot destroyed an orc). Thus, these latencies are much shorter than those reported in Experiment 1 (*Mdns* = 44 s and 9 s in Experiments 1 and 2, respectively).

For the latency analyses, latencies were again log-transformed to normalize their distribution. The best fitting model was:

$$\mathrm{Ln(Latency)} \sim \mathrm{Sex} + \mathrm{TimePressure} + \mathrm{Delay} \times \mathrm{Likelihood}$$

Across participants there were main effects of sex, $F(1, 2675) = 3.10$, $p < 0.05$, with women showing shorter latencies, and time pressure, $F(1, 2675) = 5.28$, $p < 0.01$, with time pressure producing shorter latencies. The video game experience variables were again not retained. Within-participants, there were main effects of likelihood, $F(1, 2675) = 10.29$, $p < 0.01$, delay, $F(1, 2675) = 50.52$, $p < 0.01$, and a Delay × Likelihood interaction, $F(1, 2675) = 5.56$, $p < 0.05$ (the interaction terms were again centered). There was no evidence of an effect of game level (2 through 5). The intercepts (i.e., overall speed of choice) and the level effect (i.e., slope effect of level on latency) varied across participants. The lack of a game level fixed effect along with its within-participant variance indicates that participants showed no general level trend, but some individual participants showed systematic increases in latency whereas others showed systematic decreases.

The best fitting estimates are shown in Table 4. When holding the other factors constant, women's latencies were shorter than those of the men (*Mdns* = 8.5 vs. 9.2 s), and latencies were shorter under time pressure than not (*Mdns* = 7.9 vs. 11.8 s). The effects of likelihood and delay are illustrated in the right graph of Figure 3B with untransformed labels on the right side of the plot.

We again examined smoothed distribution plots (density plots, Fig. 4B) for the latencies to determine if there were any differences in the nature of the distribution as a function of time pressure or sex. There was less variance in the latencies for men (dashed line) and in the latencies under time pressure, and women showed a more pronounced skew with a disproportionate number of very short latencies. The smoothed stair-steps in the curves are produced by participants who make decisions after the first cause-effect observation cycle (complete by the end of 4 s) or after the second observation (complete by the end of 8 s); the last two cycles are too close to distinguish on the logarithmic scale (ln(12) vs. ln(16) = 2.5 vs. 2.7).

Table 4
The parameter estimates from the best-fitting mixed effects model of Ln(Latency) for Experiment 2.

Parameter	Estimate	Std. Error
Intercept	2.32	0.15
Sex (male)	0.27	0.15
Time Pressure	-0.53	0.15
Likelihood Slope	-0.24	0.07
Delay Slope	0.14	0.02
Likelihood × Delay Slope	0.24	0.10

Note: Delays varied from 0.5 to 2.0 s and likelihoods varied from 0.50 to 1.00).

Given that the finding that women had shorter latencies was unusual, we examined the result at the individual participant level and discovered that the skew was driven by four female participants, two in the pressure condition and two in the no-pressure condition. These participants had median latencies of less than 1.5 s, a latency that generally can only be achieved by making a decision almost immediately upon entry into the building. These four participants also showed chance level performance (*Ms* = 24%, 32%, 34%, and 42%), further suggesting that their short latencies were due to inattention to the task. When these four participants were eliminated from the analysis, the differential skew and variability in the density plots was eliminated. Furthermore, rerunning the mixed effects analysis without these four participants eliminated the sex difference in latencies (the parameter estimate for being male decreased from 0.27 to 0.02, *cf*. Table 4) without producing a sex effect for accuracy even though the four worst-performing females had been dropped.

In a final analysis, we examined the other side of the latency question by assessing sex and video game play as predictors of the time it took for a participant to move from one decision point to the next. The design of Experiment 2 permitted a segregation of latencies that was not possible in Experiment 1. We only analyzed those latencies that followed a correct decision because incorrect decisions resulted in transport to the penalty box that would confuse interpretation of the measure. Using a standard regression with sex, experience amount, and experience type in a full factorial model predicting log(travel time), only sex was retained as a predictor, $F(1, 74) = 35.90$, $p < 0.01$, with men having shorter travel times than women (11.6 vs. 17.6 s, respectively). This result suggests that at least part of the sex difference in latencies in Experiment 1 might have been due to travel time disparities. Surprisingly, game experience (type or amount) again had no consistent relationship with this measure. We must be cautious in generalizing the results to Experiment 1, however, because the layout of the game environment was quite different in the two experiments.

Discussion

When we used a more explicit form of time pressure, accuracy again decreased and the expected large decrease in latencies was observed. Figure 4B includes a dashed line showing the maximum allowable latency in our time pressure condition. Given that roughly half of the latencies exceeded this threshold when there was no time pressure (as shown in the no pressure density plot in Fig. 4B), this contingency clearly disrupted the participants' normal observation behaviors.

The principal difference between the effects of the within-subject factors in Experiments 1 and 2 was the weaker effect of changing the likelihood on choice accuracy in Experiment 2 (see Fig. 3B). The maximum observation time of 15 s (four observations of the effect) appears sufficient to identify a cause that produces the effect 50%, 75%, or 100% of the time. The detrimental effect of lower likelihoods in Experiment 1 may thus have been due to the more erratic avatar movement produced by placing the avatar under fire.

Although our initial analyses indicated a possible sex difference in latencies, but not accuracies, a closer examination of the results revealed that this advantage was due to the inclusion of four women who showed inattention to the task (latencies less than 1.5 s and chance levels of accuracy). We also discovered that men showed shorter travel times between decision points. Interestingly, the average accuracies for men and women in both experiments were quite similar (*M*s = 60% and 52% for Experiment 1 and 58% and 51% for Experiment 2). The fact that this difference reached statistical significance in Experiment 1 but not Experiment 2 is due to the lower variability across participants in the first experiment, not due to sample size differences.

Finally, we observed a learning effect in accuracy, not latency, the opposite of what we observed in Experiment 1. We believe it likely that the latency decrease observed in Experiment 1 was primarily driven by faster navigation of the environment as the experiment proceeded. In Experiment 2, the decision points were all near one another and thus required fewer navigation skills (see Fig. 5). It is not clear, however, why accuracy did not increase across levels in Experiment 1 but did so in Experiment 2 except to note that likelihoods varied across levels in Experiment 1 but not Experiment 2.

General Discussion

The present study extends the findings of Young and Nguyen (2009) by generalizing their results to situations involving two different forms of pressure, (a) being under fire and (b) explicit time constraints. Furthermore, women showed lower accuracy than men both when not under time pressure (replicating Young & Nguyen, 2009) and when under time pressure, although the latter effect failed to reach significance when the pressure involved an explicit time limit. In none of our analyses did self-reported prior experience playing video games (amount or type) predict differences in accuracy or latency (*cf*. Green & Bavelier, 2003, 2006). Green and Bavelier reported that first-person-shooter (fps) game experience was the key predictor of differential performance in their visual spatial task whereas our clusters included more than just fps experience. However, our three clusters differed significantly in terms of fps play (Fig. 2) and an analysis of game play that only included fps, not cluster, likewise was not significant (not shown). Thus, our measures of pre-experimental history with these types of games failed to shed light on individual differences; this outcome may be due to the absence of a relationship or due to the inherent unreliability of self-reports. It is also possible that giving participants controlled amounts of experience with an fps game before performing the task might improve performance in the same way it did for Green and Bavelier's (2007) and Feng, Spence, and Pratt's (2007) participants.

We had hypothesized two possible outcomes for sex differences in the presence of time pressure – either women would actually perform better if the environment encouraged faster decision making, thus paralleling the behavior of men, or women might perform worse if their natural predisposition toward longer latencies was discouraged. There was no indication of a Sex × Pressure interaction in either experiment thus indicating that our time pressure manipulations had similar effects for men and women.

Because video game play was not related to performance, we remain ignorant regarding the source of the frequently-observed sex differences in our paradigm. Are men at an advantage for causal decision making due to an environmental or biological edge? If so, we would have expected more reports of sex differences in the causality literature. Does success in our task require a greater tolerance of risk, suggesting that the sex difference is related to the oft-observed propensity for men to take more risks than women (e.g., Hudgens & Fatkin, 1985; Levin, Snyder, & Chapman, 1988; Powell & Ansic, 1997) or to better strategies for handling task uncertainties (Reavis & Overman, 2001)? Or, perhaps, is first-person-shooter video game play inherently more motivating for men than for women regardless of our failure to find effects of prior video game experience? The observation that 10% of our female participants effectively seemed to give up in Experiment 2 by making decisions without observation may reflect a lack of motivation to play the game on the part of some of the women in our sample.

Our participants were met with uncertainty in the face of multiple causal candidates occurring in a complex dynamic environment. This uncertainty was created by longer cause-effect delays and lower outcome likelihoods. We believe that people faced with discriminating among possible causes will choose to minimize uncertainty by increasing their observation time, but only to a limit (*cf.* Hausmann & Läge, 2008; Johnson & Payne, 1985; Lee & Cummins, 2004). In our task, participants were allowed to make the trade-off between choosing more quickly and, potentially, less accurately versus obtaining more data in the form of additional observations of the candidate causes and the effect in an attempt to increase accuracy. Even in Experiment 2, when there was a maximum of 15 s of observation, most decisions were made near the end of this interval (see Fig. 4B). The natural motivation to complete the task in a reasonable amount of time would produce a desire to make fast decisions, but hasty incorrect decisions resulted in wasted time while a participant makes the additional shots necessary to destroy another orc (Experiment 1) or is forced to sit in the penalty box (Experiment 2). In the presence of short cause-effect delays (0.5 s), high accuracy was achievable regardless of shorter observation times, but in the presence of long delays (2.0 s) accuracy was at chance regardless of the longer observation time (see Fig. 3A and 3B). Apparently, the additional time required to obtain high certainty with long cause-effect delays must have been deemed too costly – it was easier to guess and try again.

Epilogue

The sex differences in our video game task have raised a number of questions regarding their source. A researcher's natural disposition is that these differences are due to factors other than some inherent biological difference between men and women. Perhaps our use of a video game task revealed sex differences due to the high level of uncertainty in the environment (thus penalizing women for being risk averse) or due to differences in the motivation to play. To address these issues, it is necessary to compare the sexes in other causality tasks that use different preparations. Unfortunately, despite the large body of research on causal judgment and perception, very few publications have reported the necessary data to answer this question. Although the small amount of data available (e.g., Shaklee & Hall, 1983; Wasserman et al., 1983; Wasserman & Neunaber, 1986) suggest the possible presence of sex differences in non-gaming causality tasks, we cannot discern how much of the difference in the gaming tasks might be due to a biological difference versus differences in other types of prior experience, motivation, or cultural expectations.

Our own interest in sex differences arose because of its practical implications for the types of decisions that military personnel face when in the battlefield. These pragmatic considerations extend beyond a mere cataloging of these differences in order to select personnel for differential duty to the identification of the source of these differences in order to determine if personnel improve their decision making with variations in training regimens. The practical advantages of our video game approach are that the decisions are more similar to those faced in the field, and the platform may provide a more engaging environment for extended training on causal decision making. We are anxious to see whether these benefits can be realized.

References

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

Brehmer, B. (1996). Man as a stabiliser of systems: From static snapshots of judgement processes to dynamic decision making. *Thinking and Reasoning, 2,* 225-238.

DiFonzo, N., Hantula, D. A., & Bordia, P. (1998). Microworlds for experimental research: Having your (control and collection) cake, and realism too. *Behavior Research Methods, Instruments, and Computers, 30*, 278-286.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science, 18,* 850-855.

Gray, W. D. (2000). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and microworlds in basic and applied cognitive research. *Cognitive Science Quarterly, 2*, 205-227.

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*, 88-94.

Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception & Performance, 32,* 1465–1478.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534–537.

Hausmann, D., & Läge, D. (2008). Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition. *Judgment and Decision Making, 3*, 229-243.

Hudgens, G. A., & Fatkin, L. T. (1985). Sex differences in risk taking: Repeated sessions on a computer-simulated task. *Journal of Psychology: Interdisciplinary and Applied, 119*, 197-206.

Hume, D. (1739/1969). *A treatise of human nature*. New York: Penguin Books.

Ibanez, M., Czermak, S., & Sutter, M. (2009). Searching for a better deal: On the influence of team decision making, time pressure and gender. *Journal of Economic Psychology, 30*, 1-10.

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management Science, 31*, 395-414.

Kao, S., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1363-1386.

Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review, 11*, 343-352.

Levin, I. P., Snyder, M. A., & Chapman, D. P. (1988). The interaction of experiential and situational factors and gender in a simulated risky decision-making task. *Journal of Psychology: Interdisciplinary and Applied, 122*, 173-181.

Msetfi, R. M. (2007). Depressive realism and the effect of intertrial interval on judgements of zero, positive, and negative contingencies. *Quarterly Journal of Experimental Psychology, 60*, 461-481.

Msetfi, R. M., Murphy, R. A., Kornbrot, D. E., & Simpson, J. (2008). Impaired context maintenance in mild to moderately depressed students. *Quarterly Journal of Experimental Psychology, 62*, 653-662.

Msetfi, R. M., Murphy, R. A., Simpson, J., & Kornbrot, D. E. (2005). Depressive realism and outcome density bias in contingency judgments: The effect of the context and intertrial interval. *Journal of Experimental Psychology: General, 134*, 10-22.

Mutter, S. A., DeCaro, M. S., & Plumlee, L. F. (2009). The role of contingency and contiguity in young and older adults' causal learning. *Journal of Gerontology: Psychological Sciences, 64B*, 315–323.

Mutter, S. A., Strain, L. M., & Plumlee, L. F. (2007). The role of age and prior beliefs in contingency judgment. *Memory and Cognition, 35*, 875-884.

Mutter, S. A., & Williams, T. W. (2004). Aging and the detection of contingency in causal learning. *Psychology and Aging, 19*, 13-26.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*, 190-204.

Nguyen, N., Young, M. E., & Cole, J. J. (2010). The effect of number of options on choices involving delayed causation. *American Journal of Psychology, 123,* 477-487.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*, 421-425.

Powell, M., & Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology, 18*, 605-628.

Preston, S. D., Buchanan, T. W., Stansfield, R. B., & Bechara, A. (2007). Effects of anticipatory stress on decision making in a gambling task. *Behavioral Neuroscience, 121*, 257-263.

Reavis, R., & Overman, W. H. (2001). Adult sex differences on a decision-making task previously shown to depend on the orbital prefrontal cortext. *Behavioral Neuroscience, 115*, 196-206.

Scholl, B. J., & Tremoulet, P. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*, 299-309.

Shaklee, H., & Hall, L. (1983). Methods of assessing strategies for judging covariation between events. *Journal of Educational Psychology, 75*, 583-594.

Shaklee, H., & Tucker, D. (1980). A rule analysis of judgments of covariation between events. *Memory and Cognition, 8*, 459-467.

Shanks, D. R. (1993). Human instrumental learning: A critical review of data and theory. *British Journal of Psychology, 84,* 319-354.

Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learning and Motivation, 14*, 406-432.

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory & Cognition, 19,* 174-188.

Wasserman, E. A., & Neunaber, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior, 46*, 15-35.

Young, M. E. (1995). On the origin of personal causal theories. *Psychonomic Bulletin and Review, 2*, 83-104.

Young, M. E., & Nguyen, N. (2009). The problem of delayed causation in a video game: Constant, varied, and filled delays. *Learning and Motivation, 40*, 298-312.

Young, M. E., Rogers, E. T., & Beckmann, J. S. (2005). Causal impressions: Predicting when, not just whether. *Memory and Cognition, 33,* 320-331.

Appendix

In our model fits, we used R to perform linear mixed effects modeling of the data. The lmer function in the lme4 package was used for all fits. The general form of the command is:

lmer(*dv* ~ *iv1* + *iv2* + *iv1*:*iv2* + (*iv1* + 1 | *Participant*), {family=binomial}
data=*data*)

in which the user's data has been loaded from a comma-delimited file exported from our standard statistics software into a data frame variable (designated as *data* above) using R's "read.csv" command. This data frame includes all of the experiment variables as columns – the independent and dependent variables, the participant identifier for each row of data, and any characteristics of the participants (e.g., their sex). If the dependent variable was binomial (in our case, the "Correct" column), then lmer's default Gaussian distribution was overridden using the "family=binomial" argument.

Fixed effects predictor variables are included as predictors on the right side of the tilde (note that *iv1*:*iv2* indicates an interaction between *iv1* and *iv2; iv1**iv2* is shorthand for including both the main effects of *iv1* and *iv2* and the $iv1 \times iv2$ interaction). Random effects within-subject predictor variables that are assumed to vary across participants can be included within the parenthetical expression and listed as nested within participant using the "| *Participant*" notation. In the example, *iv1* is specified as a within-subject variable with an effect magnitude that may vary across participants. Random effect coefficients represent differences between the average effect of the variable and the effect for each participant. Thus, allowing variations in the intercept ("1 | Participant") permits each participant to have their own individual adjustment in the best intercept for them relative to the best intercept for the group of participants as a whole. The same is true for the effect of an independent variable. For example, if the best fitting group slope for delay was -0.08 and intercept was -0.33 and each was allowed to vary across participants, then a participant with a random effect of delay of -0.21 and intercept of 0.20 would have a best fitting slope of -0.08 – 0.21 = -0.29 and an intercept of -0.33 + 0.20 = -0.13.

For interaction terms in regression and mixed effects modeling, it is recommended that the variables be centered (i.e., that the mean for each variable is subtracted before multiplying the predictors) to insure minimal or zero correlations between the interaction terms and the main effects. When the interaction term was included in a model, we centered the variables in that term but did not center the main effects in order to improve interpretability of the regression weights.

Linear mixed effects modeling allows the testing of a range of assumptions regarding which predictors to include and whether the effects of the within-subject predictor variables should be allowed to vary across participants. Allowing these additional degrees of freedom will always reduce error, so this reduction in error must be sufficient to offset the lack of parsimony. The Akaike Information Criterion was used to penalize model flexibility.