Memory & Cognition
2005, 33 (2), 320-331

Causal impressions: Predicting *when*, not just *whether*

MICHAEL E. YOUNG, ESTER T. ROGERS, and JOSHUA S. BECKMANN
Southern Illinois University, Carbondale, Illinois

In 1739, David Hume established the so-called *cues to causality*—environmental cues that are important to the inference of causality. Although this descriptive account has been corroborated experimentally, it has not been established why these cues are useful, except that they may reflect statistical regularities in the environment. One of the cues to causality, covariation, helps predict whether an effect will occur, but not its time of occurrence. In the present study, evidence is provided that spatial and temporal contiguity improve an observer's ability to predict *when* an effect will occur, thus complementing the utility of covariation as a predictor of *whether* an effect will occur. While observing Michotte's (1946/1963) launching effect, participants showed greater accuracy and precision in their predictions of the onset of movement by the launched object when there was spatial and temporal contiguity. Furthermore, when auditory cues that bridged a delayed launch were included, causal ratings and predictability were similarly affected. These results suggest that the everyday inference of causality relies on our ability to predict whether and when an effect will occur.

The concept of causality is fundamental to human cognition. There are two primary advantages afforded by causal attribution: prediction and control. If an organism can discover the causal nature of events, it is empowered to predict which events will follow others, thus allowing it to prepare for the arrival of events that are important to its survival, its enjoyment, and so on. Control over events can be established if the causes of those events can be produced or prevented by the organism.

Given the tremendous survival advantages afforded by a proper grasp of the "causal texture of the environment" (Tolman & Brunswik, 1935), it is not surprising that people spend a significant amount of time looking for causes or taking advantage of their own causal knowledge. Our current understanding of causal learning, however, has been limited by an over-reliance on static, point-events as causes and effects (for reviews, see Allan, 1993; Shanks, 1993; Young, 1995). Although there is still much to discover about causal learning and judgment when events are presented verbally, modern technology enables the systematic study of these processes when causes and effects are observed to unfold over time and thus require the observer to parse the event stream into its relevant components (e.g., Zacks & Tversky, 2001).

Michotte (1946/1963) laid the groundwork for such a paradigm in his exploration of the launching effect by using cleverly designed mechanical devices. Michotte performed a series of informal studies examining the "causal impressions" experienced by observers when they see one object move, see it touch another object, and see the second object begin to move along the same trajectory (see Figure 1). Michotte documented the variables that degrade this causal impression—a lack of spatial contiguity (the objects do not actually touch), a lack of temporal contiguity (the affected object does not move immediately after being struck), and changes in the relative rates of motion (when the first object is moving slower or faster than the second, the impression is altered), inter alia. But why do these variables influence causal judgments?

We are examining the thesis that causal judgments are strongly influenced by the predictive relationship between causes and their effects. Researchers have recognized that predictability is important to causality, but studies of predictability have been limited to notions of covariation, contingency, or statistical dependence (e.g., Cheng & Novick, 1992; Glymour, 2001; Shanks, 1993; Suppes, 1984). In other words, these earlier projects have focused on *whether* the effect will occur. We posit that it is important to predict both whether the effect will occur and *when* it will occur. For example, one can easily predict whether someone will have an allergic reaction to poison ivy but not when that reaction will begin. In contrast, one cannot easily predict whether a baseball player will hit the ball on any given swing, but one can predict with considerable precision when the bat will strike the ball (if it does so). The ability to predict whether and when an event will occur reflects an understanding of the causal structure of the environment, even if knowledge of the actual mechanisms is incorrect (as was the case in astronomy before

---

We thank Matthew Schlesinger, Eric Jacobs, and Michelle Ellefson for their willingness to serve as sounding boards, Rebecca Eldredge and Brent Beam for their assistance in conducting the experiments, and Timothy Hubbard for his comments on an earlier version of this manuscript. Correspondence should be addressed to M. E. Young, Department of Psychology, Southern Illinois University, Mailcode 6502, Carbondale, IL 62901-6502 (e-mail: meyoung@siu.edu).

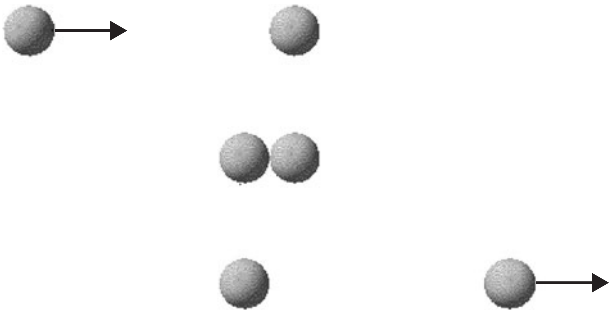

Figure 1. Figure of a direct launching. The top frame shows the position of the balls 1 sec after onset of movement, the middle frame shows the point of collision, and the bottom frame shows the position of the balls 1 sec after the collision.

Galileo, in the principle of contagion before microbiology, and in the law of effect before neuroscience).

Covariation as a cue to causality cannot explain the persistently weaker causal impression experienced when an observer watches a consecutive series of delayed launchings (in which one object strikes another that is launched after a delay; Michotte, 1946/1963) or experiences the delayed consequences of a response (Buehner & May, 2002). The observer knows that the second object will start moving some time after the first object stops because of the high covariation between the movements of the two objects, but the impression of causality is still weak or absent. Temporal contiguity has been cited as a cue to causality to explain this observation (Hume, 1739/1969), but there has been little rationale for *why* it affects causal impressions. We believe that with a lack of temporal (and spatial) contiguity, observers will be less accurate in predicting when the effect will occur, and this imprecision weakens the causal impression. This uncertainty is the direct product of people's inaccuracy in remembering the duration of a delay and the distance between objects (e.g., Gibbon, 1981; Rachlin, 1966; Worley & Markley, 1969). Furthermore, as delay and/or distance increases, the predictability decreases, and thus the strength of our causal impression wanes.

By suggesting that one's ability to accurately predict when something will occur is the basis of the importance of temporal contiguity, we also offer something that mere contiguity cannot—an explanation of why a temporally distal event is sometimes more acceptable as a cause than is a temporally proximate event. Our experience with some events may lead us to expect a delay between a cause and its outcome and thus predict a delayed effect. For example, in cases of nausea an organism is likely to attribute its condition to the food that it ate some time ago and not to the events that occurred immediately prior to the nausea (Garcia & Koelling, 1966). Contiguity may not always be necessary in causal judgments because we can attribute causality to events that are separated in time and space when the earlier events are good predictors of whether and when later events will occur (Buehner & May, 2002; Hagmayer & Waldmann, 2002) or even when the learner receives instructions to expect a delayed rather than an immediate effect (Buehner & May, 2003).

The predictability account contrasts with the analysis offered by Tenenbaum and Griffiths (2003), who suggested that the effects of time and space are derived from a top-down model of a causal interaction. The a priori expectations of the observer are represented as a sophisticated theory of force transfer that includes implicit knowledge of the effects of space and time on vibrational energy (they applied their formulation to a study involving a ball dropping on a beam that caused another ball to fly out of a trap door; cf. Anderson, 1991). However, no explanation is offered as to how this top-down knowledge is acquired. Although we do not deny that people may have acquired sophisticated physical models that can be applied in some situations (especially those with which people have extensive experience; see, e.g., Gilden & Proffitt, 1994), our interest is in both the formation of such models and in the judgments that occur in the absence of a model.

As an initial assessment of the viability of the predictability hypothesis, Experiment 1 examined the precision of an observer's predictions of the onset of the launched object's movement and their causal judgments as a function of temporal and spatial contiguity. This study sought to replicate the effects of time and space on causal judgments, to document the effects of time and space on the predictability of the effect, and to examine the relationship between judgment and predictability. This experiment also included an analysis of the psychometric properties of these two measures of causal expectations. In order to go beyond this correlational study of judgment and predictability, in Experiments 2 and 3 we manipulated the temporal predictability of delayed launchings by inserting tones with various properties during the delay; two of these tones were designed to increase the predictability of the effect, whereas one was not. The impact of these tones was assessed using both causal impression and response-based predictions tasks.

EXPERIMENT 1

Traditionally, causal impressions have been solicited in a variety of ways. For example, Michotte (1946/1963) solicited verbalizations from observers. He sometimes placed their responses into specific classes (e.g., direct launching, delayed launching, two movements) but most often relied on his own verbal summary of the responses produced by the observers. White and Milne (1999; cf. Natsoulas, 1961) formalized the classification procedure of Michotte by requiring observers to rate their agreement (on a scale of 0 to 100) with three statements that could describe an event (loosely, Object A smashed Object B, Object A popped Object B, or Object B disintegrated of its own accord; their interactions involved disintegration, not launching). Schlottmann and Anderson's (1993) participants made their judgments of the degree of causality by moving a pointer along a graphical scale. Both of these methods produced systematic data.

Because the present task involved only one type of interaction (launching rather than popping or smashing) and because there was no reason to believe that the additional resolution offered by a graphical scale was necessary, we used a Likert scale to solicit causal impressions with 1 designating that Object B moved on its own and 9 designating that Object A caused Object B's movement.

Method

Participants. A total of 44 students enrolled in an introductory psychology course at Southern Illinois University at Carbondale served as voluntary participants. They received course credit for their participation.

Materials. The participants saw one of 16 different animations: 4 gaps (0.0, 1.0, 2.0, or 4.0 cm) × 4 delays (0.0, 0.5, 1.0, or 2.0 sec). For the direct launching animation (no delay and no gap), the left object moved to the right and stopped when it was contiguous to the right object; after contact, the right object immediately began moving to the right. For the delay animations, the left object again moved to the right and stopped; the right object began moving to the right 0.5, 1.0, or 2.0 sec after the left object stopped moving. For the gap animations, the left object moved to the right and stopped when its rightmost edge was 1.0, 2.0, or 4.0 cm (0.9º, 1.9º, or 3.8º of visual angle) from the leftmost edge of the right object; after the programmed delay, the right object began moving to the right.

One additional animation was included as part of the instructions to provide an example of a clearly noncausal interaction to anchor the noncausal end of the rating scale. In this noncausal animation, the two objects began in the same positions as those in the other animations, and the motion of each individual object was identical to that observed in direct launching (no gap, no delay). But the right object moved first, and the left object moved immediately after the right object stopped. Thus, the noncausal animation lacked spatial contiguity and temporal priority (the left object moved after, not before, the right object, and they did not contact each other).

The animations were created using LightWave 3D Version 6 and were saved as QuickTime movies using Sorenson video compression. An animation light source was placed to the upper left and front (toward the observer) of the moving objects to provide realistic shading. Each ball was 1.4 cm (1.3º of visual angle) in diameter and used LightWave's "Clothing.tga" texture to provide color; objects were rendered against a black background. When moving, each object moved at a steady rate of 4.0 cm/sec (3.8 deg/sec). For every animation, the left object was 9.3 cm (8.9º) to the left of center and the right object was 0.7 cm (0.7º) to the right of center at the beginning of the movie; the right object's motion always lasted for 2.0 sec after movement onset.

Presentation of the animations and collection of response data were programmed using PsyScript Version 4.6d5 (available at http://www.maccs.mq.edu.au/~tim/psyscript/index.html). The programs were run on four 400-MHz PowerPC G4 Macintosh computers.

Procedure. After collecting the informed consent forms, the experimenter read a set of general instructions that described the procedure. The participants were then seated approximately 0.6 m from the monitor. All participants completed both the predictions and the ratings tasks, with 18 receiving the predictions task first and 26 receiving the ratings task first (the imbalance was due to experimenter error).

Predictions task. The program began with the presentation of four example animations (the four combinations of the smallest and largest delays and gaps). The participants were told to press the space bar when they believed the second ball (on the right) would start moving. For these trials, feedback was given about the accuracy of their barpresses to ensure that the participants understood what they were supposed to do (e.g., "Your response was 256 msec too slow").

After completion of the four familiarization trials, the program signaled the onset of regular trials and informed the participants that no feedback would be given during the experiment proper. The participants then observed 16 blocks of 15 consecutive presentations of each of the animations, for a total of 240 trials, with each participant receiving a different random order of the 16 blocks. A block began with the message: "You will now see a single type of movie presented 15 times. Remember to press <space> when you think the second object will start moving." Each movie in these 15-trial blocks followed the previous movie after a 1-sec delay.

Ratings task. The participants were instructed to rate the degree to which the object on the left caused the object on the right to move, using a 1–9 causal impressions rating scale. During the familiarization phase, the participants viewed two animations, the direct launching and the noncausal, as examples of a highly causal interaction that should receive a rating of 9 and a noncausal interaction that should receive a rating of 1, respectively.

Schlottmann and Anderson (1993) also began their sessions with anchors, but their anchors were the direct launching animation and the maximal gap and delay animation (with a 170-msec delay and a 2.1-mm gap). Their technique solicited responses along the entire response range but may have demanded a particular interpretation of the animations as evidenced by the very large effects of extremely small gaps and delays. Although the use of a clearly noncausal animation may produce ceiling effects on our 1–9 causal impressions rating scale, it was hoped that this technique would more closely reflect the naive judgments of the participants rather than training them to consider brief delays (170 msec) and minute gaps (2.1 mm) unacceptable.

The experiment proper presented the 16 animations in a random order for each participant, requiring a keyboard response after each animation before the next animation was displayed.

Results

We conducted three sets of analyses. In the first set, we examined participants' performance on the predictions task as a function of gap and delay. In the second set, we examined participants' performance on the ratings task as a function of gap and delay. In the final set, we examined correlations between each participant's performance on the predictions and ratings tasks to determine the extent to which the two measures were related.

Predictions task. The timing of the solicited barpresses reflects when the participant believed the to-be-launched object would begin moving. Given that there is an objective correct answer for these responses (the actual time of motion onset), our dependent variable (error score) was the amount of error for each response. We used the absolute value of the error as an index of accuracy. Although the direction of the error did vary as a function of the animation (e.g., responses for animations involving a delay were more likely to be late), those details are not provided here but are available from the authors. The absolute value was preferred because it helped differentiate behavioral profiles that generated the same mean error but differed in variability. For example, response profiles with many very early and very late responses could produce a low mean error but would produce high absolute error, whereas response profiles with consistently accurate responses would produce low absolute error.

Predictions for the first trial of a block were excluded because participants would have no prior observation on which to base their predictions. One participant in each task-ordering condition failed to complete the predictions task (one completed 123 trials, whereas the other completed 209 trials); these participants were dropped from the experiment. Initial analyses of predictions accuracy showed no statistically significant effects involving task order (predictions first vs. ratings first). Subsequent analyses omit this factor for simplicity of presentation.

Figure 2 shows the changes in mean base 10 logarithm of the error score for each animation (a logarithmic transformation was used to normalize the distribution). The smallest errors occurred for the direct launching. The size of the errors increased as both the size of the gap and the delay increased with changes in the size of the gap having a larger effect than changes in the delay.

To confirm these observations, the log-transformed error scores were subjected to a repeated measures analysis of variance (ANOVA) to identify differences in error as a function of gap and delay. There was a significant main effect of gap [$F(3,123) = 250.41, p < .01$], a significant main effect of delay [$F(3,123) = 63.84, p < .01$], and a significant gap × delay interaction [$F(9,369) = 33.54, p < .01$]. The significant interaction was followed by planned linear contrasts of the effect of delay for each of the gaps. The analyses revealed significant linear trends of delay (ps $< .01$) at each of the gaps, except for the 4-cm gap. The effect of delay was most prominent for the 0-cm (*M* slope = 0.492) and 1-cm gap animations (*M* slope = 0.248) and much weaker for the 2-cm (*M* slope = 0.072) and 4-cm gap animations (*M* slope = 0.046, $p = .053$).

As a final analysis, we examined individual differences in sensitivity to the gaps and delays in the launching animations. An analysis of intraclass correlation (Model 2 of Shrout & Fleiss, 1979) of each participant's average error for each animation revealed good agreement [$r(42) = 0.72, p < .01$]. To identify the major patterns of behavior in order to examine possible individual differences, we performed a cluster analysis (Ward's hierarchical cluster analysis of the standardized data; Ward, 1963) of the individual participant data and focused our attention on the highest clustering level (two clusters, $n = 27$ and $n = 14$). Because of the high intraclass correlation, the differences between the clusters were small. The ordinal effects of gaps and delays were nearly identical, with one exception. The larger cluster exhibited a very clear separation of the four curves shown in Figure 2, whereas the smaller cluster exhibited no sizable separation between the 0-cm and the 1-cm curves, except at the 0.0-sec delay (indeed, the 1-cm curve was above the 0-cm curve at delays longer than 0.0 sec). Overall, participant performance on the predictions task was remarkably consistent.

Ratings task. All 42 of the participants who completed the predictions task successfully completed the ratings task. Figure 3 shows the mean rating for each animation as a function of gap and delay for each of the task orderings (predictions first vs. ratings first). The direct launching animations produced predictably high causal ratings (recall that this animation was used in the instructions as an example of an interaction that should receive a rating of 9). Although increasing the size of the gap and the delay consistently produced lower ratings, there was

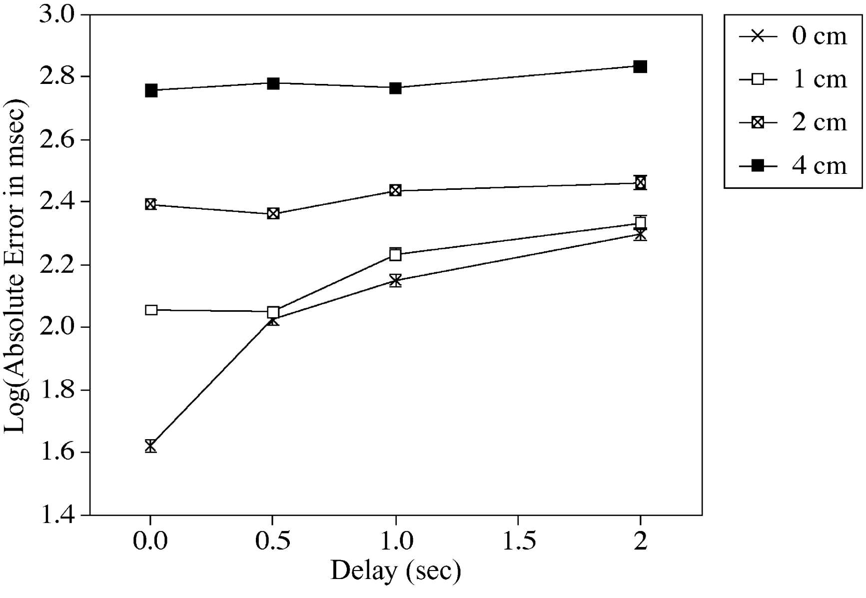

Figure 2. The mean logarithmic-transformed absolute error of response timing as a function of the size of the gap and the delay (Experiment 1). The most accurate predictions are those with low error scores. Error bars indicate ± one standard error, but the bars are so small that they are largely obscured by the line symbols.

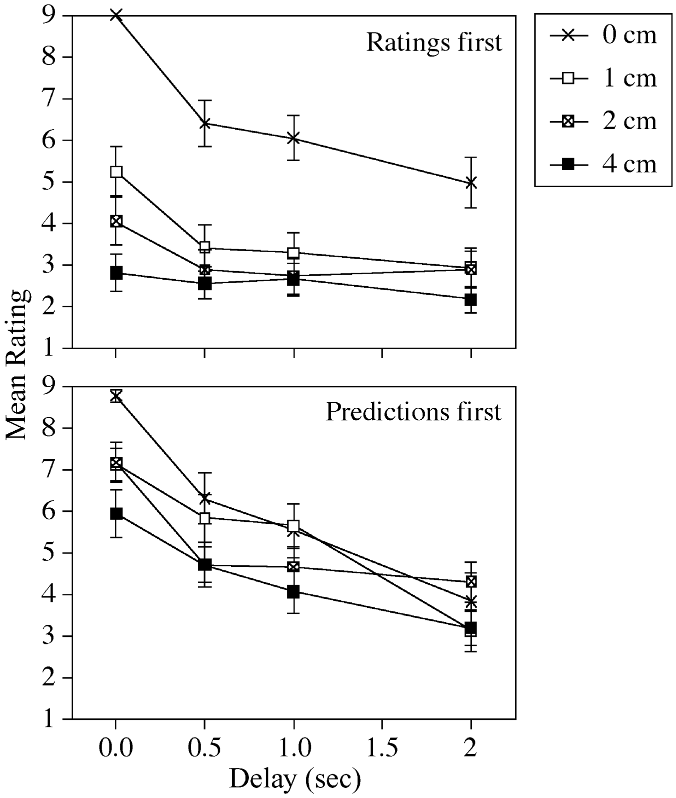

Figure 3. The mean rating of causality as a function of the size of the gap and the delay (Experiment 1) for the condition in which the ratings task preceded the predictions task (top graph) and the condition in which the predictions task preceded the ratings task (bottom graph). A rating of 9 indicates a judgment of high causality, and a rating of 1 indicates a judgment of low causality. Error bars indicates ± one standard error.

a clear effect of task ordering. The impact of gaps on causal ratings was significantly less after experience with the predictions task (i.e., the difference in ratings for the 0-cm gap and the 4-cm gap was less) than before this experience.

To confirm these observations, the log-transformed error scores were subjected to a mixed ANOVA to identify differences in ratings as a function of gap and delay for each task ordering. There was a significant main effect of gap [$F(3,120) = 48.22, p < .01$], a significant main effect of delay [$F(3,120) = 63.68, p < .01$], and a significant gap × delay interaction [$F(9,360) = 6.21$, $p < .01$]. The effect of task order was revealed by a significant main effect of task order [$F(1,40) = 7.42, p < .01$], a significant task order × gap interaction [$F(3,120) = 11.94, p < .01$], and a significant task order × delay interaction [$F(3,120) = 5.26, p < .01$].

The significant gap × delay interaction was followed by planned linear contrasts of the effect of delay for each of the gaps. The analyses revealed significant linear trends of delay ($ps < .01$). The effect of delay was most prominent for the 0-cm (*M* slope = -3.18) and 1-cm gap animations (*M* slope = -2.16) and much weaker for the 2-cm (*M* slope = -1.38) and 4-cm gap animations (*M* slope = -1.23).

As a final analysis of ratings, we examined individual differences in sensitivity to the gaps and delays in the launching animations. An analysis of intraclass correlation of each participant's rating for each animation revealed poor agreement [$r(42) = .30, p < .01$]. The intraclass correlations within each task ordering were not much higher [$r(17) = .33, p < .01$ for the predictions-first group and $r(25) = .38, p < .01$ for the ratings-first group]. Clearly, the participants did not agree on their ratings of the causality present in the 16 launching animations.

To identify the major patterns of behavior, we performed a cluster analysis (Ward, 1963) of the individual participant data and focused our attention on the level involving four clusters. The four behavioral profiles are shown in Figure 4. Because of the low intraclass correlation, the differences across clusters were substantial. For purposes of explanation, we have labeled the clusters "direct only" (because only the direct launching received a high rating; $n = 11$), "touching only" (because only the 0-cm gap animations received high ratings; $n = 4$), "delay important" (because ratings were primarily affected by the size of

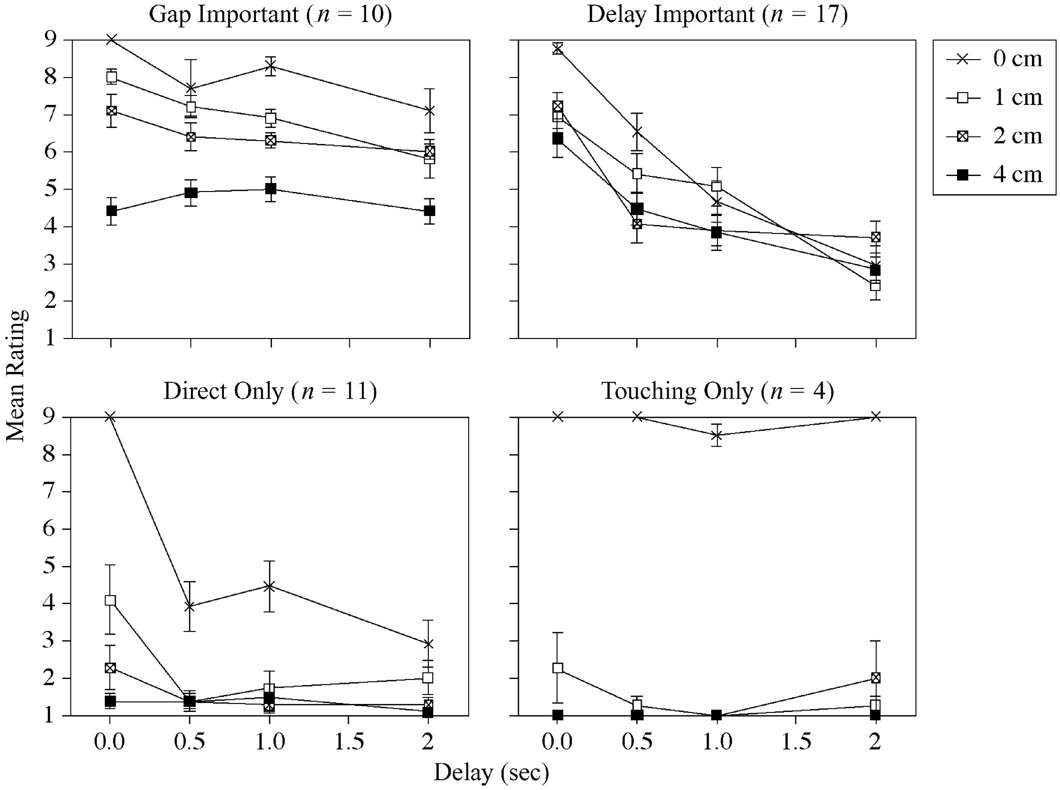

Figure 4. The mean rating of causality as a function of the size of the gap and the delay for each cluster of participants in Experiment 1 (see the text for an explanation of the cluster analysis). A rating of 9 indicates a judgment of high causality, and a rating of 1 indicates a judgment of low causality. Error bars indicate ± one standard error.

the delay; $n = 17$), and "gap important" (because ratings were primarily affected by the size of the gap; $n = 10$).

The decidedly different behavioral profiles either reflect differences in a priori expectations about causality or differences in causal perceptions. To gain insight into these differences, we examined the relative number of participants in each cluster who received the ratings task first or the predictions task first. Not surprisingly, experience with the predictions task altered the cluster profile; Table 1 reveals fewer clusters after this experience than before it. It is apparent that many (10 of 25) of our participants had a limited definition of causality at the outset of the study; they associated causality with direct contact and an immediate effect. After experience with the predictions task, only one of the participants maintained this narrow definition of causality, and none of the other participants judged contact between the interacting objects to be critical.

Comparison of predictions and ratings task performance. In our final set of analyses, we examined the relationship between performance on the predictions task and on the ratings task. We calculated individual participant Pearson's correlations between causal rating and mean predictive error for the 16 animations. The results are shown in Figure 5. None of the participants evidenced a positive correlation; the median correlation for participants who received the ratings task first was −.64 and the median correlation for those who received the predictions task first was −.39. The correlations were uniformly negative (high causal ratings for low predictive errors) and decidedly stronger for those participants who received the ratings task first [$t(40) = 2.56$, $p < .05$]. The weaker correlations for those who performed predictions before ratings was a little surprising, although this result may have been partly due to restricted range, because causal ratings were consistently higher after exposure to the predictions task.

Discussion

Experiment 1 provided a wealth of data regarding the effects of time, space, and experience on behavior involving the launching effect. It also revealed problems

Table 1
Number of Participants in Each Behavioral Cluster in the Ratings Task of Experiment 1

Cluster	Ratings First	Predictions First
Direct only	10	1
Touching only	4	0
Delay important	6	11
Gap important	5	5

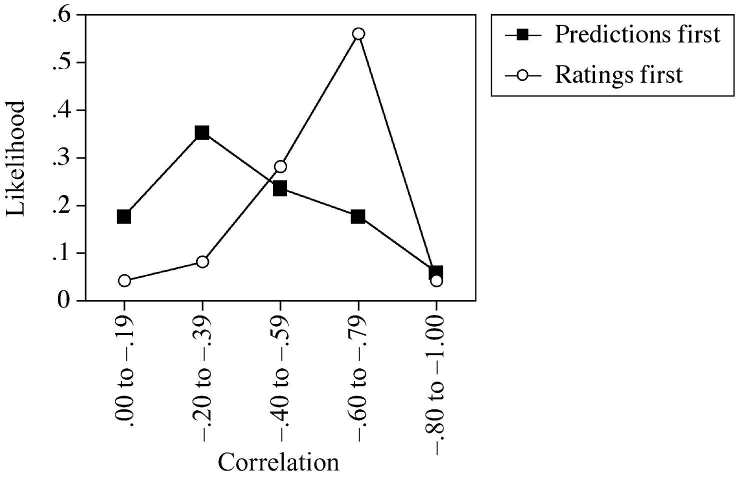

Figure 5. Frequency polygon of correlations between causal rating and mean predictive error for individual participants in each of the order conditions.

with the traditional use of explicit judgments of causality. Our original goal was to examine the relationship between predictability and causal judgments. Unfortunately, causal judgments proved to be volatile and unique to the individual (see Figures 3 and 4), thus making it difficult to form a general theory of causal judgment.

When participants were merely asked to predict the time of onset of the effect (movement of the second object), there were systematic effects of time and space on errors. These effects were uniform across individuals, and prior experience with the ratings task did not alter these effects. When participants were asked to judge the strength of the causal relation, there were systematic effects of time and space on ratings, but a clustering analysis revealed that these effects differed substantially across participants, and prior experience with the predictions task altered these effects (see Table 1 and Figure 4). Apparently, people have different naive causal theories that can be readily altered after brief exposure (less than 45 min) to predicting outcomes. Specifically, after multiple observations of the launching effect, causality at a distance was more readily accepted, whereas temporal contiguity was considered more central. In his model of causal learning, Pazzani (1991) described these naive causal theories as a "theory of causality" (a set of domain-independent principles) that influences causal judgments generated when observing a particular event or series of events.

The rapidity of change in the profile of causal judgments highlights a shortcoming in Michotte's (1946/1963) original thesis that causality in the launching effect is directly perceived. Experiment 1 requires that either (1) perceptual learning occurred during exposure, (2) experiencing perfect contingencies between distal and delayed events produced a reevaluation of one's theory of causality, or (3) participants recalibrated their use of the rating scale as a result of exposure. Hence, the causal percept as indexed by participant reports is not as constant as prior studies have assumed (see also Gruber, Fink, & Damm, 1957; Powesland, 1959).

How did our original hypothesis fare? We posited that increasing the predictability of the onset of the effect (movement of the second object) would increase the judged causal relation between the cause (the predictor) and the effect. Predictions error was moderately to strongly related to causal judgments (a median correlation of −.39 for the predictions-first group and −.64 for the ratings-first group). The weaker correlation for those participants who completed the ratings task second could reflect an issue of restricted range or any of a host of other factors. The interrater reliability of the ratings was also poorer for participants who performed the ratings task second ($r = .33$ vs. .38), which could contribute to the lower observed correlation.

EXPERIMENT 2

Thus far, we have simply assessed the impact of time and space on effect predictability and causal judgment. It is possible, however, to experimentally manipulate the predictability of the outcome through the use of exogenous cues and then examine the effects of these cues on judgments of causality. In the next set of experiments, we used auditory cues during the delay of a delayed launching in order to increase the predictability of the effect (cf. Michotte, 1946/1963, Experiments 80–82, in which a punctate "noise" was often accepted as an effect of a visual event but rarely as a cause of one; this noise was the sound of a hammer hitting a box and thus had no significant temporal extent that could improve predictability).

The predictability hypothesis suggests that an auditory cue would only be effective at increasing a judgment of causality if the onset of the cue was precisely predictable from the motion of the first object (to ensure that the first object is judged as the cause of the cue) *and* if the cue signaled the passage of time by changing in a predictable way during the delay, thus serving as a good predictor of the second object's motion onset.

To assess this prediction, we used three delay fillers—a constant tone, a tone that increased in amplitude during the interval, and a tone that decreased in amplitude during the interval. We attempted to psychophysically equate the changes in tone amplitude by increasing and decreasing the logarithm of the amplitude at a constant rate (Fechner, 1860/1966; Stevens, 1957; see Figure 6). The constant tone was predicted to have little impact on predictability and, hence, causal judgments, whereas the modulating tones were predicted to increase predictability and causal judgments. The original direct launching and unfilled delay launching animations were included as baselines to compare the effectiveness of the manipulations.

Experiment 2 used a response-based predictions task, and Experiment 3 used a causal impressions task. We used the predictions task to determine whether the insertion of various tones differentially affected predictability, and we used the causal impressions task to determine whether the tones had analogous effects on causal judgments.

Method

Participants. A total of 67 students enrolled in an introductory psychology course at Southern Illinois University at Carbondale served as voluntary participants, none of whom served in Experiment 1. They received course credit for their participation.

Materials. The direct launching and 1-sec delay animations used in Experiment 1 were supplemented with three animations that were visually identical to the delay animation. The latter animations included three different 1-sec auditory tones, one of which may have occurred during the delay (Figure 6 illustrates the changes in amplitude as a function of time for the increasing and decreasing tones).

Procedure. The basic procedure was identical to the predictions task used in Experiment 1. In the experiment proper (i.e., after the familiarization phase), each participant observed five blocks of 15 consecutive presentations of each of the animations. The assignment of animation to block was counterbalanced between participants by using a Latin square design.

Results and Discussion

Four participants were eliminated due to failure to follow directions; two failed to make a prediction on a number of trials, and the other two were at least 2 sec late for over half of their trials. The first trial of each animation was again excluded from the analysis.

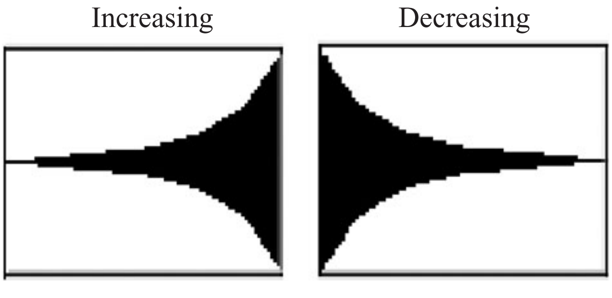

Figure 6. A graphical illustration of the changes in amplitude for the increasing and decreasing auditory tones used in Experiments 2 and 3. Although the raw amplitude increases at a changing rate throughout the interval, the logarithm of the amplitude changes at a constant rate.

Figure 7 shows the changes in mean logarithm of the error score for each animation. As predicted, the direct launching animation produced the least error, and the increasing animation produced error lower than that of the constant and delay animations. Surprisingly, the decreasing animation produced relatively high error; performance for this animation was not much different from that for the constant and delay animations.

A repeated measures ANOVA of the logarithmic-transformed error scores as a function of animation confirmed these observations. There was a significant main effect of animation [$F(4,248) = 41.03, p < .01$]. A post hoc Tukey's HSD test of the animation main effect revealed the following differences: direct launching < increasing ≤ decreasing = delay = constant (ordered from least to greatest error); the comparison between the increasing and decreasing animations did not reach significance ($p < .10$), but the comparisons between the increasing animation and the delay and constant conditions were significant (ps $< .05$).

Although the increasing tone improved the predictability of the effect as indexed by an increase in predictive accuracy relative to the empty delay, the decreasing and constant tones did not. The failure to observe a difference in the predictive efficacy of the decreasing tone may have been due to the participants' failure to identify a predictable amplitude level at which to initiate a response; the differences may have simply been too small to be helpful in our behavioral task. Given the nature of the task, participants must decide to respond at a point in time that will produce an actual response when the second ball begins moving. Producing an accurately timed response requires that perceptible changes occur at a useful time.

EXPERIMENT 3

Although our behavioral measure of predictability failed to reveal a significant facilitating effect when a decreasing auditory tone filled the delay, it is still possible that the improved predictability afforded by the predictably changing increasing and decreasing tones might improve the strength of causal judgments in delayed causation. A priori causal theories may influence responding (participants may imagine hidden mechanical events making the noise), or Experiment 2 may have lacked the statistical power to detect the increased predictability of the decreasing tone (recall that the mean error for the decreasing tone was intermediate, not statistically different from the increasing tone or from the constant or absent tones).

Method

Participants. A total of 38 students enrolled in an introductory psychology course at Southern Illinois University at Carbondale served as voluntary participants, none of whom served in Experiments 1 and 2. They received course credit for their participation.

Materials. The animations were identical to those used in Experiment 2.

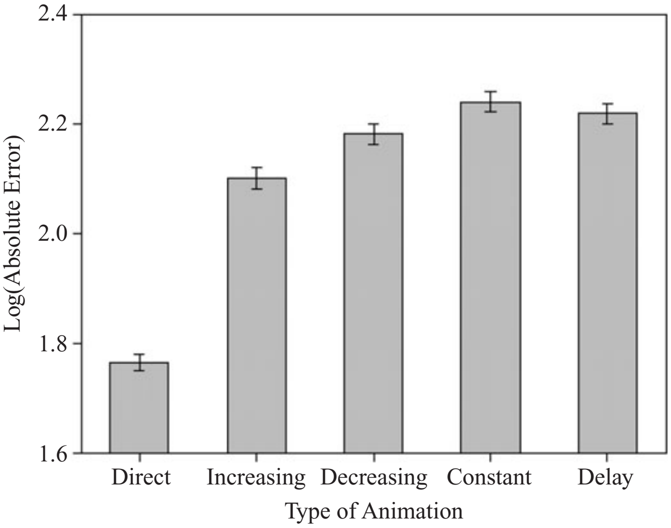

Figure 7. The mean logarithmic-transformed absolute error of response timing as a function of the type of animation (Experiment 2). The most accurate predictions are those with low error scores. Error bars indicate ± one standard error.

Procedure. The basic procedure was nearly identical to the ratings task used in Experiment 1. The experiment proper presented the five animations three times (randomly ordered). Multiple observations were used to ensure familiarity with the task.

Results and Discussion

Six participants were dropped from the study due to a failure to follow directions; four participants produced a mean rating for the direct causation animations of less than 7.0, one participant failed to respond on the majority of the trials, and one participant chose a rating of 9 on every trial. The following analyses were conducted on the remaining 32 participants.

Figure 8 shows the mean rating for each animation. The direct launching animations produced predictably high causal ratings. The increasing, decreasing, constant, and delay animations produced much lower causal ratings, with the increasing and decreasing animations producing higher ratings than the others, as originally predicted.

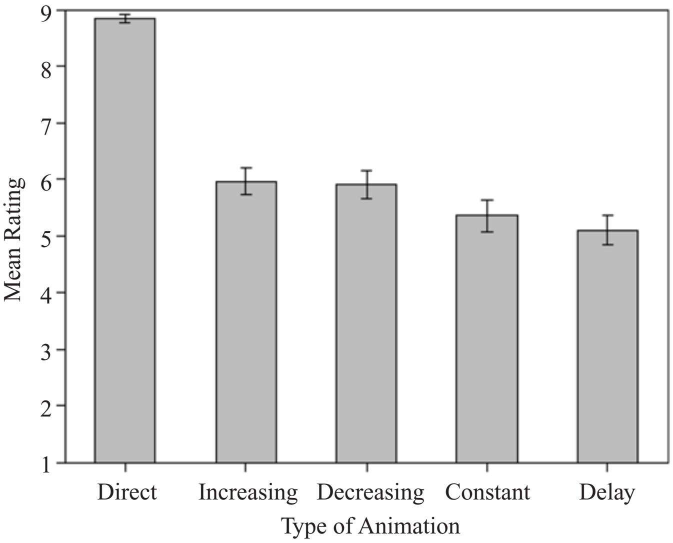

Figure 8. The mean rating of causality as a function of the type of animation (Experiment 3). A rating of 9 indicates a judgment of high causality, and a rating of 1 indicates a judgment of low causality. Error bars indicate ± one standard error.

To confirm these observations, causal ratings were subjected to a repeated measures ANOVA as a function of animation. The direct launching animations were excluded from this analysis; homogeneity of variance was significantly violated by their inclusion, and comparisons involving this animation were superfluous. The analysis revealed a significant main effect of animation [$F(3,93) = 4.30, p < .01$] A post hoc Tukey's HSD test of the animation main effect revealed the following order: increasing = decreasing ≥ constant = delay (the comparisons of the increasing and the decreasing animations to the constant animation approached significance [ps $<$ .10], and comparisons of the increasing and the decreasing animations to the delay animation reached significance [ps $< .05$]).

Increasing and decreasing tones thus succeeded in producing equivalent improvements in the judged causality of the delayed launching. We had attempted to psychophysically equate the changes in tone amplitude for the increasing and decreasing animations; this resulted in roughly similar effects of increasing and decreasing tones on judgments of causality and on predictability (the effect of the decreasing tone on predictability, however, was equivocal).

An alternative explanation of these results involves a priori causal theories; perhaps the participants drew on prior experience with motors ramping up to full power or the unwinding of springs. Although we are presently conducting studies that use different types of predictably changing auditory stimuli to address the generality of our results, we have noticed that people can quite readily generate plausible causal accounts for nearly any kind of tone filler (e.g., engines revving to describe oscillations, cartoon character special effects for a variety of auditory mediators). Unfortunately, resolving the issue will require more than a single study. Are the intermediate events effective because they complete a plausible mechanistic chain similar to one in the observer's past, or are they effective because they increase outcome predictability, which then prompts the observer to generate a mechanistic explanation (cf. Subbotsky, 2004)? It might be possible to create a situation that is so dissimilar to anything in human experience that the only recourse is to appeal to predictability. It is also possible that a sufficient number of parametric studies like Experiment 1 might prove adequate. Only time will tell.

GENERAL DISCUSSION

A visibly causal interaction, direct launching, produced high accuracy in observers' predictions of the time of onset of an effect. When temporal or spatial contiguity was absent, the accuracy in effect-onset predictions was considerably lower. The results of Experiment 1 suggest an interesting conclusion; if an observer can predict the time of onset of one event (the effect) with high precision, and another event (the putative cause) is an excellent predictor of that onset, the observer may be more likely to respond as if the interaction were causal.

Two additional noteworthy results bear on this conclusion. First, in Experiment 1 there was considerable disagreement among the participants regarding the strength of the causal relation as a function of space and time. Preexperimental causal theories did seem to have a significant impact on causal judgments (see Figure 4). Second, these differences were significantly altered after experience with the predictions task; the four distinct patterns of judgment behavior observed in those seeing the animations for the first time were largely reduced to two for those who had performed the predictions task first (see Table 1). These two results suggest that causal judgments are heavily influenced by a priori causal theories and experience in observing the relationship, even when contingencies are held constant (100% in the present study). Future research in this domain should explore additional methods for assessing people's causal expectations and examine the psychometric properties of our current and future measures.

This study represents a step toward understanding the role of temporal predictability in causal judgment. If causes are those events that serve both to predict whether and when an effect will occur, manipulations that increase the contingent *or* temporal relationship between two events should eventually result in an increase in the perceived strength of a causal relationship between those events. Our first investigations testing this prediction revealed that increasing the temporal predictability of the outcome in a delayed launching produced slightly higher causal judgments (Experiment 3). Future research will need to examine the ability of this theoretical approach to explain the diverse consequences of the various intermediate objects and events that can bridge spatial and temporal gaps (e.g., Gruber et al., 1957; Hubbard & Favretto, 2003; Michotte, 1951/1991; Young & DeBauche, 1993).

Causal impressions, however, are influenced by factors other than predictability. For example, a plausible cover story may result in an observed interaction being judged more causal without an accompanying change in effect predictability (e.g., Buehner & May, 2002, 2003). The latter attributions of causality, however, are verbally mediated and may sometimes serve the same purpose as the more obvious external ones. There are numerous examples of such verbally mediated causal judgments in the absence of personal observation of the causal relationship (e.g., smoking causes cancer, eating too much candy will make you sick). There are also, however, many examples of our ability to generate explanations for cause–effect relations that do not exist (Nisbett & Wilson, 1977). Such explicit verbal theories can override or substantially bias the conclusions reached by direct observation, but this biasing does not always serve us well.

A natural rejoinder to the proposed "predictability mediates causality" account is indoctrinated into every scientist—correlation does not imply causation. But why must we indoctrinate this into young scientists? *Because people quite readily infer causation from mere correlation.* Our research interest is in the psychological nature of causality, not its metaphysical nature. Thus, we do not mean to imply that only predictability is necessary to properly identify a causal relationship in the true, metaphysical sense. But we do mean to imply that predictability will tend to produce an attribution of causality unless such an inference is blocked by prior knowledge and experience or by a competing psychological process (e.g., the tendency to treat a sequence of claps or musical notes as a single event). Thus, predictability may not be a necessary or sufficient determiner of causal judgment, but it is likely an important component.

What about the oft-mentioned situations where predictability exists but no causality is present? Surely we all know that day does not cause night, nor does the evening news cause sunset. But let us not be so quick to judge. Many cultures throughout the ages have reached precisely such causal theories from their observation of the temporal regularities of the world. The positions of the stars in the skies were thought to cause floods, bad luck, and future fortunes (indeed, many still believe that the position of the stars and the phase of the moon can determine the future; just check your horoscope). Various oracles have interpreted chance predictive regularities as indicative of some underlying causal process (e.g., by observing the pattern of tea leaves or entrails to divine the will and intentions of higher powers), and even modern-day market analysts use data mining techniques related to regression (a prediction tool) to determine which market features might predict changes in stock prices. Although these modern seers do not necessarily believe that their trend data *cause* future changes in the market, they do believe that some form of causal relation must underlie any systematic regularity; perhaps the analyst is measuring an underlying cause or there is a common cause producing both the predictor values and the market's future. Changes in barometer readings can predict weather changes but do not cause them. However, the barometer *is* a measure of an underlying cause (air pressure) and, in the absence of knowledge, a naive observer might very well reach the conclusion that barometers *do* change the weather (modern societies call this "magical thinking"). Indeed, people attribute causal properties to the pixels on a computer monitor when speaking of the actions of one's cursor or in computer animations of the launching effect even though these pixels are merely outward signs of underlying causal mechanisms and not causal agents in themselves. What constitutes a causal explanation at one level may merely be an illusion (Subbotsky, 2004).

The extent to which we can account for apparently diverse phenomena under a single unifying theme may yet surprise us. Failure to reduce causality to predicting *whether* and *when* will also help highlight the factors that go beyond predictability—factors like prior expectations derived from naive physical theories, individual differences in the use of a judgment scale, or perceptual biases.

Final Thoughts

Given the importance of predictability to survival, organisms are motivated to use the multiplicity of environmental cues to predict both whether and when future events will occur. Causal learning may involve assessing (1) a single predictor's ability to predict both whether and when an effect of interest will occur (e.g., in the launchings of Experiment 1), (2) a chain of predictors that culminate in the effect (like the launching object and tone of Experiments 2 and 3), or (3) multiple predictors that each signal different aspects of predictability (time of occurrence and likelihood of occurrence, as may be the case in occasion setting; Young, Johnson, & Wasserman, 2000). The possibly separate effects of the likelihood and time of occurrence is nicely demonstrated by a situation in which you see someone attempt to strike a match multiple times; it often fails to light the first few times it is struck, but when it eventually does light, it happens immediately after a strike. Thus, predicting when can be as important as predicting whether.

From a psychologist's perspective, Hume may have been right that "causal" is merely a word that describes a special type of relationship in the world. We posit that this relationship importantly depends on the ability of putative causes to predict whether, when, and (perhaps) where, how, and what will happen next. If one or more events jointly predict a wealth of information about the occurrence of a future event, a special relationship between those predictors and the predicted event is implied—a relationship that our species has labeled "causal."

REFERENCES

Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin*, **114**, 435-448.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral & Brain Sciences*, **14**, 471-484.

Buehner, M. J., & May, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking & Reasoning*, **8**, 269-295.

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgement of causality: Effects of prior knowledge, experience, and reinforcement procedure. *Quarterly Journal of Experimental Psychology*, **56A**, 865-890.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, **99**, 365-382.

Fechner, G. T. (1966). *Elements of psychophysics* (Vol. 1; H. E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, **4**, 123-124.

Gibbon, J. (1981). On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology*, **24**, 58-87.

Gilden, D. L., & Proffitt, D. R. (1994). Heuristic judgment of mass ratio in two-body collisions. *Perception & Psychophysics*, **56**, 708-720.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Gruber, H. E., Fink, C. D., & Damm, V. (1957). Effects of experience on perception of causality. *Journal of Experimental Psychology*, **53**, 89-93.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, **30**, 1128-1137.

Hubbard, T. L., & Favretto, A. (2003). Naive impetus and Michotte's "tool effect": Evidence from representational momentum. *Psychological Research*, **67**, 134-152.

Hume, D. (1969). *A treatise of human nature*. New York: Penguin. (Original work published 1739)

Michotte, A. (1963). *The perception of causality* (T. R. Miles & E. Miles, Trans.). London: Methuen. (Original work published 1946)

Michotte, A. (1991). The perception of the "tool effect." In G. Thinès, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 87-103). Hillsdale, NJ: Erlbaum. (Original work published 1951)

Natsoulas, T. (1961). Principles of momentum and kinetic energy in the perception of causality. *American Journal of Psychology*, **74**, 394-402.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, **84**, 231-259.

Pazzani, M. J. (1991). A computational theory of learning causal relationships. *Cognitive Science*, **15**, 401-424.

Powesland, P. F. (1959). The effect of practice upon the perception of causality. *Canadian Journal of Psychology*, **13**, 155-168.

Rachlin, H. C. (1966). Scaling subjective velocity, distance, and duration. *Perception & Psychophysics*, **1**, 77-82.

Schlottmann, A., & Anderson, N. H. (1993). An information integration approach to phenomenal causality. *Memory & Cognition*, **21**, 785-801.

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

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420-428.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181.

Subbotsky, E. (2004). Magical thinking in judgments of causation: Can anomalous phenomena affect ontological causal beliefs in children and adults? *British Journal of Developmental Psychology*, **22**, 123-152.

Suppes, P. (1984). *Probabilistic metaphysics*. New York: Blackwell.

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 35-42). Cambridge, MA: MIT Press.

Tolman, E. C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review*, **42**, 43-77.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**, 236-244.

White, P. A., & Milne, A. (1999). Impressions of enforced disintegration and bursting in the visual perception of collision events. *Journal of Experimental Psychology: General*, **128**, 499-516.

Worley, J. K., & Markley, R. P. (1969). Distance discrimination in a reduced cue setting. *Psychonomic Science*, **17**, 237-238.

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

Young, M. E., & DeBauche, B. (1993). Causal mechanisms as temporal bridges in a connectionist model of causal attribution. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 1092-1097). Hillsdale, NJ: Erlbaum.

Young, M. E., Johnson, J. L., & Wasserman, E. A. (2000). Serial causation: Occasion setting in a causal induction task. *Memory & Cognition*, **28**, 1213-1230.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and cognition. *Psychological Bulletin*, **127**, 3-21.

(Manuscript received April 4, 2003;
revision accepted for publication June 4, 2004.)