

*Causality and continuity close the gaps in
event representations*

**Jonathan F. Kominsky, Lewis Baker,
Frank C. Keil & Brent Strickland**

Memory & Cognition

ISSN 0090-502X

Mem Cogn

DOI 10.3758/s13421-020-01102-9



Your article is protected by copyright and all rights are held exclusively by The Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Causality and continuity close the gaps in event representations

Jonathan F. Kominsky¹ · Lewis Baker² · Frank C. Keil³ · Brent Strickland^{4,5}Accepted: 17 September 2020
© The Psychonomic Society, Inc. 2020

Abstract

Imagine you see a video of someone pulling back their leg to kick a soccer ball, and then a soccer ball soaring toward a goal. You would likely infer that these scenes are two parts of the same event, and this inference would likely cause you to remember having seen the moment the person kicked the soccer ball, even if that information was never actually presented (Strickland & Keil, 2011, *Cognition*, 121[3], 409–415). What cues trigger people to "fill in" causal events from incomplete information? Is it due to the experience they have had with soccer balls being kicked toward goals? Is it the visual similarity of the object in both halves of the video? Or is it the mere spatiotemporal continuity of the event? In three experiments, we tested these different potential mechanisms underlying the "filling-in" effect. Experiment 1 showed that filling in occurs equally in familiar and unfamiliar contexts, indicating that familiarity with specific event schemas is unnecessary to trigger false memory. Experiment 2 showed that the visible continuation of a launched object's trajectory is all that is required to trigger filling in, regardless of other occurrences in the second half of the scene. Finally, Experiment 3 found that, using naturalistic videos, this filling-in effect is more heavily affected if the object's trajectory is discontinuous in space/time compared with if the object undergoes a noticeable transformation. Together, these findings indicate that the spontaneous formation of causal event representations is driven by object representation systems that prioritize spatiotemporal information over other object features.

Keywords False memory · Perception · Event representation

Processing scenes and events in real time requires a balance between accuracy and speed. We extract useful bits of information quickly and reliably, but limited processing resources make it impossible to attend to all available information in a timely fashion. A primary strategy for dealing with this

problem is to employ specialized perceptual and inferential mechanisms that serve to prioritize certain types of information to the detriment of others.

In change blindness paradigms, for example, large changes to background objects can go entirely unnoticed (Jensen, Yao, Street, & Simons, 2011), while changes to prioritized categories of objects are quite likely to be detected. For instance, changes to animate entities are easier to detect than changes to obviously inert objects (New & Scholl, 2009).

Just as *object* representations boost attention and memory in useful ways, there is also evidence that *event* representations serve the same function. Memory for token instances of events is heightened at event boundaries (i.e., the moment at which one event transitions to another; Baker & Levin, 2015; Zacks & Swallow, 2007). In addition to event *tokens*, representations of event *types*, such as containment versus occlusion, modulate dynamic attention and memory toward object properties that help predict event specific outcomes in both infants and adults (Baillargeon & Wang, 2002; Strickland & Scholl, 2015).

What principles predict how information will be prioritized and stored during the perception of events? A broad way of characterizing the processes of visual cognition is as making

Jonathan F. Kominsky and Lewis Baker contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13421-020-01102-9>) contains supplementary material, which is available to authorized users.

✉ Jonathan F. Kominsky
jonathan.kominsky@rutgers.edu

✉ Brent Strickland
brent.strickland@ens.fr

¹ Department of Psychology, Rutgers University, 101 Warren St. Rm. 301, Newark, NJ 07102, USA

² Pymetrics, Inc., New York, NY, USA

³ Yale University, New Haven, CT, USA

⁴ Ecole Normale Supérieure & Institut Jean Nicod, 29 rue d'Ulm, 75005 Paris, France

⁵ School of Collective Intelligence, UM6P, Ben Guerir, Morocco

intelligent (albeit likely unconscious) inferences regarding the nature of unfolding events (von Helmholtz, 1867). These inferences then have reflex-like consequences for attention and memory. Given the time constraints inherent to event processing and the often noisy and incomplete nature of incoming information, the mind must employ a set of heuristics that usually deliver accurate inferences but can go awry in carefully crafted laboratory settings. Here, we concentrate specifically on the perception of *causal events* as a way of exploring in detail such a set of heuristics.

Causal events are interesting from this perspective as recent work has suggested that there are highly specialized visual routines dedicated to the detection of causal events, leading to retinotopically specific visual adaptation to “causal launching” (Kominsky & Scholl, 2020; Rolfs, Dambacher, & Cavanagh, 2013). Moreover, simple causal events automatically guide visual attention towards category relevant information in both adults and infants (Kominsky et al., 2017). Thus, prior empirical evidence suggests that causal perception provides a potentially fruitful test case in which to study rapid heuristics in event processing.

In contrast to the aforementioned work, which concentrates on very simple Michottean “causal launching displays” (Michotte, 1946/1963), containing only simple geometric shapes moving in linear trajectories, the current project instead investigates the perception of more complex events, of the type we are likely to experience in our everyday lives. We employ a false-memory paradigm established by Strickland and Keil (2011) as an indirect way of asking how, in real time, heuristics are employed to create causal impressions that then trigger the creation of false memories.

Strickland and Keil (2011) showed observers simple video clips in which an agent launches an object (e.g., shoots a basketball). Crucially, these videos never actually showed the moment of contact or release in such launching events. After witnessing these “incomplete events,” the video either showed footage that implied that the launching had occurred (e.g., footage showing the resulting trajectory of a basketball toward a hoop) or did not imply causality (e.g., footage showing a person walking on a basketball court). Participants falsely reported seeing the moment of contact or release (e.g., the moment of release of the basketball) significantly more in the causal implication conditions. More recent work has gone on to demonstrate that these effects are impervious to many plausible “top down” influences, suggesting that the phenomenon is indeed driven by perceptual heuristics with specific triggers as opposed to rich background knowledge. Thus, explicit knowledge that false memory is being tested does not disrupt the effect (Papenmeier, Brockhoff, & Huff, 2019), nor does the filling-in effect vary as a function of expertise with the specific type of video being shown (Brockhoff, Huff, Mauer, & Papenmeier, 2016). For the current purposes, we use this paradigm to allow us to examine, via false-memory

creation, the elements that trigger the creation of coherent causal event representations from incomplete information.

Across three experiments, we assessed three factors that could reasonably influence these dynamic impressions of causality: (1) *Event familiarity*—Perhaps we spontaneously “fill in the causal blanks” only when there is a familiar specific event schema available in memory (e.g., shooting a basketball toward a basketball hoop). (2) *Object identity*—Perhaps we spontaneously fill in the causal blanks only when a launched object is perceptibly identical prelaunch and postlaunch (e.g., seeing the trajectory of a football may not fill in the blanks for shooting a basketball). (3) *Spatiotemporal continuity*—Perhaps we spontaneously fill in the causal blanks if and only if the trajectory of the object appears compatible with basic physical principles such as the Spelke principle (Spelke, Breinlinger, Macomber, & Jacobson, 1992), that objects should follow continuous paths through space and time.

We discuss the logic and plausibility of each factor in the relevant introductions to the individual experiments that follow. To foreshadow, the event familiarity and object identity hypotheses (Hypotheses 1 and 2) are refuted by our data, but we find evidence supporting Hypothesis 3, the spatiotemporal continuity hypothesis.

Experiment 1

Experiment 1 concentrated primarily on the event familiarity hypothesis by investigating the role of familiar event schemas. Schemas, in this context, are semantic representations in long-term memory that are used to make predictions about the outcome of an event given inferences about goals and previously observed events of the same kind (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). The stimuli used in Strickland and Keil (2011) fit into highly familiar schemas, such as “shooting a basketball toward a hoop” or “kicking a soccer ball toward a goal.” Participants’ false memory for the moment of release or contact in these events could be driven by their extensive and specific semantic knowledge about these events rather than a more general process of event representation, especially given that recent work has shown that such schemas can support filling in incomplete information from many different parts of an event (Kosie & Baldwin, 2019). With the stimuli used by Strickland and Keil (2011), one could even rely on the schemas that involve the mechanics of the human body. Past work has found that we demonstrate better memory for sequences that follow plausible bodily mechanics (Lasher, 1981), and that recognizable preparatory motions by agents draw attention and seem to support rich predictions (Cohn, Paczynski, & Kutas, 2017).

To test the event familiarity hypothesis, Experiment 1 attempted to replicate Strickland and Keil (2011) using entirely novel launching events, constructed in a three-dimensional

animated environment, with unfamiliar beginnings as well as unfamiliar outcomes, involving no human actors. While these events were novel to participants, they still had immediately recognizable causal content: They involved either straightforward launching events or “launching-by-expulsion” (Michotte, 1946/1963). In other words, even though the setting and objects were unfamiliar, the predominant causal interaction was of an abstract type that even infants reliably perceive as causal by 6 months of age (Saxe & Carey, 2006).

One could possibly argue that these “novel” events are not truly novel, as by adulthood people have ample experience seeing such events as causal. Even in the absence of identifiable agents, it may be possible to recognize the overall structure of a “preparatory action” and a “coda,” and fill in the missing link from that (Cohn et al., 2017). However, the point of these videos was not to introduce a causal relationship that was so unfamiliar that it had to be learned. Rather, it was to create novel instances of the same (implied) abstract causal relationships that are automatically extracted from the world through perceptual mechanisms (Hubbard, 2013; Kominsky & Scholl, 2020; Michotte, 1946/1963; Rolfs et al., 2013), while ensuring that there is a lack of conceptual familiarity with the basic category of event on display.

Even more precisely, the question explored in Experiment 1 was whether that abstract causal content alone, with no specific familiar specific context or schema, would be sufficient to produce the filling-in effect. A causal implication condition was contrasted with a second condition in which there was no causal implication, as in Strickland and Keil (2011).

The predicted result, following from Strickland and Keil (2011), was that participants should be more likely to fill in a moment of release or contact that they never saw when the incomplete launch is followed by a causal implication. While we argue that this is due to causal implication enabling participants to construct a complete event representation, another possibility is that the videos with the lack of causal implication actually *disrupt* memory for the events immediately preceding their onset. That could be depressed only by the disruptive nature of the non sequitur videos.

To test this alternative explanation, we added a third condition in which there was no causal implication, but the moment of contact/release was actually presented. If participants do not report seeing the moment of release in this condition, it indicates that the lack of causal implication is disruptive, rather than the presence of causal implication being constructive. If, however, participants show accurate memory for the moment of release when it is actually present, then it supports our proposal that causal implication drives this filling-in effect.

Method

Experiment 1 was conducted at the Institut Jean Nicod and ruled exempt from review by the CERES IRB board in Paris, France.

Participants

We conducted pilot experiments for Experiments 1 and 2 (see the [Supplemental Materials](#)). Based on Strickland and Keil (2011), which used six videos per participant and roughly 15 participants per condition, these pilot experiments had four videos per participant but 30 participants per group (thus double the sample size). We then conducted a power analysis based on the weakest intergroup effect on target items in Pilot Experiment 1 (see [Supplemental Materials](#)), and determined that to reach 95% power to detect that effect we would need 68 participants per group. Due to imperfect randomization, we ended up slightly exceeding this target.

Participants ($N = 206$) were recruited via Amazon's Mechanical Turk (for more information on this population, see Germine et al., 2012; Paolacci, Chandler, & Ipeirotis, 2010) for \$0.75 compensation for an approximately 5-minute task. Participants were randomly assigned to one of three conditions, described below. An additional 11 participants were recruited but not included in the final sample due to violating preregistered exclusion criteria (see Results).

Materials and procedure

Four novel movies were animated using the 3D editing software, Blender (v2.66; The Blender Foundation, www.blender.org, 2015). Stimuli are available to view (<https://osf.io/mjwkd/>). The animations depicted novel (and thus unfamiliar) event types, in each case following the same progression of three shots (depicted in Fig. 1): The first shot showed all objects and items in the environment, the second initialized the movement of a ball leading to a launching action, and the third showed the consequences of the launching event on objects on the other side of the environment.

Participants watched four videos from one of three randomly assigned conditions (each participant only saw videos from one condition). The *causal condition* depicted an implied object release that cut to a causally consistent shot of the launched object continuing on an expected trajectory towards a target and having some impact on an object on the other side of the space. Importantly, the moment of release was never actually shown in this condition. The *non sequitur condition* did not show the launched object in the second shot of the video, but instead showed the occurrence of an unlikely event. For example, as illustrated in Fig. 1, one video implied a ball getting hit by a bar. In the causal condition, the following shot showed the ball hitting cylinders on a platform and knocking one of them over (an effect of the ball's trajectory), whereas in the *non sequitur* condition, the ball was not present, and the video instead depicted cylinders moving up and down like pistons. Neither the causal nor the non sequitur conditions actually showed the launching action. As

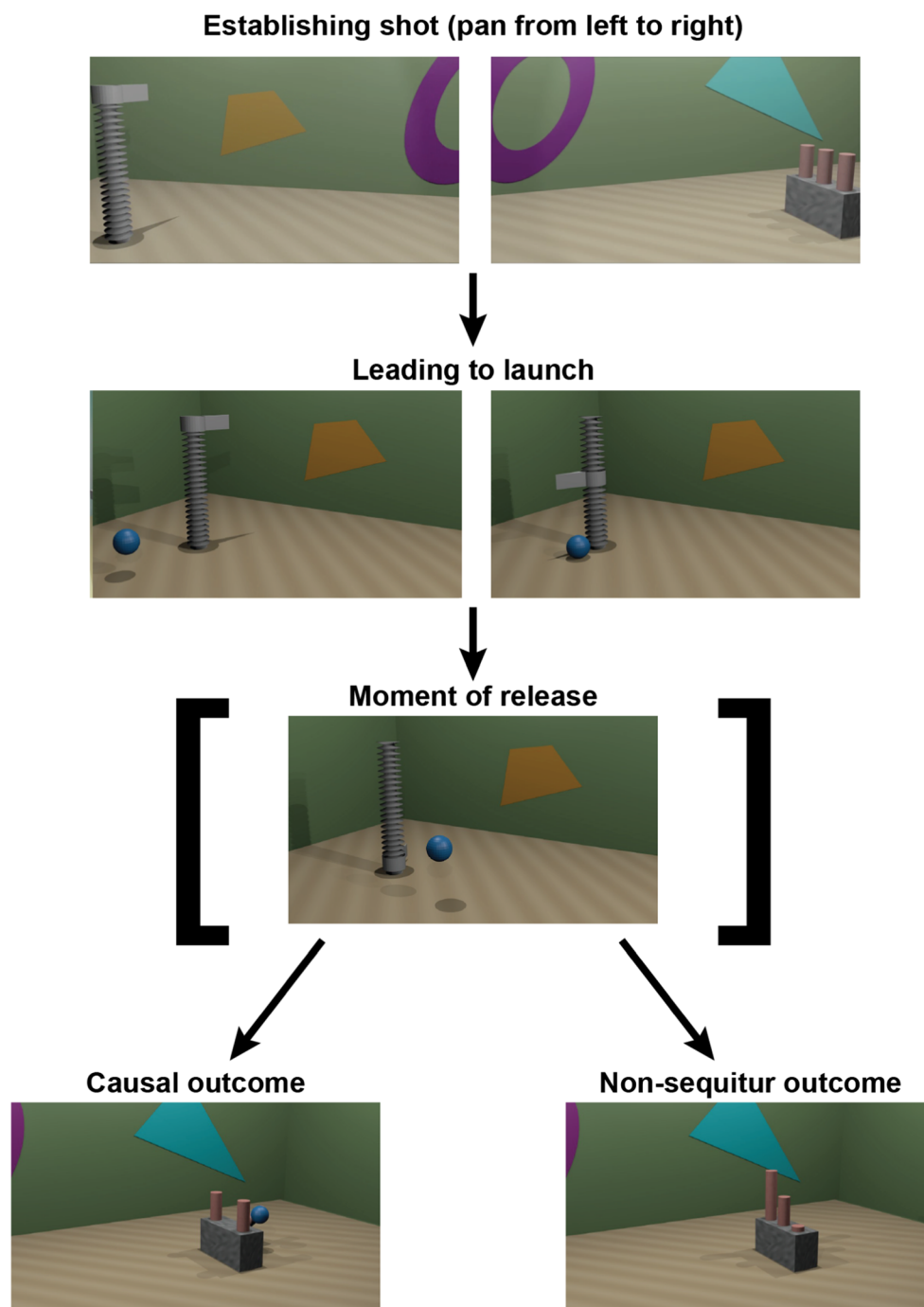


Fig. 1 Schematic depiction of stimuli used in Experiment 1. The video starts with an establishing shot that shows the whole scene (top images), and then a movement sequence prior to the launch (next pair). The moment of release was shown in the non sequitur/complete condition,

otherwise it was cut from the video. Whether or not it was shown, the video immediately cut to the outcome, either causal (bottom left) or non sequitur (bottom-right; see text)

a control to verify that the non sequitur conclusion was not simply disrupting attention to the end of the first half of the video, a third group of participants saw a *non sequitur complete* condition, where participants actually saw the launching action (the “moment of release” or “moment of contact”) leading to the causally unpredictable conclusion.

After viewing each video, participants saw 10 images and were asked to indicate whether each image had appeared in the preceding video. There were three image types: images of the implied but unseen moment of release (*target moment-of-release images*; one per video), images taken from the preceding video for which the correct answer was “yes” (*seen-*

action images; six per video), and images taken from a video with highly salient changes for which the correct answer was “no” (lures; e.g., a picture of the scene in which a central object was a different color; three per video). All video and picture orders were randomized.

Results and discussion

A total of 11 participants were removed and replaced with new recruitment for reaching less than 50% accuracy across all nontarget recognition items (computed as the average of the average accuracy for seen-items and the average for lures, to compensate for the uneven number of items of each type): six from the causal/ incomplete condition, two from the non sequitur (NS)/incomplete condition, and three from the non sequitur/complete condition. In addition, due to imperfect randomization, participant assignment was slightly unbalanced, with two extra participants in the non sequitur/incomplete condition. This left 68 participants in the causal/incomplete condition, 70 in the non sequitur/incomplete condition, and 68 in the non sequitur/complete condition.

The key dependent variable (DV) was the proportion of “yes” responses to the test items averaged across all four events for each participant. Our preregistered analyses started with a 3 (condition: incomplete vs. NS incomplete vs. NS complete; between-subjects) \times 3 (item type: target vs. seen-image vs. lure; within-subjects) mixed-model analysis of variance (ANOVA) using R’s afex package (Singmann et al. 2019). This analysis found main effects of condition $F(2, 203) = 22.96, p < .001$, and item $F(2, 406) = 359.83, p < .001$, as well as a significant interaction, $F(2, 406) = 30.71, p < .001$. We conducted separate preregistered one-way between-subjects ANOVAs examining the effect of condition for each item type.

We first examined the item type of primary interest, the target moment-of-release image. Note that “yes” (recognition) responses in the incomplete conditions were false memory of an implied image, whereas recognition in the non sequitur complete condition was an accurate memory of an action seen by the participants. The results can be found in Fig. 2.

A one-way ANOVA of the effect of condition on average moment-of-release recognition responses was significant, $F(2, 203) = 46.80, p < .001, \eta_p^2 = .316$. Post hoc Tukey HSDs confirmed the impression provided by Fig. 2: There were significant differences between the non sequitur/Incomplete condition ($M = 38.2\%, SD = 33.2$) and the causal/incomplete condition ($M = 71.3\%, SD = 27.1$), as well as between the non sequitur/Incomplete condition and the non sequitur/complete condition ($M = 81.6\%, SD = 21.0$), $ps < .001$. However, there was no significant difference between

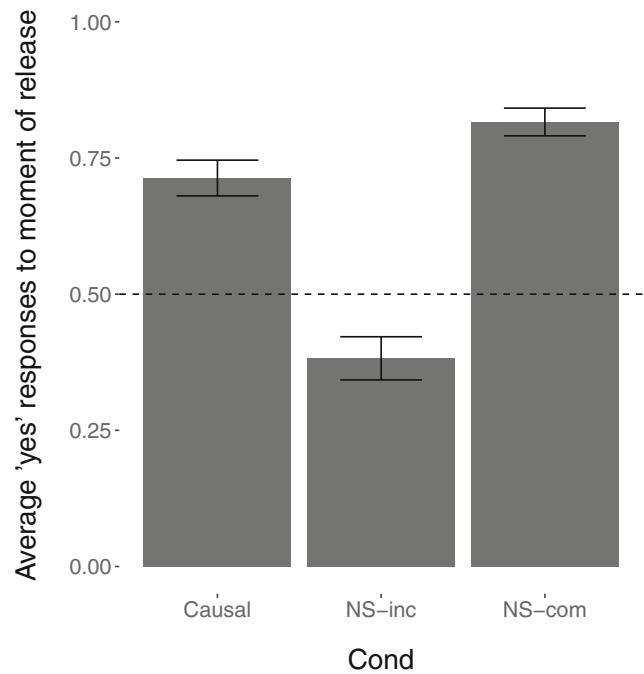


Fig. 2 Average frequency of “yes” responses to moment-of-release images in Experiment 1. Error bars represent ± 1 SEM. The dashed line at 50% represents chance responding

the causal/incomplete condition and the non sequitur/complete condition, $p = .078$.¹

In short, participants were highly likely to correctly recognize that they saw the moment of release in the non sequitur/complete condition and to falsely recognize an implied moment of release in the causal/incomplete condition, but much less likely to make the same error in the non sequitur/incomplete condition. (These findings were nearly identical to what we observed in the pilot experiment.)

The corresponding ANOVAs for the seen-image and lure item types found no significant effects of condition on the rate of “yes” responses, $F(2, 203) = 1.78, p = .17, \eta_p^2 = .02$ and $F(2, 203) = 0.11, p = .9, \eta_p^2 = .001$, respectively. The full pattern of responses can be found in Table 1. Experiment 1 extends the results of Strickland and Keil (2011) to completely novel events that are not supported by familiar schemas, or even somewhat abstracted schemas having to do with bodily motion (Lasher, 1981). Participants “filled in” the moment of release for events that they had never seen before, in completely unfamiliar contexts, provided there was spatiotemporal continuity and a causal consequence. Furthermore, we were

¹ The magnitude of this contrast in the pilot experiment, which was also nonsignificant ($p = .094$), was the basis of our power analysis for the sample size of this experiment and Experiment 2. While this difference was nonetheless nonsignificant here, despite 80% power to detect it, it is also not critical to our account either way. Our account predicts that filling in should be stronger when there is a causal implication, and our results leave no question as to that. While it would be intriguing to ask whether filling-in yields a memory as strong as actually seeing the event, it does not bear on the role of causality, regardless of whether there is a familiar schema available or not.

Table 1 Average percentage “yes” responses to moment-of-release images, seen action images, and lure images in all conditions in Experiment 1

	Release	Seen action	Lures
Casual/Inc.	71.32% (27.07)	79.84% (15.13)	31.74% (19.75)
Non sequitur/Com.	81.62% (20.99)	84.31% (12.18)	30.27% (18.32)
Non sequitur/Inc.	38.21% (33.17)	81.25% (14.94)	31.07% (18.11)

Note. Numbers in parentheses are standard deviations. Inc. = incomplete (moment of release not shown), Com. = complete (moment of release shown)

able to rule out the deflationary explanation that the non sequitur event was simply distracting and thus disrupted memory around the moment of release: Participants had no difficulty recognizing that they had seen the moment of release when it was actually presented, even when followed by a non sequitur event. These findings support the hypothesis that false recognition of an implied action relies upon causal inferences (likely guided by spatiotemporal information), but not upon highly specific semantic event schemas.

Experiment 2

In the causal implication condition of Experiment 1, the perceived motion after (implied) contact/release was always of the causally relevant (i.e., launched) object interacting further with the scene (e.g., knocking over a cylinder). That is, in addition to the trajectory of the object, the relevant object was involved in a further causal interaction in the causal condition of Experiment 1. In the non sequitur condition, participants instead saw a causally irrelevant event that could not be caused by the launched object or the launching event (e.g., pistons pumping up and down). Thus, the absence of a causally relevant object was confounded with the presence of a causally irrelevant event. It is therefore impossible to determine whether causal impressions were created by seeing causally relevant object motion in the second half of the video (i.e., the launched object having some further causal interaction), or inhibited by the presence of an event that could not be caused by the launched object in any way.

Experiment 2 examined this issue explicitly by replicating and extending the findings from Experiment 1. In this experiment, we crossed the presence/absence of the object's motion with the presence/absence of a secondary event that could have been a consequence of the object's subsequent trajectory. By presenting the effect of the ball separate from its movement, we created a case which would allow us to assess more precisely the types of causal information required to trigger filling in: If any causal schema is enough, then the causal consequence should be the factor that determines the filling-

in effect. However, under the object identity and spatiotemporal continuity hypotheses, the presence or absence of the object in the second half of the event should be the determining factor.

Method

Participants

To be as conservative as possible, we based our power analyses for Experiments 1 and 2 on the weakest effect with $p < .1$ across both pilot experiments (see [Supplemental Materials](#)), which was the difference between the causal incomplete and non sequitur complete conditions found in Experiment 1, as described above. Therefore, we once again aimed to recruit 68 participants per condition in Experiment 2. Participants ($N = 272$) were recruited via Amazon's Mechanical Turk for \$0.75 compensation for an approximately 5-minute experiment, and randomly assigned to one of four groups.

Materials and procedure

The animated stimuli from Experiment 1 were modified into four conditions. We manipulated two orthogonal features of these stimuli: Whether the causally consistent outcome (e.g., the cylinders getting knocked over) or non sequitur outcome (e.g., the cylinders pumping up and down) occurred in the final shot of the video (causal/non sequitur), and separately whether the ball was visible in the second shot of the video (ball visible/ball invisible). These four conditions are illustrated in Fig. 3.

Thus, the causal and non sequitur incomplete conditions of Experiment 1 were the *causal-visible* and the *non sequitur-invisible* conditions in the current experiment, respectively, and the videos used in those conditions were simply the same ones as were used in Experiment 1.

In the *causal-invisible* condition, the ball was simply removed from the final shot of the video, but the causal consequence still occurred (e.g., the cylinder on the platform still fell over). In the *non sequitur-visible* condition, the ball was present and moved in a plausible trajectory (though a different one from the causal video), while the non sequitur event occurred without the ball ever making contact (e.g., the pistons went up and down while the ball soared overhead).

In this experiment, the “lure” items were constructed by taking images from other conditions. For example, the “lure” items in the non sequitur-invisible condition were simply the seen-action items from the second half of the causal-visible condition.

Importantly, none of the four conditions actually depicted the moment of release. If causal inference about typical causal interactions (e.g., knocking over a cylinder) is sufficient on its own to elicit false recognition of a launching event, then we

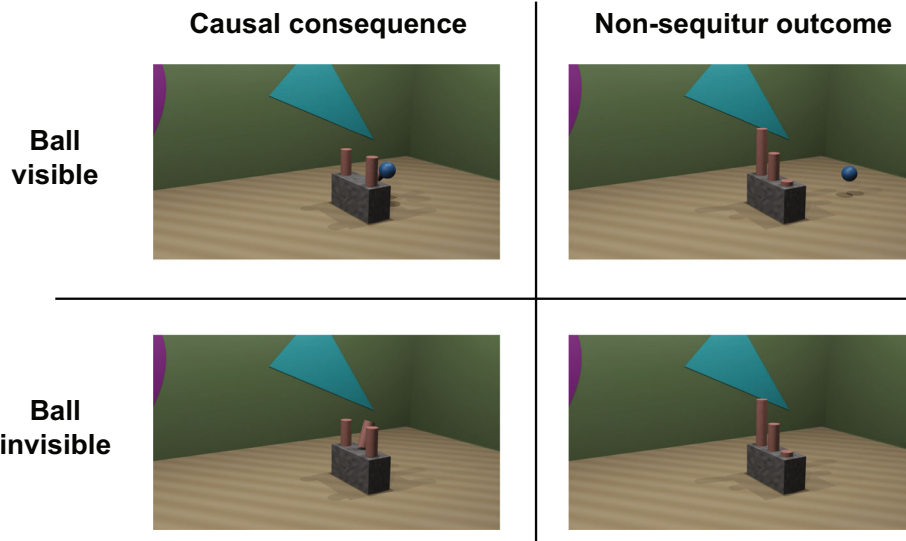


Fig. 3 Example stimuli from the second half of the videos in Experiment 2, corresponding to the two pictures at the bottom of Fig. 1. The causal consequence in this case was a cylinder being knocked over. The non

sequitur outcome involved the pistons moving up and down. Separately, the ball was or was not visible in the second half of the video

should see equally high false recognition of the moment of release in the visible and invisible causal conditions, and equally low false recognition in the visible and invisible non sequitur conditions. If the non sequitur outcome disrupts the formation of a coherent event representation, then we should find high false recognition only in the causal-visible condition. However, if spatiotemporal continuity alone is sufficient and necessary to form a causal event representation of the object being launched, then we should find high false recognition rates in both causal- and non sequitur-visible conditions, but in neither of the invisible conditions.

Results and discussion

We used the same exclusion criteria as Experiment 1 (<50% accuracy across all nontarget items, weighted by the number of lure and seen items), removing and replacing participants until we had 68 in each condition. This removed a total of 28 participants—six from the causal-visible condition, seven from the causal-invisible condition, eight from the non sequitur-visible condition, and seven from the non sequitur invisible condition.

We conducted a 2 (causality: causal consequence vs. non sequitur outcome; between-subjects) × 2 (ball visibility: visible ball vs. invisible ball; between-subjects) × 3 (item type: target vs. seen-action vs. lure; within-subjects) mixed-model ANOVA, which found main effects of ball visibility, $F(1, 268) = 3.95, p = .048$, and Item type, $F(2, 536) = 611.54, p < .001$, and an interaction between the two, $F(2, 536) = 73.19, p < .001$. There was no significant main effect of causality, $F(1, 268) = .04, p = .83$, no significant interaction of causality and visibility, $F(1, 268) = .31, p = .58$, and there was a

significant interaction between causality and item type, $F(2, 536) = 4.50, p = .01$. The three-way interaction was not significant, $F(2, 536) = 2.90, p = .055$. To test the specific hypotheses of interest for target items, we conducted separate preregistered 2 (causality: causal consequence vs. non sequitur outcome) × 2 (ball visibility: visible ball vs. invisible ball) fully between-subjects ANOVAs for each item type.

The rate of "yes" responses for the target moment-of-release images can be found in Fig. 4. The 2 × 2 ANOVA revealed a significant main effect of visibility, $F(1, 268) = 51.62, p < .001, \eta_p^2 = .162$, but no effect of causality, $F(1, 268) = 1.49, p = .22$, and no interaction, $F(1, 268) = .02, p = .96$. Participants generated significantly more false alarms to the target item in the visible conditions ($M = 64.9\%, SD = 29.2$) than the invisible conditions ($M = 37.9\%, SD = 32.6$). In other words, the filling-in effect was entirely contingent on whether the trajectory of the ball was visible in the second half of the video, regardless of whether the other events that

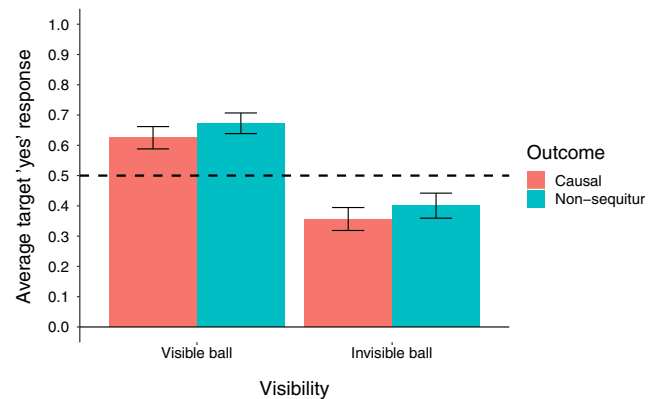


Fig. 4 Results of Experiment 2. Error bars represent ± 1 SEM. The dashed line represents chance responding

occurred implied a further causal interaction with the ball. Full means can be found in Table 2.

For seen-action items, there was a significant effect of visibility, $F(1, 268) = 5.99, p = .015, \eta_p^2 = .022$, no effect of causality, $F(1, 268) = .007, p = .94$, and no interaction, $F(1, 268) = 2.94, p = .088$. The effect of visibility for these items is precisely the opposite of the effect on target items: Participants were *less* likely to say “yes” to an item they actually saw in the visible condition ($M = 82.8\%, SD = 12.4$) than the invisible condition ($M = 86.5\%, SD = 12.4$), so this was not indicative of an overall “yes” bias.

The analysis of Lure items indicates the effect on target items was also not an overall drop in accuracy in the visible conditions. This analysis found a main effect of visibility as well, $F(1, 268) = 26.67, p < .001, \eta_p^2 = .090$, but in this case participants were *less* likely to falsely report seeing one of the lure items in the visible condition ($M = 18.9\%, SD = 19.2$) than the invisible condition ($M = 31.2\%, SD = 20.5$). However, this analysis also revealed a main effect of causality, $F(1, 268) = 5.64, p = .018, \eta_p^2 = .020$, and a significant interaction, $F(1, 268) = 5.40, p = .021, \eta_p^2 = .020$. Post hoc Tukey HSD-corrected pairwise comparisons found that the effect of visibility was only significant in the causal conditions ($p < .001$) and not in the non sequitur conditions ($p = .18$). Critically, none of the effects observed for seen-action or lure items can explain the effect of visibility on target items: There is neither evidence for an overall “yes” bias, nor for an overall drop in accuracy that could explain why participants reported seeing the moment of release more often in the “visible” conditions.

These results (which also closely matched the results of the corresponding pilot experiment) confirm that implying a causal consequence in the second half of the video alone is not sufficient to induce false recognition of an implied action. However, spatiotemporal continuity of the central object was necessary to induce false recognition of the moment of release with or without additional causal implication. Whereas some work, particularly in language development, has argued that the goal or end point of a movement plays a particularly

critical role in encoding events in memory (Lakusta & Landau, 2005), our results indicate that information about object trajectory itself is critical to forming a coherent event representation (see also Liao, Flecken, Dijkstra, and Zwaan, 2020).

Experiment 3

The results of Experiments 1 and 2 indicated that the filling-in effect relies on the presence of a plausible object trajectory in the second half of the video, but not event familiarity or exposure to other types of causal information. This led us to wonder how robust this process of event construction truly was. Therefore, in Experiment 3, we delved deeper into the question of how the mind establishes correspondence between the object in the second half of the movie with the object being launched in the first. In particular, we were interested in whether event perception prioritizes spatiotemporal object tracking over object-intrinsic properties in ways that mirror infant and primate object cognition (thus, hypotheses 2 and 3 from the Introduction).

Particularly relevant here are findings that infants and primates prioritize spatiotemporal continuity over object-intrinsic features (e.g., category, color, shape) in object individuation tasks (Flombaum, Kundey, Santos, & Scholl, 2004; Xu & Carey, 1996). For example, if a 10-month-old infant witnesses a duck turn into a truck (and can be empirically shown to have noticed this change), the infant will nevertheless respond in their looking behavior as if they believe there to be only a single object in the scene. However, when shown a display in which objects move in such a way that the pattern of movement could only have been produced by a single object “popping in and out of existence,” infants will respond in their looking behavior as if they expect for there to be two objects. Evidence for this “tunnel effect” has been found in both research on primates (Flombaum et al., 2004) and adult perceptual processing in demanding visual environments (such as crowded search displays; Flombaum & Scholl, 2006).

Thus, prior evidence on object representation suggests that in resource limited populations and contexts, spatiotemporal object features (e.g., trajectory) are prioritized over object-intrinsic category features for the purposes of object individuation. It would stand that causal impressions, insofar as they depend on object representations as *input*, may also show a similar prioritization of spatiotemporal over object-intrinsic features.

This possibility is explicitly tested below by assessing false memory in incomplete events (e.g., someone throwing a dart at a dartboard, but not showing the moment of release) in which subsequent video footage either contains an object that has visibly changed to a new object category (e.g., the dart has

Table 2 Average “yes” responses to moment-of-release images, seen action images, and lure images in all conditions in Experiment 2

	Release	Seen action	Lures
Visible ball			
Causal	62.50% (30.38)	84.13% (12.56)	19.00% (19.57)
Non sequitur	67.28% (28.08)	81.43% (12.14)	18.87% (19.02)
Invisible ball			
Causal	35.66% (31.24)	85.23% (13.22)	36.76% (19.44)
Non sequitur	40.07% (34.05)	87.68% (11.57)	25.61% (20.22)

Note. Numbers in parentheses are standard deviations

turned into a balled-up piece of paper), but maintains spatiotemporal continuity, or shows an object that has not changed category but no longer maintains continuity (e.g., the dart is significantly further along its trajectory than it “should” be).

We conducted two in-lab pilot experiments (Pilot Experiments 3a and 3b), testing a new set of more naturalistic stimuli and attempting to gauge the effect of the planned manipulations of object identity and spatiotemporal continuity. These experiments are discussed in the [Supplemental Materials](#), and data can be found at <https://osf.io/mjwkd/>.

Experiment 3 used the videos validated in these pilot experiments to test the impact of two types of disruption: category violations (in which the object transforms from the first half of the video to the second, as described above) or continuity violations (in which the object appeared “too far” along its trajectory in the second half of the video).

Pilot Experiments 3a and 3b suggested that in a causal context, spatiotemporal continuity alone was sufficient to drive the filling-in effect, and that the effect was not disrupted by drastic changes to the features of the object itself. However, these pilot experiments were underpowered. Therefore, we conducted Experiment 3 to provide a well-powered examination of these issues using a full 2×2 design, allowing us to tease apart the effects of category and continuity violations.

Method

Experiment 3 was conducted at the Institut Jean Nicod and ruled exempt from review by the CERES IRB board in Paris, France.

Participants

Based on a power analysis of the effects observed in Pilot Experiment 3b, we found that for the contrast between continuity and category violation to reach 80% power in an independent-samples t test, we would need 25 participants per group. In order to detect any possible interactions in our 2×2 design, we doubled this estimate, and therefore preregistered a sample of 50 participants per group, or 200 participants total. This gave us 80% power to detect a Cohen's f^2 effect size of .040, corresponding to an η^2 effect size of .038. The preregistration can be found at <https://osf.io/g9nyu>.

We recruited 200 participants from Amazon Mechanical Turk, 50 in each group. Based on the same exclusion criteria as Experiments 1 and 2, we excluded and replaced nine participants.

Stimuli and procedure

Each participant saw eight different videos. Each video depicted a person who began an action as if they were going

to release or launch an object (e.g., shooting a basketball or kicking a soccer ball). Just before the moment of release (e.g., shooting or kicking), the video switched perspectives so that the moment of release was never seen.

We created a simple 2×2 between-subjects design, category violation (no violation vs. violation) \times continuity violation (no violation vs. violation), yielding four conditions: no violations, category violation only, continuity violation only, and both violations. In the category violation conditions, in the second half of the video, the object was replaced by a different object (e.g., a dart in the first half might be replaced by a ball of paper in the second half). In the continuity violation conditions, the second half of the video picked up ~ 1 second later into its trajectory than in the corresponding no-continuity-violation condition (i.e., the same object as the first half when there was no category violation, and the different object in the both violations condition). An example item is presented in Fig. 5, and the full stimuli can be found at <https://osf.io/t3gh5/>.

The first 200 participants were randomly assigned to one of these four conditions. The final nine participants were randomly assigned to conditions that did not have a full sample of 50 after the first round of exclusions. Only one round of exclusions was needed.

There were nine or 11 test images per video, but not all of them were the same as in Pilot Experiments 3a and 3b (see [Supplemental information](#)). In particular, in an effort to make the test items more balanced, excluding the target moment-of-release images, half of the images were seen action, and half were lures. For the “lure” items, three (or for one of the videos, two) were images that appeared in no video at all (e.g., in which the actor had different clothing), while one (or for one of the videos, two) was a still image from the second half of video in which the object was the object from the *other* category violation condition. That is, participants saw images of both the second half of the video that they actually observed, and the second half of the video in the other category violation condition. Using Fig. 5 as an example, a participant in the no-violation condition would see one test image of, for example, the dart hanging from the dartboard (a seen-action image), as well as an image of the ball of paper bouncing off the dartboard (a lure). For a participant in the category violation condition, they would see the same images, but which one was analyzed as a seen-action image versus a lure would be reversed. This allowed us to establish whether the category violation was detected, independent of any impact on the filling-in effect.

Results and discussion

We once again analyzed the proportion of “yes” responses to each item type. Our initial analysis was a 2 (category

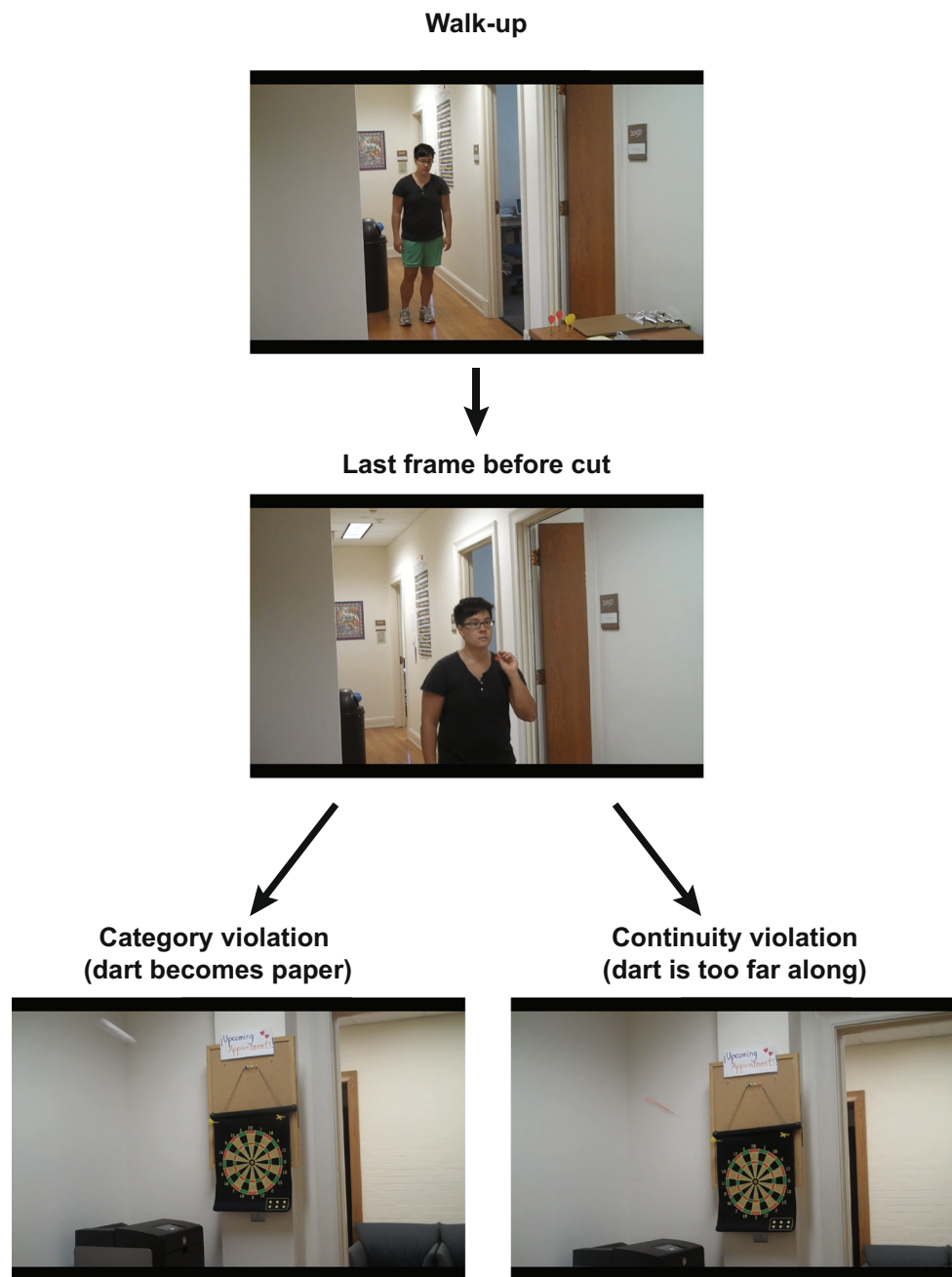


Fig. 5 Schematic depiction of stimuli manipulations used in Experiment 3. The video starts with a "walk-up" of a person approaching and interacting with the object. The moment of release was never shown. The video immediately cut to the outcome, which could involve a category violation (bottom left) or a continuity violation (bottom right). Notably, bottom-right image was seen in the continuity violation

conditions, but was used as a "lure" item in the conditions without a category violation, to verify that participants noticed the object transformation in the category violation condition (see text). Because each type of violation was manipulated independently, there were also videos (not shown) in which neither violation occurred, and in which both occurred (i.e., the paper shown "too far along")

violation; between) \times 2 (continuity violation; between) \times 3 (item type; within) mixed-model ANOVA. This revealed a main effect of continuity violation, $F(1, 196) = 4.33$, $p = .04$, an interaction between continuity violation and item type, $F(2, 392) = 4.38$, $p = .01$, and an interaction between feature violation and item type, $F(2, 392) = 4.17$, $p = .02$. No other effects were significant,

$ps > .1$. Following our preregistered analysis plan, we then conducted separate 2×2 ANOVAs for each item type. The average rate of "yes" responses for each item type in each condition can be found in Table 3.

The primary analysis of interest was of course the analysis of the target moment-of-release images. The rate of "yes" responses to target items is depicted in Fig. 6. A 2

Table 3 Average “yes” responses to moment-of-release images, seen action images, and lure images in all conditions in Experiment 3

	Release	Seen action	Lures
No category vio.			
No continuity vio.	80.83% (27.52)	90.71% (11.03)	23.71% (19.40)
Continuity vio.	69.00% (31.36)	83.01% (12.18)	28.59% (18.06)
Category vio.			
No continuity vio.	75.00% (28.57)	87.48% (14.34)	34.34% (19.42)
Continuity vio.	66.00% (30.62)	88.24% (12.96)	32.70% (20.26)

Note. Numbers in parentheses are standard deviations. “vio.” = violation

(category violation) × 2 (continuity violation) ANOVA² revealed only a main effect of continuity violation, $F(1, 196) = 6.21, p = .014, \eta_p^2 = .031$, no effect of category violation, $F(1, 196) = 1.12, p = .29$, and no interaction, $F(1, 196) = .11, p = .74$. Participants were significantly less likely to fill in the moment of release when there was a violation of spatiotemporal continuity ($M = 67.5\%, SD = 30.8$) than when there was not ($M = 77.9\%, SD = 28.1$). While the effect of the continuity violation is notably weaker than that observed in Pilot Experiment 3b (which found a nearly 30% drop in filling in from a continuity violation), it is still reliable. However, even radical changes to the features of the object had no detectable impact on the filling-in effect, and no interaction with continuity, indicating that spatiotemporal continuity alone is the primary driver of the filling-in effect. Notably, an analysis of the specific “lure” and “seen-action” items that captured the category violation found that participants were 77% accurate (yes for seen action, no for lures) in the category violation conditions, well above chance responding of 50%, one-sample $t(99) = 16.12, p < .001$. Thus, while the intrinsic object properties were represented at some level, and participants could recognize that a transformation had occurred, such transformations had no detectable impact on the rapid construction of causal event representations.

The 2 × 2 analyses of all seen-action and lure items found significant effects as well, but none of them align with the results of the target images and therefore cannot explain the effect of continuity on target image. For the seen-action items, there were no main effects of category violation, $F(1, 196) = .26, p = .61$, or continuity violation, $F(1, 196) = 3.597, p = .059$, but there was a significant interaction, $F(1, 196) = 5.88, p = .016, \eta_p^2 = .029$. We conducted post hoc pairwise comparisons of the effect of continuity violation at each level of category violation, and found that there was no effect when there was a category violation, $t(98) = .35, p = .73$, but

² Each ANOVA had 80% power to detect an effect size of $\eta_p^2 \geq .038$, equivalent to Cohen's $f^2 \geq .04$ for main effects and interactions, which would correspond to a small effect size (Cohen, 1988).

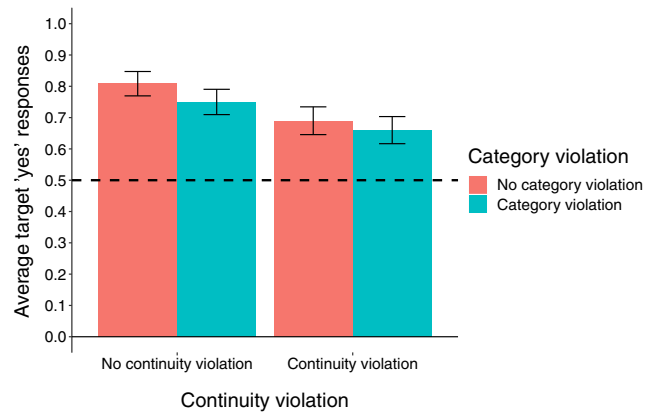


Fig. 6 Results of Experiment 4. Error bars represent ± 1 SEM. The only significant effect was a main effect of continuity violation. The dashed line represents chance responding

significantly worse performance (i.e., fewer “yes” responses) in the continuity-violation-only condition ($M = 83.0\%, SD = 12.2$) than in the no-violations condition ($M = 90.7\% SD = 11.0$), $t(98) = 3.31, p = .001, d = .66$.

For lure items, there was a significant main effect of category violation, $F(1, 196) = 7.29, p = .008, \eta_p^2 = .036$, such that participants were more likely to falsely report seeing a lure item when there was a category violation ($M = 33.5\%, SD = 19.8$) than when there was not ($M = 26.2\%, SD = 18.8$). There was no effect of continuity violation, $F(1, 196) = .35, p = .55$, and no interaction, $F(1, 196) = 1.42, p = .23$. The effects on seen-action and lure items were unexpected and cannot explain the primary result of interest, and so we do not investigate them further. However, there is certainly room for future work to explore the effects of these violations on memory for aspects of these events that were observed, or that were unobserved, but also not relevant to their causality.

General discussion

In three experiments we found that causal event representations are formed spontaneously in response to apparent spatiotemporal continuity regardless of the familiarity of the event or changes to the intrinsic properties of the object involved. Experiments 1 and 2 showed that no specific schema is required at all, by using novel computer-animated events and demonstrating that filling in occurred anytime the launched object’s trajectory was shown. Experiment 3 showed that the *identity* of the launched object is less relevant to the filling-in effect than spatiotemporal continuity of object trajectory: The object can undergo drastic changes to its surface features, but as long as its trajectory is plausibly continuous with the launching event, causal filling in will occur.

These results show that the causal filling-in effect originally reported by Strickland and Keil (2011) is the result of a process of constructing event representations that is flexible,

spontaneous, and sensitive to a specific set of perceptual cues. The spontaneity of these processes is particularly evident in Experiment 3: Participants in the category-violation condition accurately recognized when the identity of the object changed from the first half of the video to the second, which, intuitively, one would expect to indicate that the second half of the video was unrelated to the first. Nonetheless, in Experiment 3, the filling-in effect only detectably responded to continuity violations, showing that it occurred regardless of recognizing the change in the object's identity.

It is also worth noting the impressive robustness of the basic effect: Between these three experiments and the two reported by Strickland and Keil (2011), this filling in of the moment of release has now been found with over 20 different stimulus videos, and similar effects have been found with a further set of different stimuli by other researchers (Papenmeier et al., 2019). While each participant only sees four to eight videos, and there is some partial overlap in stimuli between a few of these experiments, this effect has now been shown across a variety of different scenarios, and in both live video and computer-generated animations. The effect is also generally unobtrusive. When the event representation is not disrupted, filling-in rates exceeded 60% in every experiment reported here.

As striking as the cases in which the filling-in effect emerges are the conditions in which it does not, and the degree of difference between the two (anywhere from a difference of 10% in Experiment 3 to 30% in Experiment 2). People failed to fill in blanks if the launched object was simply absent from the second half of the event (Experiments 1-2; see also Strickland & Keil, 2011, and Pilot Experiments 1-3a), if the video order is scrambled such that the "cause" event is not immediately followed by the trajectory of the object (Strickland & Keil, 2011), or if the object appeared "too far" along its trajectory (Experiment 3). Together with the filling-in effect in the category violation conditions and with novel events, the evidence to date indicates that the formation of these event representations is primarily affected by the impression of spatiotemporal continuity of movement of a discrete object, which is most strongly disrupted when the object is absent, but also diminished if it is in the wrong place.

While we can safely rule out the role of specific familiar schemas driving this filling-in effect, it is an open question whether these filling-in effects would emerge with familiar *or* unfamiliar events that are not automatically perceived as causal, as with the launching or launching-by-expulsion events used throughout these experiments. Future work could explore, for example, whether you find the same "filling in" for the flipping of a light switch, or if training participants on a novel "blicket-detector" causal system (e.g., Gopnik & Sobel, 2000) leads them to fill in a moment at which a causal object is put into contact with the device.

While we leave such questions to future work, based on the current findings we predict that such events will not generate the same effects. The results of Experiment 3 emphasize the role of apparent spatiotemporal continuity of movement even in the absence of what we might call semantic coherence (e.g., Davenport & Potter, 2004). This is consistent with work demonstrating that spatiotemporal continuity is a defining feature of events (Zacks & Swallow, 2007), and that violations of spatial continuity induce the perception of new events and increase awareness of changes in a scene (Baker & Levin, 2015). Combined, this implies a system that relies on automatic processing of dynamic events in order to form event representations.

The timing of our videos largely rules out the possibility of these event representations forming from Michottean causal perception phenomena like "causal capture" (Scholl & Nakayama, 2002), because we cut the video more than 300 ms before the moment of release (11 frames at 30 fps = 333ms) and previous work has shown a signature temporal integration window for causal perception that is 200 ms or less on either side of the event (Choi & Scholl, 2006; Newman, Choi, Wynn, & Scholl, 2008). However, other work has shown that we make automatic predictive simulations of events as they occur, or postdictive simulations as they counterfactually could have occurred (Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017). We propose that the process underlying the filling-in effect is related to these predictive and postdictive simulations: Either predictive simulations that are disrupted by the discontinuity of movement, or postdictive simulations of the moment of release driven by the apparent continuity of movement and implication of causality.

However, the nature of this "continuity" is worth further examination. In all of the cases that generated the filling-in effect, there is still an abrupt change in viewpoint from the first half of the event to the second half. Our results provide an intriguing contrast to work examining the effect of changes in viewpoint on multiple object tracking (MOT), which has typically found that large changes in viewing angle disrupt tracking (Huff, Jahn, & Schwan, 2009), even more so if the objects' features change during the cut (Papenmeier, Meyerhoff, Jahn, & Huff, 2014). The latter finding in particular is an interesting contrast to our Experiment 3, in which a similar transformation did not detectably disrupt the filling-in effect.

These MOT failures do not necessarily indicate that participants did not have the impression of continuity of motion before and after the cut. Rather, the fact that changing the surface features of the objects impairs MOT performance suggests that continuity may be preserved, but mis-assigned to incorrect objects. In our experiment, with only one projectile object before and after the cut, no such confusion is possible, even when there is substantial feature change. The robustness of the tracking of individual objects has, to our knowledge,

never been demonstrated in this way before. It aligns most closely with the “tunnel effect” wherein cues to spatiotemporal continuity lead to the impression that a briefly obscured object has undergone a dramatic transformation (Spelke, Kestenbaum, Simons, & Wein 1995). Notably, the tunnel effect is not simply a matter of failing to recognize that a category violation has occurred, something that one might reasonably consider given the similarity between our stimuli in Experiment 3 and classic “change blindness” paradigms (e.g., Levin & Simons, 1997). Rather, adults typically recognize that the transformation has occurred, but the visual system nonetheless tracks it as one object for the purposes of attentional allocation (Flombaum & Scholl, 2006). The tunnel effect has been found to be influenced by causal perception in adults (Bae & Flombaum, 2011), but only in contexts of brief occlusion. Here, we may have demonstrated a tunnel effect in the absence of a tunnel—that having only one moving object before and after a drastic shift in perspective prompts the visual system to treat it as the same object regardless of featural similarity. The role of perspective shifts on the tracking of individual, rather than multiple, objects is deserving of further investigation.

This view also highlights an aspect of these results that might, at first glance, seem to minimize the role of causality in the first half of these videos: One can imagine a video containing the trajectory of an object already in motion that cuts to the same trajectory from a different angle, and it is possible that viewers would “fill in” parts of this trajectory that were not actually shown in a similar way. In fact we would find this unsurprising; it would be a sort of “representational momentum” effect (Freyd & Finke, 1984). We find it more remarkable that, in the *absence* of any such trajectory, people fill in the onset of an object’s motion.

Conclusion

Our environment is complex and chaotic. In order to form coherent event representations, the human mind employs a suite of sophisticated and automatic systems that connect disparate information and readily “fill in” anything our perceptual apparatus may have failed to capture. Using these filling-in effects, we have shown that these systems rely on spatiotemporal continuity and implied causality, but are not reliant on specific event schemas in memory or information about object identity. The minimal nature of the information required and the apparent automaticity of this filling-in effect implicate a system that sits at the interface of cognition and perception, and provides exciting opportunities for future investigations of both.

Acknowledgements J.F.K. was supported by NIH grant F32HD089595. L.J.B. was supported during a portion of this work by NSF GRFP #2013139545. FK was supported by NSF grant #1561143. B.S. was

supported by funding from the European Research Council (ERC) via two grants awarded to Philippe Schlenker: under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC (grant agreement no 324115, FRONTSEM); and under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 788077, Orisem). He was also supported by ANR-10-IDEX-0001-02 PSL, ANR-10-LABX-0087 IEC.

Open practices statement All data and materials can be found at <https://osf.io/mjwkd/>. Experiments 1 and 2 were preregistered at <https://osf.io/jt87a> and <https://osf.io/rgtp8>, respectively. Experiment 3 was preregistered at <https://osf.io/g9nyu>.

References

- Bae, G. Y., & Flombaum, J. I. (2011). Amodal causal capture in the tunnel effect. *Perception, 40*(1), 74–90. doi:<https://doi.org/10.1068/p6836>
- Baillargeon, R., & Wang, S.-H. (2002). Event categorization in infancy. *Trends in Cognitive Science, 6*(2), 85–93. doi:[https://doi.org/10.1016/s1364-6613\(00\)01836-2](https://doi.org/10.1016/s1364-6613(00)01836-2)
- Baker, L. J. & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition, 136*, 14–29. doi:<https://doi.org/10.1016/j.cognition.2014.11.030>
- Brockhoff, A., Huff, M., Maurer, A., & Papenmeier, F. (2016). Seeing the unseen? Illusory causal filling in FIFA referees, players, and novices. *Cognitive Research: Principles and Implications, 1*(7), 1–12. doi:<https://doi.org/10.1186/s41235-016-0008-5>
- Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception, 35*(3), 385–399. doi:<https://doi.org/10.1068/p5462>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Hillsdale, NJ: Erlbaum.
- Cohn, N., Paczynski, M., Kutas, M. (2017) Not so secret agents: Event-related potentials to semantic roles in visual event comprehension. *Brain and Cognition, 119*:1–9
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564. doi:<https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- Flombaum, J. I., Kundey, S. M., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychological Science, 15*(12), 795–800. doi:<https://doi.org/10.1111/j.0956-7976.2004.00758.x>
- 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 and Performance, 32*(4), 840–853. doi:<https://doi.org/10.1037/0096-1523.32.4.840>
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 126–132.
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*(5), 847–857. doi:<https://doi.org/10.3758/s13423-012-0296-9>
- Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological Science, 28*(12), 1731–1744. doi:<https://doi.org/10.1177/0956797617713053>

- Gopnik, A., & Sobel, D. M. (2000). Detectingblickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*(5), 1205–1222. doi:<https://doi.org/10.1111/1467-8624.00224>
- Hubbard, T. L. (2013). Phenomenal causality I: Varieties and variables. *Axiomathes, 23*(1), 1–42. doi:<https://doi.org/10.1007/s10516-012-9198-8>
- Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition, 17*(3), 297–306. doi:<https://doi.org/10.1080/13506280802061838>
- Jensen, M. S., Yao, R., Street, W. N., & Simons, D. J. (2011). Change blindness and inattentional blindness. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 529–546. doi:<https://doi.org/10.1002/wcs.130>
- Kominsky, J. F., & Scholl, B. J. (2020). Retinotopic adaptation reveals distinct categories of causal perception. *Cognition, 203*, 104339. doi:<https://doi.org/10.1016/j.cognition.2020.104339>
- Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science, 28*(11), 1649–1662. doi:<https://doi.org/10.1177/0956797617719930>
- Kosie, J. E., Baldwin, D. (2019) Attentional profiles linked to event segmentation are robust to missing information. *Cognitive Research: Principles and Implications, 4*(1)
- Lakusta, L., & Landau, B. (2005). Starting at the end: the importance of goals in spatial language. *Cognition, 96*(1), 1–33. doi:<https://doi.org/10.1016/j.cognition.2004.03.009>
- Lasher, M. D. (1981) The cognitive representation of an event involving human motion. *Cognitive Psychology, 13*(3):391–406
- Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4*, 501–506. doi:<https://doi.org/10.3758/BF03214339>
- Liao, Y., Flecken, M., Dijkstra, K., & Zwaan, R. A. (2020). Going places in Dutch and mandarin Chinese: Conceptualising the path of motion cross-linguistically. *Language, Cognition and Neuroscience, 35*(4), 498–520. doi:<https://doi.org/10.1080/23273798.2019.1676455>
- Michotte, A. (1963). *The perception of causality*. New York, NY: Basic Books. (Original work published 1946)
- New, J. J., & Scholl, B. J. (2009). The functional nature of motion-induced blindness: Further explorations of the perceptual scotoma hypothesis. *Journal of Vision, 9*(8), 253–253. doi:<https://doi.org/10.1167/9.8.253>
- Newman, G. E., Choi, H., Wynn, K., & Scholl, B. J. (2008). The origins of causal perception: Evidence from postdictive processing in infancy. *Cognitive Psychology, 57*(3), 262–291. doi:<https://doi.org/10.1016/j.cogpsych.2008.02.003>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411–419. doi:<https://doi.org/10.10630a/jdm10630a>
- Papenmeier, F., Brockhoff, A., & Huff, M. (2019). Filling the gap despite full attention: The role of fast backward inferences for event completion. *Cognitive Research: Principles and Implications, 4*(1), 3. doi:<https://doi.org/10.1186/s41235-018-0151-2>
- Papenmeier, F., Meyerhoff, H. S., Jahn, G., & Huff, M. (2014). Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 159–171. doi:<https://doi.org/10.1037/a0033117>
- Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology, 23*(3), 250–254. doi:<https://doi.org/10.1016/j.cub.2012.12.017>
- Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica, 123*(1/2), 144–165. doi:<https://doi.org/10.1016/j.actpsy.2006.05.005>
- Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science, 13*(6), 493–498. doi:<https://doi.org/10.1111/1467-9280.00487>
- Singmann, H., Bolker, B., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., ... Christensen, B. (2019). afex: Analysis of Factorial Experiments (R Package Version 0.23-0) [Computer software].
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*(4), 605–632. doi:<https://doi.org/10.1037/0033-295X.99.4.605>
- Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13*(2), 113–142. doi:<https://doi.org/10.1111/j.2044-835X.1995.tb00669.x>
- Strickland, B., & Keil, F. (2011). Event completion: event based inferences distort memory in a matter of seconds. *Cognition, 121*(3), 409–415. doi:<https://doi.org/10.1016/j.cognition.2011.04.007>
- Strickland, B., & Scholl, B. J. (2015). Visual Perception Involves Event-Type Representations: The Case of Containment Versus Occlusion. *Journal of Experimental Psychology: General, 144*(3), 570–580. doi:<https://doi.org/10.1037/a0037750>
- von Helmholtz, H. (1867). *Handbuch der physiologischen Optik 3* [Manual of physiological optics, Vol. 3]. Leipzig, Germany: Voss.
- Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30*(2), 111–153. doi:<https://doi.org/10.1006/cogp.1996.0005>
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293. doi:<https://doi.org/10.1037/0033-2909.133.2.273>
- Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*(2), 80–84. doi:<https://doi.org/10.1111/j.1467-8721.2007.00480.x>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.