Causation, causal perception, and conservation laws

CHARLES R. TWARDY
Monash University, Clayton, Victoria, Australia

and

GEOFFREY P. BINGHAM
Indiana University, Bloomington, Indiana

We investigated the perception of causation via the ability to detect conservation violations in simple events. We showed that observers were sensitive to energy conservation violations in free-fall events. Furthermore, observers were sensitive to *gradually* perturbed energy dynamics in such events. However, they were more sensitive to the effect of decreasing gravity than to that of increasing gravity. Displays with decreasing gravity were the only displays in which the energy profile was dominated by (apparent) potential energy, leading to an asymmetric trajectory.

When we see a soccer player kick a goal, we can see immediately what set the ball in motion. When a cue ball strikes another billiard ball and that second ball heads for the corner pocket, we can see that the motion of the cue ball has caused the once stationary ball to move. In simple cases such as this, we perceive causation to take place. But what is it to perceive causation? What is causality as a perceptible property of events, and what is the information that specifies causality?

In the study of event perception, the first step is to determine what properties of objects and events can be perceived and how they are perceived. David Hume (1739–1740/1973, 1748/1955) is well known for his discussions of the conditions under which causation is perceived. His method was introspection, rather than experimentation, and he chose billiards as his fundamental example. His answer was that our sense of causation is evoked by events that are constantly conjoined in our experience, where one is prior to the other and the two are contiguous in time and space.

Michotte (1963) provided the first experimental study of the perception of causation, using animated two-body collisions. By manipulating the timing of the collision, Michotte found that he could elicit or fail to elicit a judgment of causation in his observers. Contrary to Hume's analysis, repetition of some events did not suffice to make them look like causation, and others were judged as causal on the first try, even though they were unusual. This was a

first step toward discovering the underpinnings of the perception of causation. However, Michotte's work was limited by subjective measures and experimental manipulations that were not founded in a theory of the perception of the dynamic variables he was manipulating (Runeson, 1977/1983).

Others have extended Michotte's (1963) study of the perception of causation (see, e.g., Kruschke & Fragassi, 1996; Schlottman & Anderson, 1993; Schlottman & Shanks, 1992; Weir, 1978), but likewise have refrained from a dynamical analysis of the collision events and so have not shown what physical properties are actually perceived when observers see causality. On the other hand, the literature of event perception (see, e.g., Gilden & Proffitt, 1989, 1994; Hecht & Proffitt, 2000; Kaiser & Proffitt, 1987; Merfeld, Zupan, & Peterka, 1999; Proffitt & Gilden, 1989; Proffitt, Kaiser, & Whelan, 1990; Runeson, 1977/1983, 1995; Runeson & Frykholm, 1983; Runeson & Vedeler, 1993) has addressed the perception of dynamic quantities (e.g., relative mass, elasticity) but has not directly addressed the perception of causality. To do so, we must determine what property (or properties) in events correspond to perceived causality. Once this is established, we can then investigate sources of information about that property and whether observers are indeed sensitive to that information.

Dowe's (2000) conserved-quantity (CQ) theory is concerned with demonstrating what aspects of the physical world satisfy the causal relations. According to the CQ theory, *causal interactions* are marked by the exchange of CQs between or among *causal processes*. A causal process is a (space–time) trajectory of an object, be it a photon or a baseball. CQs are whatever quantities are actually conserved in nature, and these are taken to be those indicated by current physics, such as charge, energy, momentum, and angular momentum.

For example, a moving billiard ball has energy and momentum in proportion to its speed. It will also have other CQs, such as angular momentum. A stationary billiard ball also has energy and momentum and is still a process, even though (in our reference frame) the values may be zero. During a collision, the trajectories of two billiard balls briefly intersect, and the objects (processes) exchange energy and momentum. That constitutes a causal interaction. Two ocean waves possessing energy may also "collide," but they do not exchange energy, so their interaction is not causal in that way.

The CQ theory provides a principled tie between event perception and the perception of causality. It is not a replacement for Runeson's *kinetics specify the dynamics* (KSD) theory of event perception.[1] It is a philosophical account of causation that happens to tie in very nicely with something like KSD, which is not in itself a theory of causation. Event dynamics consist of exchanges of energy, momentum, and other CQs; the study of event perception has provided evidence that observers are, in many cases, sensitive to event dynamics (e.g., Bingham, Schmidt, & Rosenblum, 1995; Gilden & Proffitt, 1994; Hecht & Proffitt, 2000; Kaiser & Proffitt, 1984; Pittenger, 1990; Proffitt & Gilden, 1989; Runeson & Vedeler, 1993). Therefore, if physical causal interactions are just exchanges of CQs, then in perceiving events, observers also perceive some important aspects of causation itself, contrary to Hume.

This is not to suggest that observers have privileged access (Bingham et al., 1995) to Newtonian ideas of space, time, mass, energy, or momentum. Neither does it suggest that observers can always identify causes correctly. Observers cannot see most exchanges of charge. Even when they feel a shock, they cannot tell positive from negative. Observers have difficulty judging angular momentum (Proffitt & Gilden, 1989; Proffitt et al., 1990), although they can become reasonably skilled in certain domains, such as billiards (Hecht & Proffitt, 2000).

CQ theory does suggest that there are specific physical determinants to causal interactions and that these determinants are scientifically known quantities that have specific effects on the resulting kinematics. Event perception claims that the data in real observational situations are rich enough for observers to perceive at least the qualitative dynamics involved. If observers are sensitive to the kinds of dynamic variables (such as energy and momentum) involved in conservation laws, then on the CQ theory, observers are sensitive to causal interactions.

That ability alone implies nothing about the phenomenological manifestations of such a sensitivity. Observers need not be aware of CQs as such or be able to report violations of CQs explicitly, even if they are sensitive to such violations. So it will clearly not do to ask, "did you see energy conserved?" Furthermore, a falling object dissipates some of its energy to air friction and, thus, obeys the principle of conservation of energy only approximately. Similarly, a normal bounce is imperfectly elastic:

Some energy is lost. We do not expect observers to attune to perfect conservation, but only to that which they experience. Hume was right on that count: Experience is relevant to the perception of causation. What we are looking for is sensitivity to deviations from the normal exchanges of CQs.

Of course, observers can discriminate only between events for which they can discriminate the forms of motion: the trajectories. The form of the trajectories by which objects exchange energy and momentum over time constitutes the structure of an event. A particular kind of deviation from an "expected" trajectory should specify an appropriate hidden cause (e.g., it's not a bouncing ball, it's animate (Bingham et al., 1995). In cases in which the specified cause is *not* recognizable, the event should just look unnatural.

Indeed, when Pittenger (1990) used a naturalness rating task, he found that observers detected deviations as small as 0.1–0.2 sec from the natural period of a pendulum. Pittenger made the pendulum swing faster or slower than normal by driving it with a second, hidden pendulum. In free response, participants did not spontaneously report forces in the anomalous displays, but they did spontaneously describe the natural motion as "free of force." Apparently, people do not see gravity as a force at all, but as a natural part of (vertical) object behavior. They are attuned to this invariant in their environment. On the other hand, occasional experimenter errors resulting in out-of-place accelerations were detected immediately and were recognized as external forces or hidden causes, exactly as we would expect from the CQ theory.

OVERVIEW

In our experiments, we chose a lawful kind of violation that, in the context of the experimental displays, did not specify any kind of naturally occurring energy source or sink. We chose unmistakably inanimate events, since observers are known to be able to recognize animate motion (the most famous demonstration is the point-light walkers in Johansson, 1973). We chose a relatively simple event—namely, a single-body free-fall and bounce under gravity. We chose it, in part, because the perception of this event had been studied extensively before. McConnell, Muchisky, and Bingham (1998) found that observers shown displays of balls falling at various distances could tell how far away the balls were, even though the balls had the same optical size, indicating that the observers were using dynamical information or, at least, the average velocity (Hecht, Kaiser, & Banks, 1996; see also Saxberg, 1987). Warren, Kim, and Husney (1987) found that observers could judge the elasticity of a ball by viewing bounces. Bingham, Schmidt, and Rosenblum (1995) found that observers could distinguish real bounces from hand-driven bounces with the same period, suggesting, again, that the trajectory form is the source of information. Finally, Muchisky and Bingham (2002) found that observers are indeed sensitive

to such variations in trajectory forms (see also Wickelgren & Bingham, 2001).

We showed observers both possible and impossible motions and asked them to rate the naturalness. Assuming the CQ thesis, if observers are sensitive to violations of conservation laws and if we have truly avoided the conflation of our events with other common natural events, observer responses on naturalness rating tasks should indicate the extent to which observers are sensitive to causal relationships.

EXPERIMENT 1

We used displays similar to those in Muchisky and Bingham (1992) and McConnell, Muchisky, and Bingham (1998). Observers viewed simulations of balls falling freely and then bouncing on a hard surface. The ball was always simulated at the same distance and was always the same size and mass. The bounce event occurred against a ground texture, with proper occlusion of background elements by foreground elements, so that the relative distance was specified by ground contact.

The main goal was to test observer sensitivity to a gradually introduced conservation violation achieved by allowing gravity to vary over time. We compared the perception of this kind of conservation violation with that of simple sudden-onset violations achieved by setting the elasticity of the ball greater than one. We also examined whether observers were equally sensitive to both additions and deletions of energy. There were therefore two kinds of displays: elasticity and varying gravity.

In the elasticity displays, the ball's elasticity was varied, from $e = 0.6$ to $e = 1.2$, with gravity remaining earth-normal.

In the varying-gravity displays, the gravitational field g was made to increase or decrease during the course of the display, following the equation $g \rightarrow g + \gamma t$, where γ is just the constant by which gravity changed over time, so that $\gamma = 0$ for normal displays. Five values of γ were chosen: -0.2, -0.1, 0, $+0.2$, and $+0.4$. The maximum rate of decrease ($\gamma = -0.2$) was chosen so that gravity approached, but did not quite reach, zero by the end of 4.5 sec. The maximum rate of increase ($\gamma = +0.4$), twice the magnitude of the maximum rate of decrease, was chosen so that successive bounces were never higher (unlike for $e \geq 1.1$), eliminating one obvious sign of unnaturalness. In the varying-gravity displays, elasticity was always 0.9, the bounce that looked most natural in an informal pilot study. An elasticity of 0.9 corresponds to steel-on-steel or super-ball bounces.

Although the displays were intermixed, these manipulations were carried out separately. Each manipulation had five levels. In addition, in order to control for duration effects, all the displays were constrained to last 2.5, 3.5, or 4.5 sec. In the elasticity experiment, the ball was always dropped from the same maximal height[2] of 3.5 m.

In the varying-gravity displays, once we fixed the duration, we had a choice. We could maximize the number of bounce cycles or the distinctness of each bounce. More cycles would give the observer greater variation in bounce period. However, once duration was fixed, to show more cycles we would have to lower the drop heights and reduce the trajectory information. Alternatively, we could maximize the drop height and make larger, more distinct cycles but sacrifice the information from variations in bounce period. We did not know whether it would be more important for observers to witness a greater number of cycles or fewer but larger and more distinct cycles. Therefore, there were two conditions: First, we kept the drop height constant and maximal at 4.5 m and cut off the display whenever the time was up; second, we varied the drop height so as to guarantee that each display had exactly five ground strikes in the allotted time[3] (fixed number of bounces condition). As we will see, the fixed-height condition was much easier for observers and was the only condition used in follow-up studies.

Method

Observers. Twelve graduate and undergraduate students at Indiana University participated in the experiment. All had normal or corrected-to-normal vision. The observers were paid $5/h.

Display generation. A large (0.54 m) heavy (6.8 kg) ball was simulated at a viewing distance of 50 m. The ball appeared as a two-dimensional black outline with gray shading and a black radial line to indicate orientation and rotation. The background was a sagebrush texture gradient (see Figure 1). QuickTime movies of simple planar events were created using event-dynamic models generated by the Interactive Physics II application (1992). All models included components for gravity and air resistance. A Runge-Kutta algorithm was used to calculate positions at successive time intervals of $\Delta t_{calc} = 0.0015$ sec. Every 10th such position was used to generate an animation frame, so that $\Delta t_{display} = 0.015$ sec, or 66.7 Hz. The Macro-Mind Accelerator (1989) application was used to lock the frame rate of the simulations to the refresh rate of the monitor (66.7 Hz), resulting in smooth animation without aliasing. Interactive Physics II does not simulate compression during bounce, so this was approximated by cutting off the bottom 2 pixels of the QuickTime movie, resulting in a flattening of the ball upon impact.

Animations were generated in advance for each of the 45 conditions investigated. In each display, the ball appeared stationary at its starting location and began moving when observers clicked the mouse. At the end of the movement, the ball disappeared. A trial consisted of a bounce sequence repeated three times in response to observer mouse clicks, followed by a judgment (written on paper). The whole set of 45 trials was presented in a predetermined random order, broken up into two blocks of 22 and 23 trials separated by a small break. Each observer saw three repetitions of the 45 trials. For each repetition, a different ordering for the trials was used.

The displays used the full 24.8 × 18.5 cm (640 × 480 pixel) display of the Macintosh II, and the observers sat approximately 0.5 m from the display.

Instructions. Instructions were presented one sentence at a time on screen in response to observer mouse clicks, in a slide-show format. The observers saw the following (where semicolons separate slides and each sentence was given a bullet "•"):

You will view simulations of a simple inanimate event: a ball bouncing. The simulated ball is almost 2 feet in diameter, and about 15 pounds. The ball appears as if it were about 165 feet away. Some bounces will be natural throughout, some will not. You are to judge how natural the events appear. Base your judgment on the way the ball moves. ;
Things you do not have to worry about: The balls start at different heights, from almost 0 to offscreen. They all start moving to the right at
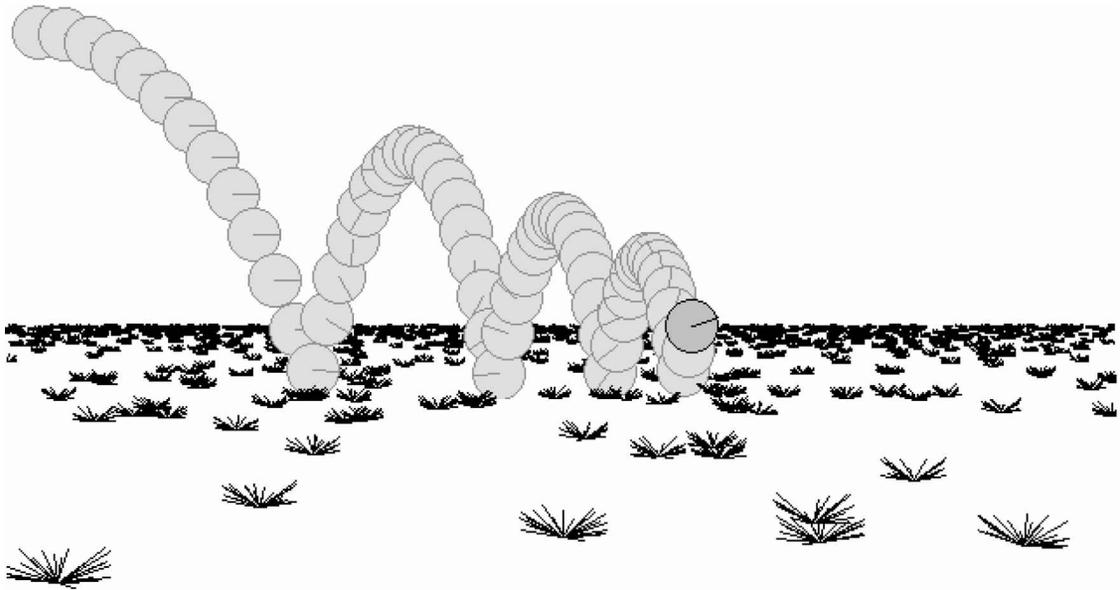
Figure 1. Display for Experiment 1: an elapsed-time view of the bounce trajectory for normal gravity ($\gamma = 0$), normal elasticity ($e = 0.9$), and air friction proportional to velocity. Initial height, 4.5 m; 300 frames sampled every 8 frames.

the same speed. They will all disappear suddenly, sometimes soon, sometimes late. The spinning is always natural and is there to help you. When the ball hits, it may get a little flat on the bottom: That's OK. The sagebrush is for perspective only. The ground is actually quite hard. ;

OK, here's the setting: Imagine you are sitting in a chair at ground level. Imagine this monitor is a window to the ground outside. The sagebrush is about 6 inches high. Look at the form of motion. Look at the timing of the motion. Ask yourself if that could be a real event's motion. ;

Final instructions: Each event will repeat (exactly) 3 times, when you click. Try to use all 3 times to make your judgment. Rate the event from 1 (unnatural) to 5 (natural) on your scoresheet. If it looks unnatural, write down why, if you can. You will have 6 blocks of 22 or 23 events each. You may take breaks between the blocks of trials. ;

Ask now if you have any questions. . . .

In short, the observers were told to watch all three repetitions and to rate the naturalness, paying special attention to the form and timing of the motion. The first 5 observers were asked to follow their rating with a short free-response description of what, if anything, seemed unnatural.

Because it took so long to describe what seemed unnatural, Observers 5–12 were told to perform the free-response task for only one of the sets of 45 trials.

Procedure. The observers were seated in front of the computer with a mouse, an answer sheet, and a pencil in front of them. An architect's lamp illuminated the answer sheet, and the overhead lights were shut off to reduce glare on the screen and help with the illusion that the observer was looking out through the monitor. The observers read the instructions, looked over the response sheets, and had the opportunity to ask the experimenter any further questions. They then worked through a practice block of six trials in order to become familiar with the mechanics of the program and the range of naturalness of the displays. No feedback was given during the practice block. After answering any further questions, the experimenter then left the room. Experimental sessions lasted about 1 h.

Results

Elasticity data. We performed an analysis of variance (ANOVA) on the data from the 12 observers. There was a

main effect of duration [$F(2,20) = 3.54, p < .05$], indicating that at shorter durations, the observers had more difficulty discriminating natural and unnatural displays. More important, there was a statistically significant effect of elasticity [$F(4,40) = 18.76, p < .001$].

Figure 2 illustrates that the observers detected $e = 1.1$ and $e = 1.2$ as being very unnatural. Although perceived naturalness appears to show a peak at $e = 0.9$, a Tukey HSD post hoc analysis on the data combined for all durations shows that the two points ($e = 0.6, e = 1.0$) were not significantly different from $e = 0.9$. Likewise, the points $e = 1.1$ and $e = 1.2$ were not significantly different from each other, although both were significantly different from all of the other points (using $p = .05$ as the significance level).

There was a statistically significant two-way interaction between duration and elasticity [$F(8,80) = 3.19, p < .005$]. Figure 2 shows that this interaction was due chiefly to the unnatural (high-elasticity) events' looking even more unnatural at longer durations, as one would expect.

Because the data were response frequencies, we compiled two frequency tables for the responses (Table 1) to verify the results of the ANOVA. The left side of Table 1 shows the total response frequencies by elasticity, for all durations. The right side shows the response frequencies for the longest duration (4.5 sec).

From both sides of Table 1, we can see that the observers were most confident that $e = 0.9$ was natural. They were beginning to find $e = 1.0$ unnatural, but they definitely rated $e \geq 1.1$ as unnatural. The bimodal distribution at $e = 0.6$ indicates some confusion: On the one hand, it generated more 5 ratings than any other category, but on the other hand, there were about as many 1 as 5 ratings.[4]
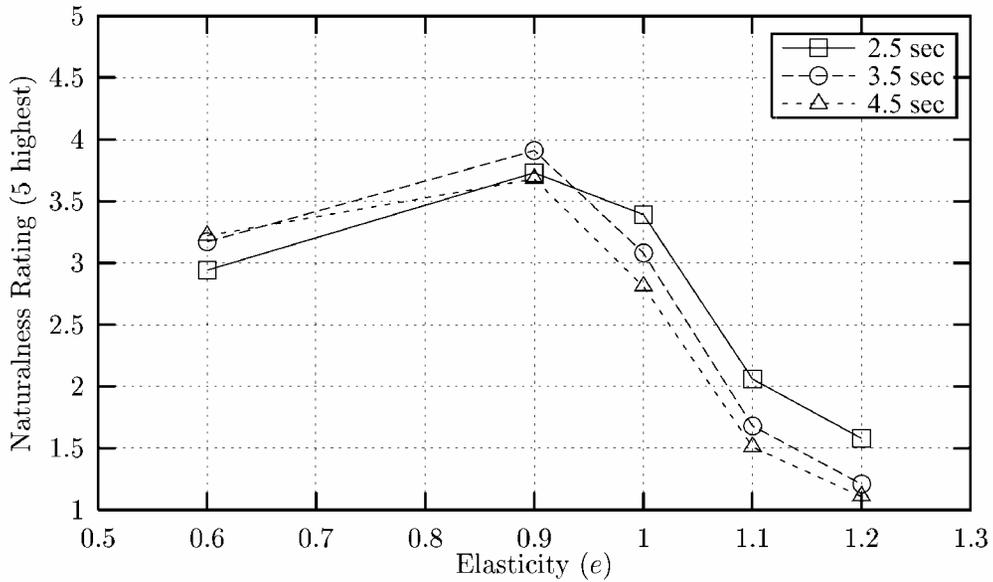
Figure 2. Ratings of naturalness according to level of elasticity.

Gravity study. An ANOVA with order as a between-subjects variable yielded main effects of gravity [$F(4,40) = 29.71, p < .001$], duration [$F(2,20) = 14.67, p < .001$], and fixed-condition [whether number of bounces or initial height was held fixed; $F(1,10) = 8.71, p < .05$]. There were statistically significant interactions between gravity and fixed-condition [$F(4,40) = 22.74, p < .001$] and between gravity and duration [$F(8,80) = 3.65, p < .005$]. A three-way interaction between gravity, duration, and order of presentation was also statistically significant [$F(8,80) = 2.28, p < .05$].

The main effect of gravity reflected the fact that displays simulating decreasing gravity appeared unnatural, although the participants did not discriminate between normal gravity and increasing gravity (using a Tukey test, $p > .5$ for all conditions). The main effect of duration indicates that, once again, shorter durations made discrimination more difficult. The main effect of fixed condition indicates that the observers were better able to discriminate when the initial screen height was large, resulting in sizable and well-defined bounces.

The two-way interaction between gravity and fixed condition (see Figure 3) confirms that when the number of bounces was held fixed (making some bounces quite small), discriminability dropped almost to zero [a Tukey HSD test revealed that in the upper curve, only the extreme left point ($\gamma = -0.2$) could be discriminated, and only in the 4.5-sec condition]. In contrast, on the curve depicting a fixed initial height, the observers discriminated all three leftmost points.

The two-way interaction between gravity and duration falls naturally out of those two main effects. At short durations, the participants were less sensitive to the gravity manipulations. At longer durations, they became more sensitive to decreasing gravity, but not to increasing gravity.

One potential worry was that the number of bounces (which varied according to initial height and duration, in addition to gravity) might be a better predictor than gravity. However, a multiple regression of gravity, duration, and number of bounces on naturalness rating was statistically significant [$F(3,176) = 29.9, p < .001; R^2 = .30$] and yielded statistically significant effects for gravity (partial $F = 15.5, p < .001$) and duration (partial $F = 5.9, p < .02$), but no statistically significant effect for number of bounces ($p > .2$).

The two frequency tables for the responses to the gravity manipulations are given in Table 2. The left side shows the total response frequencies by gravity for all durations.

Table 1
Elasticity Displays: Response Frequencies

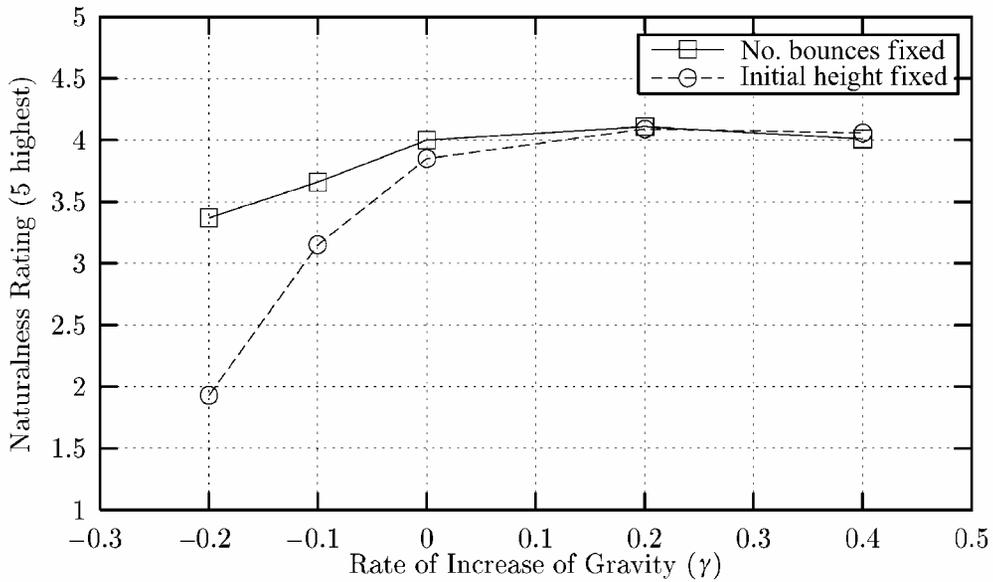|	Response Rating: All Durations					Response Rating: 4.5 sec				
e	1	2	3	4	5	1	2	3	4	5
0.6	28	15	10	23	30	7	7	2	7	11
0.9	1	10	25	46	24	1	4	8	16	7
1.0	6	29	34	27	12	4	11	12	6	3
1.1	50	42	9	6	1	21	12	2	1	0
1.2	86	14	2	2	2	34	1	0	1	0
Total	171	110	80	104	69	67	35	24	31	21

Figure 3. Two-way interaction between gravity and fixed condition.

The right side shows the response frequencies for the longest duration (4.5 sec). In both cases, they were computed by using all the data, not just those of fixed initial height, so if anything, they should underestimate statistical significance.

As in the elasticity analysis, multiple χ^2 analyses support the intuitive observation that the rows in the tables are different from each other and from a uniform distribution.[5] As a whole, both tables show statistically significant relationships well below the $p = .001$ level. Row-by-row analyses confirm what you see by inspection: The observers thought that decreasing-gravity bounces looked less natural than a normal bounce and that increasing-gravity bounces looked as natural as (or perhaps more natural than) normal bounces.

As a final sanity check, we computed χ^2 for nonmatched samples to see whether there was a statistically significant difference between the elastisty data for the $e = 0.9$ displays and the gravity data for the $\gamma = 0.0$ displays. Since these displays were generated with the same parameters, there should have been no significant difference. There was no significant difference ($p > .1$).

Discussion

Observer responses to the elasticity displays showed sensitivity to unnatural increases of energy during the bounce. It will be helpful to think of this as the negative-entropy condition. Other studies (Bingham, 1995; Bingham et al., 1995; Gibson & Kaushall, 1973; Pittenger, 1990; Wickelgren & Bingham, 2001) on event dynamics, time-reversed motions, and animate versus inanimate dynamics have shown similar sensitivity to negative-entropy motion. There was no statistically significant sensitivity to normal but relatively inelastic bounces.[6]

In the varying-gravity displays, gravitational energy leaked into or out of the system smoothly and continuously, making it harder to perceive a distinct energy source or sink than in the elasticity displays. Also in contrast to the elasticity displays, there were two ways for an event to be unnatural: It could have an energy source strong enough to overcome entropic losses, making the local motion at least negatively entropic, or it could have an energy sink in addition to regular entropy.

If we thought only in terms of entropic or negative-entropic motion, as in the elasticity displays, we would pre-

Table 2
Gravity Displays: Response Frequencies

γ	Response Rating: All Durations					Response Rating: 4.5 sec				
	1	2	3	4	5	1	2	3	4	5
-0.2	53	59	34	42	26	43	21	2	3	1
-0.1	10	41	59	60	44	9	17	22	13	10
0.0	0	17	49	83	67	0	9	16	24	23
$+0.2$	1	9	38	86	81	1	1	16	25	28
$+0.4$	2	19	34	76	84	1	7	10	26	28
Total	66	145	214	347	302	54	55	66	91	90

dict that observers should be sensitive to negative-entropic motion (increasing gravity), but not to entropic motion (decreasing gravity). However, the effect of decreasing gravity is *not at all like* the effect of increasing entropy. With decreasing gravity, the bounces do get slower, but they also remain quite high, and the time course of the bounce is, therefore, much different. The bounces have different space–time trajectories.

So, following the lead of the CQ hypothesis, we predicted that both increasing and decreasing gravity should specify a hidden cause (which, of course, they did), but because that cause was so far outside the observers' experience, instead of seeing a changing gravitational field, they would see *both* manipulations as cases of unnatural motion.

In fact, this is not what happened. The observers were quite sensitive to the cases of decreasing gravity ($\gamma < 0$), but not at all to cases of increasing gravity ($\gamma > 0$).

It is not really surprising that the observers considered the cases of decreasing gravity to be unnatural. What is puzzling is why they did not also consider the cases of increasing gravity to be unnatural. The first question to ask is whether the observers could even *detect* any difference between increasing and normal gravity.

We ran a follow-up experiment to find out whether the observers could discriminate between normal cases and increasing-gravity cases (even though they might not correctly interpret a detected difference).

EXPERIMENT 2

We reasoned that a fairly straightforward *same–different* task would answer the question. Rather than rating naturalness, the observers had only to say whether two displays were of the same kind (the discrimination task). The observers were instructed not to latch onto any peculiarities (such as in which weed a ball landed), but to judge on the basis of the overall motion. They were given feedback after each trial. After a session in that condition, the observers returned on a different day and were then given a two-alternative forced-choice (2AFC) task without feedback. On the 2AFC task they had to decide which of two displays was the more natural (the naturalness task). The idea was to train the observers to discriminate, using feedback, and then later to add the task of deciding which was more natural. If the observers could discriminate but not interpret correctly, their performance should have been good on the discrimination task and poor on the naturalness task.[7]

Method

Observers. Five undergraduate students at Indiana University participated in the experiment. All had normal or corrected-to-normal vision. The observers were paid $7/h.

Display generation. The displays were generated like the varying-gravity displays in Experiment 1, with a few exceptions. First, only the fixed-duration 4.5-sec (300-frame) displays were used. Second, we randomized the initial heights' coming from either the low range (3.3–3.8 m) or the high range (4.3–4.8 m), so that each pair had one *high* and one *low*, in random order. Likewise, we randomized the

initial horizontal velocity, also in two groups: low (1.0–2.0 m/sec) and high (3.0–4.0 m/sec). In addition, we randomized the direction of motion for each pair (from the left edge of the screen moving right, or conversely). Again, all the displays were generated and sequenced in advance. Finally, the ball was given a textured surface.[8]

Instructions. Instructions were presented slide-show style, as before. On the discrimination task, the observers were instructed to check *same* or *different* after watching a pair of bounces and to indicate how sure they were. On the naturalness task, they were told to choose which one looked more normal. As before, they were told to pay attention to the form and timing of the motion, and not to fixate on minor details. In addition, they were told that the initial positions and velocities were slightly randomized, to prevent them from being able to cue off of specific conditions. In the discrimination task, they were told that they would receive feedback. In the naturalness task, they were told that they would not. After the naturalness task, the observers completed a short free-response form.

Procedure. The observers were seated in front of the computer and made their answers on printed answer forms, as before. The overhead lights were left on this time. Alternate observers were counterbalanced for order of presentation. Twelve blocks were presented in about an hour: 6 blocks of decreasing gravity and 6 of increasing gravity. Each block within a gravity condition had 12 trial pairs, in different orders. Letting "N" stand for *natural* and "U" for *unnatural*, each block had 4 <NU>, 4 <UN>, and 4 <NN> pairs. The <UN> and <NU> pairs were divided equally between the two levels of unnaturalness (e.g., $\gamma = -0.1$, $\gamma = -0.2$). Therefore, each block yielded four data points for each level of γ.

Results

The discrimination d' data are presented in Table 3. Recall that in Experiment 1, the observers rated decreasing gravity as unnatural, but not increasing gravity. The purpose of Experiment 2 was to find out whether the observers, nevertheless, could distinguish hypergravity cases from normal cases, when the task did not involve explicit naturalness judgments. If we examine the increasing-gravity portion of Table 3 (the bottom half), we see that the observers *do* appear to be able to distinguish these hypergravity displays from the ordinary ones.

However, oddly, the top half of Table 3 indicates that the observers were unable reliably to distinguish decreasing-gravity displays from normal ones, which is the opposite

Table 3
Percentage Correct (%) and d' Data for the Discrimination Task (*Same–Different*)

Observer	Hits	False Alarms	%	±	d'	±
Hypogravity (Collapsed Over $\gamma = -0.1$ and -0.2)						
1	.67	.58	.59	.66	0.79	0.62
2	.71	.58	.61	.64	0.97	0.63
3	.54	.50	.53	.68	0.51	0.61
4	.73	.54	.64	.63	1.19	0.63
5	.77	.50	.68	.59	1.47	0.64
Hypergravity (Collapsed Over $\gamma = +0.2$ and $+0.4$)						
1	.81	.67	.65	.62	1.01	0.66
2	.85	.42	.76	.50	1.99	0.66
3	.67	.33	.67	.60	1.62	0.64
4	.60	.46	.58	.66	0.98	0.62
5	.85	.33	.79	.45	2.23	0.68

Note— $n = 24$ <NN>, $n = 48$ <UN> or <NU>. Error values (±) are 95% confidence intervals.

Table 4
Percentage Correct (%) and d' Data for the Naturalness Task
("Which is more natural?", 2AFC)

Observer	Hits	False Alarms	%	±	d'	±
Hypogravity (Collapsed Over $\gamma = -0.1$ and -0.2)						
1	.71	.29	.71	.84	0.78	0.74
2	.94	.11	.92	.21	1.99	0.86
3	.75	.01	.83	.38	1.46	0.74
4	.83	.06	.89	.27	1.81	0.83
5	.81	.31	.75	.51	0.97	0.64
Hypergravity (Collapsed Over $\gamma = +0.2$ and $+0.4$)						
1	.38	.75	.31	.89	−0.70	0.75
2	.47	.36	.56	.67	0.20	0.59
3	.42	.69	.36	.63	−0.51	0.60
4	.03	.94	.04	.11	−2.48	1.07
5	.44	.61	.42	.66	−0.30	0.58

Note—$n = 36$. Error values (±) are 95% confidence intervals.

of what we found in Experiment 1. In fact, none of the observers achieved more than threshold (75% correct) in the hypogravity condition, and all the observers were at chance for recognizing <NN> displays as *same* on the discrimination task in the hypogravity condition and were only slightly above chance for the hypergravity condition.

The naturalness task data show a different pattern (see Table 4). There was a clear difference in discriminability for hypogravity and hypergravity, and the observers could do the hypogravity task reasonably well but could not do the hypergravity properly, as we saw in Experiment 1.

Discussion

The data are encouraging but suggest that the observers did not understand part of the *same–different* task. If we look just at the hypergravity data, the experiment has done exactly what we wanted: The observers discriminated hypergravity from normal gravity when asked whether they were the same or different, but judged the hypergravity displays to be at least as natural as the normal ones, showing that the asymmetry in Experiment 1 was due not to an inability to *discriminate*, but to a genuine impression of naturalness.

On the discrimination task (*same–different*), the observers were able to distinguish hypergravity trials from normal trials, with d's ranging from 1.0 to 2.2. However, that very solid discriminability performance fell apart on the naturalness task (2AFC). Four of the observers had d's not significantly different from 0, whereas 1 observer exhibited nearly perfect discriminability, but with the semantics completely reversed: Observer 4 clearly identified the hypergravity trials each time they occurred and thought that they looked more natural than the normal trials.

However, for hypogravity displays, the observers did worse on the easier *same–different* questions than on the harder *which is more natural* task. In fact the *same–different* task was quite ambiguous, because *same* did not mean *identical*—since no displays were identical—but *same kind*. As was noted, discriminability performance was un-

expectedly low on the *same–different* task, a task we already knew (from Experiment 1) that observers could do. Much of the poor performance is explained by the fact that the observers remained at chance on judging <NN> trials: They did not understand what was being asked and had to guess.

Consequently, hypogravity performance was better on the no-feedback naturalness (2AFC) task, which means that for *these* displays at least, the separate tasks did not perform their intended functions. Therefore, to make sure that the hypergravity results of Experiment 2 were not adversely affected by these issues, we ran a short follow-up that addressed the shortcomings of Experiment 2 by making the instructions and the task more straightforward.

EXPERIMENT 3

Method

Observers. Two observers were run, a graduate student and a Bloomington resident in his mid-twenties.[9] The observers were paid $6/h, with a $20 incentive reward for best performance. All the observers had normal or corrected-to-normal vision.

Display generation. The displays for Experiment 3 were a subset of those from Experiment 2. We removed the intermediate values of γ from the stimuli, so the observers saw only $\gamma = +0.4$, $\gamma = 0$, and $\gamma = -0.2$. An *example* slide show was added to the instructions. The examples showed several cases that were to be called *same* and contrasted them with extreme cases of *different*, so that the observers would know the kind of overall differences that interested us.

Procedure. The experiment proceeded much as Experiment 2. The order of presentation for both of the observers was the same on both days. Six blocks of decreasing gravity were followed by six blocks of increasing gravity. Each block within a gravity condition presented the same 24 trial pairs, in different orders. Letting "N" stand for *natural* and "U" for *unnatural*, each block had 6 <NU>, 6 <UN>, and 12 <NN> pairs. Therefore, each <NN> cell had 12 × 6 = 72 data points, as did the combined <UN> and <NU> cells.

Results

The results were very clear. On the *same–different* task, both observers had very high d's for both decreasing and increasing gravity conditions. On the 2AFC naturalness task, the observers had high positive d's for the hypogravity cases and high negative d's for the hypergravity cases (see Figure 4). Therefore, the proper way to interpret the asymmetry in Experiment 1 is that the observers were or would have been able to distinguish hypergravity displays from natural displays but that, nevertheless, these displays did not appear unnatural. The observers were sensitive to the difference but unaware of its semantic content. This result confirmed Experiment 2, without Experiment 2's unfortunate glitch in observer response to the hypogravity condition.

Looking at the free responses, both Observers 1 and 2 found the *same–different* task hard and the naturalness task easy. Both indicated attending to the overall timing and shape of the motion but did not discover any rules. Their performances resembled those of the observers in the previous experiments.
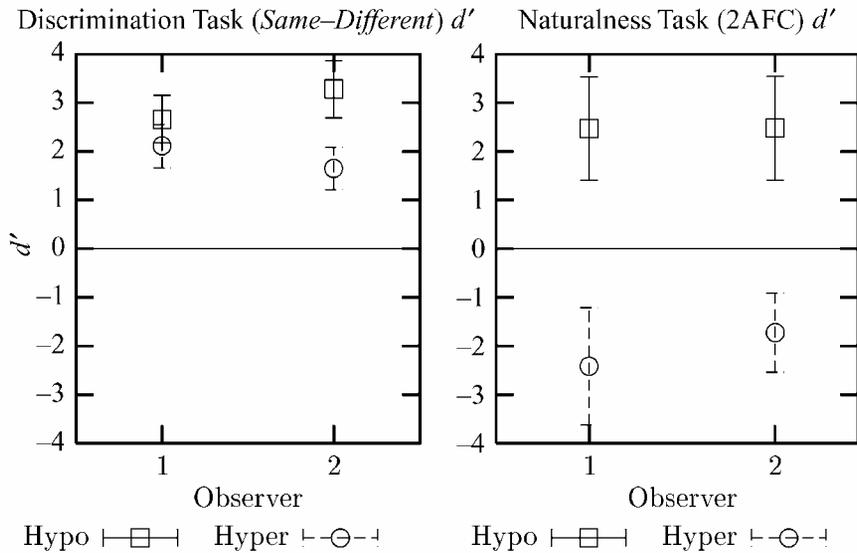
Figure 4. Experiment 3 d' data. The discrimination task on the left shows that both observers could distinguish both hypo- and hypergravity from normal. The naturalness task on the right shows that hypergravity looked very natural, whereas hypogravity looked unnatural.

GENERAL DISCUSSION

In Experiment 1, the observers showed an asymmetric sensitivity to energy violations (see Figure 3). To determine why, we first had to establish whether the apparent insensitivity to increasing gravity was a true indistinguishability or merely a question of what looked natural or unnatural. In Experiments 2 and 3, we determined that the observers could discriminate between the hypergravity bounces and the normal gravity bounces but that the hypergravity bounces did not look unnatural. In fact, the hypergravity bounces looked *more* natural than the reference case. This is puzzling. Can the results be reconciled with the CQ theory or even with the general sensitivity to dynamics supposed by KSD?

Figure 5 shows the graph of energy versus time for three bounces: quickly decreasing gravity ($\gamma = -0.2$), normal gravity ($\gamma = 0$), and quickly increasing gravity ($\gamma = +0.4$). The top graph displays true energy, taking the changing gravity into account. The bottom graph displays apparent energy: the energy computed from the kinematics while assuming that gravity is constant at earth-normal ($\gamma = 0$).

Within each series, there are discontinuous losses when the ball hits the ground, because $e < 1$. In normal gravity ($\gamma = 0$), energy loss *between* bounces is due only to air friction, which is highest near the bounces, when velocity is highest. For $\gamma = 0$, total energy decreases monotonically. For the two unnatural cases, what happens depends on whether we consider true or apparent energy.

If we consider true energy, the data suggest that the observers thought rapidly *decreasing* energy was more un-

natural than rapidly *increasing* energy—a puzzling result indeed, considering the elasticity data from Experiment 1. However, if we look at *apparent* energy in the bottom graph, the distinction is not so neat. The salient difference between $\gamma = -0.2$ and $\gamma = +0.4$ then is not the overall increase or decrease in apparent energy, which is about the same, but *when* it increases and *when* it decreases.

In the case of an increase of the gravitational field ($\gamma = +0.4$), when the ball reaches the ground, it is going faster than it should, given its starting height, so apparent energy increases during the descent. After the bounce, as the ball leaves, it is traveling quite quickly, but does not get as high as it should, given its rebound speed: Apparent energy decreases during the ascent, hitting a local minimum at the apex of the ball's trajectory.

In the case of a decrease of the gravitational field ($\gamma = -0.2$), the ball falls more slowly than it should, reaching the ground with less speed than one would expect: Apparent energy decreases during the descent. After the rebound, the ball rises higher than it should, given its rebound speed: Apparent energy increases during the ascent.

The asymmetries in observer sensitivity mimic asymmetries in the entropy dynamics of real events. In a real bounce, the potential energy at apex could, at most, almost equal the kinetic energy exiting the previous bounce, but it could assume almost any value below that, depending on how much the ball is affected by friction. So, from experience, observers should have a tight upper bound on expected height but a broad lower bound, favoring the hypergravity bounces.

To analyze the descent, it will be convenient to split the energy into kinetic and potential energy. Figure 6 breaks
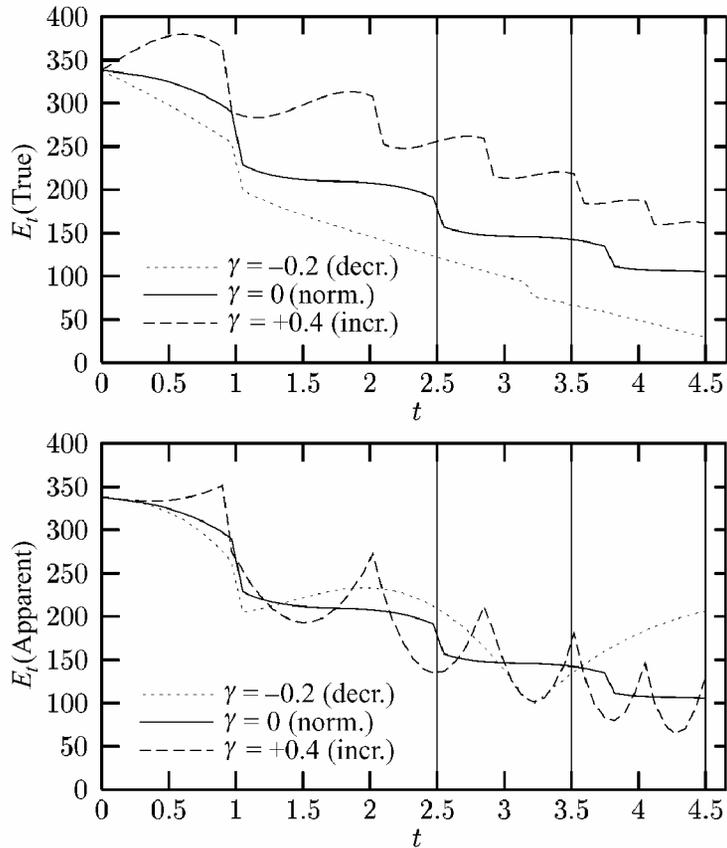
Figure 5. Comparative energy profile of total energy ($E_t = E_K + E_P$) versus t for three principal values of γ, where $g \rightarrow g(1 + \gamma t)$. In the top graph, energy is measured in the objective, or simulation, frame of reference. In the lower graph, *apparent* energy is measured, as if gravity were constant at earth-normal. The vertical lines show the cutoff times for the 2.5-, 3.5-, and 4.5-sec durations.

down the energy profile into kinetic and potential energy, so that we may see how the energy is exchanged over time, rather than just looking at the total energy, E_t. The right column provides the true values, but we are interested in the apparent values, which are in the left column.

The middle left figure shows the normal energy profile: E_P and E_K exchange dominance of the energy budget about halfway through each arc, and the curve for E_t combines both and resembles neither.

For increasing gravity ($\gamma = +0.4$, the top left figure), E_t is dominated by and resembles in form the curve for E_K. However, it is still true that E_P and E_K alternate about midway through each arc, and the E_K arcs are roughly symmetric about their minima.

For decreasing gravity ($\gamma = -0.2$), the curves differ more strongly. In the lower left graph, we can see that now E_P accounts for most of E_t most of the time, especially after the first bounce, and that the E_K curves are highly skewed (asymmetric) about their minima. One symptom of the asymmetric E_K curve is the laconic descent of the ball, a motion described by many observers as "floaty." In the lower left graph, E_K increases too slowly during de-

scent, meaning that the ball falls much too slowly when compared with the height lost, given its relatively rapid ascent.

Both increasing and decreasing gravity are conservation violations. In particular, they upset the normal balance between the *exchange* of E_K and E_P. Both increasing and decreasing gravity differ in form from the normal case. Unlike normal gravity and increasing gravity, however, only decreasing gravity results in markedly *asymmetric* trajectories, reflected in E_K profiles asymmetric about their minima at the bottom of Figure 6.

Since the observers were sensitive only to decreasing gravity, we tentatively conclude that observers detected this asymmetry. The result was that they preferentially detected an excess of potential energy for the amount of kinetic (or total) energy, rather than the other way around.

A potentially simpler explanation is that the observers picked up on the progression of bounce frequencies. In a natural bounce, the interbounce interval shortens, meaning that the bounce frequency increases with each successive bounce. Bounces under increasing gravity also have increasing bounce frequencies. However, bounces under
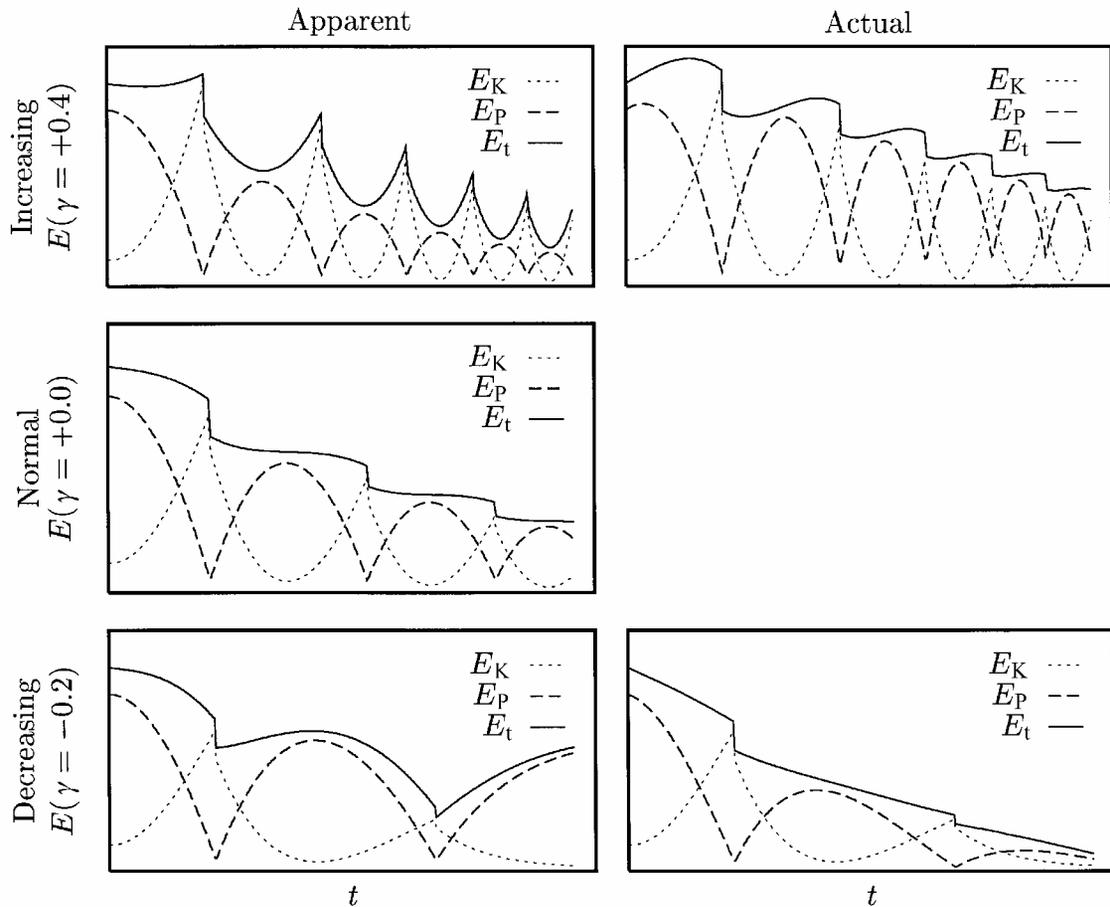
Figure 6. Kinetic and potential energy: Each row shows the relationship between E_K and E_P for one level of γ. The left column shows apparent E versus t. The right column shows actual E versus t (for row 2, the actual is the apparent, so there is a blank).

decreasing gravity (as well as hyperelastic bounces) have decreasing frequencies, marking them as unnatural.

Figure 7 shows that bounce frequency does divide neatly on either side of $\gamma = +0.0$, and indeed, 1 participant in Experiment 1—a psychology graduate student and a musician—specifically mentioned that, after awhile, he began to judge timing between bounces, using increasing frequency as a criterion for naturalness.

Nevertheless, we do not think that bounce frequency is the whole story. Even the participant noted above came to that criterion by trying to discern why some bounces looked more or less natural. In addition, other participants' free-response data pointed to properties of the overall trajectory: The observers frequently described the decreasing-gravity bounces as too "floaty," an apt description of the ball's tendency to rise too high for its speed and to be rather laconic in its subsequent descent. Even the observer noted above also mentioned that the unnatural bounces seemed to have oddly asymmetric trajectories.

Finally, despite our instructions and, indeed, the information specified in the trajectories, the observers may not have regarded the stimulus as a large heavy ball very far

away. After all, what they saw was a 1-cm disk about 0.5 m in front of them. However, although it is true that a small ball seen up close will match the hypergravity bounce more closely than any others, there is still a huge difference. By the time the hypergravity ball has had one bounce, the 1-cm ball has had six.

CONCLUSION

First, we have demonstrated that observers are sensitive to gradually perturbed energy dynamics. Second, we have demonstrated that observers are more sensitive to the effect of decreasing gravity than to that of increasing gravity. Third, we have begun to show how observers combine information about kinetic and potential energy in their assessments of the naturalness of a free-fall and bounce. Fourth, we have thereby provided some support for the idea that perceptions of causation are tied to exchanges of energy and momentum.

The observers could discriminate only event types that had different trajectory forms. Both increasing and decreasing gravity generated significant changes in the en-
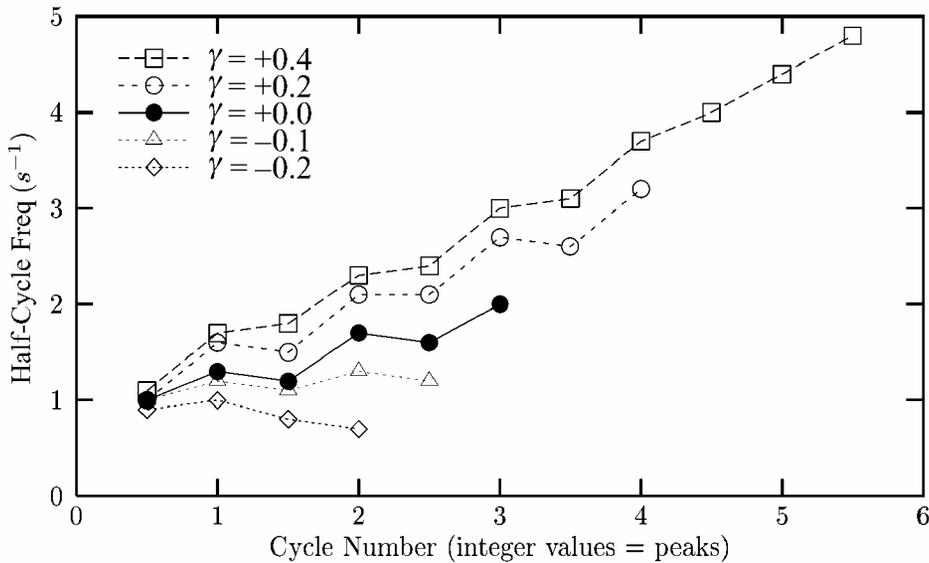
Figure 7. Graph of bounce (and apex) frequency between successive extrema (bounce or apex) for various values of γ.

ergy profile of a bounce. However, under decreasing gravity, potential energy dominated the trajectory, and the kinetic energy trajectory was asymmetric. The observers thought such trajectories unnatural.

The observers did not find the trajectories under increasing gravity to be unnatural, even though they could discriminate them from normal trajectories in a *same–different* task. Such trajectories were unnaturally dominated by kinetic energy but still had basically symmetric exchanges of kinetic and potential energy.

Since actual causation in this world is tied to the exchange of conserved quantities, such as energy and momentum (Dowe, 2000), we expect that the *perception* of causation is likewise tied to such exchanges of physical quantities. If so, we can extend Runeson's (Runeson, 1977/1983) KSD theory to the perception of causation.

In a natural motion, the proper exchange of energy and momentum among visible objects reveals that these objects are, in fact, the causes of the resultant motion. Motions that require a cause that can neither be seen nor rapidly inferred will be judged to be unnatural. An otherwise "unnatural" motion can be explained, or made sense of, by adding an appropriate and unobserved cause. In our experiment, it was very difficult to hypothesize the actual hidden cause: a dynamic change of gravity. Therefore, some of the displays did not look natural.

That displays with increasing gravity did look natural suggests that humans give more weight to excesses of potential energy (given the kinetic or the total energy) when detecting unnatural motion or causal anomalies. We would not expect humans to be perfect conservation detectors, because humans do not observe perfect conservation in the world. Therefore, motion with negative entropy should look unnatural, in part because it specifies either reversed-

time motion or backward causation. However, in our experiment, the observers detected motion with negative entropy only when it resulted in an asymmetric energy profile.

Human ideas of causality are more general than the kinds of causes that can be so perceived. However, accounts of "higher order" causation, such as the more reasoning-and-judgment approach to causation pursued by Cheng and Spellman (Cheng, 1997; Cheng & Holyoak, 1995; Glymour & Cheng, 1998; Spellman, 1996) and the causal modeling framework of Judea Pearl (Pearl, 2000), all rely on some prior, primitive notion of causal *mechanism.* In related studies, Ahn and colleagues (Ahn & Bailenson, 1995; Ahn, Kalish, Medin, & Gelman, 1995) have shown that humans prefer mechanism-style explanations to even more predictively powerful covariational accounts. We think that this prior notion of mechanism stems from the perception of actual physical causation, which is governed by the exchange of conserved quantities among objects.

REFERENCES

Ahn, W.-K., & Bailenson, J. (1995). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology,* **31**, 82-123.

Ahn, W.-K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition,* **54**, 299-352.

Bingham, G. P. (1995). Dynamics and the problem of visual event recognition. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 403-448). Cambridge, MA: MIT Press, Bradford Books.

Bingham, G. P., Schmidt, R. C., & Rosenblum, L. D. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception & Performance,* **21**, 1473-1493.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review,* **104**, 367-405.

CHENG, P. W., & HOLYOAK, K. J. (1995). Complex adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In J.-A. Meyer & H. Roitblat (Eds.), *Comparative approaches to cognition* (pp. 271-302). Cambridge, MA: MIT Press.

DOWE, P. (2000). *Physical causation*. New York: Cambridge University Press.

GIBSON, J. J., & KAUSHALL, P. (1973). *Reversible and irreversible events* [Film]. (Psychological Cinema Register, Pennsylvania State College, University Park)

GILDEN, D. L., & PROFFITT, D. R. (1989). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 372-383.

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

GLYMOUR, C., & CHENG, P. (1998). Causal mechanism and probability: A normative approach. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 296-313). Oxford: Oxford University Press.

HECHT, H., KAISER, M. K., & BANKS, M. S. (1996). Gravitational acceleration as a cue for absolute size and distance? *Perception & Psychophysics*, **58**, 1066-1075.

HECHT, H., & PROFFITT, D. (2000). *What cues do billiard experts use? Conceptual and perceptual judgments of spin*. Unpublished manuscript.

HUME, D. (1955). *An inquiry concerning human understanding*. Indianapolis: Bobbs-Merrill. (Original work published 1748)

HUME, D. (1973). *A treatise of human nature*. London: Oxford University Press, Clarendon Press. (Original work published 1739–1740; reprinted from Selby-Bigge's 1888 edition)

Interactive physics II [Computer software]. (1992). San Francisco: Knowledge Revolution.

JOHANSSON, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**, 201-211.

KAISER, M. K., & PROFFITT, D. R. (1984). The development of sensitivity to causally relevant dynamic information. *Child Development*, **55**, 1614-1624.

KAISER, M. K., & PROFFITT, D. R. (1987). Observers' sensitivity to dynamic anomalies in collisions. *Perception & Psychophysics*, **42**, 275-280.

KRUSCHKE, J. K., & FRAGASSI, M. M. (1996). The perception of causality: Feature binding in interacting objects. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 441-446). Hillsdale, NJ: Erlbaum.

MacroMind Accelerator [Computer software]. (1989). San Francisco: MacroMind.

MCCONNELL, D. S., MUCHISKY, M. M., & BINGHAM, G. P. (1998). The use of time and trajectory forms as visual information about spatial scale in events. *Perception & Psychophysics*, **60**, 1175-1187.

MERFELD, D. M., ZUPAN, L., & PETERKA, R. J. (1999). Humans use internal models to estimate gravity and linear acceleration. *Nature*, **398**, 615-617.

MICHOTTE, A. (1963). *The perception of causality* (T. Miles & E. Miles, Trans.). London: Methuen. (C. A. Mace, Ed.; originally published as *La perception de la causalité*, 1946 and 1954.)

MUCHISKY, M. M., & BINGHAM, G. P. (1992). Perceiving size in events via kinematic form. In J. Kruschke (Ed.), *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 1002-1007). Hillsdale, NJ: Erlbaum.

MUCHISKY, M. M., & BINGHAM, G. P. (2002). Trajectory forms as a source of information about events. *Perception & Psychophysics*, **64**, 15-31.

PEARL, J. (2000). *Causality*. New York: Cambridge University Press.

PITTENGER, J. B. (1990). Detection of violations of the law of pendulum motion: Observers' sensitivity to the relation between period and length. *Ecological Psychology*, **2**, 55-71.

PROFFITT, D. R., & GILDEN, D. L. (1989). Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 384-393.

PROFFITT, D. R., KAISER, M. K., & WHELAN, S. M. (1990). Understanding wheel dynamics. *Cognitive Psychology*, **22**, 342-373.

RUNESON, S. (1983). *On visual perception of dynamic events*. (Acta Universitatis Upsaliensis: Studia Psychologica Upsaliensia, No. 9). Uppsala: Academiae Upsaliensis. (Original work published 1977)

RUNESON, S. (1995). Support for the cue-heuristic model is based on suboptimal observer performance: Response to Gilden and Proffitt (1994). *Perception & Psychophysics*, **57**, 1262-1273.

RUNESON, S., & FRYKHOLM, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, **112**, 585-615.

RUNESON, S., & VEDELER, D. (1993). The indispensability of precollision kinematics in the visual perception of relative mass. *Perception & Psychophysics*, **53**, 617-632.

SAXBERG, B. V. (1987). Projected free fall trajectories: I. Theory and simulation. *Biological Cybernetics*, **56**, 159-175.

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

SCHLOTTMAN, A., & SHANKS, D. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology*, **44A**, 321-342.

SPELLMAN, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science*, **7**, 337-342.

WARREN, W. H., JR., KIM, E. E., & HUSNEY, R. (1987). The way the ball bounces: Visual and auditory perception of elasticity and control of the bounce pass. *Perception*, **16**, 309-336.

WEIR, S. (1978). The perception of motion: Michotte revisited. *Perception*, **7**, 247-260.

WICKELGREN, E. A., & BINGHAM, G. P. (2001). Infant sensitivity to trajectory form. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 942-952.

NOTES

1. KSD states that the kinematics specify the dynamics. That is, some dynamical properties are fully specified by the kinematics of the event and, furthermore, can be directly perceived.

2. Any higher and the bounce with elasticity = 1.2 went off screen.

3. We could not do this for the elasticity experiment because, at low elasticities, only two or three bounces are distinguishable even for maximal drop heights. After that, the ball just rolls.

4. Several χ^2 analyses added little to the ANOVA: The patterns visible in the table were statistically significant.

5. A uniform distribution is the most obvious null hypothesis (really a *nil* hypothesis), but by no means the only one that should be investigated. We also checked against two other nulls: ($H2$) = all rows are no different from $\gamma = 0.0$ and ($H3$) = all rows are no different from the average row.

6. It is certainly possible to perceive low elasticity as unnatural. One can buy from teaching supply stores a pair of small, apparently identical rubber balls, one of which bounces quite nicely, whereas the other, although it feels the same to the touch, hits the floor with a solid thud and bounces almost not at all. The effect is quite surprising.

7. One reviewer quite sensibly worried about confounding presence/absence of feedback with different tasks: Observers could just do better with feedback than without. Fortunately, the data answer this worry. The fact that the observers split not so much on task but on hypo- versus hypergravity shows the difference to be due not to feedback, but to semantics. Experiment 3 demonstrated this quite clearly.

8. While testing the upgrade to the dynamics package, we noticed that the earlier version had used only one half the cross section in approximating the air resistance. The main effect was that the balls in Experiment 1 traveled horizontally a bit farther than they should have (a couple of ball widths after 4.5 sec). The effect on bounce height was small. Using the correct cross section is unlikely to have changed the general result. In Experiments 2 and 3, we used the correct value, rather than matching the previous error.

9. A 3rd observer, who was not naive about the experiment design, is not included. She performed at ceiling on both tasks.