

Waiting to decide helps in the face of probabilistic uncertainty but not delay uncertainty

Michael E. Young · Steven C. Sutherland ·
James J. Cole · Nam Nguyen

Published online: 16 November 2010
© Psychonomic Society, Inc. 2010

Abstract A first-person shooter video game was adapted for the study of causal decision making within dynamic environments. Participants chose which of three potential targets in each of 21 groups was producing distal explosions. The source of the explosion effect varied in the delay between the firing of its weapon and its effect (0.0, 0.5, 1.0, and 2.0 s), the probability that the weapon caused the effect (50%, 75%, and 100%), and the occurrence of auditory events that filled the delay. In [Experiment 1](#), participants' choice accuracy was highest with short delays but was not affected by probability; participants often compensated for lower probability by increasing their latencies, and thus the number of outcomes sampled. In [Experiment 2](#), a broad range of delays (0–2 s) and probabilities (20%–100%) were randomly sampled for each cause; the results largely replicated those of the prior experiment. The experiments demonstrate people's ability to successfully modulate their environmental sampling in the face of uncertainty due to lower cause–effect probabilities, but not in the presence of longer cause–effect delays.

Keywords Casuality · Delay · Probability · Inference · Video game

In our latest research on the nature of causal decision making, we chose a preparation that allows for the experimenter to retain some control over the presentation of candidate causes and their effects but also cedes some of this control to the participant. Allowing variation within an experiment has recently been demonstrated to improve the

generalizability of results across replications (Richter, Garner, Auer, Kunert, & Würbel, 2010). Traditional experimental design has emphasized high degrees of experimental control in order to increase internal validity while sacrificing external validity. At some point in a research program, however, it is necessary to determine whether commonly accepted laboratory results generalize to other situations (Young, 2010). For example, it is widely accepted that causal judgments between stimuli are adversely affected by cause–effect delays or by a decreased probability of the effect given the cause. In less structured environments, will these relationships continue to hold? Might an organism adapt its behavior to successfully compensate for some types of uncertainty, but not others?

In most studies of causal choice, the flow of information is manipulated by the experimenter. The data are provided to the decision maker in a series of mini-experiments in which candidate causes are or are not present, occur at programmed times relative to the effect, and have probabilities designed such that the target probability is achieved at the end of the experiment (see, e.g., Allan, 1993; Shanks, Darby, & Charles, 1998; Van Hamme & Wasserman, 1994; White, 1995; Young, Wasserman, Johnson, & Jones, 2000). These approaches are analogous to the study of classical conditioning, in which conditioned and unconditioned stimuli follow preestablished programs of presentation (Pavlov, 1927). Other approaches have adopted an operant conditioning design in which the participant determines whether to make a response and subsequently observes the consequences of the range of responses available (e.g., Dickinson & Shanks, 1995; Shanks, 1993; Wasserman & Neunaber, 1986). Giving up control over the occurrence of the candidate causes, however, creates different experiences across participants, limits the study of multiple causation (because most participants will attempt to isolate the effects of individual causes), and complicates interpretation when

M. E. Young (✉) · S. C. Sutherland · J. J. Cole · N. Nguyen
Department of Psychology, Southern Illinois University,
281 LSH, Mailcode 6502,
Carbondale, IL 62901, USA
e-mail: meyoung@siu.edu

delays are present (e.g., Buehner, 2005; Hammond & Paynter, 1983; Shanks & Dickinson, 1991; Shanks, Pearson, & Dickinson, 1989; Wasserman & Neunaber, 1986).

The purpose of this study was to address the questions both of generalizability of prior results to a less-controlled environment and of possible compensatory behavior in the face of uncertainty due to delayed and probabilistic causation. Our participants were allowed to decide how long to observe the requisite events, from which angles, in which order, and from what distance (cf. Richter et al., 2010; Richter, Garner, & Würbel, 2009). Once satisfied, a participant could then make a choice among the causal candidates. How participants chose to solicit information—specifically, how long they observed before making a decision—was thus a crucial determinant of causal candidate selection.

In many causal decisions, there exist two sources of uncertainty—the delay between the cause and its outcome and the probability of the outcome given the cause. The presence of significant delays or lower effect probabilities decreases causal judgments (for reviews, see Shanks, 1993; Young, 1995). Delays increase the probability that a spurious event will be more proximal to the effect of interest than the cause. Lower effect probabilities decrease the number of effect occurrences that follow the cause. Both situations make it more difficult to identify the connection between the cause and its effect. If, however, an observer could modulate their observation time to compensate for these increases in uncertainty, would this increased time moderate the detrimental effects of long delays and/or decreased effect probabilities? When observing discrete events, longer observation allows a larger sampling of the event characteristics, which can provide a stronger inferential basis for a correct decision regarding possible cause–effect relations.

In our instantiation of probability as a form of uncertainty, we have restricted ourselves to the type of probabilistic outcomes that occur when a cause fails to produce the desired effect 100% of the time. Examples include striking a match, asking someone out on a date, changing the incentives for a behavior, or shooting a weapon. This type of uncertainty must be contrasted with the broader notion of contingency that has been commonly studied in psychology (e.g., Allan, Tangen, Wood, & Shah, 2003; Mandel & Lehman, 1998; Shanks, 1987). In contingency studies, uncaused effects are possible and weaken the inferred causal relationship between a cause and an effect. Although our paradigm could be extended to include uncaused effects or effects produced by alternative causes, addressing these questions is outside of the scope of this report.

In the presence of the type of probabilistic uncertainty that we studied, those trials that did not involve an outcome

would provide no information helpful to choosing among the candidate causes, and thus might encourage longer observation in order to compensate. We predicted that the participants might effectively bracket their observations to focus only on those trials involving the effect and would wait longer to ensure that there were sufficient observations on which to base their decision. In contrast, waiting longer in the face of *delay* uncertainty would not just permit the observation of more occurrences of the cause followed by its delayed effect but would also include more observations of spurious relationships between the foils and the effect. If participants are indeed sensitive to these differences in the efficacy for waiting behavior, we might observe longer waiting due to lower probabilities but not to longer delays.

In our study, we were interested in addressing three questions. First, would prior results documenting the detrimental impact of delays and probabilistic uncertainty on causal judgment generalize to our video game environment? Second, would participants increase their observation times in the presence of delay or probabilistic uncertainty? Third, would increasing observation time in the presence of uncertainty actually improve choice accuracy? To answer these questions, in **Experiment 1** we manipulated the delay and the outcome probability across causal decisions within the context of a video game preparation and measured both accuracy and latency between decisions. In **Experiment 2**, we more fully explored the functional relationship between our independent variables (delay and probability) and our dependent variables (accuracy and latency) by choosing to randomly sample from a range of effect delays and probabilities for each choice (cf. the notion of representative stimulus sampling presented by Brunswik, 1947). For example, one choice might involve a 1.23-s delay and a 38% probability of producing the effect, whereas the next choice might involve a 0.56-s delay and a 74% probability. This richer sampling of the stimulus domain allows for better estimates of the nature of the relationship between these variables within a more complex environment.

Experiment 1

We adapted a video game containing numerous decision points at which the participant must observe three possible causes (weapon discharges by three targets) of an event (an explosion), with each of the candidate causes randomly occurring within a 3-s interval, followed by a 1-s interval during which none of the weapons would discharge. Participants were simply asked to choose which candidate was the most likely cause of the effect of interest. The choice approach is common in nonhuman animal studies and developmental studies of causation because nonverbal animals and young children cannot accurately report on

their causal beliefs except by making a choice (e.g., Killeen, 1981; Shultz, 1982; Siegler, 1975).

For **Experiment 1**, we operationalized uncertainty regarding when the outcome would occur by creating various delays (0.0, 0.5, 1.0, and 2.0 s) between an enemy's weapon's firing motion (a recoil that jostles the shooter) and its effect (an explosion at a building), and uncertainty regarding whether the outcome would occur by creating various outcome probabilities (100%, 75%, and 50%). In other work using our first-person-shooter video game environment, we discovered that men and women sometimes respond differently in this environment (Young & Nguyen, 2009). Because these differences might be due to differences in their prior video game experience (cf. Feng, Spence, & Pratt, 2007; Green & Bavelier, 2003), we collected demographic data on sex and on both the amount and types of prior video game experience.

Method

Participants A total of 40 introductory psychology students (20 men and 20 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 the top of Fig. 1). Using the terrain editor, a perimeter of mountains was created around a village that contained trios of orcs, each firing at buildings in the village (see the bottom of Fig. 1). The critical variables modified the consequences of the orcs' weapon fire. The weapon projectiles were not visible because this would provide an all-too-obvious causal link between the cause and its effect.

The game world included four levels—when all of the choice points in a level were exhausted, the game was reset to randomly assign conditions to choice points, thus creating a new level (the game environment disappeared and was replaced with a “loading” screen between levels). Each level contained seven separate regions, with each region populated by three visually identical orcs (an example of three orcs in a region is shown in the top of Fig. 1; the seven regions are shown in the bottom of Fig. 1 from a view above the terrain). For simplicity, the orcs were stationary and oriented toward a target 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 target trio. Every 4 s (on average), each orc fired its weapon (an orc's firing was noticeable



Fig. 1 Screen shots from the video game. Top: One trio (three potential targets) is shown as well as the outcome of the true target's weapon (an explosion on the surface of a building). Bottom: A top-down view of the relative positions of the seven trios in the game's landscape

from the recoil of the weapon and an audible click, although it might or might not produce an explosion). Specifically, every 4 s each orc was programmed to fire once within the next 3 s, with the precise delay chosen from a [0 s, 3 s] uniform distribution. For the enemy orc, its firing was followed by an explosion at the programmed delay and with the programmed probability. Thus, although explosions could occur at any time within these 4-s trials, weapons' fire could only occur during the first 3 s. This restriction was chosen to decrease the likelihood that an explosion would carry over into the next 4-s interval while still allowing some degree of overlap.

In each three-orc region, the firing of one of the orcs (the *enemy orc*) produced contingent explosions in the target region. 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 (a fixed-ratio 8 [or FR8] schedule), since our pilot studies revealed that participants

showed greater discernment under these conditions because poor choices would lengthen the game; when only one shot was required, many of the participants simply shot indiscriminately, because they could complete the experiment in about 10–15 min by doing so (cf. Williams, 1971, who documented minimum ratio requirements for pigeons). For a video clip showing a participant observing a trio, destroying a target (i.e., making a choice), and observing the consequences of that choice, see <http://bcs.siuc.edu/facultypages/young/Research/VideoGameClip.mov>. This clip also shows a bird's-eye view of the entire game region present in every level.

The enemy orcs in the seven trios of Level 1 had no weapon delays and 100% probabilities and were used as a method for orienting the participant to the game's requirements and assessing their understanding. Although Level 1 performance was noted, it was tangential to the main variables of interest. The enemies in the seven trios of Levels 2–4 had either no delay and a 100% probability (the control trio) or a 0.5-, 1.0-, or 2.0-s delay (two trios each; one of the trios had auditory sounds that followed each orc's firing for the length of the programmed delay, and the other trio had no sounds) and the programmed outcome probability (50%, 75%, or 100%) for that level. The assignment of each outcome probability to a level was counterbalanced across participants, with each receiving one of the following orders for Levels 2–4: 50/75/100, 75/100/50, or 100/50/75. The left-to-right location of the target in each trio was randomly determined for each trio for each participant.

Procedure The participants were seated at one of four identically configured 1.25 GHz Mac Mini computers. The participants received specific verbal instructions, including a description of the task, using a screen shot of what they would see once the experiment was started. The instructions included the following:

Today you will be participating in a computer experiment. You will be placed in a virtual village being attacked by sets of orcs. The orcs will be firing timed grenades. Only one of the orcs has a working weapon; the others are duds. Your task is to identify and destroy the orc causing the damage, the enemy, without shooting the duds. If you choose the right orc, the explosions will stop and the blue bar will drop on the screen. If you choose the wrong orc, the explosions will continue and the blue bar will persist. Continue firing on one of the remaining orcs that is causing the damage until you discern the enemy.

In addition to the task instructions, participants were advised how to navigate within the environment. When

they indicated an understanding of the procedure, the experimenter started each of the programs.

Once a participant had destroyed the enemy in each trio on each level, the video game ended. After finishing the game, the participants completed a demographic questionnaire asking their sex, self-rated video game experience during elementary, middle, and high school and college (Likert scale with 0 indicating *none*, 6 indicating *daily*), and the types of video games that they played. Because a principal components analysis revealed no distinct role for periods of experience, we summed across the four periods. For ease of presentation and to ensure adequate sample sizes, we used Ward's hierarchical clustering to identify three distinct clusters of participants, as determined by the types of video games played. The analysis included data from both experiments; we summarize their characteristics in Fig. 2, which shows the percentage of participants in each cluster who self-reported playing each type of video game.

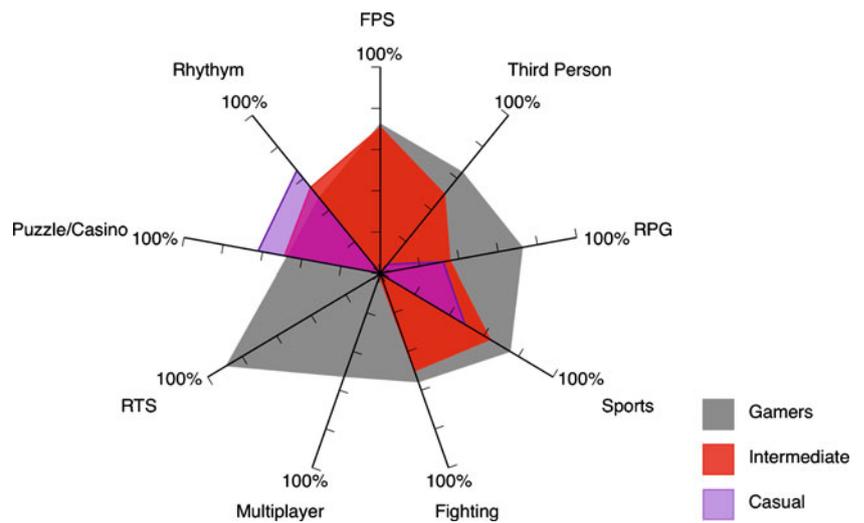
Results

All 40 participants successfully completed the task. Although 5 participants showed accuracies that did not exceed chance on Levels 2–4, 4 of the 5 were women and thus were retained to avoid differential attrition. Accuracy increased across the critical levels (53%, 55%, and 57%, for Levels 2, 3, and 4, respectively) and decision times decreased (51, 46, and 45 s, respectively); thus, counterbalancing probability conditions across level orders was clearly necessary.

The initial choice accuracy and log(latency) as a function of outcome delay and programmed probability are shown in Fig. 3.¹ Lower accuracy for longer delays is consistently evident, but there was no obvious effect of outcome probability on accuracy. Latencies were generally longer for the longer delays and for lower probabilities. Men produced slightly higher accuracies that were not due to a speed–accuracy trade-off, because men showed much shorter latencies. In none of our analyses did the 0.5-, 1.0-, and 2.0-s auditory events have an effect on performance (Young & Nguyen, 2009, found minimal and

¹ Throughout the text we report and analyze the data on the basis of the programmed probabilities. Because of the limited sampling of these probabilities, each participant experienced small variations on these values. Unfortunately, we do not know what they experienced, because we cannot know whether a participant was able to see (due to distance) or was looking in the correct direction when each orc fired its weapons. For shots that occurred during the latency period (which included travel from trio to trio), the actual probabilities were highly correlated with the programmed probabilities for both experiments ($r = .94$ for Experiment 1, $r = .97$ for Experiment 2). Although it is certain that the participants were not able to observe all of these shots, the limited sampling of the probabilities would not produce any bias in their experienced values.

Fig. 2 A star (or spider) plot of the results of a cluster analysis of self-reported types of video game played across both experiments. The point at which a polygon crosses each axis indicates the percentage of people in that cluster who indicated that they played the type of game represented by the axis. Note—A total of 24 people were in the gamers cluster (12% women), 55 in the intermediate cluster (47% women), and 32 in the casual cluster (84% women). FPS, first-person shooters; RPG, role playing games; RTS, real-time strategy games



idiosyncratic effects of these types of events), so this factor was ignored in our presentation of the results.

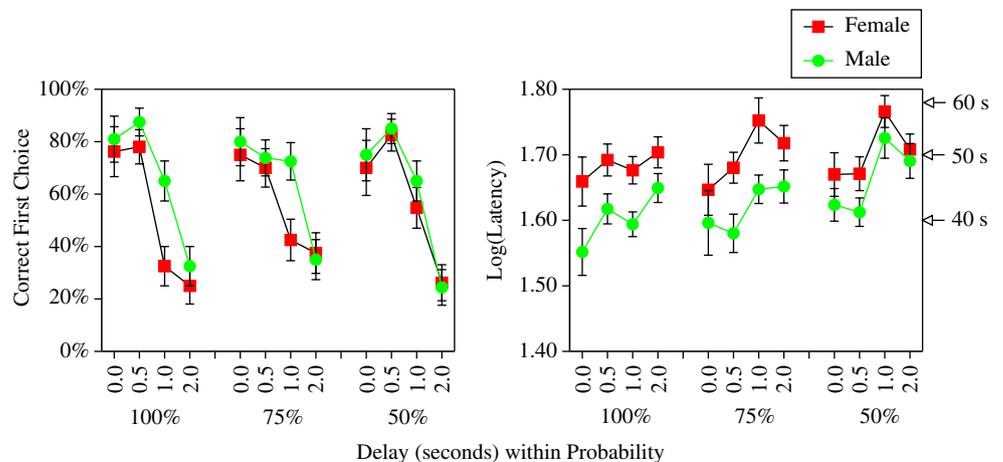
In our analyses of the results, we used the R statistical environment (available at www.R-project.org) to fit a range of linear mixed-effects models (Pinheiro & Bates, 2004) to identify the best-fitting model (assessed using the Akaike information criterion; Akaike, 1974) that described our participants' behavior. Each of the between-subjects factors and interactions was systematically included or excluded until the best model was identified. This method was necessary because of multicollinearity among the between-subjects variables (men were more likely to fall into the gamers cluster and to play more often). A linear relationship between delay/probability and accuracy produced the best fits; neither log transforming the predictors nor treating them as categorical variables produced better models.

A repeated measures logistic regression (mixed-effects model specifying a binomial error distribution) revealed a

main effect of delay, $z = 11.05, p < .01$, and a main effect of sex, $z = 2.55, p < .05$, but no main effect of probability, $z = 0.42$, nor a Delay \times Probability interaction, $z = 0.15$ (the analysis generates z s due to our specification of a binomial error distribution). The best-fitting model allowed intercepts (overall accuracy) to randomly vary across participants. The effects of delay and probability translate to Cohen's f^2 values of .210 and .000, respectively.

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 included 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 analyses revealed no systematic differences among firing rates as a function of our within-subjects independent variables; once a participant started firing, he or she continued to shoot at the same target until it was

Fig. 3 Experiment 1's accuracy (left figure) and latency (right figure) from the last accurate decision to the initial choice in the next trio as a function of delay, probability, and sex. Error bars indicate ± 1 standard error



destroyed—exceptions were quite rare. Travel time could not be influenced by delay and likelihood, because the participant has no anticipatory knowledge of the behavior of the targets being approached. Thus, we are confident that any latency differences across delays and likelihoods are primarily a function of differences in observation time. In contrast, we cannot make those same assurances regarding our between-subjects variables, so sex and cluster latency differences must be treated with caution (cf. Young, Sutherland, & Cole, 2010).

The best-fitting model of log latencies revealed main effects of delay, $F(1, 850) = 18.33$, $p < .01$, probability, $F(1, 850) = 10.60$, $p < .01$, and sex, $F(1, 850) = 7.47$, $p < .01$, but no Delay \times Probability interaction, $F(1, 850) = 0.01$. Note that the denominator degrees of freedom for mixed-effects models are quite different from those using standard ANOVA. The best-fitting model allowed intercepts (overall latency) to randomly vary across participants. The effects of delay and probability translate to Cohen's f^2 values of .022 and .013, respectively.

Discussion

The detrimental impact of delays on both accuracy and latencies replicates Young and Nguyen's (2009) results. The absence of an effect of outcome probability on accuracy was somewhat surprising. The longer latencies for lower probabilities suggest that participants were simply compensating for the uncertainty in whether the outcome would occur by increasing their viewing time. Although participants also showed longer latencies for longer delays, these longer observation times did not fully compensate for the effect of delay on choice accuracy.

The analyses also revealed that men produced stronger performance, as reflected by higher accuracies (11% higher) and shorter latencies (7 s shorter). This basic result replicates Young and Nguyen (2009), but in their study sex differences in accuracy were accounted for by differences in the self-reported amount of prior video game experience. In *Experiment 1*, we found no effects of amount of self-reported video game play nor self-reported types of video games played; only sex was retained as a contributor to each analysis.

In conclusion, waiting longer in the presence of lower probabilities appears to have offset any potential effect on choice accuracy. In contrast, both men and women had longer latencies in the presence of longer delays, but this increase did not offset the effect of delays on accuracy. Besides the natural desire to maximize accuracy, the time required to complete the task appears to be a key factor in participant decision making. Given that a choice requires completing an FR8 (taking about 7–10 s for most players), a correct choice would require an average of 7–10 s to

destroy the enemy, whereas choosing a target at random would require an average of 14–20 s to destroy the enemy. Thus, if the observer concludes that less than 7–10 s of additional observation time will ensure a high probability of choosing correctly, then longer observation is an efficient strategy. Indeed, the average increase in latencies from the 100% to the 75% condition was only 1.6 s, and from the 75% to the 50% condition was an additional 2.7 s. This difference corresponds to watching roughly one more shot by the target for the 50% weapons than for the 100% weapons.

Experiment 2

The results of *Experiment 1* replicated prior published effects of delay on accuracy and choice latency, provided further evidence that men generally produce higher accuracies and shorter latencies on the task that cannot be explained by self-reported prior video game play, and provided the first evidence that decreasing the probability that the cause would produce the effect from 100% to 50% had no effect on accuracies but instead increased choice latencies.

The failure to find an effect of probability on accuracy was surprising, given prior results using systematic designs (e.g., Allan et al., 2003; Kao & Wasserman, 1993; Mandel & Lehman, 1998), but was consistent with our prediction that observers would simply wait longer to ensure a sufficient sampling of effect occurrences when allowed to do so. In our final experiment, we considered two possible explanations of the failure to find an effect of probability on accuracy. First, perhaps we simply had experienced a restricted-range problem, and thus testing even lower probabilities would reveal an effect. Second, probabilities varied *across* game levels, whereas delays varied *within* a game level. This design may have allowed participants to adapt to each probability because it was identical across the seven trios present within a game level.

Thus, in *Experiment 2* we made three changes to the design to more fully explore how changing the outcome delay and probability affect choice accuracies and latencies: (1) Probability varied between 20% and 100%, (2) probability and delay both varied from trio to trio within and across game levels, and (3) for each trio, the particular probability and delay values were randomly sampled from the chosen range. The third change leveraged an innovation in our laboratory intended to more richly sample the entire range of a predictor in order to better identify the functional relationship between a predictor and the dependent variables. The use of random stimulus sampling was inspired by Brunswik's (1947, 1955) notion of representative sampling of stimuli.

Method

Participants A total of 61 introductory psychology students (30 men and 31 women) at Southern Illinois University at Carbondale received course credit for their voluntary participation.

Game environment and design The landscape and trio locations were identical to those used in [Experiment 1](#). Level 1 comprised the seven trios, each with a 0.0-s delay and 100% outcome probability. This level was intended to familiarize players with the game and was not analyzed. For each trio in Levels 2–4, however, the particular delay for each target was a constant value randomly chosen from a 0.0- to 2.0-s range, and the particular probability was a constant value randomly chosen from the 20% to 100% range, using sampling with replacement from continuous uniform distributions. For example, participant f03d encountered a target in Level 2 with a 0.83-s delay and a 34.6% probability in her first trio, followed by a target with a 1.96-s delay and a 71.4% probability in her second trio, a target with a 1.77-s delay and 95.2% probability in her third trio, and so forth. Once a game level was finished, the level was reset with a new set of randomly chosen delays and probabilities. Thus, each participant experienced 21 different randomly chosen delays and 21 different randomly chosen probabilities.

Procedure The procedure was identical to that from [Experiment 1](#).

Results and discussion

We retained all participants because the use of a performance threshold would have produced differential attritions across men and women. To visualize the relationship between the two continuous predictors and the outcome variables, we first performed a multidimensional spline fit using generalized additive modeling (R's *gam* function). This technique produces a smoothed surface using a nonparametric approach to help identify any unmodeled nonlinear relationships that should be included in our subsequent analyses. The results are shown in [Fig. 4](#). Accuracy was much lower for longer delays and slightly lower for lower probabilities. The obvious nonlinearity in the relationship was a reflection of the logistic relation between the continuous predictor and our binomial dependent variable. Latencies were somewhat longer for longer delays and much longer for lower effect probabilities, with both relations approximately linear.

A repeated measures logistic regression revealed main effects of delay, $z = 7.08$, $p < .01$, and cluster, $z = 2.37$,

$p < .05$, but no main effect of probability, $z = 1.05$, nor a Delay \times Probability interaction, $z = 1.44$. The best-fitting model allowed intercepts (overall accuracy) to randomly vary across participants. The effects of delay and probability translate to Cohen's f^2 values of .039 and .001, respectively. The cluster main effect was due to a statistically lower accuracy for the casual cluster (see [Fig. 2](#)), which included those who frequently self-reported as nonplayers of various types of video games, including first-person shooters like the one used in this study ($M_s = 43\%$, 55% , and 57% for the casual, intermediate, and gamer clusters, respectively).

The best-fitting model of log latencies revealed main effects of delay, $F(1, 1327) = 2.37$, $p < .05$, and probability, $F(1, 1327) = 32.32$, $p < .01$, but no Delay \times Probability interaction, $F(1, 1327) = 0.06$. None of the between-subjects variables (sex or video game play) were retained. The best-fitting model allowed intercepts (overall latency) to randomly vary across participants. The effects of delay and probability translate to Cohen's f^2 values of .001 and .025, respectively.

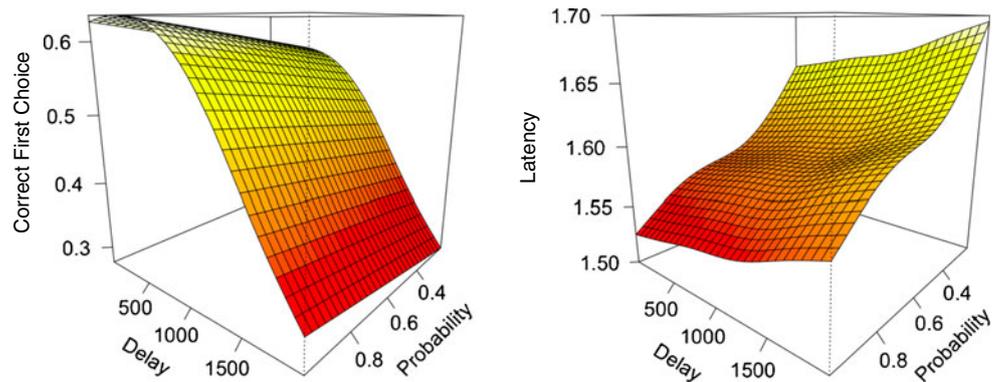
The absence of an effect of outcome probability on accuracy but the presence of an effect on latencies replicated [Experiment 1](#). Although there was a small trend toward lower accuracies for lower probabilities (see [Fig. 4](#)), the trend was not significant despite the much greater range in probabilities (20%–100%). Thus, even when we sampled a wider range in probabilities, allowed them to vary from trio to trio within a level, and used a random sampling design, participants still failed to show a significant impact of probability on initial choice accuracy, and instead compensated by waiting longer, thus obtaining larger observation samples, before making their decisions.

Sex had no effect on accuracy or latency in our analyses, unlike in [Experiment 1](#), but self-reports of prior types of video games played did predict accuracy in [Experiment 2](#). Given the strong confounding between sex and types of video games played (e.g., only 12% of those in the gamer cluster were women, but 84% of those in the casual cluster were), the emergence of one or the other factor as significant may be idiosyncratic given the tested sample sizes.

General discussion

Participants demonstrated the ability to discriminate among possible causes of an environmental event (an explosion) by integrating information regarding time of weapon discharge, delay between the discharge and the outcome, and the probability of the outcome following the discharge. Delays decreased accuracy and often increased decision latencies, a common finding in our laboratory and one that

Fig. 4 A nonparametric fit of Experiment 2 describing the relationship between accuracy (left figure) and log-transformed latency (right figure) from the last accurate decision to the initial choice in the next trio as a function of delay (measured in milliseconds) and probability



we fully expected. In contrast, decreasing the probability of the outcome surprisingly had little or no effect on accuracy, but increased latencies. Given the ubiquity of previous demonstrations of the detrimental impact of outcome uncertainty on causal judgments (e.g., Allan, 1993; Wasserman, Elek, Chatlosh, & Baker, 1993; Young, 1995), the lack of an effect in the present experiments suggests that participants were compensating for the probabilistic uncertainty by delaying their decisions (an option that was generally not available in prior studies of outcome uncertainty).

Using a wider range of probabilities (20%–100%), rapid changes in probability across trios, and random stimulus sampling still resulted in no effect of probability on accuracy. Although there may yet exist conditions under which participants will not compensate for lower effect density (due to lower probabilities) by increasing their observation time, a failure would have to occur at outcome probabilities even lower than 20%. In contrast, delay had a robust effect on accuracy, with performance near chance for 2.0-s delays whether we used systematic sampling of delays (Experiment 1), in which the longest delays represented 14% of the decisions, or random sampling of delays (Experiment 2), in which the longest delay was the endpoint of a continuum.

Finally, in our analysis of sex as a predictor variable, it was a significant predictor of accuracy and latency differences in Experiment 1 but not in Experiment 2 (although cluster was a significant predictor of accuracy). An open question regards what is producing these differences. Because men comprise a greater proportion of people who self-report playing a lot of action-oriented video games (see Fig. 2), the natural confounding of these variables may produce sex differences. In follow-up research, we will record a much greater range of participant behavior to identify the source of these differences between men and women.

Compensating for uncertainty

Our participants were met with two sources of uncertainty in the face of multiple causal candidates occurring in a

complex dynamic environment: uncertainty due to delays and uncertainty due to lower effect probabilities. The presence of uncertainty creates a significant challenge to the decision maker. The present study suggests that participants may compensate for uncertainty in whether the effect would occur by simply increasing their viewing time and thus obtaining a larger sample of the critical events. In contrast, although the participants also increased their viewing time in the presence of greater uncertainty due to delays (i.e., waited longer for situations involving longer delays), delay uncertainty still had a large impact on decision accuracy.

Note, however, that we made no attempt to directly manipulate latency by limiting observation time for some observations or participants (cf. Young et al., 2010). Although this could be done, it is self-evident that limiting observation time for low-probability effects would decrease accuracy, because doing so would increase the number of observation periods involving no effects (thus requiring participants to guess). For example, if participants were limited to viewing only two shots by the target before making a decision, then for a 50% outcome condition, 25% of the two-shot observations would involve no effects, and for a 20% outcome condition, 64% of the two-shot observations would involve no effects. This restricted observation would clearly produce lower accuracies for very low outcome probabilities than for higher outcome probabilities.

In our task, participants were allowed to make the trade-off between making a decision quickly and potentially less accurately versus obtaining more data (in the form of additional observations of the candidate causes and the effect) to try to increase their accuracy. The natural motivation to complete the task in a reasonable amount of time (our students apparently had better things to do than to play our video game) produced a desire to make some decisions more quickly, but hasty incorrect decisions meant that time would be wasted while they made the additional eight shots necessary to destroy another target. People revealed choice behavior that indicated that some situations were so complex (e.g., those with long cause–effect delays)

that a choice was made despite existing uncertainty. 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.

Increasing generalizability

Our ability to identify the way in which our participants attempted to compensate for different types of uncertainty was a direct product of our willingness to cede some control to the participants (cf. Richter et al., 2010; Richter et al., 2009). Although we retained the ability to control the unfolding of environmental events, we permitted the participants to control how much environmental sampling they required before making a choice. Unlike in many other observations of speed–accuracy trade-offs, our participants could collect more data on how discrete events unfolded across time and not merely decide how long to observe a static stimulus. The ability to adapt to environmental uncertainty is central to successful decision making, whether this adaptation involves longer observation, acquiring additional cues to outcomes, requesting expert advice, or judging that the outcome is so unpredictable that rapid guessing is the best strategy. Furthermore, prior results using more systematic investigations of a discrete set of delays nicely generalized to our less-structured environment, further establishing their robustness. The study of causal judgment may thus be well informed by using designs that parallel the unstructured and dynamic nature of the real world.

Acknowledgments N.N. is now at Beth Israel Deaconess Medical Center, Harvard University. This research was supported by AFOSR Grant FA9550-07-1-0429 and NSF Grant SES-0720588. We thank Garren King, Tara Webb, Joel Smith, James McCoy, Victoria Babb, Andrew Vaz, and Amber Blacharski for their assistance in running these experiments. Please address correspondence to M.E. Young, Department of Psychology, Mailcode 6502, Southern Illinois University, Carbondale, IL 62901-6502 (e-mail: meyoung@siu.edu).

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin*, *114*, 435–448.
- Allan, L. G., Tangen, J. M., Wood, R., & Shah, T. (2003). Temporal contiguity and contingency judgments: A Pavlovian analogue. *Integrative Physiological and Behavioral Science*, *38*, 214–229.
- Brunswik, E. (1947). *Systematic and representative design of psychological experiments; with results in physical and social perception*. Berkeley: University of California Press.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.
- Buehner, M. J. (2005). Contiguity and covariation in human causal inference. *Learning & Behavior*, *33*, 230–238.
- Dickinson, A., & Shanks, D. R. (1995). Instrumental action and causal representation. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate. Symposia of the Fyssen Foundation* (pp. 5–25). Oxford: Oxford University Press, Clarendon Press.
- Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*, 850–855.
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534–537.
- Hammond, L. J., & Paynter, W. E. (1983). Probabilistic contingency theories of animal conditioning: A critical analysis. *Learning and Motivation*, *14*, 527–550.
- 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.
- Killeen, P. R. (1981). Learning as causal inference. In M. L. Commons & J. L. Nevin (Eds.), *Quantitative analyses of behavior: Discriminative properties of reinforcement schedules* (Vol. 1, pp. 89–112). Cambridge, MA: Ballinger.
- Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General*, *127*, 269–285.
- Pavlov, I. P. (1927). *Conditioned reflexes* (G.V. Anrep, Trans.). London: Oxford University Press.
- Pinheiro, J. C., & Bates, D. M. (2004). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Richter, S. H., Garner, J. P., Auer, C., Kunert, J., & Würbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nature Methods*, *7*, 167–168.
- Richter, S. H., Garner, J. P., & Würbel, H. (2009). Environmental standardization: Cure or cause of poor reproducibility in animal experiments? *Nature Methods*, *6*, 257–261.
- Shanks, D. R. (1987). Acquisition functions in contingency judgment. *Learning and Motivation*, *18*, 147–166.
- Shanks, D. R. (1993). Human instrumental learning: A critical review of data and theory. *British Journal of Psychology*, *84*, 319–354.
- Shanks, D. R., Darby, R. J., & Charles, D. (1998). Resistance to interference in human associative learning: Evidence of configural processing. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 136–150.
- Shanks, D. R., & Dickinson, A. (1991). Instrumental judgment and performance under variations in action–outcome contingency and contiguity. *Memory & Cognition*, *19*, 353–360.
- Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of Experimental Psychology*, *41B*, 139–159.
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*(1, Serial No. 194).
- Siegler, R. S. (1975). Defining the locus of developmental differences in children's causal reasoning. *Journal of Experimental Child Psychology*, *20*, 512–525.
- Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*, 127–151.
- 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, and 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.

- White, P. A. (1995). Use of prior beliefs in the assignment of causal roles: Causal powers versus covariation-based accounts. *Memory & Cognition*, *23*, 243–254.
- Williams, B. A. (1971). Color alternation learning in the pigeon under fixed-ratio schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *15*, 129–140.
- Young, M. E. (1995). On the origin of personal causal theories. *Psychonomic Bulletin & Review*, *2*, 83–104.
- Young, M. E. (2010). Contemporary thought on the environmental cues that determine causal decisions. In T.R. Zentall & E.A. Wasserman (Eds.), *Handbook of comparative cognition*.
- 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., Sutherland, S. C., & Cole, J. J. (2010). Individual differences in causal judgment under time pressure: Sex and prior video game experience as predictors. *International Journal of Comparative Psychology*.
- Young, M. E., Wasserman, E. A., Johnson, J. L., & Jones, F. L. (2000). Positive and negative patterning in human causal learning. *The Quarterly Journal of Experimental Psychology*, *53B*, 121–138.