Categories and constraints in causal perception

Jonathan F. Kominsky*[1], Brent Strickland*[2], Annie E. Wertz[3], Claudia Elsner[3], Karen Wynn[4], & Frank C. Keil[4]

[1]Harvard University
[2]Ecole Normale Superieure/PSL/Institut Jean Nicod
[3]Max Planck Institute for Human Development, MPRG Naturalistic Social Cognition
[4]Yale University

*Authors contributed equally to this project

Running Head: Categorical constraints on causal perception

Address for reprints and correspondence	:	Jonathan F. Kominsky Department of Psychology, Harvard University William James Hall room 1154 33 Kirkland St. Cambridge, MA 02138
Email	:	jkominsky@g.harvard.edu
Phone/Fax	:	(617)-384-7930

Word Count	:	1999 (excluding all methods and results except footnotes)

Abstract (148 words)

When a moving object (A) moves adjacent to a stationary object (B), and in that instant A stops and B starts moving, it is irresistibly seen as an event in which A causes B to move. Real-world collisions are subject to Newtonian constraints on the relative speed of B, but here we show that *perceptual* constraints (which imprecisely align with Newtonian principles) define two categories of causal events in perception based on the relative speed of B. Using performance-based tasks, we show that “triggering” events in which B moves noticeably faster than A are treated as being categorically different from “launching” events in which it does not, and that these categories are unique to causal events (Experiments 1 & 2). Furthermore, we show that 7- to 9-month-old infants are sensitive to the same distinction, suggesting that this boundary may be an early-developing component of causal perception (Experiment 3).

Keywords:

Causality, Perception, Infant development, Naïve Physics

Our minds evolved in an environment that contains certain physical regularities. Some of those regularities are reflected in the adult perceptual system as well as in 'core knowledge,' i.e., early-developing sets of expectations about the world that shape perception, learning, and cognition from infancy (Carey, 2009; Spelke & Kinzler, 2007). This is most obvious for the domain of physical objects. For example, from early infancy we expect objects to obey certain physical principles, such as spatiotemporal continuity, i.e. not moving between two locations without traversing the space in between, and we expect individual objects to be cohesive by maintaining single bounded contours (Spelke, Breinlinger, Macomber, & Jacobson, 1992). These same principles that seem to be part of our core knowledge of objects also constrain the perceptual processing of objects in adulthood (e.g., Scholl & Pylyshyn, 1999; vanMarle & Scholl, 2003). For example, moving objects which obey the principle of continuity (but not those which do not) serve as natural "units" of attention (Scholl, 2001; Flombaum & Scholl, 2006).

Notably, these results need not imply the existence of some kind of 'physics engine' in perception, but rather that the input the visual system has received from the world over the course of evolution has been constrained in particular ways by physics (McIntyre, Zago, Berthoz, & Lacquaniti, 2001). However, there are many physical constraints on our environment that are much more sophisticated than simple continuity or cohesion. In particular, there are physical constraints on interactions between objects that may have shaped the sensitivity of the visual system to certain specific properties of those events. Here, we propose that a perceptual constraint combined with a real-world physical regularity creates a categorical distinction in the perception of causal collision events.

Causal perception.

Imagine a simple event involving two objects like the one rendered schematically in Fig. 1a (available animated at jfkominsky.com/CategoricalConstraints.html). In this event, one object (A) moves until it is directly adjacent with the second object (B), at which point A immediately stops and B begins moving in the same direction. As long as certain spatiotemporal constraints are satisfied (cf. Figs. 1b-c), we irresistibly perceive this event as containing a causal relationship, that is, A causes B to move (Michotte, 1946/1963; Scholl & Tremoulet, 2000). Importantly, we truly *perceive* causality in these events. While this is not the venue for a comprehensive review of the extensive debate on this point, causal judgments and causal perception can be dissociated empirically (Schlottmann & Shanks, 1992; cf. Rips, 2011), and launching events are subject to uniquely perceptual effects (e.g., Moors, Wagemans, & de-Wit, 2017), including retinotopically specific visual adaptation (Rolfs, Dambacher, & Cavanagh, 2013). Causal perception is also early-developing, emerging by six months of age (Leslie & Keeble, 1987) or earlier (Mascalzoni, Regolin, Vallortigara, & Simion, 2013).

Real-world collision events obey Newtonian constraints. One little-known consequence of Newton's 3rd law is that no matter the relative masses of objects A and

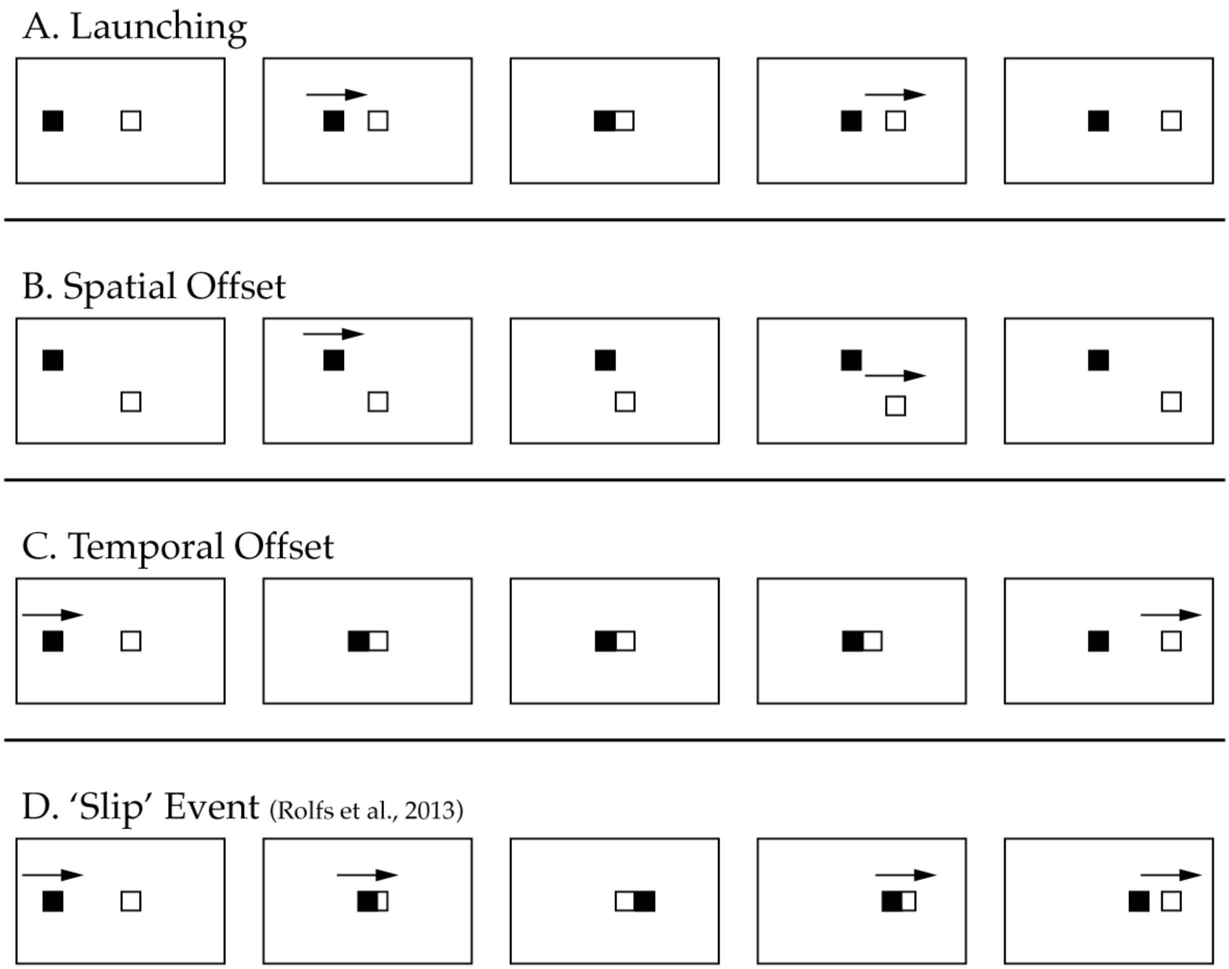

Fig. 1. Causal perception. The classic example of causal perception is the so-called 'launching' event (a). In this event, observers automatically and irresistibly perceive that the first object causes the second object to move. This percept can be disrupted by introducing a spatial offset (b) or temporal offset (c), or having the objects appear to 'slip' past each other without making contact (d).

B, B can never move at more than double the speed of A based on the force of the collision alone (see supplemental materials for a mathematical proof). This rule provides an absolute limit to B's speed, but due to air resistance, friction, and imperfect collisions, events in which object B moves any faster than object A are physically (and empirically) unlikely in the natural environment, except in cases in which B is selfpropelled. For similar reasons, events in which B moves *slower* than A are very likely (Runeson, 1983). Thus, events in which B moves faster than A are both unexpected and an indication that some unseen forces are acting on B.

Early work on causal perception suggested sensitivity to this asymmetry. Collision events in which B moved detectably faster than A after launching were not described as 'launching', but rather as 'triggering' or 'releasing', while events in which A moved even three times faster than B were still often labeled as 'launching' (Boyle, 1960; Michotte, 1946/1963; Natsoulas, 1961, cf. Sanborn, Mansinghka, & Griffiths, 2013). However, these explicit report tasks indicate the existence of a distinction between launching and triggering in causal *judgment*, but they do not directly address whether there is actually a distinction in causal *perception*.

If perception is sensitive to this categorical distinction between causal events, we should expect that perception would treat events in which B moves detectably faster than A as qualitatively different from other causal events. For example, events with a speed ratio A:B of 1:3 should be seen as categorically different from 1:1 events. However, no such boundary should exist between events with equally different speed ratios in which B moves slower than A (1:1 vs. 3:1). Such a distinction may not only be present in adult causal perception, but an aspect of core causal perception that is present from infancy. We designed three experiments to test these hypotheses with adult observers (Experiments 1 & 2) and preverbal infants (Experiment 3).

In Experiment 1, we devised a visual search task, on the logic that this proposed categorical boundary should lead to 'oddball' effects. We tested this under two conditions: First, among an array of 1:1 events, a 1:3 event should be easily detectable because it is an oddball belonging to a different perceptual event category, while a 3:1 event should be less detectable (Experiment 1a). This sensitivity should only apply to causal events, and not to minimally matched non-causal events in which two objects move independently (and therefore the 'speed limit' does not apply). Second, if this advantage is genuinely an oddball effect indicating a categorical boundary rather than 1:3 events simply standing out on their own, then a 1:1 event should be easier to detect in an array of 1:3 events than in an array of 3:1 events (Experiment 1b).

Experiment 1a

Methods.

Stimuli and procedure. All stimuli were presented on a 2010 11″ MacBook Air running MatLab and the Psychophysics Toolbox (Brainard, 1997).

To test whether perception distinguishes events in which B moves faster than A from events in which it does not, we designed a visual search task in which the search array consisted of sets of two-object events (like those in Fig. 1). If the target event violates this constraint while the distractors do not, then the target event should stand out as an oddball, resulting in quicker reaction times to detect the event. So, if the distractor events in the search array are all symmetric 1:1 speed ratio events that adhere to this Newtonian constraint, and the target event is an asymmetric event that violated this constraint (e.g., a 1:3 event), then the target event should be easier to find, compared to an equally asymmetric target event that does not violate the Newtonian constraint (e.g., a 3:1 event). However, this advantage should only hold for cases in which Newtonian limits could apply (i.e., causal events), and not in any cases where both objects in the event appear to move independently.

We designed four conditions, run entirely between-subjects: causal, temporal offset, spatial offset, and 'slip' event. In every condition participants saw three pairs of discs, separated by vertical lines. Fig. 2 shows a not-to-scale diagram of a display sequence. Each disc subtended 0.6° of visual angle (assuming 60cm viewing distance), and at start the two discs in the pair were separated by 2.4° of visual angle. The mid-point of each pair was separated from the others by 10° of visual angle.

In all conditions except 'slip', one disc in each pair began moving toward the x-coordinate of the other disc in that pair, until they were adjacent on the x-axis. In the spatial offset condition, the bottom of the left disc in each pair was vertically offset from the top of the right disc by 0.6° of visual angle. In the 'slip' condition, the first disc moved until it was fully overlapping with the second, and then continued moving until it was adjacent to the disc *on the opposite side* (see Fig. 1d). In the contact, spatial offset, and slip conditions, the first disc stopped and other disc in the pair immediately began moving the same direction (in the slip condition this meant passing through

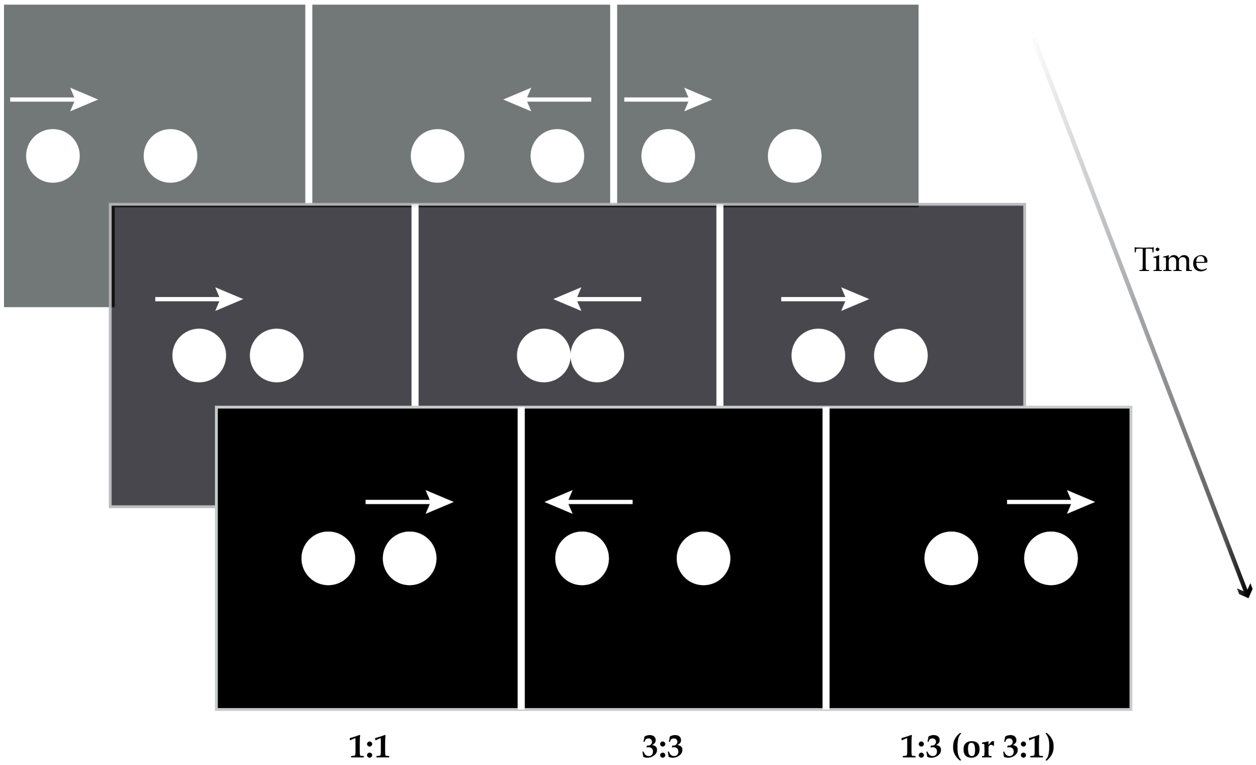

Fig. 2. Schematic illustration of a trial in the causal condition of Experiment 1. All three events looped until participants responded. Participants had to find the event where the two discs moved at different speeds relative to each other. In half of the trials, the target event was a 1:3 event, in the other half it was a 3:1 event.

the first disc). In the temporal offset condition, both discs were stationary for 300ms before the second began moving. The second disc then moved until the two discs were once again horizontally separated by 2.4° of visual angle, at which point it stopped and both discs were stationary for 200ms. After this pause, the second disc started back towards the first and the animation repeated.

In every condition, for two of the three pairs the approaching disc and the receding disc moved at the same speed. In one pair, both moved at 9°/sec. In the second pair, both moved at 3°/sec. In the third pair, one disc moved at 9°/sec and the other at 3°/sec. On slow/fast (1:3) trials, the approaching disc moved at 3°/sec, and on fast/slow (3:1) trials, it moved at 9°/sec. All three events started at the same time, so the collision happened sooner for events where the first object moved at 9°/sec, and after the first collision the three events were completely decoupled. (This fact of the design actually works against our hypotheses, since participants are exposed to the speed difference in a 3:1 event *before* they are exposed to the difference in a 1:3 event.)

Participants were instructed to press the spacebar as soon as they detected the pair in which the two discs were moving at different speeds (i.e., an asymmetric event). After pressing the spacebar, the animation paused and they used the mouse cursor to select which of the three pairs was the asymmetric event, thereby ensuring that they had to locate the target event before pressing the spacebar.

There were a total of 96 trials in the causal, temporal offset, and spatial offset conditions, 8 repetitions of each possible combination of target event type (1:3 vs. 3:1), target event location, and the location of each distractor. The slip condition added two more repetitions of each, for a total of 120 trials, because it was designed after the other three had started and we felt participants could complete the additional trials in the allotted time (analyses using only the first 96 trials of the 'slip' condition yield qualitatively identical results). All trials in all versions were presented in fully random order.

Participants. Based on effect sizes observed in in-lab piloting, we estimated that each of our four conditions would require approximately 12 participants, and so we recruited until we had that many who passed our exclusion criterion in each condition (see below), though due to recruiting before being able to check exclusion criterion we ended up over-recruiting in some conditions, and elected to include all valid data rather than arbitrarily exclude participants. This involved recruiting a total of 85 participants from the New Haven, CT area. All subjects were over 18 years old and gave informed consent, and were compensated with either $5 or a half-hour of course credit for a roughly 30-minute study.

Results.

Exclusion criteria. Participants were excluded if they failed to select the correct target event on more than 50% of trials. Across all four conditions this excluded 28 participants, plus 1 additional participant who failed to complete the experiment and 3 who participated in more than one version of the experiment due to experimenter error, excluding a total of 32 participants (roughly 38% of all participants). Failure to identify the correct event on at least 50% of trials was an easy and objective test of whether participants understood the task and were able to complete it successfully (and stricter than mere chance responding, which would be 33%). However, it is still somewhat surprising that there were so many exclusions. We can offer no definitive explanation, but present two likely contributing factors. The first is that some participants failed to understand the instructions (which can be found at osf.io/k8t4b), and therefore did not know what kind of target event they were looking for. We endeavored to address this in Experiment 1b by adding practice trials (see below). The other possibility is that some participants genuinely could not detect the asymmetric events ("they all look the same"; see Experiment 2).

This resulted in a final total of 13 participants in the causal condition, 14 each in the temporal and spatial offset conditions, and 12 in the slip condition. In addition, prior to analyzing group effects, individual trials were excluded if the participant selected the incorrect event or their RT on that trial was more than 2.5 standard deviations from that participant's average RT for accurate trials.

Reaction times. We analyzed average reaction times for each target event by participant. The results by condition and event can be found in Fig. 3. A 4 (causal/space offset/time offset/slip; between) x 2 (1:3 vs. 3:1; within) mixed-model ANOVA revealed a significant interaction between causal condition and speed ratio, $F(3, 49) = 3.48$, $p = .023$, $\eta_p^2 = .176$. We then analyzed the effect of 1:3 vs. 3:1 separately in each condition using paired-sample *t*-tests. As predicted, participants in the causal condition were significantly faster to detect the 1:3 (triggering) event ($M = 4.13$, $SD = 2.31$) compared to the 3:1 event ($M = 4.86$, $SD = 2.70$), $t(12) = 3.751$, $p = .003$, $d = .253$, 95% CI of d = [.094, .413]. (Cohen's *d* for *t*-tests and CIs calculated using the R package metafor [Viechtbauer, 2010].) In contrast, participants showed no significant effect of event type

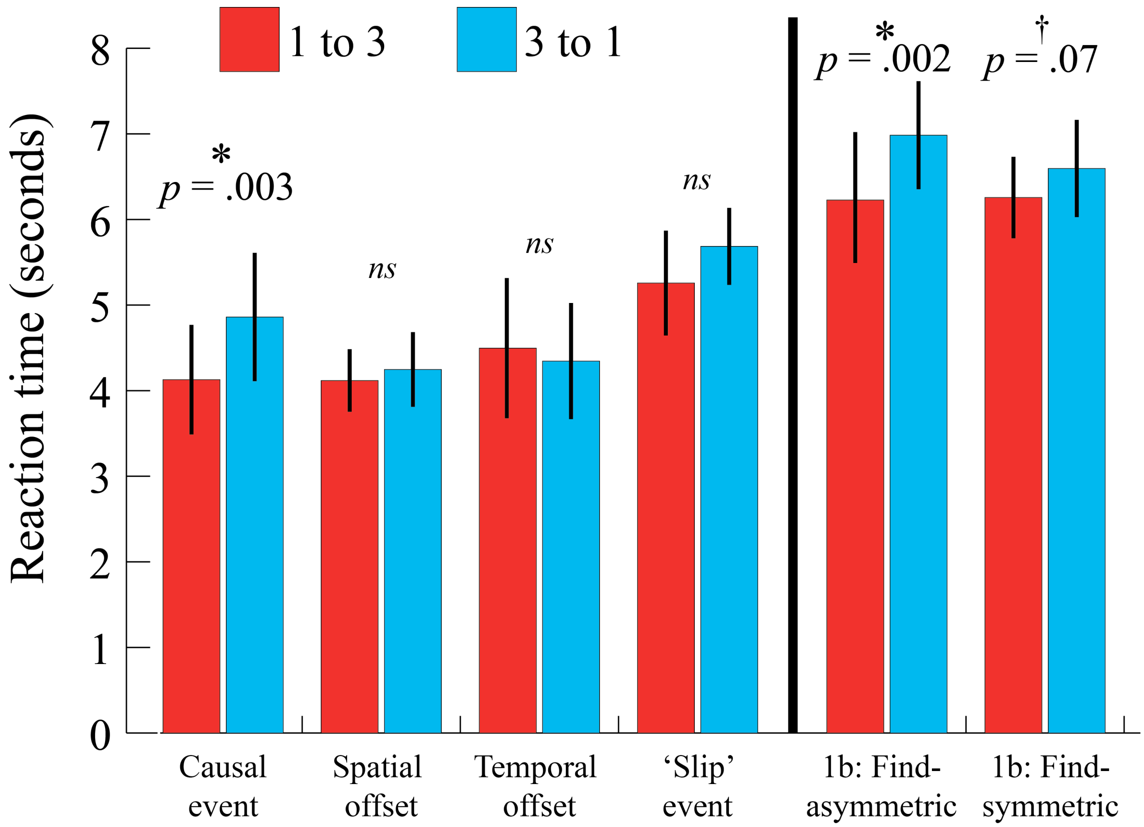

Fig. 3. Results of Experiment 1. Participants in the causal conditions were faster to respond to 1:3 events in Exp. 1a and 1b, but in non-causal conditions responded equally quickly to 1:3 and 3:1 events. * = $p < .05$, error bars represent +/- 1 SEM.

in the spatial gap condition (Slow/fast: $M = 4.12$, $SD = 1.37$; Fast/slow: $M = 4.25$, $SD = 1.63$), $t(13) = .779$, $p = .45$, temporal gap condition (Slow/fast: $M = 4.50$, $SD = 3.07$; Fast/slow: $M = 4.35$, $SD = 2.54$), $t(13) = .833$, $p = .42$, or slip condition (Slow/fast: $M = 5.26$, $SD = 2.12$; Fast/slow: $M = 5.69$, $SD = 1.56$), $t(11) = 1.52$, $p = .158$.

We further tested whether there were significant interactions between the causal condition and each non-causal condition, using 2 (causal vs. non-causal) x 2 (1:3 vs. 3:1) mixed-model ANOVAs. The analysis of the causal and spatial offset conditions found a significant interaction, $F(1, 25) = 5.63$, $p = .026$, $\eta p^2 = .184$, as did the analysis of the causal and temporal offset conditions, $F(1, 25) = 11.04$, $p = .003$, $\eta_p^2 = .306$. However, the analysis of the causal and slip conditions found a non-significant interaction, $F(1, 23) = .80$, $p = .379$.

To further understand the inconclusive results of the comparison between the causal and slip conditions, we conducted additional post-hoc 2 x 2 mixed-model ANOVAs comparing the slip condition to the other non-causal conditions. We found no significant interaction when comparing the slip condition to the spatial offset condition, $F(1, 24) = .90$, $p = .353$, or the temporal offset condition, $F(1, 24) = 3.15$, $p = .089$.

In short, the slip condition shows no significant advantage for 1:3 vs. 3:1 target events on its own, but the magnitude of the raw (non-significant) RT difference does not differ significantly from that of *any* other condition, neither the causal condition that shows the advantage nor the other non-causal conditions that do not. While inconclusive, we take these results as indicating that, if there is truly any advantage for locating 1:3 over 3:1 slip events, it is at least less reliable than the advantage we find for causal events, even if it is not significantly different in magnitude. However, the properties of the slip event may be worth more thorough investigation in future work, as to our knowledge this is only the second time it has been used in studies of causal perception (Rolfs et al., 2013).

Accuracy. We also analyzed accuracy by participant, to ensure that the RT results did not simply reflect a speed/accuracy trade-off. We analyzed raw accuracy, with no trials excluded based on RT, in a 4 (causal condition) x 2 (speed ratio) mixed-model ANOVA. We found no main effects and no interaction, all *p*s > .5. Participants were equally accurate in the causal condition (*M* = .88, *SD* = .13), spatial offset condition (*M* = .91, *SD* = .13), temporal offset condition (*M* = .90, *SD* = .13), and slip condition (*M* = .91, *SD* = .11), and equally accurate for 1:3 (*M* = .90, *SD* = .13) and 3:1 events (*M* = .90, *SD* = .11). To further demonstrate that a speed/accuracy trade-off could not account for our results, we conducted a separate by-trial analysis of accuracy by speed ratio for each condition. In the causal condition, the only condition with an RT effect, responses were equally accurate for 1:3 (*M* = .87, *SD* = .34) and 3:1 events (*M* = .88, *SD* = .32), $t(12) = .43$, $p = .67$. There was also no difference for spatial offset (1:3: *M* = .91, *SD* = .10; 3:1: *M* = .91, *SD* = .09), $t(13) = .53$, $p = .61$, temporal offset (1:3: *M* = .90, *SD* = .15; 3:1: *M* = .90, *SD* = .12), $t(13) = .26$, $p = .797$, or slip events (1:3: *M* = .92, *SD* = .09; 3:1: *M* = .90, *SD* = .09), $t(11) = 1.24$, $p = .24$. Put simply, the difference in reaction times we find between 3:1 and 1:3 events, and the interaction with causal condition, cannot be accounted for by a speed/accuracy tradeoff.

Experiment 1b

Methods.

Stimuli and procedure. To investigate whether the results of Experiment 1a indicate a categorical boundary rather than 1:3 causal events simply being more prominent in any circumstance, we tested whether 1:1 causal events are easier to detect in an array of 1:3 events than in an array of 3:1 events. The stimuli were identical to those used in Experiment 1a, but adapted to run in a web browser using the Qualtrics online survey system (*Qualtrics*, 2005) and the Greensock TimelineMax javascript animation library (GreenSock, Inc., 2015). The only visual difference was that the events involved black discs on a white background instead of white discs on a black background, and the discs were slightly larger (though variation in the monitor resolution of participants' computers and viewing distance means that the size participants saw may have been bigger or smaller than in Experiment 1a, and it was not expected to matter).

There were only two conditions, a find-asymmetric condition, identical to Experiment 1a's causal condition, and a find-symmetric condition. In the find-symmetric condition, participants were instructed to find the *symmetric* event among two *asymmetric* events. In this condition, the target symmetric event could be either a 1:1 or 3:3 event, and the asymmetric distractor events were either both 1:3 or both 3:1. To prevent the asymmetric events from syncing up and giving an impression of common motion, all three events in each trial of both the find-asymmetric and find-symmetric conditions were started at a (separately determined) random point in their animation.

Participants still responded using the spacebar, but instead of clicking the target event, they pressed a number key (1, 2, or 3) to indicate which event was the target event. In an attempt to compensate for the fact that there would be no experimenter reading the instructions to the participants, four training items with feedback were added to the start of the experiment, in which participants were not allowed to proceed until they had selected the correct option.

Participants. Experiment 1b was conducted online using Amazon Mechanical Turk. All participants were Amazon Mechanical Turk workers over the age of 18, located in the US, and with prior HIT acceptance rates of >90%. Because we anticipated noisier RT data from MTurk due to the implementation of the study to run in a web browser and the inherent variability of allowing people to complete the study in their own home rather than a controlled lab environment, we doubled our target N. Recruitment continued until there were 24 participants in each condition who passed the exclusion criterion (48 total). This required recruiting a total of 67 participants, of whom 17 were excluded for failing to meet the accuracy cutoff, and two additional participants excluded due to an unexpected and inexplicable technical glitch wherein Qualtrics failed to record their reaction times, for a total of 19 exclusions (~28%). Participants were paid $2 for a study that took most of them less than 20 minutes to complete.

Results.

Individual trials were excluded using the same criteria as in Experiment 1a. The results can be found in Fig. 3, right panel. Preliminary analyses indicated that there was no effect of target event in the find-symmetric condition (1:1 vs. 3:3), so we collapsed across this variable for the primary analysis. This allowed us to conduct a 2 (task; find-asymmetric vs. find-symmetric) x 2 (asymmetric speed ratio; 1:3 vs. 3:1) mixed-model ANOVA.

Regardless of whether the asymmetric event was the target or the distractors, participants were significantly faster to respond when the asymmetric speed ratio was 1:3 (M = 6.24, SD = 3.09) than 3:1 (M = 6.79, SD = 2.92), $F(1, 46) = 11.319$, $p = .002$, $\eta_p^2 = .197$. However, there was no main effect of task, $F(1, 46) = .044$, $p = .835$, and no interaction, $F(1, 46) = 1.647$, $p = .206$.

We conducted planned paired-samples t-tests of the effect of asymmetric speed ratio in each task. Replicating Experiment 1a, there was a significant effect of asymmetric speed ratio in the find-asymmetric task, such that participants were faster to locate 1:3 target events (M = 6.23, SD = 3.74) than 3:1 target events (M = 6.98, SD = 3.09), $t(23) = 2.787$, $p = .01$, $d = .237$, 95% CI = [.085, .389]. In the find-symmetric task, there was no significant advantage for finding 1:1 (or 3:3) events among 1:3 events (M = 6.26, SD = 2.33) compared to finding them among 3:1 events (M = 6.60, S = 2.78), $t(23) = 1.885$, $p = .072$, $d = .063$, 95% CI = [-.239, .365]. Looking at RT alone, it would appear that there might be no oddball effect (or only a marginal effect) in the find-symmetric task, but the analysis of accuracy suggested otherwise.

Accuracy. The two participants for whom Qualtrics failed to record RT data still achieved above-threshold accuracy, so we included them in our accuracy analysis (one in each condition). (Analyses of accuracy excluding these two participants yielded qualitatively identical results.) We conducted a 2 (task) x 2 (1:3 vs. 3:1) mixed-model ANOVA, which found no effect of task, $F(1, 48) = 1.70$, $p = .198$, but a main effect of 1:3 vs. 3:1, $F(1, 48) = 11.58$, $p = .001$, $\eta p^2 = .194$, and a significant interaction, $F(1, 48) = 10.01$, $p = .003$, $\eta p^2 = .173$.

To examine this interaction further, we conducted paired-sample *t*-tests examining the effect of 1:3 vs. 3:1 in each task. In the find-asymmetric condition, as in Experiment 1a, participants were not significantly more or less accurate when finding 1:3 ($M = .88$, $SD = .16$) or 3:1 events ($M = .87$, $SD = .15$), $t(24) = .19$, $p = .85$. However, in the find-symmetric condition, participants were significantly better at finding the symmetric event when the distractors were 1:3 events ($M = .86$, $SD = .15$) than when the distractors were 3:1 events ($M = .77$, $SD = .17$), $t(24) = 4.22$, $p < .001$, $d = .532$, 95% CI = [.245, .819].

This is the only instance in our search task data of a significant effect on accuracy, and it indicates two things. First, it supports the hypothesis that there is a genuine categorical boundary between 1:3 events and symmetric events, rather than 1:3 events simply “standing out” on their own. Second, and more decisively, it demonstrates that 3:1 events are not easily distinguished from symmetric events, since it was clearly more difficult to distinguish symmetric events from 3:1 events. It is unclear why we should find this effect on accuracy in only this condition, but in fact it is even stronger evidence that there is a *perceptual* distinction between triggering events and launching events, but not between symmetric launching events and 3:1 events.

Discussion.

Experiment 1 found evidence that adult causal perception distinguishes between 'triggering' and 'launching' events. Causal events with 1:3 speed ratios were easily distinguished from those with 1:1 speed ratios, but 3:1 events were not as easily distinguished from 1:1 events, despite being equally different in objective terms. Critically, there is no such asymmetry for non-causal events. These performance-based results provide initial evidence that causal perception, rather than judgment or reasoning, is sensitive to a distinction between launching and triggering.

Experiment 2

While Newtonian physics imposes a 1:2 limit under ideal conditions, the real world is very rarely ideal. Rather, as Michotte put it, the triggering impression likely emerges whenever "the speed of B [becomes] noticeably greater than that of A" (Michotte, 1946/1963, pg. 109). As such, events in which B moves noticeably faster than A, but below the 1:2 limit, may still be perceived as triggering.

However, we do not yet know the perceptual constraints on detecting speed differences in launching events. Research with moving Gabor patches would predict that speed ratios as low as 1:1.06 could be distinguished from 1:1 (Brown, 1931; Orban, Van Calenbergh, De Bruyn, & Maes, 1985; Traschütz, Zinke, & Wegener, 2012;

Werkhoven, Snippe, & Alexander, 1992), but research on single moving objects suggests a range of detection thresholds anywhere between 1:1.4 and 1:4 for a given observer (Calderone & Kaiser, 1989; Watamaniuk & Heinen, 2003). Yet, no studies have explicitly examined speed discrimination in events involving two objects.

The goal of Experiment 2 was therefore two-fold. First, starting from Michotte's assertion that triggering only requires a noticeable increase in B's speed, we wanted to establish what changes are "noticeable" in this context. Understanding these perceptual constraints then provides clear predictions about the speed ratios that might produce an advantage for slow/fast causal events like the one we found in Experiment 1. For example, it would not be worth investigating whether 1:1.5 events are seen as triggering if the change in speeds is not detectable in isolation.

Second, we wished to rule out a low-level perceptual differences account of the results of Experiment 1. For example, causal events may have been easier to process than non-causal events, or speed information easier to extract from them, thus producing differing performance on slow/fast vs. fast/slow events. Rather than a categorical boundary, this low-level account of Experiment 1 predicts a difference in the ability to detect changes in speeds in isolated causal and non-causal events.

Therefore, we directly tested sensitivity to changes in speeds for causal and non-causal events, for events in which A moved faster than B (fast/slow X:1 events) and events where A moved slower than B (slow/fast 1:X events), and for speed multipliers above and below 2. We designed a task in which participants judged the relative speed of two objects in serially presented single events.

Methods.

Participants. This experiment was run online using Amazon Mechanical Turk, excluding workers who had participated in Experiment 1b but otherwise using the same criteria. We aimed to recruit 24 participants who passed our exclusion criteria (see below). This required recruiting 34 participants, who were paid $6.50 for a task that took approximately 40 minutes.

Stimuli and procedure. The stimuli in these experiments were constructed the same way as in Experiment 1b, and were very similar except that there was only one event on the screen at any given time instead of three.

All of the events were two-object events, in which the two objects (A and B) could either move at the same relative speeds, or at different relative speeds, and this varied across trials (but not within a trial). Participants were told that they would see one event at a time, and they would have to determine whether it was a 'match-speed' (i.e. symmetric) event, in which both objects moved at the same speed, or a 'different-speed' (i.e. asymmetric) event, in which each object moved at a different speed. They responded by pressing the "F" key for "Match" events and the "J" key for "Different" events. The events looped until participants made a response, and as soon as a key-press was detected the experiment advanced to the next trial.

There was a 'causal' block and a 'non-causal' spatial offset block, with the order of presentation counterbalanced across participants. In the causal block, all trials were collision events like those in Exp. 1b. The non-causal offset block was identical to the causal block except that, in every event, there was a vertical offset between the closest edges of the two objects of one diameter, corresponding to the spatial offset condition of Experiment 1a.

In each block there were three categories of trials, match-speed trials, slow/fast trials, and fast/slow trials. These three categories were crossed with five speed multipliers, 1.5, 1.75, 2, 2.25, and 2.5. Thus, there were five different types of slow/fast trials, (1:1.5, 1:1.75, 1:2, 1:2.25, and 1:2.5) and a corresponding set of five fast/slow trials (1.5:1, 1.75:1, etc.) and match-speed trials (1.5:1.5, 1.75:1.75, etc.). In addition, there was a set of 1:1 events. This was done to ensure that participants could not immediately determine whether a trial was a match-speed trial or a different-speed trial based on the speeds of either object alone. There were eight repetitions of each of these 16 trial types, yielding a total of 128 test trials in each block. In addition, there were eight each of 1:3, 3:1, and 3:3 trials in each block that were used as exclusion criteria: if participants were less than 50% accurate on these trials (in *either* block), they were excluded from analyses and replaced. Note that this excluded 10 of 34 participants (29%), suggesting that some individuals might not easily detect speed differences even as high as 1:3 or 3:1. This suggests that some of the exclusions in Experiment 1 may be the result of the same inability to detect changes in speeds, even of this magnitude.

Trials within each block were presented in fully random order, and each block was preceded by four training trials (two match, one slow/fast, one fast/slow) in which participants received feedback on their answers.

Results.

Analysis strategy. Our primary dependent variable was participants' sensitivity to changes in speeds, using the d′ sensitivity index from signal detection theory. We calculated d′ separately for each asymmetric event type for each participant. Computing d′ requires hits, misses, and false alarms, which we defined as follows: 'hits' were asymmetric trials that participants correctly categorized as different-speed events, 'misses' were asymmetric trials that participants inaccurately categorized as match-speed trials, and 'false alarms' were match-speed trials *of the same speed multiplier* that participants categorized as different-speed events. For example, to calculate a given participant's d′ for 1:1.5 causal events, we would compute false alarm rate as the proportion of 1.5:1.5 causal events that the participant classified as different-speed events.

One issue with d′ is that it becomes indeterminate when there is extreme sensitivity or insensitivity, that is, if either the hit rate or false-alarm rate are exactly 0 or 1. If these values occurred, they were corrected by .0625 in the appropriate direction (i.e., half the effect of getting one trial accurate or inaccurate), which is a common correction for this sort of data (Brown & White, 2005; Murdock Jr. & Ogilvie, 1968).

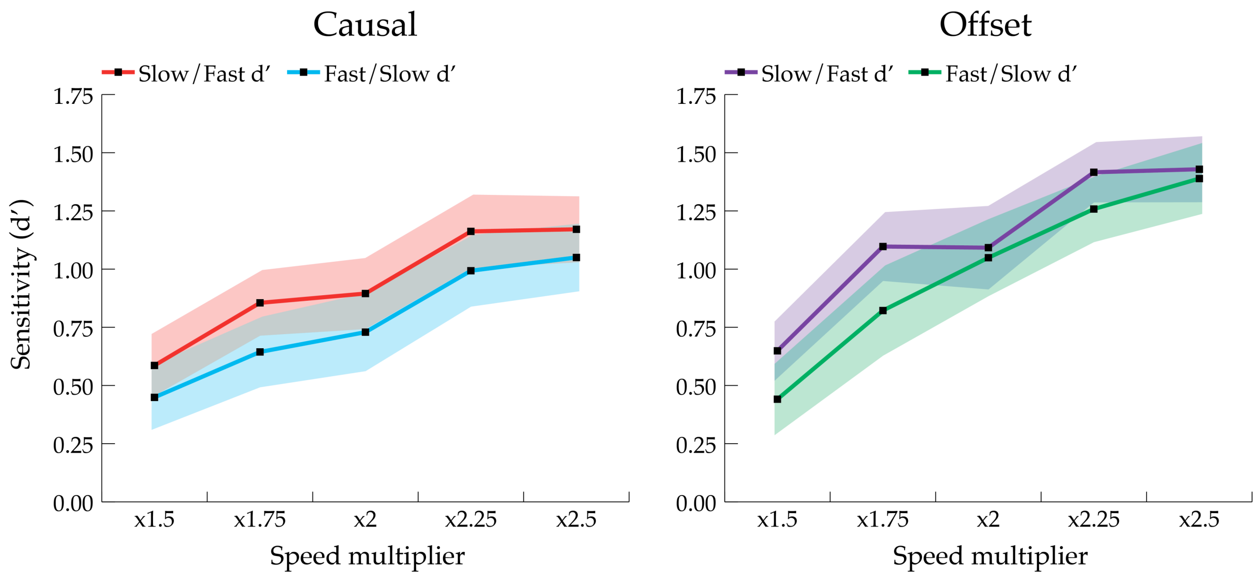

Fig. 4. Results of Experiment 2. Sensitivity to changes in speeds dropped off steadily between 2.5 and 1.5. Sensitivity overall was slightly greater for slow/fast events compared to fast/slow events, and offset events compared to causal events. Shaded areas represent +/- 1 SEM.

'Noticeable' changes in speed. The results are depicted in Fig. 4. We conducted a 2 (causal vs. non-causal) x 2 (slow/fast vs. fast/slow) x 5 (speed multiplier) repeated measures ANOVA. For this initial question, the speed multiplier is the factor of greatest interest. We found a main effect of speed multiplier, $F(4, 92) = 28.34$, $p < .001$, $\eta p^2 = .552$, but no interaction between speed multiplier and causality, $F(4, 92) = 1.03$, $p = .39$, or speed multiplier and slow/fast vs. fast/slow, $F(4, 92) = .80$, $p = .53$, and no three-way interaction, $F(4, 92) = .51$, $p = .72$, indicating that the effect of speed multiplier was consistent across causal and non-causal events, and across slow/fast and fast/slow

events. Therefore, to analyze the effect of speed multiplier, we collapsed across causality and slow/fast vs. fast/slow and conducted post-hoc comparisons on the resulting average d′ values.

Bonferroni-corrected pairwise comparisons revealed that participants were significantly more sensitive to events with speed multipliers of 2.5 (M = 1.26, SD = .71) than those with speed multipliers of 2 (M = .94, SD = .81), of 1.75 (M = .85, SD = .78), and of 1.5 (M = .53, SD = .67), ps < .001, but not events with speed multipliers of 2.25 (M = 1.21, SD = .71), $p > .9$. All other differences between speed multipliers were significant at $p <= .01$, except that there was no significant difference in sensitivity between speed multipliers of 2 and 1.75, $p > .9$.

So what counts as a 'noticeable' change in B's speed? Because d′ is a dimensionless statistic it is a little difficult to tell. A d′ of 0 would be purely at chance (equal false alarm rate and hit rate). One-sample t-tests of the average d′ values given above against 0 found that all of them were significantly higher than that, ps < .001. However, the drop-off indicates that people become significantly less consistent in their ability to detect the change in speed at lower multipliers. Since this experiment did include a set of items at the speeds used in Experiment 1 (the 1:3 and 3:1 events used as exclusion criteria), we computed d′ for these items as well and conducted an additional set of Bonferonni-corrected pairwise comparisons against the other five speed multipliers. We found that sensitivity was high for speed multipliers of 3 (M = 1.33, SD

= .53), and significantly greater than sensitivity to speed multipliers of 2 and below (*p*s < .001), but not 2.5 or 2.25 (*p*s > .6). This suggests that events with speed multipliers of 2 and below are, most likely, perceptually categorized as triggering significantly less often than those of 2.25 and above, but it is not a "hard" boundary. For some individuals, events with speed ratios of 1:2 and below may still be seen as triggering some of the time.

This is in contrast to earlier work by Natsoulas (1961), who found that reports of triggering were over three times more frequent for speed ratios of 1:2 compared to those of 1:1. There are two possible explanations for this difference. The first possibility is purely methodological: while we cannot directly compare our stimuli to Natsoulas's (as we did not have control over participants' viewing distance or monitor size in this experiment), we modeled our online stimuli on the stimuli we used in Experiment 1a, and in comparison to Experiment 1a, the objects in Natsoulas's stimuli were roughly 3 times smaller and moved about 2.5 times faster. This may have made differences in speed more consistently detectible to his participants. The second possibility is that the mechanisms of speed discrimination and causal perception are surprisingly independent: causal perception may be affected by changes in speeds that are so subtle they cannot be explicitly detected. In other words, a modular system of causal perception may have an internal threshold for detecting changes in speeds that is below the threshold of explicit speed discrimination. This would be somewhat surprising but the current evidence cannot rule it out.

Ruling out low-level alternatives. If the results of Experiment 1 are due to low-level differences in speed perception between causal and non-causal events rather than causality per se, we should see corresponding differences in performance across causal and non-causal conditions in this experiment. By examining speed perception in individual events separately from the search task, we can see whether there are differences in observers' ability to detect changes in speeds that mirror differences in performance in the search task. For example, a low-level account might predict better performance detecting speed changes in isolated causal events than non-causal events, or speed changes in slow/fast causal events being more detectable than those in fast/slow causal events with no such difference for non-causal events.

Returning to the 2 x 2 x 5 ANOVA, participants were more sensitive to changes in speeds for non-causal offset events (M = 1.06, SD = .80) than for causal events (M = .85, SD = .75), $F(1, 23) = 6.43$, $p = .019$, $\eta_p^2 = .218$, and more sensitive to slow/fast events (M = 1.04, SD = .75) than fast/slow events (M = .88, SD = .81), $F(1, 23) = 10.59$, $p = .003$, $\eta_p^2 = .315$, but there was no interaction between these factors, $F(1, 23) = .05$, $p = .83$. So, while participants were more sensitive to speed changes in slow/fast than fast/slow events overall, this advantage in sensitivity was equal for causal and non-causal events. Thus this difference cannot explain the results of Experiment 1: if a difference in sensitivity alone was driving the RT advantage for 1:3 causal events in Experiment 1, we would have found the same advantage in the non-causal spatial offset condition of Experiment 1 as well, but this was clearly not the case. Moreover, we cannot explain the difference in performance between causal and non-causal events from Experiment 1 by postulating that speed differences in causal events were simply *easier* to process, since the current results show exactly the opposite.

Because there was an overall difference in sensitivity to changes in speeds between causal and non-causal offset events, we conducted two additional post-hoc analyses to verify that this difference could not explain the results of Experiment 1. First, we tested whether there was an effect of slow/fast vs. fast/slow for the causal and non-causal conditions considered separately, using separate 2 (slow/fast vs. fast/slow) x 5 (speed multiplier) repeated-measure ANOVAs, both of which found a significant advantage for slow/fast events (Causal: $F(1, 23) = 9.15$, $p = .006$, $\eta_p^2 = .285$; Non-Causal offset: $F(1, 23) = 4.91$, $p = .037$, $\eta_p^2 = .176$), indicating that there is a significant advantage for seeing speed changes in slow/fast events in both types of events independently, not just overall.

Second, we computed a slow/fast - fast/slow difference score in the causal and non-causal conditions for each participant, and conducted a Bayesian paired-sample *t*-test using the JASP implementation (JASP team, 2016) of R's ttestBF function from the BayesFactor package (Morey & Rouder, 2015). This allowed us to compute a Bayes factor for the null hypothesis that the difference scores for the causal and offset conditions are the same (BF_0). This analysis yielded a BF_0 of 4.56, indicating that the magnitude of the difference of differences was 4.56 times more likely to occur if the null hypothesis were true (i.e., if there was no significant difference between the causal and non-causal conditions).

Discussion.

We found a relatively linear drop-off in sensitivity to changes in speeds at lower speed multipliers. While a minority of participants have some ability to detect changes in speeds in events below the Newtonian speed limit of 1:2, their sensitivity is significantly reduced. Assuming the ability to perceive a speed distinction is a prerequisite for perceiving triggering, this sensitivity reduction would make it difficult to conduct a version of Experiment 1 with speed ratios at or below 1:2 (but see Natsoulas, 1961). Greater difficulty finding either target event would severely decrease accuracy and add variability to RTs. Therefore, we conclude that the category boundary between launching and triggering is more likely to involve constraints on perceptual discrimination than Newtonian constraints on real-world events, even if Newtonian constraints may explain why slow/fast causal events were relevant in the environment in which our visual system evolved.

We furthermore found that the differences between causal and non-causal events in Experiment 1 cannot be attributed to low-level differences in sensitivity to changes in speeds. Observers were overall more sensitive to speed changes in *non-causal* events than causal events, thus ruling out the possibility that the results of Experiment 1 are due to causal events being generally easier to process. Observers were more sensitive to speed changes in slow/fast events than fast/slow events, but critically this asymmetry did not differ significantly between causal and non-causal events. This suggests that the results of Experiment 1 were due to a categorical distinction between 1:3 and 1:1 causal events that does not exist between 3:1 and 1:1 causal events, and that non-causal events have no such asymmetry.

Experiment 3

Here, we examine whether sensitivity to this categorical boundary is a reliably early-developing component of human cognitive architecture using a classic dishabituation paradigm with preverbal infants (e.g., Colombo & Mitchell, 2009): Infants habituated to 1:1 causal events should dishabituate strongly to 1:3 causal events, but not to 3:1 causal events. Similar to Experiment 1, there should be no such difference for non-causal events.

Methods.

Participants. Based on the sample sizes of earlier causal perception research with infants (Leslie & Keeble, 1987), but not knowing the magnitude of the effect, we conservatively aimed to recruit 34 participants in each of four conditions. A total of 136 infants (67 female) aged 6 months 15 days to 10 months 0 days recruited from the greater New Haven and Berlin areas. Preliminary ANOVAs found no significant effects of age or data collection site, and only a marginally significant effect of sex such that male infants tended to look at the display longer than the female infants did in all conditions. Therefore, the analyses below are collapsed across these factors. An additional 25 babies were tested but excluded due to fussiness/distraction (6), procedural error (13), parental interference (1), and test trial looking times greater than 3 standard deviations from the mean looking time in their condition (5). A further 28 babies were excluded because of live looking time coding errors that impacted infants' habituation times, however including these babies in the analyses does not substantially change the pattern of results (see supplemental materials).

Stimuli and procedure. Infants' sat on their parent's lap throughout the session. Parents were instructed not to direct their infant's attention during the testing session. Additionally, parents closed their eyes during the test trial so that they would not know whether their infant was seeing a 1:3 or 3:1 test event. This ensured that infants could not be influenced by their parent's reactions to the stimuli. During the testing session, infants were shown animated videos of two identical red squares modeled closely on the stimuli used by Leslie & Keeble (1987). Each video was 2 seconds long and presented in a continuous loop on a large flat screen monitor at 30 frames per second. Infants' looking times were recorded live by a trained coder, who was blind to condition, using jHab (Casstevens, 2007); a second independent coder subsequently evaluated all looking times from videos of the sessions, and the two coder's looking times were highly correlated (*r* = .97). Sufficiently large disagreements that resulted in changes to the computed habituation criterion would lead to replacement of that participant. For the analyses reported below, the first coder's data were used, except in cases in which re-coding process uncovered an error by the live coder (that did not affect the habituation criterion), in which case the second coder's data were used.

For all infants, each trial began with a short attention-getting noise. When infants looked at the screen, the trial began and the animation started to play. The trial ended when the infant looked away for 2 contiguous seconds, or 60 seconds had passed since the start of the trial, whichever came first.

In the causal condition, infants (N = 68, 35 female) were shown a launching event at a 1:1 speed ratio until they habituated to the presentation (i.e., once their total looking times over three consecutive trials decreased to less than half of the sum of their first three trials). After the habituation phase, infants were shown a single test trial. Half of the infants (N = 34) were shown a causal event with a 1:3 speed ratio, while the other half (N = 34) were shown a causal event 3:1 speed ratio. Parents were instructed to keep their eyes closed during the test trial. The presentation in the non-causal condition (N = 68, 32 female, divided evenly between 1:3 and 3:1 test events) was identical, except that the animations in both the habituation and test phases included a 0.5 second pause when the two squares came into contact. This manipulation has been previously shown to disrupt preverbal infants' perception of causality in such events (Leslie & Keeble, 1987).

Results.

Average test trial looking times can be found in Fig. 5. A 2 (condition: causal vs. non-causal) x 2 (test trial speed ratio: 1:3 vs. 3:1) ANOVA indicated a significant condition x speed ratio interaction, $F(1, 132) = 5.56$; $p = .02$, $\eta^2_p = 04$. As predicted, infants in the causal condition looked longer at the 1:3 events than the 3:1 events during the test trial, $t(66) = 2.29$, $p = .025$, $d = .55$, 95% CI = [.064, 1.033] , while infants in the non-causal condition showed no significant difference in looking times between 1:3 and 3:1 test events, $t(66) = -.88$, $p = .38$. Analyses using log-transformed looking time yielded similar results (see supplemental materials). These results suggest that, like adults in Experiment 1, preverbal infants are sensitive to a categorical boundary between launching (1:1 and 3:1) and triggering (1:3).

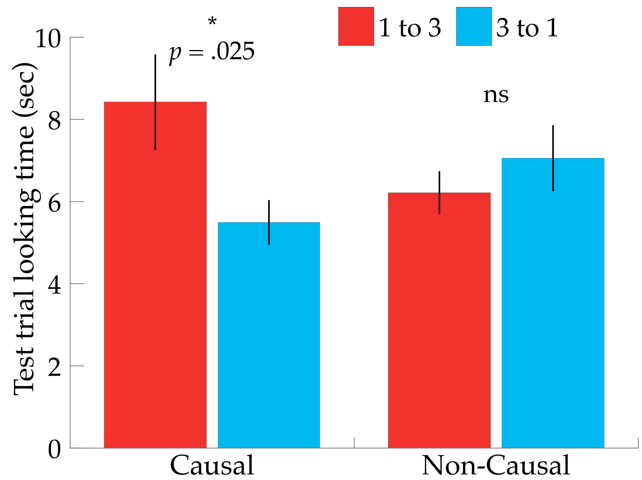

Fig. 5. Results of Experiment 3. Infants who were habituated to causal launching events dishabituated more to 1:3 causal events than 3:1 causal events. Infants habituated to non-causal events dishabituated equally to 1:3 and 3:1 non-causal events. Error bars represent +/- 1 SEM.

General Discussion

Our three experiments reveal categorical boundaries within causal *perception*, defined by an interplay of physical and perceptual constraints. In Experiment 1, adults' performance on a search task indicated that causal events with speed ratios of 1:3 are categorically different from symmetrical launching events, but events with speed ratios of 3:1 are not. In Experiment 2 we found that adult observers have difficulty detecting changes in speeds for lower speed ratios (particularly 1:2 and below), and further evidence that this categorical distinction cannot be attributed to low-level differences between causal and non-causal events. Finally, Experiment 3 provided evidence that this categorical distinction is present in 7- to 9-month-old infants, raising the possibility that it is a reliably early-developing feature of causal perception and perhaps core knowledge (Carey, 2009).

While real-world collisions are constrained by Newtonian mechanics, our results suggest that this categorical boundary is more directly determined by constraints on perception. Those perceptual constraints are in the vicinity of the Newtonian limit on collision events, but seem to be rough approximations rather than a precise reflection of Newtonian physics. It makes sense that causal perception should define this boundary flexibly, given that there are many features of both objects and the environment that could drop the limit below 1:2. However, our performance-based and infant methods strongly suggest that causal perception distinguishes causal events that would likely indicate some hidden force acting on B (internal or external) from those that do not.

Author Contributions

Experiments 1 and 2 were designed by JFK, BS, and FCK, and executed and analyzed by JFK and BS. All authors contributed to the design of Experiment 3, which was executed and analyzed by AEW, CE, and KW. The manuscript was primarily written by JFK and BS, with the methods and results of Experiment 3 written by AEW and CE, and revision by all authors.

References

Boyle, D. G. (1960). A contribution to the study of phenomenal causation. *Quarterly Journal of Experimental Psychology*, *12*(3), 171-179.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.

Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods*, *37*(3), 436-449.

Brown, J. F. (1931). The visual perception of velocity. *Psychological Research*, *14*(1), 199-232.

Calderone, J. B., & Kaiser, M. K. (1989). Visual acceleration detection: Effect of sign and motion orientation. *Perception & Psychophysics*, *45*(5), 391-394.

Carey, S. (2009). *The origin of concepts.* Oxford; New York: Oxford University Press.

Casstevens, R. (1997) *jHab* [Computer Software]. http://woodwardlab.uchicago.edu/resources/

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225-234. doi:10.1016/j.nlm.2008.06.002

GreenSock, Inc. (2015). *TimelineMax* [Computer Software]. https://greensock.com/timelinemax.

Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception & Performance*, *32(4)*, 840-853.

JASP Team (2016). JASP (Version 0.8.0.0)[Computer software].

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265-288.

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental Science*, *16*(3), 327-335.

McIntyre, J., Zago, M., Berthoz, A., & Lacquaniti, F. (2001). Does the brain model newton's laws? *Nature Neuroscience*, *4*(7), 693-694.

Michotte. (1963). *The perception of causality.* New York, NY: Basic Books. (Original work published 1946)

Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster than non-causal events. *PeerJ*, *5*, e2932.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factors for Common Designs [computer software manual]. https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf (R package version 0.9.12-2).

Murdock Jr, B. B., & Ogilvie, J. C. (1968). Binomial variability in short-term memory. *Psychological Bulletin*, *70*(4), 256-260.

Natsoulas, T. (1961). Principles of momentum and kinetic energy in the perception of causality. *The American Journal of Psychology*, *74*(3), 394-402.

Orban, G. A., Van Calenbergh, F., De Bruyn, B., & Maes, H. (1985). Velocity discrimination in central and peripheral visual field. *Journal of the Optical Society of America A*, *2*(11), 1836-1847.

Qualtrics. (2005). [Computer Software]. Provo, UT: Qualtrics.

Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science*, *6*(1), 77-97.

Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology*, *23*(3), 250-254.

Runeson, S. (1983). *On visual perception of dynamic events.* Uppsala: Academiae Ubsaliensis; Stockholm, Sweden.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), 411-437.

Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *44*(2), 321-342.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1-2), 1-46.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*(2), 259-290.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299-309.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89-96.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605-632.

Traschütz, A., Zinke, W., & Wegener, D. (2012). Speed change detection in foveal and peripheral vision. *Vision Research*, *72*, 1-13.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, *14*(5), 498-504.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48.

Watamaniuk, S. N., & Heinen, S. J. (2003). Perceptual and oculomotor evidence of limitations on processing accelerating motion. *Journal of Vision*, *3*(11), 698-709.

Werkhoven, P., Snippe, H. P., & Alexander, T. (1992). Visual processing of optic acceleration. *Vision Research*, *32*(12), 2313-2329.