Causation Without Realism

Christos Bechlivanidis, Anne Schlottmann & David A. Lagnado

University College London

Corresponding Author: Christos Bechlivanidis, 26 Bedford Way, London, WC1H 0AP, Email: ucjtcbe@ucl.ac.uk, tel: +447810355946

Abstract

Current theories of causality from visual input predict causal impressions only in the presence of realistic interactions, sequences of events that have been frequently encountered in the past of the individual or of the species. This strong requirement limits the capacity for one-shot induction and thus does not sit well with our abilities for rapid creative causal learning, as illustrated, for example, by the effortless way we adapt to new technology. We present four experiments (N=720) that reveal strong causal impressions upon first encounter with collision-like sequences that the literature typically labels “non-causal”. Our stimuli include both the commonly employed computer-based animations and edited video sequences. Besides direct reports, we present evidence based on goal-oriented behaviour that makes sense only in the presence of strong causal assumptions. Finally, we document impressions of causality in highly unrealistic sequences involving, for example, instantaneous shape or size change. In the case of the more realistic clips used in the past, causal ratings abruptly decline and approach the findings of previous work, only after a canonical collision (launch event) is presented. We argue that previously used experimental procedures conceal order effects due to participants adapting to the task and re-interpreting its demands. We discuss ways to account for this adaptation whereby people either focus on experiences of perceptual causation or take realism into account even when asked for impressions of causality.

Introduction

Imagine a rock rolling down a hill, heading for a house. Right at the time when the rock contacts the house, the house transforms into a red double decker bus. You would clearly be dumbfounded, but would your surprise be due to the co-occurrence of two incredible but independent events or to some type of supernatural causal link? Would you think that it was the rock that caused or triggered the bus transformation? If a friend was bored of his old immobile house and yearned for a British lifestyle, would you suggest searching for a big rock? Or does the surprising, unrealistic nature of a sequence imply the absence of causality? In the absence of prior knowledge, can two events appear causally related or are there rigid constraints on what can be causal, making causality sparse, at least according to our view of the world?

Although, to our knowledge, there have been no studies investigating impressions of causality in the case of rolling rocks and magic buses, the issue has been discussed in its more abstract form in philosophy of science. For example, to account for surprising quantum phenomena such as those implied by the EPR experiment[1], physicists and philosophers had either to reject the presence of causality or to rethink its definition (Bell, 1964). If the metaphysical view of causality demands spatiotemporal contiguity, then one has either to view the distant correlations in quantum experiments as non-causal or relax the requirements for a causal interaction (Chang & Cartwright, 1993).

From a psychological perspective, the question is whether sequences of events must appear realistic, i.e. congruent with past experiences, in order for people to have impressions of causality. Despite different starting points, all psychological theories answer positively. Thus, according to a classic definition, causal impressions are driven by habit, by repeated experiences that lead to the development of strong expectations of the effect in the presence of the cause (Hume, 1748). Almost tautologically, people are not habituated to uncommon sequences and, therefore, these do not result in causal impressions. Schema-matching approaches (Dittrich & Lea, 1994; Tenenbaum & Griffiths, 2003; Weir, 1978; White, 2006), consistent with the Humean tradition, postulate the presence of stored representations against which stimuli are compared in order to reach causal conclusions. Unrealistic sequences are by

---

[1] The Einstein–Podolsky–Rosen thought experiment implied that measuring a particle affects the state of another "entangled" particle by transmitting information faster than the speed of light. For Einstein the implication of "spooky action at a distance" meant that quantum theory is incomplete or incorrect.

definition not encountered in the environment and, thus, do not become part of stored schemata. White observes that “stimuli that are unrepresentative of real interactions between objects in ways other than incompleteness will not give rise to visual impressions of causality because they would not be matched against any schema” (White, 2006, p. 179). Similarly, if people embody a semi-Newtonian simulation engine that assesses the presence of causality (Sanborn, Mansinghka, & Griffiths, 2013), then unrealistic sequences that deviate from Newtonian principles should once more fail to invoke causal responses.

In contrast to the Humean tradition, Michotte (1963) followed Kant (1781) in arguing that prior experience is not always required when establishing causal relationships: In so-called “launching” or “entraining” events, humans perceive causality in a purely bottom-up, stimulus driven way, in the same way as they perceive colour or depth. For example, in the “launching” sequence, an object A approaches an object B and comes to a halt at the time when object B starts moving[2] in the same direction and at approximately the same speed as A (see Figure 1A). Michotte postulated the presence of an innate causal detector which can explain how individuals perceive causation even when they lack prior knowledge, e.g. in infancy (Leslie & Keeble, 1987; Mascalzoni, Regolin, Vallortigara, & Simion, 2013; Schlottmann, Ray, & Surian, 2012) or when they know that physical causation is impossible (Michotte, 1963 exp. 28). While Michotte’s work focused on collision-like interactions, others have since extended his approach to events such as triggering (Boyle, 1960; Kominsky et al., 2017), bursting/disintegration (White & Milne, 1999), and also action-and-reaction events seen by adults as social causality (Kanizsa & Vicario, 1968; Schlottmann & Surian, 1999; Schlottmann et al., 2002, 2006, 2013).

Nevertheless, even on Michotte’s account, realism remains a requirement. Even if no prior experience is needed at the level of individuals, there is still the requirement of exposure at the level of the species. For the theory to make sense, the detector must respond to configurations that correspond to stable invariances in our evolutionary environment. Thus, the properties of launch sequences approximate those of mechanical collisions. Conversely, since a sequence with a detectable delay between A stopping and B starting to move or a

[2] We follow Michotte (1963) in using the term “object A” to refer to the launcher, i.e. the object that moves first, and the term “object B” to refer to what is often called the launchee or the target.

spatial gap between the two objects, has never been encountered by our ancestors, it will not activate the causal detector and will not lead to causal impressions.

To clarify, a Michottean sequence is realistic if it matches real-world collisions in terms of its core kinematic properties, e.g. spatial and temporal distance between the two objects at the time of interaction, direction of movement before and after the interaction etc. Other, peripheral aspects of the sequence may dramatically diverge from reality. Thus, according to Michotte (1963), human perception has evolved a sensitivity to realistic core features of collisions and a susceptibility to visual illusions of causality when presented, for example, with unrealistic, two-dimensional displays (see Figure 1A). Such visual illusions have been the focus of empirical research ever since Michotte's original studies, and they highlight that perceptual causation, despite requiring realistic core features, it is not strictly derived from individual experience (Kant, 1781).

This apparent agreement of theories is rather astonishing since the requirement for realism severely constrains causal learning from perceptual data. For Humean accounts, the only path to discovering genuinely novel causal relations goes through contingency learning, repeatedly observing the co-occurrence of two events in order to form an association (Hume, 1748) or to develop a causal schema (Weir, 1978; White, 2006). Although more Kantian, perceptual accounts a la Michotte allow for one-shot causal learning, they apply to a very limited set of stimuli with very particular properties. Such capacity might be useful for bootstrapping causal learning in infants but is not of much help to adults who are already familiar with mechanical collisions.

The demand for causal learning from a single or a small set of examples, however, is not withdrawn after infancy. Consider, for example, the Gestalt psychologist Karl Duncker observing that even if the light turned on exactly after the door closed, "one knew ever so well that no causal connection exists" (Duncker, 1945, p. 66). While Duncker's observation would be uncontested at his time, if his granddaughter were to see the lights turning on as she entered her friend's house, she wouldn't be so adamant. Today's automation technologies establish causal links between doors and lights, sunlight and window blinds, humidity and sprinklers, some of which will become unquestioned causal pairs in the future. Similarly, our ability to fluidly interact with digital computers and touchscreens, to adapt, with minimal instruction, to a continuously shifting technological environment testifies for our capacity for causal discovery from very limited exposure.

However, when the requirement for realism is imposed, one-shot learning, which has been hailed as one of the areas of human excellence (Lake, Salakhutdinov, Gross, & Tenenbaum, 2011) is nearly ruled out in the causal domain. What is the evidence in support of this requirement and the resulting sparsity of causation from perceptual data? The starting point for this enquiry is Michotte's work, nearly 100 experiments that documented the precise conditions necessary for the generation of what he saw as a causal percept (Michotte, 1963). Indirectly, but crucially for the purposes of the current work, Michotte also described in detail the conditions that destroy the launching effect, i.e. where causality is absent.

Since then, there has been an intense debate in philosophy and a wide array of experiments in psychology both for and against Michotte's claims. In every discussion, the focal point has invariably been whether, in dynamical displays, the route to causal impressions involves high-level inference, i.e. whether or not it depends on experience and background knowledge. Here, we follow Rips (2011) in using "causal impression" as a neutral term referring to phenomenally immediate experiences of causation, irrespective of whether these are results of perception or inference, as this issue remains open.

In support of the purely perceptual route, many of Michotte's experiments have been replicated, with participants reporting causal impressions in canonical sequences[3] but not in sequences that even slightly violate Michotte's ideal conditions. There has also been evidence for sensitivity to causal structure in infancy (Leslie & Keeble, 1987; Oakes, 1994) for the involvement of low-level visual processes (Rolfs, Dambacher, & Cavanagh, 2013), and for dissociations between causal impressions and clearly inferential forms of causality (Schlottmann & Shanks, 1992; Schlottmann, 1999). Michotte's adversaries demonstrated large individual differences (Beasley, 1968; Schlottman & Anderson, 1993), in contrast to claims of universality and irresistibility of the effect (Scholl & Tremoulet, 2000), and presented proof-of-concept algorithms capable of inferring causality from lower-level perceptual data through schema-matching and similar devices (Sanborn et al., 2013; Weir, 1978; White, 2006).

Nevertheless, and while the debate is still open (Rips, 2011), all parties agree on an indirect finding from Michotte's experiments: irrespective of how we detect causation in canonical

[3] By the term "canonical" or "ideal" we refer to Michottean sequences without spatial or temporal gap between object A stopping and object B starting to move.

collisions, when the ideal conditions are violated, we neither perceive nor infer causality; non-canonical collisions are void of causation. In Michotte's own words, "the presence of the interval [200ms or more] makes the causal impression disappear completely…two events that are obviously separate, which arise successively, and which on their own give no impression of causality…Not only is there no causal impression, but there is no tendency towards a causal 'interpretation' in these cases" (Michotte, 1963, p. 22).

As a result, it has become commonplace in the literature to freely and routinely use the term "non-causal" when referring to Michottean sequences with a 20mm spatial gap between the resting position of object A and the starting position of object B (see Figure 1B) or a 250ms temporal delay between the motions of the two objects (see Figure 1C) (Cohen & Amsel, 1998; Leslie, 1986; Leslie & Keeble, 1987; Moors, Wagemans, & De-Wit, 2017; Newman, Choi, Wynn, & Scholl, 2008; Oakes & Cohen, 1990; Roser, Fugelsang, Dunbar, Corballis, & Gazzaniga, 2005; Schlottmann & Surian, 1999; Scholl & Nakayama, 2004; Scholl & Tremoulet, 2000; Wagemans, van Lier, & Scholl, 2006; Weir, 1978). This consensus in psychology has led philosophers to argue for causality as a categorical concept (Butterfill, 2009) and neuroscientists (Blakemore et al. 2001; Fugelsang, Roser, Corballis, Gazzaniga, & Dunbar, 2005; Straube & Chatterjee, 2010) to look for brain patterns that correlate with the experience of causal but not with non-causal events. If a quarter of a second delay suffices to rob a sequence of its causal status, less common sequences are undoubtedly ruled out. Theories of causation were developed to account for this assumption, thus excluding causal impressions in unrealistic sequences unless supported by contingency data.

A closer look, however, reveals that there is hardly any direct evidence bearing on this consensus, as there are no studies looking at how observers see individual unrealistic sequences. Rather, indirect evidence comes from experiments on Michottean events that all use very similar procedures: participants watch a large number of clips, with canonical collisions interspersed among deviant sequences. All displays are followed by identical causal questions. Rating scales that range from causal to non-causal (or independent) are most commonly employed. When, during the analysis, these ratings are collapsed across different presentation orders, what is commonly reported is that canonical sequences are rated high, while deviant sequences receive very low causal ratings (e.g. Beasley, 1968; Fugelsang et al., 2005; Sanborn et al., 2013; Schlottmann, Ray, Demetriou, & Mitchell, 2006; Scholl & Nakayama, 2002; Straube & Chatterjee, 2010)

The clarity of this picture, however, is disturbed by a set of studies showing that people's causal attributions are not always stable and are affected by prior exposure (Brown & Miles, 1969; Gruber, Fink, & Damm, 1957; Powesland, 1959; Woods, Lehet, & Chatterjee, 2012). Brown and Miles (1969), for example, split participants into three groups exposed to 12 Michottean sequences differing in the range of delays between object A stopping and object B starting to move. The range of delays was 60-210ms for the "short" group, 150-300ms for the "medium" group and 240-390ms for the "long" group. Subsequently, all groups watched sequences with delays covering the full range (60-390ms) and reported their causal impression for each sequence. The causal reports varied as a function of prior exposure: Participants in the "short" condition were less likely to report causality in sequences with long delays than participants in the "long" group. More recently, Woods et al (2012) showed that the more canonical collisions participants see, the less likely they are to attribute causality to deviant sequences.

Such results show that the order in which participants watch the various sequences affects the reported causal ratings. In particular, unrealistic, deviant events seem to become less causal after exposure to canonical events. This might be due to low-level adaptation where, for example, delays simply appear longer when preceded by shorter delays. More worryingly, however, the observed flexibility may be due to semantic adaptation, with participants changing what they rate as 'causal' during the course of the experiment. If causality were experienced in all sequences, then participants, in an effort to provide meaningful responses that somehow differentiate between the observed sequences, might, after a few examples, either narrow their definition of causality (from causality in general to mechanical causation, for example) or completely switch to ratings along another dimension. Thus, while experimenters might think they are capturing the conditions that give rise to impressions of causality, participants could be responding to different or additional properties of the stimuli. Then, theories that try to account for that experimental evidence could unwarrantedly turn these other properties into determinants of causal impressions from perceptual input.

Our primary aim in this paper, is to investigate whether causal impressions are bounded by realism. We will, for the first time, provide direct evidence on whether deviant, unrealistic sequences fail to generate causal impressions, as assumed in the literature, or conversely whether the low causal ratings observed in the past reflect a property of the universally employed experimental procedure (Schlottmann, Allen, Linderoth, & Hesketh, 2002). Our second goal, is to investigate the effects of that procedure by studying how watching ideal collisions affects the way the deviant sequences are being rated and, consequently, to what extent participants adapt and re-interpret the task.

We will first use a standard rating measure to evaluate causal impressions of commonly tested deviant sequences both before and after participants are exposed to ideal collisions (Experiment 1). Then, we will switch to a behavioural measure to assess potential differences in the causal predictions generated by canonical and deviant sequences (Experiment 2). Subsequently we will ask for causal impressions in highly unrealistic sequences that have rarely been studied in the past (Experiment 3). Finally, we will use edited video sequences in place of computer animations, to investigate how observers evaluate violations of core kinematic properties in the presence of photorealistic peripheral features (Experiment 4).

Experiment 1

In this experiment[4], we re-evaluate the established claim that small deviations from ideal collisions are seen as non-causal. We also investigate how impressions of such deviant sequences are affected by the exposure to and comparison with a canonical Michottean sequence. For that reason, we expose participants to animations of deviant sequences before and after viewing a canonical Michottean collision (panel A in Figure 1). In particular, participants are asked to watch and causally evaluate three animations. The first is a deviant sequence, an exemplary case of what the literature calls a "non-causal" interaction. As shown in Figure 1, it features either a 250ms temporal delay (panel B), a 30mm spatial gap (panel C) between red square A stopping and blue square B starting to move, or what is known as a "non-causal pass" in which A completely overlaps object B before stopping (panel D).

[4] All experiments are available at https://goo.gl/ef43f2 and the data can be viewed at https://osf.io/2wjrh/. All experiments were approved by the UCL research ethics committee: CPB/2009/013

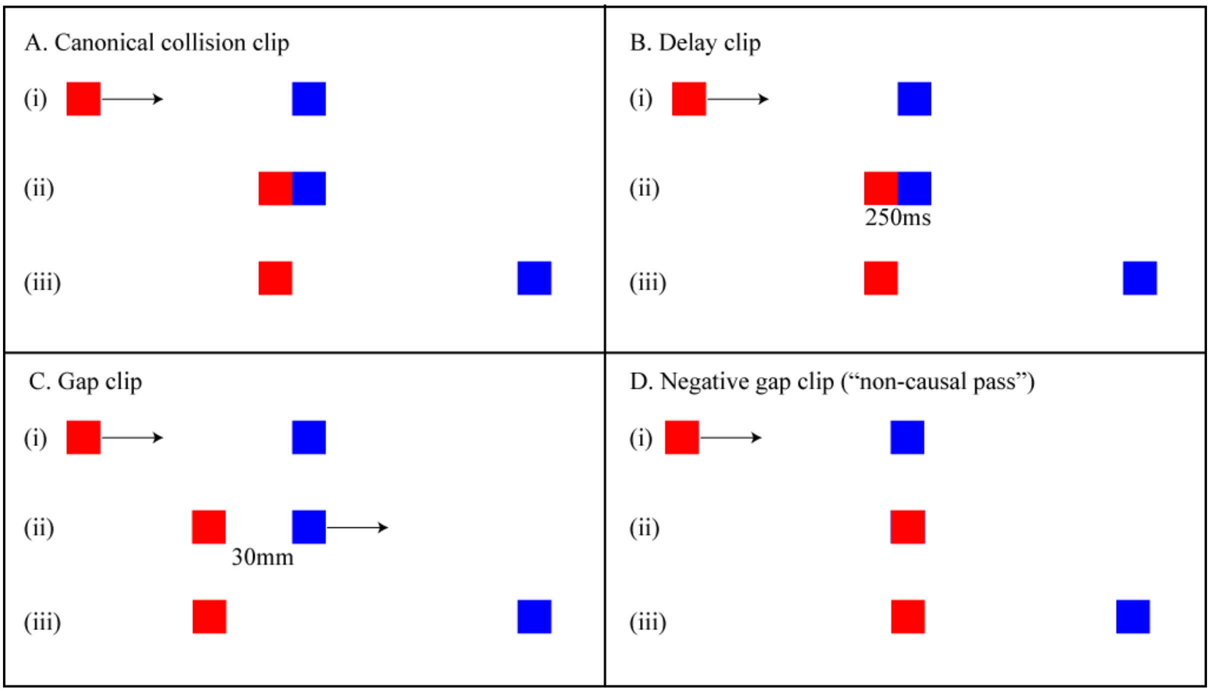

Figure 1: The canonical Michottean collision used in all experiments and the three deviant clips used in Experiments 1 and 2. Each panel shows (i) the initial configuration, (ii) the positions at the time of interaction and (iii) the final positions. The difference between the canonical collision (A) and the delay clip (B) is a 250ms delay between the time when the red square stops and the blue square starts moving. In the gap clip (C) the red square stops 30mm to the left of the blue square and the blue square starts moving immediately. The opposite is the case for the negative gap or "non-causal pass" clip (D) where the red square stops 10mm later, completely overlapping the blue square (the arrows represent the direction of motion and were not visible during the experiments).

The first two deviant sequences were chosen because they are the most frequently investigated deviations from ideal collisions. Moreover, the particular parameters (250ms delay and 30mm gap) are among the most extreme values used in the literature[5] – for more details see Appendix. Even when less extreme values were used in the context of the usual repeated measures procedure, the majority of participants responded with very low causal ratings (Fugelsang et al., 2005; Guski & Troje, 2003; Michotte, 1963; Sanborn et al., 2013; Schlottmann & Anderson, 1993; Straube & Chatterjee, 2010; White, 2014; Yela, 1952). The "non-causal pass" clip is the opposite of the gap sequence in the sense that object A stops spatially beyond the position dictated by Newtonian principles. As its name implies, it is thought to be a paradigmatic case of a non-causal sequence since it features the violation of

[5] Yela (1952) experimented with gaps up to 90mm, while Schlottmann et al (2006), discussed later, presented participants with a clip featuring a 1300ms delay.

object boundaries [6] (Bae & Flombaum, 2011; Choi & Scholl, 2006; Hubbard, 2013; Rolfs, Dambacher, & Cavanagh, 2013; Scholl & Nakayama, 2002, 2004).

If a causal relationship is present in these sequences and given the different types of violations featured (temporal, spatial, boundary violation), a wide variety of mechanisms may potentially mediate the causal link. Therefore, our dependent measure attempts to capture impressions of causality while leaving the underlying mechanism explicitly open ("Do you have the impression that red somehow made blue move?").

Under the view that one-shot induction of causal relations should be possible, it would seem ideal to focus on a single encounter with the deviant event before asking for a causal impression. However, each sequence is only a few seconds long and it is very easy to miss some of it (crucially, to miss the point of interaction) while familiarising oneself with the environment. Moreover, even if participants notice the core unrealistic aspects of the sequence (e.g. the delay or the gap), they might discount them as the result of computer error, or misperception. Michotte (1963) himself stated that observers were often initially somewhat confused, finding it difficult to organise their experience. From this point of view, it might be important to ensure adequate exposure and rule out such factors, by repeating the sequence a few times. This, however, opens the door for another issue, in that repeated encounters might allow a very rudimentary form of statistical learning. Under a standard Humean account, observing a succession of events repeatedly might support causal induction. Although these possibilities are routinely contemplated in the literature, we are not aware of any data. Accordingly, to evaluate the initial development of a (non-)causal impression empirically, we decided to vary exposure, i.e., the number of initial repetitions of the deviant sequence, displaying it 1, 3 or 5 times before asking for an initial rating of causality.

Subsequently, to assess the role of a canonical collision, participants watch a standard version of such sequence (also 1, 3 or 5 times) before returning to watch and evaluate the deviant sequence one more time. This experiment is, thus, a condensed version of the usual repeated measures design, using extreme deviations from Newtonian expectations but also controlling for the order of clip presentation. If such deviations destroy causal impressions we will observe low ratings for the deviant sequences throughout the experiment. Otherwise, if past

[6] This sequence has been used with objects either of the same colour (Rolfs et al., 2013) or of different colours (Scholl & Nakayama, 2002). This may be an important difference but here, we will follow the original choice by Scholl and Nakayama (2002) to have objects of different colours.

results are at least partially due to this repeated measures procedure, we should expect high causal ratings before the canonical sequence, which drop significantly after its presentation.

Despite the plethora of studies in the Michottean tradition, there is only a single experiment, to our knowledge, that reports participants' impressions before the exposure to the canonical Michottean collision. Although the overall design of Schlottmann et al. (2006) involved the usual repeated viewing and evaluation of sequences, the researchers did report participants' judgments following the first exposure. Despite large deviations from Newtonian principles (delay=1300ms, gap 30mm), participants still reported moderate impressions of causality in one case (delay) and were indecisive in the other case (gap). Consistent with our current predictions, causal impressions were greatly reduced only after exposure to the other sequences that included canonical collisions. However, this was a within subjects experiment and the order of sequence presentation was counterbalanced between participants, leading to a very small sample size (N=6) for first-time viewings.

Materials and Design

The overall design was mixed factorial with type of clip ("gap", "delay" or "pass") and number of repetitions (1, 3 or 5) varied between participants and time of causal report (before or after viewing the canonical collision) varied within.

The clips were created using Adobe Flex 4.7[7]. They all featured a red and a blue square of size 10x10mm[8]. The red square was positioned 67.5mm to the left of the blue one (fig. 1 - panels i). The two objects faded in and 2 seconds later the red square started moving to the right towards the blue square at a speed of 100mm/sec. In the "delay" condition the red square stopped directly to the left of the blue square, in the "gap" condition it stopped 30mm earlier (fig. 1C – panel ii), while in the "pass" condition it stopped 10mm later, thus completely overlapping the blue square (fig. 1D – panel ii). In the "gap" and "pass" conditions the blue square started moving immediately after the red object stopped[9], while in the delay condition it started 250ms later. The blue square moved horizontally to the right at the same speed as the red (100mm/sec) and came to a halt 67.5mm later (fig. 1 - panels iii).

---

[7] It is customary for such experiments to be conducted in the lab and developed with programming languages that allow for tighter control especially over presentation times. With that in mind we added a special procedure at the beginning of the experiments to ensure consistent presentation. The procedure is described in detail in Appendix A of (Bechlivanidis & Lagnado, 2016).

[8] Apart from the colour of the objects, all other settings were copied from Sanborn et al (2013).

[9] It actually started moving in the next frame, which at 30fps was displayed approximately after 33ms.

The animation lasted for approximately 1100-1650ms depending on the clip. The two objects remained static for 1000ms and finally faded out. In the “canonical” clip which was the same in all conditions, the red square stopped directly next to the blue square and the blue square started moving immediately (fig. 1A). Each animation cycle (fade-in, animate, fade-out) lasted for 4100-4650ms and, depending on condition, the whole sequence was about 5 seconds in the shortest (1 repetition) to approximately 24 seconds in the longest (5 repetitions).

Participants

The experiment was conducted over the Internet using Amazon Mechanical Turk. There were 270 participants in total, 30 per condition. The mean age was 36.7 (SD=11.5). 141 subjects were male and 129 were female. Each participant was paid $0.40 for taking part.

Procedure

After successfully completing the calibration session (Bechlivanidis & Lagnado, 2016), providing consent and some basic demographics, participants were informed that they would see a short clip either a single time or being repeated for 3 or 5 times. On the next screen, they saw one of the three deviant clips for 1, 3 or 5 times depending on condition.

Then they were shown a static image of the initial configuration of the clip (fig. 1, panels i) and were asked for their causal impression. The exact wording was copied from Schlottmann et al (2006; see also White, 2012) with a couple of changes to reflect the different colours used and the method to respond (a software-based slider rather than a scale): “Do you have the impression that red somehow made blue move”? The causal question was further qualified with the following, also from Schlottmann et al (2006): “If you feel strongly that red made blue move, set the slider below at the left end of the scale. If you feel that red made blue move, but this impression is not very strong, set the slider towards the left, but not all the way. If you feel strongly that red did not make blue move, set the slider at the right end of the scale, etc. If you do not know or cannot decide, set the slider at the middle. Use all of the scale to mark the strength of your impression”. Participants marked their answer by dragging a slider on a scale that was marked from left to right with the following statements: “red made blue move”, “don’t know”, “red did not make blue move”.

After responding, participants were informed that they would see another clip and, irrespective of condition, they saw a canonical collision repeating 1, 3 or 5 times.

Subsequently, they were asked for a causal impression in the way just described. Finally, after another information screen they saw the same condition-dependent deviant clip for 1, 3 or 5 times and were asked for a causal impression.

Results

Figure 2 shows the mean causal rating for each deviant clip the first time participants watched it, as a function of number of repetitions. Clearly, none of these supposedly non-causal events are rated as non-causal on first encounter. Observers give causal ratings, beyond the midpoint (which corresponds to indecision), even if there is only a single encounter, though the ratings show a further small increase as a function of repetition.

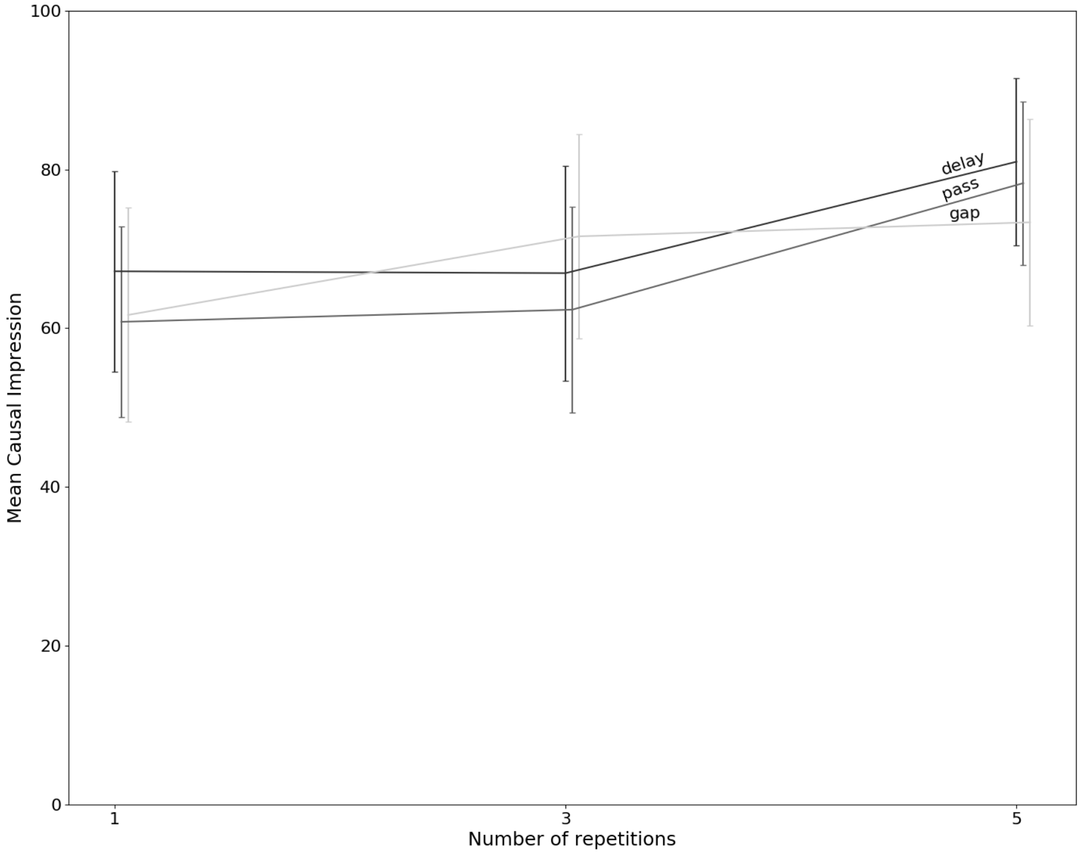

Figure 2: Mean reported causal rating for the first viewing only, by clip type and number of repetitions in Experiment 1 (error bars represent 95% CIs).

The analysis of variance shows only a significant main effect of repetitions ($F(2,261)=4.113$, $p=.017$) but not of type of event ($F(2,261)=.395$, $p=.674$) and no interaction effect ($F(4,261)=.402$, $p=.807$). Even when watching the deviant clip for a single time, the majority of participants (60/90 or 66%) placed the slider to the left of the midpoint towards the causal statements and causal ratings were significantly higher than the midpoint (which corresponds to indecision), $mean=63.21$, $t(89)=3.553$, $p<.001$, $d=.374$. The results were very similar when participants saw the deviant clips three times, with 65/90 (72%) placing the slider towards the causal statement and a significant difference to midpoint, $mean=66.94$, $t(89)=4.408$, $p<.001$, $d=.465$. When participants watched the clip 5 times, there was a further increase, with 74/90 (82%) electing to place the slider closer to the causal statement and a significant difference to midpoint: $mean=77.52$, $t(89)=8.280$, $p<.001$, $d=.873$. In sum, while the impression does appear to develop slightly with exposure, there was no evidence at all that these deviant events appear non-causal even after a single presentation.

Turning now to the influence of canonical collisions, Figure 3 shows that, while ratings for all three clip types were, as already discussed, positive after the first viewing, they significantly drop, towards non-causal, after the interpolation of the canonical launch, which received very high causal ratings. A repeated measures ANOVA shows a corresponding main effect of time of causal report (before/after the canonical collision), $F(1,261)=73.035$, $p<.001$, and a significant effect of interaction between time and clip type $F(2,261)=4.605$, $p=.011$, reflecting a stronger drop in the delay event. There was no main effect of clip type $F(2,261)=.609$, $p=.545$ or repetitions, $F(2,261)=2.390$, $p=.094$, or other interaction effect, $F<1.797$. The drop of ratings was significant for all clip types: Delay: $t(89)=7.361$, $p<.001$, $d=.796$, Gap: $t(89)=4.492$, $p<.001$, $d=.460$ and Pass: $t(89)=3.112$. $p=.002$, $d=.365$. Thus, low, non-causal ratings for deviant events appear only after exposure to a canonical event.

Finally, the canonical launching sequence received high ratings in all conditions ($mean=93.31$), significantly higher than the ratings for the deviant clips that preceded it: $t(269)=11.637$, $p<.001$, $d=.878$).

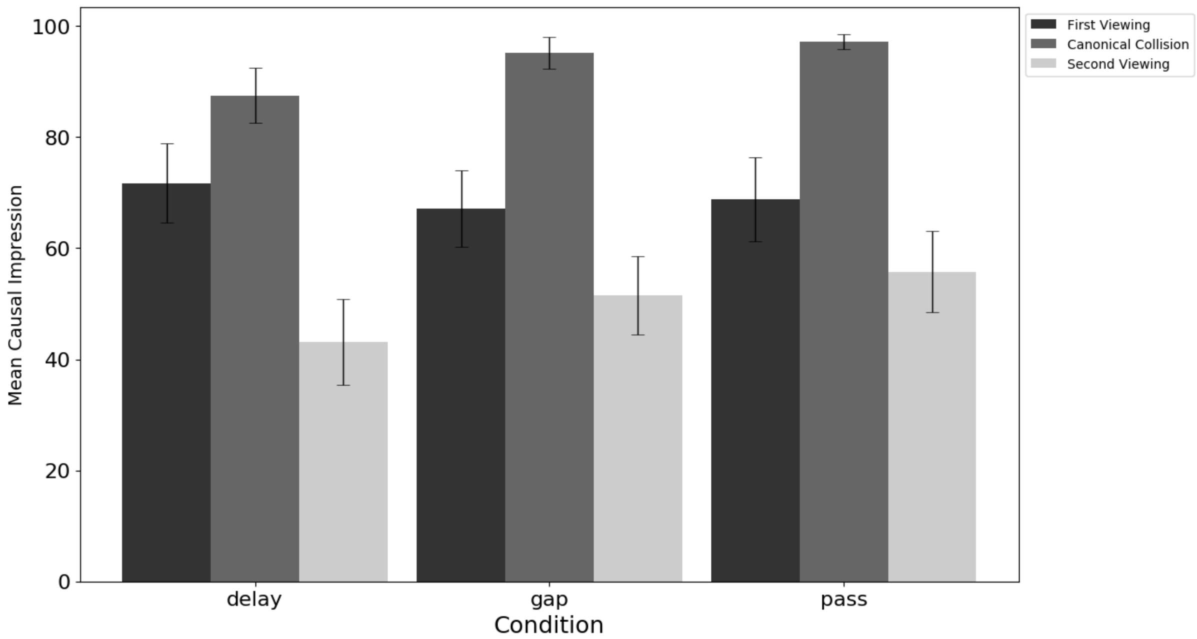

Figure 3: Mean reported causal ratings across repetitions, by clip type and viewing order in Experiment 1 (error bars represent 95% CIs)

Discussion

In Experiment 1, three sequences that are traditionally labelled "non-causal", received clearly causal ratings, with a very small proportion of participants judging that the motion of object B (launchee) was not causally related to the motion of object A (launcher). This is a first indication that impressions of causality from visual evidence are far more inclusive and far less sensitive to deviations from canonical launch events than commonly assumed. Realism or adherence to Newtonian principles does not appear to be a requirement for "seeing" a causal link.

Observers' causal ratings of the deviant events are not due to statistical learning. While the ratings increased slightly if repeated 3 or 5 times, causal impressions were present even after a single exposure. Although contingency information definitely has a role in causal learning (Cheng, 1997; Shanks & Dickinson 1987; White, 2003), it is not sufficient to explain our findings. Contingency learning amplifies, rather than generates the causal impression.

It is not entirely clear, in any event, that the observed increase in ratings with repetitions reflects Humean learning. The alternative is that some participants are confused the first time they watch the deviant sequence, and that this confusion clears up with repetition. As Michotte observes, after the first exposure participants are "all 'mixed up' and do not realise what is going on at all, and their impression is chaotic and unorganised" (Michotte, 1963, p. 20). On this view, the repetitions give participants time to organise their experience.

However, while we cannot strictly decide between these views, the finding that causal impressions were observed even after a single repetition does not seem to agree with an "all mixed up" view.

While our results from observers' ratings of the unrealistic clips stand in stark contrast to the non-causal ratings reported in prior research (see Appendix), our finding that the ratings drop significantly after viewing a canonical launch provides an explanation for the discrepancy between current and past findings: Our procedure here, which is a condensed version of the usually employed repeated measures design, reveals a strong influence of viewing and evaluating a canonical Michottean collision. Participants in this experiment and probably in previous ones, change the way they interpret the task requirements after watching the realistic clip. Under the observers' new interpretation, the same deviant clips receive significantly lower ratings, approaching the values reported in the past.

This leaves open the question of the nature of this change, a question critical for the interpretation of the 60-year old literature upon which all theories of phenomenal causation are based. We will consider this issue in the general discussion, after three further experiments that probe for causal impressions while varying the stimuli, the contextual cues and the method of eliciting responses, to help constrain our interpretation.

The most pressing issue perhaps, addressed in the next experiment, is whether the high initial causal ratings for unrealistic events in Experiment 1 reflect the observers' genuine causal impressions or whether they are simply a response to task demands. By asking directly whether the deviant events are causal, we are raising for participants a possibility that they might not have spontaneously considered. Thus, it could be objected that the high initial ratings follow from the verbal instruction and articulation of the causal question, not from the perceptual input. As noted elsewhere, "one of the most serious concerns is that verbal reports reflect not only what subjects are seeing but also their higher-level interpretations and judgments" (Choi & Scholl, 2006, p. 93). The aim of the next experiment is to validate our findings with a radically different response measure immune to this criticism. We investigate whether observers not only report causality, but whether they also act on it.

Experiment 2

This experiment differs from Experiment 1 only in the response measure. Rather than asking for ratings of causality, it uses a purely behavioural response measure. Specifically, after watching the same delay, gap or pass sequence as before, participants are shown the screen depicted in Figure 4, and are asked to imagine that the clip they watched would be played again. Their task, this time, is to use the black rectangle (“obstacle”) in order to prevent the blue square from moving.

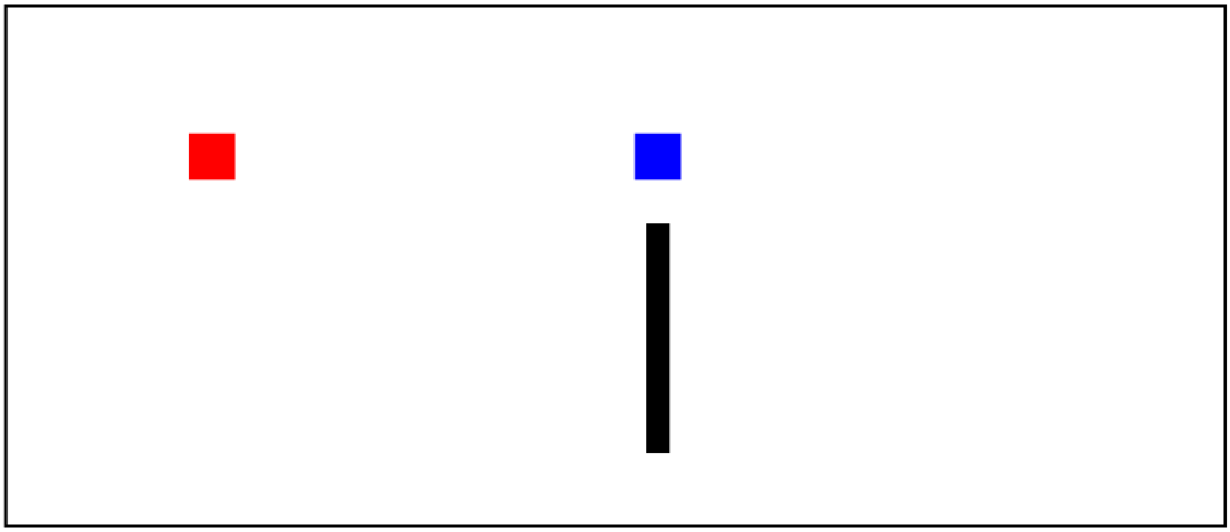

Figure 4: Screenshot of the dependent measure used in Experiment 2. Participants had to drag and drop the black rectangle ("obstacle") anywhere in the screen to prevent the blue square from moving (the direction of motion in the clip that participants had observed earlier was left-to-right)

Confronted with this task and given that the direction of movement in the original clip was left-to-right, there are two options: placing the obstacle either to the right of the blue square or to its left, between the two objects. The most reasonable choice is to place the obstacle to the right, to prevent its motion directly, regardless of how this is caused. But if the possibility of the blue square moving on its own does not even cross a participant’s mind, then obstructing the red square is also an effective way to prevent blue from moving. In that case, one would place the obstacle between the two squares in order to break the causal link.

Thus, a between placement of the obstacle is a stringent behavioural measure of causal impressions. In many ways, it is a superior measure compared to the rating scale because it captures causal impressions uncontaminated by additional assumptions and interpretations (Choi & Scholl, 2006). Moreover, this response should be independent of exactly how observers think the link between the two motions is mediated. Regardless of whether they think that the link is, for example, a form of mechanical causation or that B is programmed to move when A approaches it, as long as they think that A’s motion causes B’s motion, blocking it would be an effective strategy.

Our blocking measure also has weaknesses. It is a coarse, categorical response that does not allow observers to express the strength of their causal impression. Moreover, a negative result, i.e. placing the obstacle to the right of the blue square, is inconclusive and does not equate to having the impression of two independent motions: observers may place the obstacle on the right because they see two independent motions, or because they are uncertain of a causal relation, or because they take a generally cautious attitude despite their causal impression, or even because they make a random choice between two equally viable strategies. Thus, we expect some right placements even for canonical launch events, for which observers were close to ceiling in Experiment 1 and past studies. In sum, if one sees a causal link between the two motions, then both placements, to the right and to the left of the blue square, are equally effective. Thus, the number of left placements, on which we focus here, is a highly conservative measure of observers' causal impressions, while the number of right placements is inconclusive.

Nevertheless, the blocking measure is well suited for present purposes, to assess whether Experiment 1 demonstrated genuine causal impressions of unrealistic events or artefacts of our verbal method. If the former is true, we expect that after watching one of the deviant clips, the majority of participants will have no second thoughts about the causal link between the motions of two objects and a considerable proportion of them will spontaneously attempt to break it by placing the obstacle between the objects.

In Experiment 2, as in Experiment 1, observers evaluate the unrealistic events before and after seeing canonical launch events. In Experiment 1 we observed a drop in the causal ratings for unrealistic events after the canonical event, which, in our view, reflects a change in observers' interpretation of the causal question. In Experiment 2, however, there is no causal question that could change in interpretation. Rather, the causal impression is implicit in observers' behaviour, and there is no reason to expect a similar reduction in between placements.

Materials and Design

The stimuli and design were identical to Experiment 1, but a new response measure was used. In particular, after viewing one of the deviant clips (delay, gap or pass) for 1, 3 or 5 times, depending on condition, participants saw the screen depicted in Figure 4. Apart from the objects in their starting positions, there was a black rectangle measuring 5mm by 50mm, placed 20mm directly below object B. By clicking and dragging this rectangle participants could move it anywhere in the screen, except from a position overlapping one of the squares; if an overlapping position was chosen the rectangle was returned to its starting position and an error message appeared, before inviting participants to try again.

Participants

We recruited 269 participants in total, however 32 were removed from the analysis for failing to provide reasonable responses (see below). Of the remaining participants, 117 were male and 120 female. Their mean age was 37.62 (SD=12.67) and each was paid $0.40 for taking part.

Procedure

After the calibration, consent and demographics screens, participants were informed that they would watch a clip with moving objects that would be played 1, 3 or 5 times depending on condition, and were asked to pay attention. Also depending on condition, they then watched the delay, gap or pass clip (Figure 1 B, C or D). In the following screen, participants were shown the squares in their starting positions together with a black rectangle (Figure 4). They were asked to imagine that the clip would be played again and told to use their mouse to "drag and drop the black rectangle anywhere in the screen to make sure that the blue square will remain at its position when the clip is played again". In order to continue, participants could place the rectangle anywhere in the screen except from its starting position or a position overlapping one of the two squares.

In the next screen, participants were informed that they would watch another clip playing for 1,3 or 5 times and, irrespective of condition, they saw the canonical Michottean collision clip (Figure 1A). Again, they had to use the obstacle to stop the blue square from moving. Finally, they saw the deviant clip for a second time followed by the obstacle placement task.

Results

We did not include in the analysis participants who placed the obstacle at an unreasonable location, for any of their three responses: namely to the left of the red square A or above/below the two squares. Given the invariably left-to-right direction of movement, these positions could not reasonably be expected to obstruct the trajectory of either square. The majority of participants (237/269 or 88%) chose a reasonable location and that was not affected by the number of repetitions, another indication that Michotte's claim for an initially chaotic and unorganised experience (Michotte, 1963) does not apply.

In Figure 5, we can see that, on the initial trials, the majority of participants placed the obstacle between the objects, so as to prevent A's movement in both the delay (49/81 or 60% across repetitions) and the pass conditions (54/82 or 66%), while this proportion was lower in the gap condition (26/74 or 35%). It is important to remember here that, unlike the rating scales in Experiment 1, the midpoint (50%) in these graphs does not reflect random choosing or uncertainty about the causal relationship. As explained, a causal impression is compatible with both middle and right obstacle placements (if A caused B, one can obstruct either to prevent B from moving) but uncertainty about the causal link should lead to predominantly right placements.

The effect of repetitions is reversed compared to Experiment 1, with the probability of between placements decreasing as the number of repetitions increases. These observations are confirmed by a logistic regression which revealed significant effects of clip type ($p<.001$) and repetition ($p=.050$) on the probability of placing the obstacle between the objects after the first observation.

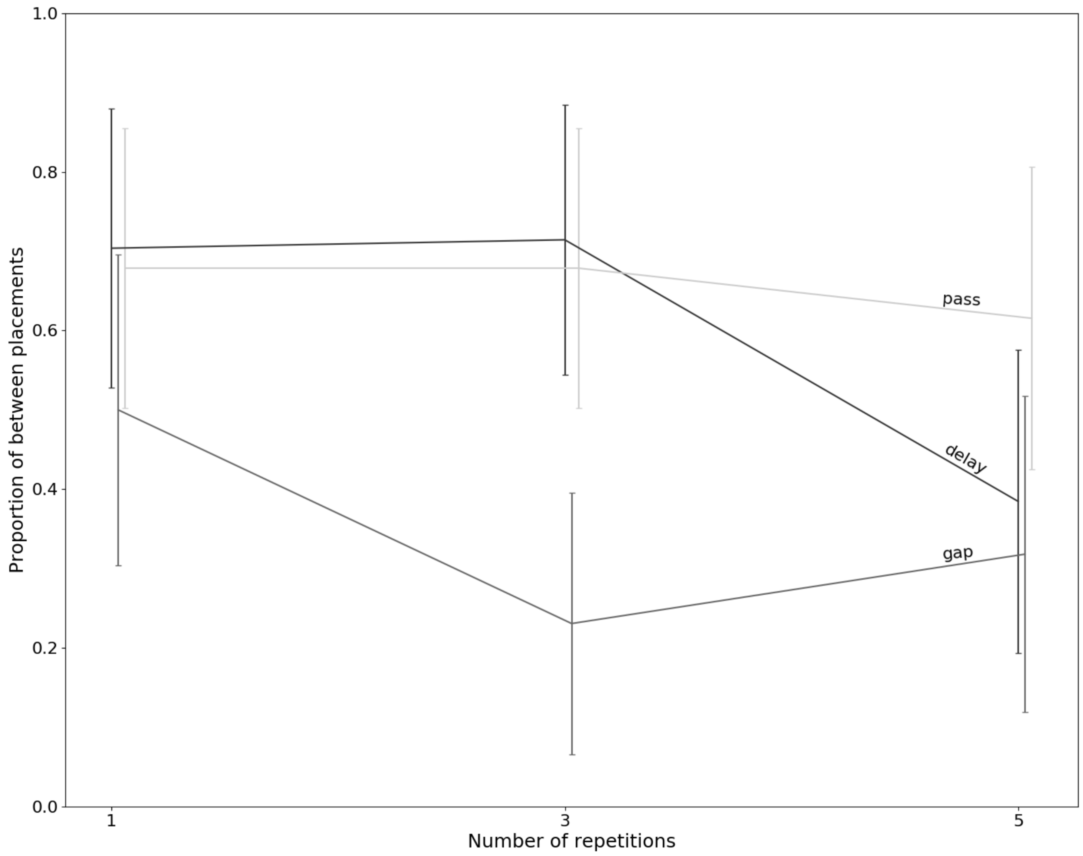

Figure 5: Proportion of participants placing the obstacle between objects A and B, after the first viewing of the deviant clip in Experiment 2, per clip type and number of repetitions (error bars represent 95% Cis).

Figure 6 shows the obstacle placements for first viewings of the unrealistic events, the interpolated canonical event, and for the second viewing of the unrealistic events. Clearly, the proportion of between placements for the canonical launch event and for the delay and pass events are almost identical (~60%). This proportion therefore, seems to be ceiling level with this method, even for Michotte's ideal events. Figure 6 also shows, as already noted, that the gap condition has a different pattern from the other sequences, with a majority of right placements after the first viewing (48/74 or 65%) and the last viewing (47/74 or 64%). While observers of gap events clearly more often acted ambiguously, in a way that allowed for alternative causes of B's movement, nevertheless, a substantial group still gave conclusive evidence of their causal impression.

Finally, and of equal importance, Figure 6 shows that in all three conditions, the number of left placements, indicative of a causal impression, remained the same after the interpolated canonical collision. This was in contrast with the findings of Experiment 1 that used verbal report measures. A logistic regression predicting the probability of switching (from a between placement to a left placement or vice versa) showed no effect of clip type (p=.530) or number of repetitions (p=.412).

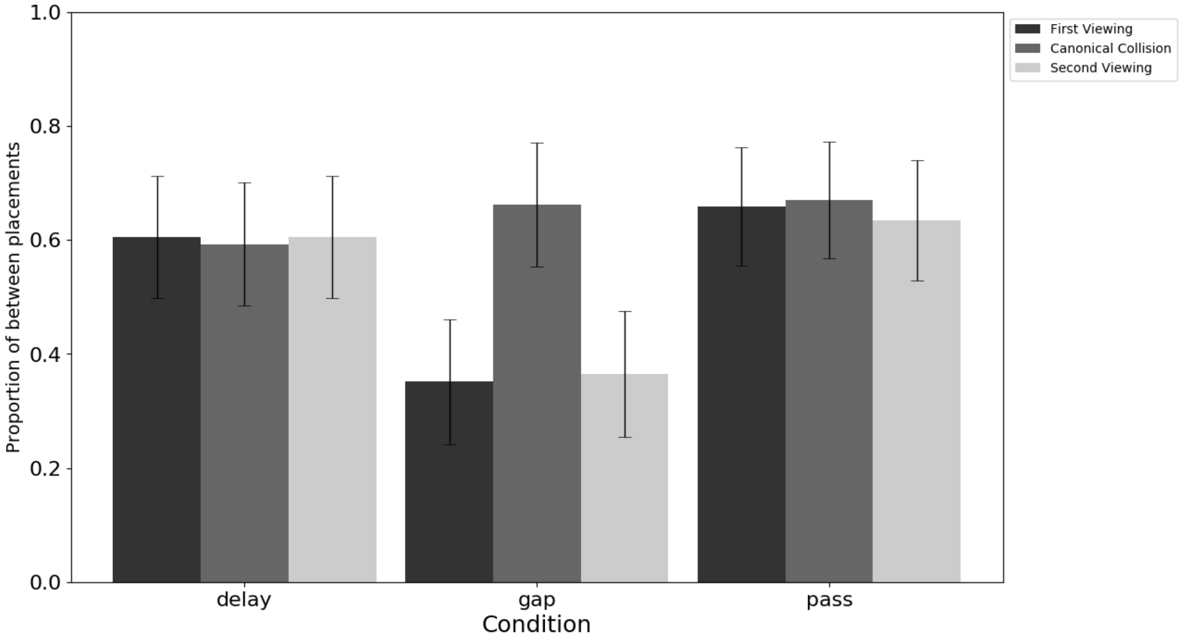

Figure 6: Proportion of participants placing the obstacle between objects A and B in Experiment 2, per clip type and viewing order (error bars represent 95% Cis).

Discussion

Experiment 2 validated the findings of Experiment 1: Observers do not merely report causal impressions of unrealistic events, but they also act on these impressions. Using a behavioural measure, we have found clear evidence that unlike what is commonly claimed, people spontaneously postulate causal links, at least in the case of temporal delays and negative spatial gaps (“passing”).

Since, in our task, there was the safe option to place the obstacle in the trajectory of the outgoing object B, thus avoiding any causal commitment, our results show that for the majority of participants there was no doubt as to the existence of the causal link. This is more impressive if we consider that both here and in Experiment 1, object A actually starts moving on its own, thus providing some evidence for the possibility of spontaneous motion in these displays. Nevertheless, it appears that most participants in the delay and pass conditions did not even consider the possibility of object B moving on its own.

In the gap condition, the evidence was inconclusive. While it is possible that the presence of a gap weakens the causal impression, the lower level of obstacle placements between the objects might occur despite a causal impression, due to the location and the assumed properties of the gap[10]. As reported elsewhere (Michotte, 1963; White, 2016), some people see the gap in a Michottean launching sequence as containing an invisible mediating substance. This raises the question of how the obstacle would interact with that mediator. Observers may assume that the obstacle cannot be placed in a location already occupied by the mediator, or they might have thought that the invisible material transmits the force exerted to it even if the obstacle is placed in it. Generally, if the properties of the gap are unknown, the safer option is to avoid it altogether (e.g., consider the case of magnets).

Regardless of this difficulty, the divergent result for the gap condition serendipitously counters a potential criticism of our behavioural paradigm: One might object that the between placements here do not correspond to causal impressions, because observers merely react to the symmetry of the static visual display, with the shapes A and B bounding a target area on the left. The right placements in the gap condition, with identical static visual display, rule this out. Our results thus illustrate that participants' obstacle placements result indeed from considerations of the specific way objects interact in these dynamic events.

Regarding the effect of interpolated canonical events on obstacle placement, here we found no drop in the number of placements that correspond to unambiguous causal impressions. This is in stark contrast with the finding of Experiment 1, which found a large drop in causal ratings. This supports our view that the drop in ratings reflects a change in interpretation of a verbal question about causality, which is not observed when a behavioural measure without explicit mention of causality is used.

Finally, regarding the role of repeated experience on observers' causal impressions we recorded contrasting effects in Experiments 1 and 2. In neither case were causal impressions dependent on repetitions, since they already occurred after a single viewing of the unrealistic events. The amplification in Experiment 1 and the slight decrease observed in Experiment 2 as a result of repeated exposure, could be reconciled by allowing the simultaneous operation of multiple processes with potentially opposing effects. While more experiences of an event

---

[10] Note that since the distance between the two objects was 67.5mm and object A stops 30mm before object B, approximately half the area to the left of the of object B is "taken up" by the gap.

might strengthen causal impressions through some form of contingency learning, more time available to reflect on the situation might lead to the acknowledgement of the potential fallibility of first impressions or the realisation that placing the obstacle to the right of object B is the safest option. This is the view typically taken on people's experience of visual illusions, e.g., that a straw looks bent in the glass but we know on reflection that the perception deceives. While this requires further study, it is a side issue here. In our subsequent experiments we bypass the issue, and restrict ourselves to showing observers only a single repetition of each event. This is the more conservative approach and yet sensitive enough for present purposes, as minimal exposure is sufficient for generating causal impressions of unrealistic events and does not leave observers, at least 21^{st} century observers, in the substantial state of confusion postulated by Michotte (1963).

In the next experiments we also return to recording causal impressions using the rating scale. Despite the merits of the behavioural measure, it can't be used with any imaginable sequence. In the next experiment, for example, the motion of object B will be replaced by even more unrealistic events such as size and shape change, in an effort to test the boundaries of our findings. It is hard to imagine how such unrealistic events can be stopped with an obstacle and, thus, we will be returning to the rating scale that has now been cross-validated behaviourally.

Experiment 3

In Experiment 3, we ask participants for causal impressions in events that differ even more strikingly from canonical launch events and have rarely been used in prior literature. Specifically, as can be seen in Figure 7, while object A still approaches the stationary object B from the left as before, the behaviour of the latter upon contact is surprising: B will change colour by taking the colour of object A (fig. 7A), or it will change shape by becoming a circle (fig. 7B), or it will change size by increasing its surface four-fold (fig. 7C) or it will move in a 90-degree angle compared to the trajectory of object A (fig. 7D). While the angle manipulation may be seen as another, more extreme, quantitative deviation from the kinematic features of a collision between two movable objects, the shape, size, and colour change manipulations present more qualitative deviations.

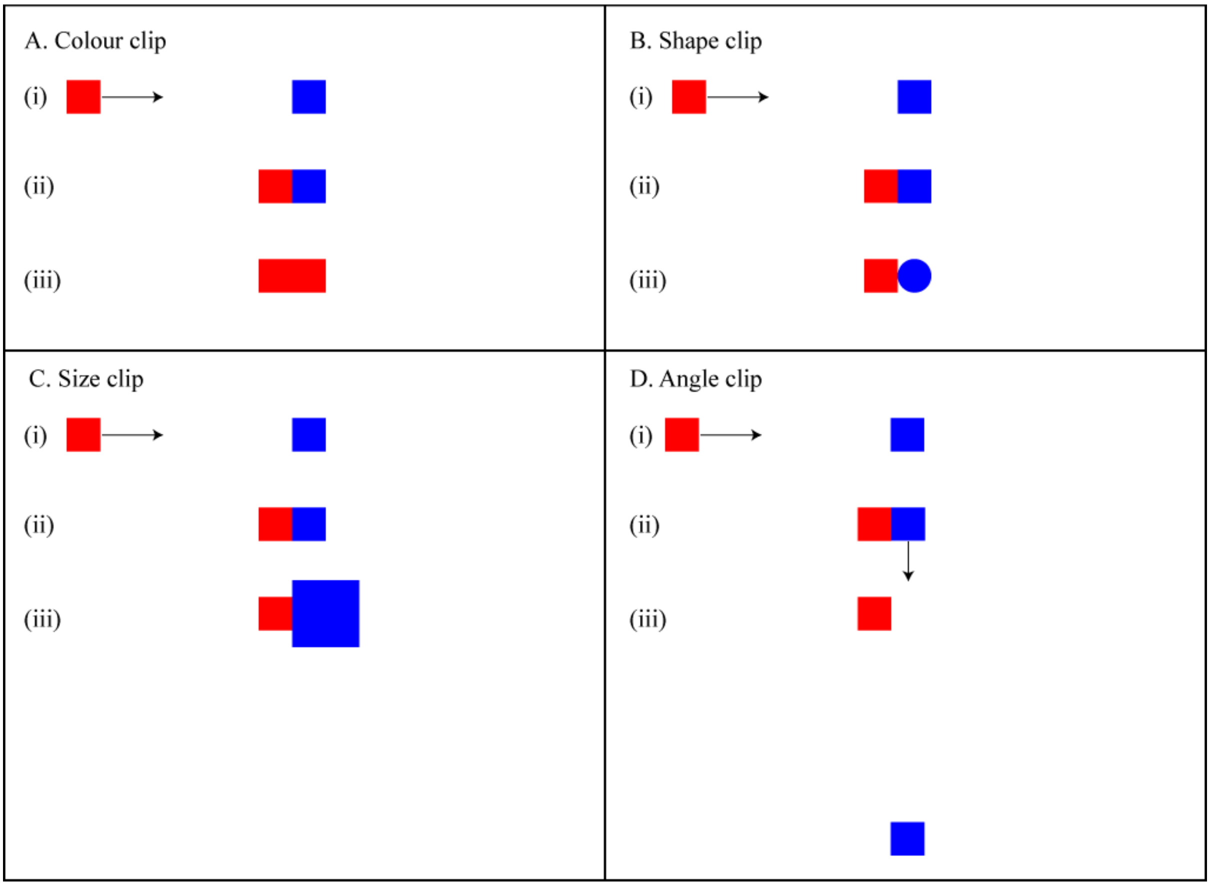

Figure 7: The four clips corresponding to the four conditions in Experiment 3. Each panel shows (i) the initial positions, (ii) the configuration at the time of interaction and (iii) the final positions (the arrows represent the direction of motion and were not visible during the experiments).

Our behavioural measure of causality used in Experiment 2 is, unfortunately, not a viable option here: although a between placement could still potentially block the effect, there is no alternative candidate placement to compare against and no rationale under which a placement to the right of B would stop it from changing shape, size or colour. In the angle condition participants would have to rotate the obstacle, thus introducing further complications. Thus, to assess these more unrealistic events, we revert to the verbal rating from Experiment 1.

To the extent that causal induction from visual input is not constrained by realism, as we have argued, we expect consistently high causal ratings after a single exposure to all these highly unrealistic sequences. Among the various stimuli, the angle variable has received the most attention, albeit not at the level of the spatiotemporal properties discussed earlier. Michotte (1963) informally reported that the sharper the angle the fewer the causal reports, completely disappearing at 90^o. In contrast, 45% of participants in Beasley (1968) did describe seeing a causal interaction in similar conditions. More recently, Straube & Chatterjee (2010) following the usual repeated-measures procedure reported that sequences with angles averaging 31.53^{o} were judged as non-causal while participants in White (2012) gave low causal ratings to sequences with an 80^{o} angle. White (2006), in his argument that unrealistic sequences do not lead to causal impressions because they do not match any stored schema, also offers the case of a sharp angle as an example of such sequences. Consequently, clips with 90^{o} sequences have been used as non-causal control stimuli in various studies (Hubbard, 2013).

Several studies considered colour change, however, typically this appeared as a causal factor, when we use it here as an effect. Using stimuli very similar to ours, Michotte (1963, exp. 75) reported that only 2/11 naïve participants had the impression of A making B change colour. Based on findings from another study, Michotte (1963, exp. 78) argued that an object changing colour is not seen as cause of another object's movement, confirmed by Schlottmann and Shanks (1992). However, in the latter study, the colour change coincided with the motion of object B, and this competed with A’s motion as a cause. Two further studies without strong alternative causes found the opposite effect (White, 2005; Young and Falmier, 2008). Finally, Guski & Troje (2003) found that B changing colour increased causal impressions when it acted as a filler during a delayed launch event and White (2016) found similar results. The repeated-measures procedure was used in all these studies, but Young and Falmier (2008) did not include canonical launch events. To our knowledge, there are no studies looking at visual impressions of causality in the case of size or shape change but, in any case, it is clear that the literature would not predict causal impressions in any of the stimuli discussed here, especially after a single exposure.

Among the four clips, the angle one is the most realistic in the sense that it comes closer than the others to a canonical Michottean collision by having one motion followed by another upon contact. On the other hand, the colour clip is a case of property transmission (White, 2009), since object B takes the colour of object A upon contact. In Michotte’s view, a causal impression does result from such transmission, so-called “ampliation”, though he limits this to the transmission of motion properties. The shape and size clips should receive very low causal ratings, not only because of their improbable nature but also due to the absence of any property transmission.

In order to test for causal impressions in such unlikely sequences, we follow the approach of Experiment 1, except that, as discussed, we do not vary event repetitions, for efficiency reasons. Participants watch the target sequence twice, once before and once after viewing and evaluating a canonical clip. Similar to Experiments 1 and 2, this allows us to study how the evaluation of unrealistic clips is affected by interpolated realistic sequences. In Experiment 1, we argued that causal ratings for sequences with delays and gaps decreased only after these sequences were compared to the canonical realistic event, reflecting the findings of previous research. By using a wider range of events in Experiment 3, we assess whether there are limits to such comparison processes. Do participants report comparative judgements for any two events for which explicit causal ratings are requested, or do the events need to be sufficiently similar? This question arises because in other domains, science reasoning, for instance, it is argued that presentation of similar cases facilitates reasoning, because minimal contrasts highlight specific dimensions for comparison (Schwarz, Chase, Opezzo & Chin, 2011). Here, we explore how the distance of a sequence from the causal exemplar affects observers' evaluations of causality.

Materials and Design

The clips were created using Adobe Flex 4.7 and are depicted in Figure 7. Up to the point when object A (red square) contacts object B (blue square) the clips are identical to the canonical Michottean collision (Figure 1A). Immediately after contact, however, object B starts moving vertically towards the bottom at the same speed as object A (angle condition), instantaneously changes colour to red (colour condition), doubles in size to 20x20mm (size condition) or changes into a circle with a 10mm diameter (shape condition).

As in the other experiments, the design was mixed factorial with the type of clip ("angle", "colour", "shape", "size") varied between participants and the time of causal report ("before" or "after" watching the canonical collision) varied within.

Participants

We recruited 123 participants through Amazon Mechanical Turk. There were 30 participants in the angle condition and 31 in the other three conditions. The mean age was 36.61 (SD=11.53) and there were 57 female participants. Each participant was paid $0.40.

Procedure

The procedure was almost identical to Experiment 1, except that the clip in each phase of the experiment was shown only a single time. The causal question was slightly adapted to account for the behaviour of object B. Specifically, depending on condition, participants were asked: “Do you have the impression that red somehow made blue move/change colour/change size/change shape?” After rating the unrealistic event, participants in all conditions watched the standard canonical collision and were asked whether they had the impression that the red square made the blue one move. Finally, participants watched the unrealistic clip a second time and answered the condition-dependent causal question again.

Results

As can be seen in Figure 8, participants gave high causal ratings to all clips the first time they watched them. Across conditions the mean rating was 85.85/100 (SD=23.00). In fact, only 13 out of the 123 participants across conditions gave a causal rating equal or below the midway point (50) upon first viewing and in all conditions the mean ratings were significantly higher than chance (50), Colour: t(30)=7.837, p<.001, d=1.408, Shape: t(30)=7.937, p<.001, d=1.426, Size: t(30)=16.577, p<.001, d=2.977, Angle: t(29)=6.608, p<.001, d=1.207.

As in Experiment 1, the canonical collision clips received significantly higher ratings (mean=96.55%, SD=8.98) than the deviant clips, F(3,119)=31.990, p<.001, $\eta^2_{partial}$=.212, without an effect of condition or an interaction effect. Causal ratings for canonical clips did not differ from the respective clips in Experiment 1 (p=.113) but the ratings for the striking qualitative deviations used here were noticeably higher (mean=85.85%, SD=23.00) than the ratings for clips with quantitative deviations from realistic collisions used in Experiment 1 (mean=63.21%, SD=35.27), t(212)=5.663, p<.001).

Regarding the effect of interpolated realistic collisions, in most conditions and unlike Experiment 1, watching and evaluating the canonical clip did not diminish the ratings of the critical clip the second time it was shown. A repeated measure ANOVA showed no significant main effect of the time of causal report (before-after viewing the canonical collision), F(1, 119)=2.993, p=.086, $\eta^2_{partial}$=.025, or of condition F(3, 119)=2.610, p=.055, $\eta^2_{partial}$=.062 but did show an interaction effect, F(3, 119)=3.081, p=.030, $\eta^2_{partial}$=.072. Statistical tests support the visual impression (Figure 8) that this interaction effect is due to the angle condition. Paired t-tests showed non-significant differences between the first and the second viewing of the deviant clip in the colour condition (t(30)=-0.772, p=.446), the shape condition (t(30)=.263, p=.794) and the size condition (t(30)=.742, p=.464). In contrast, there was a decrease in rating after seeing the canonical collision clip in the angle condition (t(29)=2.472, p=.019, d_z=.451).

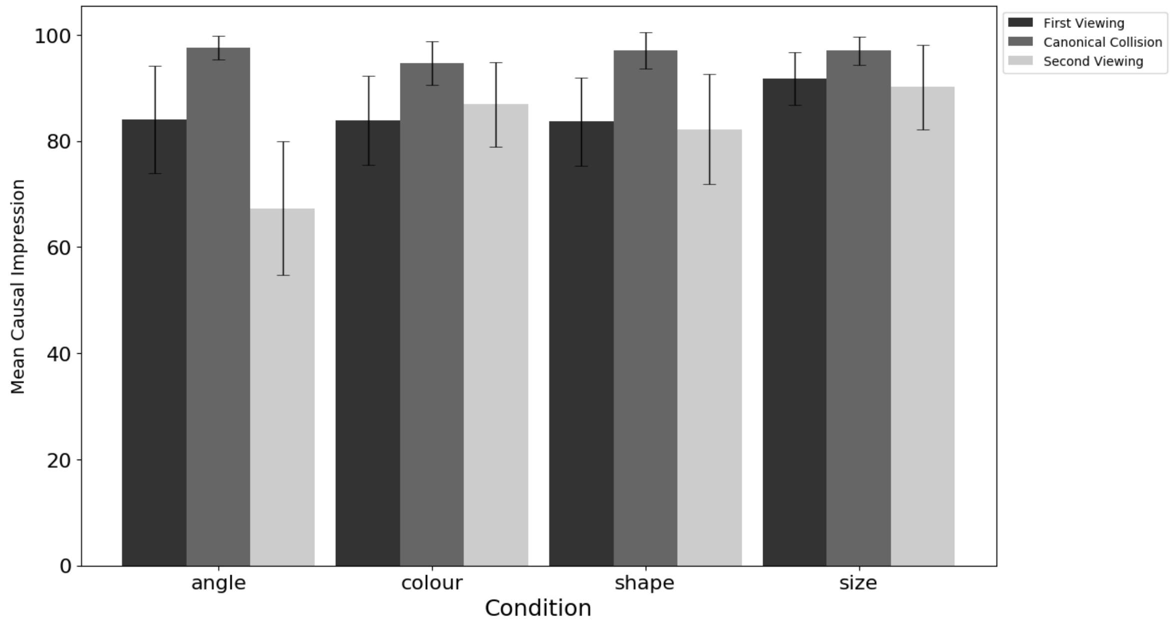

Figure 8: Mean reported causal rating per condition and viewing order in Experiment 3. In every condition, the middle clip was a canonical Michottean collision (error bars represent 95% CIs).

Discussion

Experiment 3 found very high ratings of causality for very unrealistic sequences that our participants are unlikely to have experienced in the past. Their causal impressions were strong and immediate in the sense that they resulted from a single exposure.

These results provide further evidence that, unlike what most theories predict, realism is not a requirement for the generation of causal impressions. This particular assumption has limited researchers to looking for phenomenal causality only in realistic sequences: besides launching, Michotte documented causal impressions in what is known as “entraining”, in which object A, after arriving next to B, continues its motion, as if carrying B (Michotte, 1963 exp. 48) or “triggering” where object B moves noticeably faster than object A (Hubbard, 2013; Kominsky et al., 2017; Michotte, 1963). White and Milne (1999) have shown impressions of causality in “bursting”, in which object B appears to be smashed by object A, and “disintegration” in which object B apparently explodes upon contact with object A (see also White, 2017), while Kanizsa and Vicario (1968, see also Schlottmann et al., 2006) have shown impressions of social causaliy in action and reaction events. The point of the current findings is not to add more types of interactions to this catalogue, but rather to highlight the need to change focus. If realism ceases to be a requirement, causality in visual events is far more ubiquitous than hitherto assumed. The question should no longer be what interaction types produce causal impressions but rather what interactions fail to do so and why.

With the exception of the angle condition, viewing the canonical collision had no effect on the causal ratings of the unrealistic clips, which were roughly identical for the two viewings. Unlike what was observed in Experiment 1, participants this time did not change their interpretation of the causal question in most conditions. This suggests that participants do not switch their interpretation of the task merely because they have now seen a canonical clip, but rather this depends on the relationship between the canonical and the deviant clip. Participants may compare the events and begin to evaluate them relative to each other mainly if the target event and the canonical collision are fairly similar in the sense of belonging to the same event category. In the canonical collision, object motion causes object motion. The angle event was the only unrealistic event here falling under such description, the only one for which we could ask the same verbal question (did A make B move?) and only there did we observe a drop in ratings. The qualitative change events (shape, size, colour) and the canonical collision, in contrast, may not appear as instances of the same category at all, nor could we ask the same question (did A make B change shape/colour/size?). This may have prevented comparison, and kept the ratings for realistic and unrealistic events independent.

A parallel argument can explain why the ratings for the unrealistic clips in this experiment were higher compared to the ratings for sequences in Experiment 1, even though the latter featured much milder deviations from reality. Even before experiencing a canonical collision, some participants (or all participants to some degree) may spontaneously compare the target against a good realistic collision, if the target sequence is sufficiently similar to the canonical collision to trigger such a memory. When the unrealistic target sequences do not resemble a collision, as is the case with the state change events in Experiment 3, such intuitive comparisons are less likely, and participants respond solely to the degree of causality present in the sequence. Clearly, though, this interpretation requires further empirical verification.

Experiment 4

A general objection that could be raised against our findings is that they apply only to limited, artificial animation displays. In such displays, the lack of realism in the context of the motions, or in what we earlier called the peripheral features of the sequence, might be interpreted by participants as a signal to relax their requirements for causal attribution. Thus, our finding that participants report causal impressions in sequences with unrealistic core features might be due to the unrealistic properties of the peripheral features of the sequence.

We previously assumed, following Michotte (1963), that only the core kinematic properties of the event sequence matter for the generation of causal impressions. If there is a modular causal detector, it responds only to core properties, sometimes producing illusions, e.g. when a wooden ball sets a shadow in motion (Michotte 1963, exp. 11) or when computer generated objects appear to collide. This type of selective visual mechanism is not unusual. Mice respond to two-dimension visual looming patterns by freezing, as they would to a looming predator (Yilmaz & Meister, 2013)[11]. Baby birds open their beak for food to an artificial stimulus, a red spot on an abstract 2D drawing of the mother's head but stop responding when the spot is not-present or has any other colour (Tinbergen & Perdeck, 1950). Such visual mechanisms respond to stimuli of evolutionary importance and a similar mechanism for the detection of causality certainly could explain the remarkably efficient manner of everyday human causal thinking. From Michotte's displays to the stimuli used here, the peripheral properties of the animated sequences were thus assumed to be irrelevant to whether causal impressions are generated or not.

However, from a very different perspective, it might be exactly those peripheral features that explain our results. The flat, 2-dimensional appearance of the shapes and the absence of shadows or any other environmental features are clear indications of the artificial nature of the animations. Participants recognising this might respond to representations of causal interactions rather than reporting causal impressions as in the real world. The argument is that artificial animations lack ecological validity, and that observers would not act in the way we are reporting here in more ecologically valid circumstances. Although this is a general

[11] We would like to thank an anonymous reviewer for bringing our attention to this example.

problem potentially affecting the whole research field, we are not aware of any discussion of the role of peripheral features in the perception of causality.

In our final experiment, we will thus replace our animated stimuli with video-based sequences. We will keep the experimental procedure of Experiments 1 and 3 and will use edited video sequences that match as closely as possible some of the animations from those experiments. In particular, we will use the delay and gap clips from Experiment 1, as examples of events that are unrealistic in their core kinematic features, as well as the colour sequence of Experiment 3 as an example of an unlikely qualitative state change event. For all sequences, the unrealistic peripheral features (CG graphics) are replaced by realistic ones: as shown in Figure 9, the objects are now two balls rolling and colliding on a table, with the video carefully edited to produce the deviant features.

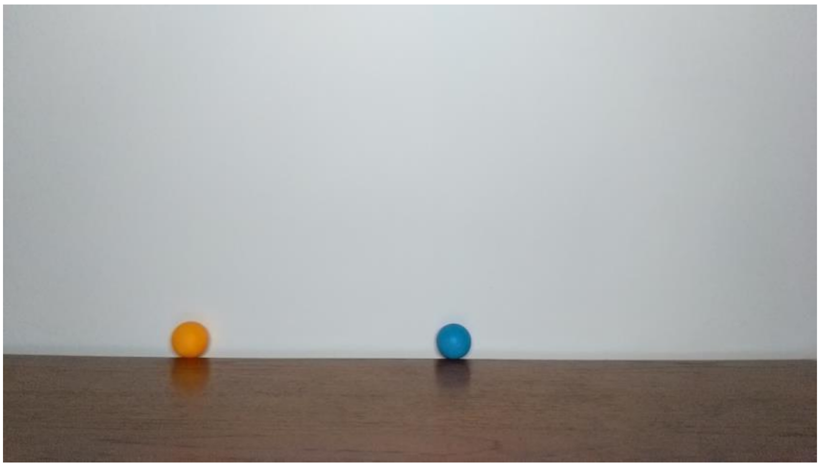

Figure 9: The starting frame in all video-based stimuli of Experiment 4

Materials and Design

The experiment was created with HTML5 and the clips were processed using Adobe Premiere Pro CC 2018. An example frame is shown in Figure 9 and the clips can be viewed at https://goo.gl/ef43f2. We produced four types of clips (delay, gap, colour and canonical) whose properties (spatial and temporal) matched as close as possible the respective animations from experiments 1 and 3 (figures 1 and 7). Two noteworthy differences were that the squares were replaced by table tennis balls and the colour of object A was orange instead of red.

As before, the design was mixed factorial with the type of clip (“delay”, “gap”, “colour”) varied between participants and the time of causal report (“before” or “after” watching the canonical collision) varied within.

Participants

We recruited 90 participants through Amazon Mechanical Turk. 30 participants were randomly assigned to the delay condition, 31 to the gap condition and 29 to the colour condition. The mean age was 38.23 (SD=13.75) and there were 49 female participants. Each participant was paid $0.40.

Procedure

The procedure was identical to Experiments 1 and 3 with each clip being displayed a single time. For the delay and gap conditions we used the causal question from Experiment 1, adjusted for the new colours (“Do you have the impression that orange somehow made blue move”?) and for the colour condition we used the phrasing from Experiment 3 (“Do you have the impression that orange somehow made blue change colour?”).

Results

Across conditions, participants rated the initial deviant clip significantly higher than the midpoint towards the causal statement ($t(89)=2.461$, $p=.016$, $d=.260$) and there was a significant drop after the observation of the canonical clip, ($t(89)=3.430$, $p<.001$, $d=.321$). There were, however, differences between conditions, as can be seen in Figure 10.

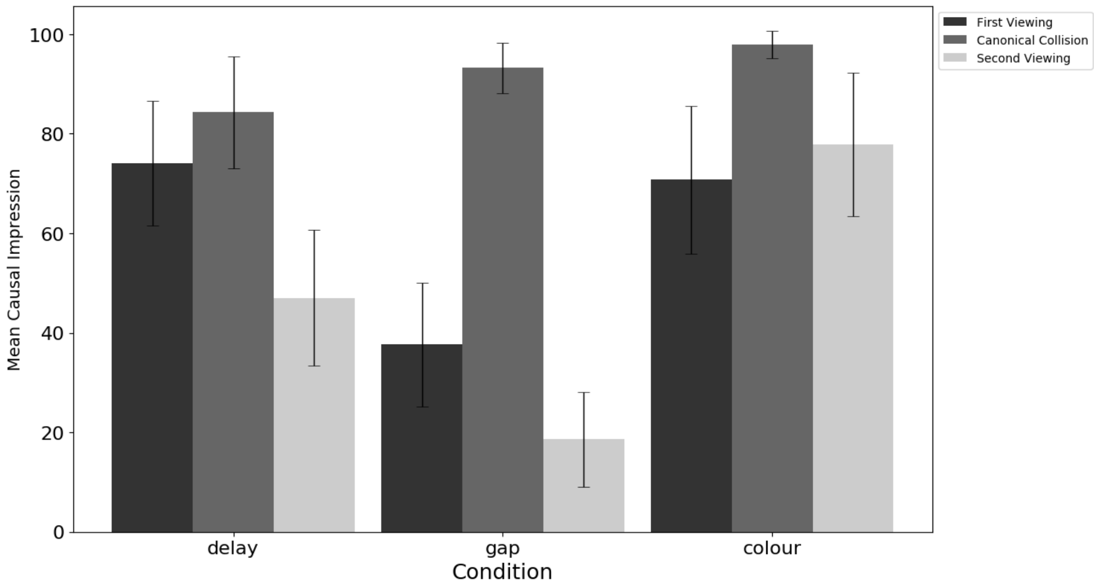

Figure 10: Mean reported causal rating per condition and viewing order in Experiment 4. In every condition, the middle clip was a canonical Michottean collision (error bars represent 95% CIs).

Participants in the delay and colour conditions behaved almost identically to the participants in the respective conditions in Experiments 1 and 3 (compare Figure 10 with Figures 3 and 8). In both cases the rating for the first viewing of the deviant clip was significantly higher than the midpoint towards the causal statement (delay: $t(29)=3.767$, $p<.001$, $d=.688$, colour: $t(28)=2.744$, $p=.010$, $d=.509$). In the case of delay there was a significant drop after the canonical collision ($t(29)=4.011$, $p<.001$, $d=.740$), similar to Experiment 1, while in the case of colour there was no such change, ($t(28)=-1.012$, $p=.320$) similar to Experiment 3.

In contrast to Experiment 1, however, the majority of participants did not report a causal impression in the gap condition. In fact, the mean rating was almost significantly lower than the midpoint ($t(30)=-1.954$, $p=.060$, $d=-0.351$). However, similar to Experiment 1, even in the gap condition, there was a significant decrease following the observation of the canonical clip ($t(30)=3.867$, $p<.001$, $d=.605$)

The comparisons against our previous experiments are verified formally: The only difference against earlier results occurs in the first viewing of the gap clip, the ratings of which are significantly higher in the animation version (Experiment 1, single repetition) compared to the video version ($t(60)=2.627$, $p=.011$, $d=.673$). The ratings for the first viewing of the other two clips did not show any effect of peripheral features ($p>.130$) and the same was true regarding the change in ratings before and after the canonical clip for all three conditions ($p>.34$).

Discussion

Experiment 4 replaced the commonly used animation stimuli with realistic video captures, to investigate whether the peripheral features of Michottean sequences affect causal impressions. Overall, our results concur with Michotte (1963), in that impressions of causation depend predominantly on the core features and not the peripheral features of event sequences. In both the delay and the colour clip, participants report strong causal impressions, very similar to the ratings observed when animations were used. In addition, the role of the canonical collision was the same as in our earlier experiments: no effect in the case of qualitative departure from an ideal collision (colour) and a very pronounced effect when the deviation is quantitative (delay and gap).

Regarding the gap sequence, this time the majority of participants did not think that there was a causal relationship between the motions of two balls. This was an unexpected result and the explanation is not clear. Remember also that in Experiment 2, participants in the gap condition were less prepared compared to the other conditions to obstruct the motion of object A as a means to stop object B from moving. In that case, we discussed the possibility that participants are uncertain about the location and properties of the gap, thus choosing to avoid it, especially since another effective response was available. Given the current results, it would be tempting to discard that explanation and assume, more generally, that the presence of a gap weakens causal impressions from first encounter. However, the results of our Experiment 1 and previous findings that show that causal impressions are less, not more affected by gaps compared to other deviations (see Appendix), seem to argue against this.

A more mundane explanation is that the particular way we generated the video clip for the gap sequence accidentally introduced properties that led to the different pattern of responses. In general, the process of creating animations is very different from producing video clips, with much coarser control in the latter case. This was made even harder by the need for the core properties of the video clips to closely match each other and, especially, the properties used in the animations. Simplistic solutions, like animating the picture of a ball against a static background would fail to produce the realistic peripheral features we were looking for here, e.g. shadows on the surfaces and on other objects. However, using recordings of actual moving objects led to reduced control over the precise speed of the objects, which were simply knocked into motion. To achieve the intended speed, we used post-processing techniques, like holding a frame visible for longer (slowing down) or shorter (speeding up), but this inevitably introduced artefacts (e.g. less than perfectly natural motion). As a result, our aim was to achieve a balance between matching the core properties used earlier and keeping artefacts to a minimum. One can assess the degree of our success for the various clips by reviewing the stimuli (https://goo.gl/ef43f2). In any case, the variable findings that we are observing here point to the need for more focused work regarding the way spatial deviations affect causal impressions.

This was only a first attempt to study Michottean collisions in realistic environments and, clearly, further work is needed to establish the precise role of non-core features. What is critical for now is that the causal impressions in deviant sequences that we observed in Experiments 1-3 are not due to participants relaxing their requirements for causality in ecologically invalid circumstances. The absence of realism in the core kinematic features of the animation does not destroy causal impressions, even if all other aspects of the animation are realistic.

General discussion

In four experiments we examined whether non-realistic visual sequences, i.e. dynamical sequences whose abstract properties grossly violate Newtonian mechanics and are unlikely to have real-world counterparts, generate causal impressions. There are five key findings. First, people give high ratings of causality to a variety of unrealistic, never before seen visual sequences. This is in contrast to what standard theories of causality predict, but in line with what one-shot causal induction requires[12]. Second, this does not reflect contingency learning. Third, it is not just a response to an implicit verbal task demand, but it also appears in a behavioural paradigm. Fourth, it is not limited to artificial, animated displays, but is also found if unrealistic events are shown as photorealistic video sequences. We conclude that these ratings reflect observers' spontaneous causal impressions. The discrepancy with the previous literature is explained by our fifth finding, that ratings of non-causality appear for these deviant launch events after comparison with canonical launch events. These results are discussed in turn.

In Experiments 1 and 3, we presented participants with Michottean sequences featuring either the quantitative deviations from realistic collisions commonly discussed in the literature

---

[12] In a later section, we speculate about the possibility of either a more abstract perceptual module, or some mechanism-based account explaining our data.

(Experiment 1) or even more unrealistic state-changing events (Experiment 3). In every case, there were clear impressions of causality after a single encounter: the majority of participants attributed the behaviour of object B to object A, regardless of whether there was a 250ms temporal delay or a 30mm gap between the cause and the effect, whether object A completely overlapped object B in what is known as a “non-causal pass” (Scholl & Nakayama, 2002) or whether object B instantaneously changed shape, size or colour upon contact, or, finally, whether it moved in a highly unrealistic 90 degree angle. Across conditions (N=424), 78% of participants gave ratings higher than the indecision midpoint closer to the causal statement, with the average rating being 74%.

Note that we are not, at this time, committing to a particular process basis for these causal impressions but our findings present a major puzzle to current theories of phenomenal causation. For direct perception theories (Michotte, 1963; Scholl & Tremoulet, 2000), the hypothesised input analyser should not be responding to sequences with extreme deviations from the spatiotemporal parameters of canonical Michottean collisions, let alone sequences that involve unrealistic state changes. A modular process shaped by evolution cannot be tuned to stimuli not found in the natural environment. Schema-matching and simulation-based theories (Sanborn et al., 2013; Weir, 1978; White, 2006), on the other hand, although more flexible, are still constrained by realism: for a sequence to be matched against a stored representation or to be simulated, the individual must have had prior experiences of sequences with similar properties.

To avert the challenge, proponents of these views may argue that their theories do not apply to the work presented here, because participants in our experiments do not report “true” causal impressions from visual input. Our findings, the critic argues, are constrained to the particular experimental settings and would not appear in the real world. Michotte (1963) already attempted to disqualify initial impressions in artificial environments, attributing them to confusion, though this is hardly compatible with the consistency of our findings here.

One concern along these lines is that participants in our experiments did not report causal impressions relying on the visual properties of the sequences but rather inferred a relationship based on the co-occurrence of events in the experiment. Indeed, in Experiment 1 (but not Experiment 2), when different groups of participants watched the critical sequences for a variable number of times before reporting their impressions, we observed a mild repetition effect: causal ratings were higher, the more times the sequence was displayed. However, if statistical learning has a role to play here, that role is to strengthen the causal impression rather than generate it: the majority of participants reported a causal impression in all four experiments even when the deviant sequence was displayed for a single time. A single piece of data is clearly inadequate to support statistical learning. This is the second major implication of our findings.

The third major point is that our results are not artefacts of the verbal response measure used. Although we agree that asking participants for ratings of causation might lead to various misinterpretations, Experiment 2 shows that the reported impressions are stable enough to support spontaneous goal-oriented action. After viewing the deviant delay or pass clips in Experiment 1, the majority of participants chose to obstruct object A as a means to stop object B from moving. This was the case, despite the fact that the option to obstruct object B directly was available and was, in fact, the safest response option. Thus, the behavioural evidence indicates that deviant sequences are treated as causal, while cross-validating the verbal response measure.

The fourth major conclusion is that our results generalize beyond the confines of computer generated animations. According to a very general objection, the lack of ecological validity due to the use of computer animations also deprives our findings of their validity. The proponent of this view insists that in real-world situations, the absence of prior experience would still lead to the absence of causal impressions. The consideration of this objection has led to Experiment 4, one of the few studies in this tradition to use edited video sequences (see also Oakes & Cohen, 1990). Nevetheless, in two of the three sequences that we used (delay and colour), we found strong causal impressions and no difference in ratings between animated and video-based stimuli.

One might insist, nevertheless, that these days, people are quite familiar with the possibility of deception through photo or video editing. An edited video sequence of a collision may still be experienced as a representation and participants might again relax their requirements for causal interactions just as they might do for animations. Although we believe that the presented work has moved the discussion forward, we concur that there is still no conclusive evidence against this form of the ecological objection. Albeit, we think that the examples of filmmaking or of (successful) magic shows, where observers have convincing visual impressions of causality for unrealistic events, including events known to be impossible, yet experienced “as if real”, also support our case. Special effects in cinema are effective because, even in the apparent absence of realism, events appear to be causally connected. Bad movies, in contrast, are ridiculed for their special effects, because we can read the intention of causality, yet fail to see or otherwise experience it. Similarly, we know that the dove was not transported to the magician's hat because she tapped it with the stick three times but we, nevertheless, applaud the illusion of a direct causal link. In contrast, we smile indulgently at our children performing magic tricks, because we know what visual impression is intended, yet all we see is how the trick is performed. We know of no formal evidence on this type of dual experience, but hope that the examples ring familiar to readers, as they point to examples of visual causal illusions for unrealistic events "in real life" and to their limits.

This overall conclusion leaves us with two major questions: (1) What is the source of causal impressions in non-realistic events and (2) what explains the discrepancy between our findings and the data that has led to the consensus formed in the last 60 years according to which observers treat non-realistic events as non-causal?

Regarding the first question, although we are not committed to a particular process, one can speculate that our high initial ratings for non-realistic events might still reflect the presence of a perceived causality module, albeit one that is sensitive to far more abstract properties compared to the module that Michotte (1963) envisaged. If such a module exists, it reflects causal invariances in the environment, which clearly go beyond the conditions identified by Michotte for launch and related events.

Alternatively, at least for some of our sequences, the observed initial causal impressions might have a more inferential basis (Danks, 2017; Waldmann, 2017; White, 2014). When discussing the gap sequence in the context of Experiment 2, for example, we suggested that participants postulate the presence of an invisible substance between the two objects, thus avoiding placing the obstacle on that substance. In case the postulated mechanisms require autonomous agents, participants might also have impressions of social causality as discussed earlier (Kanizsa & Vicario, 1968; Schlottmann & Surian, 1999; Schlottmann et al., 2002, 2006, 2009, 2012). It is not clear, however, whether, the dynamic relationships featured in our sequences resemble known mechanisms thus producing the causal impressions here or, conversely, whether other low-level properties generate the causal impressions, which then prompt a search for mediating mechanisms.

One concrete suggestion in that direction comes from White (2014) who suggests 14 clues that are used as heuristics in single instance causal detection. In White's view, these clues are generalised from and originate in the experience of action on objects. Although many of those clues are present in the sequences presented here, the theory has been tested only with textual descriptions of realistic events. It is not clear which of these clues are strong indicators, which are weaker, and which are, in fact, irrelevant for causal detection. In addition, the author has elsewhere (White, 2006) argued against the possibility of causal impressions in the absence of realism. Nevertheless, we hope that our findings will widen the scope of research on phenomenal causation, in search for those visual properties that lead to such spontaneous causal inferences or trigger the "abstract" causal module.

Regarding the other open question, the second component of our experiments attempted to shed some light on the causes of the puzzling discrepancy between the current and prior results: in every case, we presented the same clips and asked the same causal question before and after displaying a canonical collision, thus generating a shortened, controlled version of the usually employed repeated measurement experimental procedure. In Experiment 1, where the clips and the dependent measures closely matched previous work, participants' ratings decreased significantly after watching Michotte's ideal sequence. In contrast there was no change before and after the canonical clip either in Experiment 2 where we inferred causal impressions from the behaviour of participants or in Experiment 3 where most clips differed qualitatively from the canonical collision. When the animations were replaced with edited video clips in Experiment 4, the observed patterns matched closely the respective sequences from Experiment 1 (delay, gap) or from Experiment 3 (colour).

This decrease in ratings for identical clips in Experiment 1, and we would argue, in most past studies, shows that participants change what they rate as 'causal' during the course of the experiment. Observers naturally compare successive displays and, if sufficiently similar, align them by their commonalities and differences. The presentation of the launch event brings into focus properties that vary between the sequences and which initially receive little attention. We argue that observing a canonical collision after a deviant one produces such a new focus and leads participants to change the dimension of judgment.

What is the new dimension that is made salient after comparing the deviant clip to a canonical collision? One possibility is that observers initially make a general causal inference, as described above, but the canonical launch triggers the perceived causality module (Michotte, 1963). From then on, observers narrow their judgments to report only low-level perceptions of causation. This requires the additional assumption that people can distinguish causal impressions resulting from their modular input analyser against those inferred from the dynamical features of a sequence. Scholl and Tremoulet (2000) argue, for instance, that perceptions of causality are particularly compelling and irresistible, in contrast to other forms of causal impressions. We know of no study providing direct evidence for this phenomenal aspect of observers' impressions. It would be difficult to get convincing experimental data on such a subtle aspect of perceptual experience, but we made a similar argument regarding our real-life ability to distinguish between illusions of causality and perceived intentions to represent causal links in magic shows. The continued popularity of Michotte's approach despite its well-rehearsed ambiguities perhaps attests to the power of the idea.

According to an alternative account, all sequences produce the same perceptual or inferential impression of causality. However, the presentation of the canonical collision triggers a separate inferential process: participants either start reporting the level of realism present in a sequence, in other words, they start responding to whether they have experienced a sequence of this type before, or they narrow the type of causality they respond to, in other words they start reporting the degree to which the mechanism that mediated the causal relationship adheres to Newtonian laws. The canonical launch is an idealization of a Newtonian sequence, similar to what people have frequently experienced in the past, at least in terms of its core properties. The comparison against a deviant sequence, which is clearly non-Newtonian or just perceptually very uncommon, highlights the lack of realism in the latter. The question of whether "square A made square B move" starts being interpreted as meaning whether the sequence was realistic or whether A made B move through a particular known mechanism, e.g. by transferring its impetus or its kinetic energy (Hamrick, Battaglia, & Tenenbaum, 2011; McCloskey, 1983). Similar to the perceptual account discussed before, this view depends on the assumption that people can distinguish between impressions of causality and impressions of realism, or, in other words, between impressions of general and mechanical causation.

Clearly, our data do not suffice to decide between the competing theories. We note, however, that each account has quite different implications for theories of perceptual causation. On the view that the launch leads to distinguishing perceptual from inferential judgments of causation, the findings justify the standard repeated measurement procedure, as perceptual causality is not reported from the beginning of the study. From this point of view, the present results highlight that while a Michottean module may help with spontaneous identification of core prototypical instances of causality, this does not rule out other forms of equally spontaneous causal impressions for events that do not trigger the module. These are, as we have argued, equally important for causal learning.

If, however, the launching sequence triggers judgements of realism, then in previous studies that interspersed canonical collisions among deviant ones, participants made judgements of realism while experimenters assumed they were capturing impressions of causation. This resulted in theories of causation from perceptual data that are far too narrow, building-in a requirement for such realism. Our results indicate that impressions of realism and impressions of causation do not coincide, and thus, special care should be taken to verify that participants are indeed responding only to the latter.

To conclude, in four studies we found that it takes a single observation for people to form strong causal impressions about unrealistic and novel events. This is not what most theories of causal thinking predict but it is what one should expect if humans are prolific causal discoverers who need to adapt rapidly to changing environments. Clearly, many such initial impressions will turn out to be false alarms, so the first induction (a hypothesis from observation), needs to be followed by more critical causal reasoning. Thus, if you actually see your house becoming an English bus after a rock hits it, our data suggests that you would, in fact, be dumbfounded at the apparent causality – even if you would also begin to search for alternative explanations. The idea that we can perceive causality from perceptual cues, not just prior knowledge, has continuing appeal, and it is time to extend the perspective to study how phenomenal causality is a tool for learning.

References

Bae, G. Y., & Flombaum, J. I. (2011). Amodal causal capture in the tunnel effect. *Perception*, *40*, 74–90. http://doi.org/10.1068/p6836

Beasley, N. A. (1968). The extent of individual differences in the perception of causality. *Canadian Journal of Psychology*, *22*(5), 399–407. Retrieved from http://psycnet.apa.org/journals/cep/22/5/399/

Bechlivanidis, C., & Lagnado, D. A. (2016). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition*, *146*, 58–66. http://doi.org/10.1016/j.cognition.2015.09.001

Blakemore, S.-J., Fonlupt, P., Pachot, M., Darmon, C., Boyer, P., Meltzoff, A. N., Segebarth, C., Decety, J. (2001). How the brain perceives causality: an event-related fMRI study. *Neuroreport, 12(17)*, 3741–3746. https://doi.org/10.1097/00001756-200112040-00027

Bell, J. S. (1964). On the Einstein Podolsky Rosen Paradox. *Physics*. http://doi.org/10.1002/prop.19800281202

Brown, H. V., & Miles, T. R. (1969). Prior stimulation and the perception of causality. *The Quarterly Journal of Experimental Psychology*, *21*(2), 134–136. http://doi.org/10.1080/14640746908400205

Butterfill, S. (2009). Seeing causings and hearing gestures. *The Philosophical Quarterly*, *59*(236), 405–428. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9213.2008.585.x/full

Chang, H., & Cartwright, N. (1993). Causality and realism in the EPR experiment. *Erkenntnis*, *38*(2), 169–190. http://doi.org/10.1007/BF01128978

Cheng, P. W. (1997). From covariation to causation: A causal power theory. Psychological Review, 104(2), 367–405. http://doi.org/10.1037/0033-295X.104.2.367

Choi, H., & Scholl, B. J. (2006). Measuring causal perception: connections to representational momentum? *Acta Psychologica*, *123*(1–2), 91–111. http://doi.org/10.1016/j.actpsy.2006.06.001

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development*, *21*(4), 713–731. http://doi.org/10.1016/S0163-6383(98)90040-6

Cohen, L. B., Amsel, G., Redford, M. A., & Casasola, M. (1998). The Development of Infant Causal Perception. In A. Slater (Ed.), Perceptual development: Visual, Auditory and Speech perception in Infancy (pp. 167–209). Hove, England: Psychology Press. Retrieved from https://spl.uoregon.edu/wp-content/uploads/2011/05/LBC-GA-MAR-MC_InfantCog_98.pdf

Danks, D. (2017). Singular Causation. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 201–215). Oxford: Oxford University Press.

Dittrich, W. H., & Lea, S. E. G. (1994). Visual Perception of Intentional Motion. *Perception*, *23*(3), 253–268. http://doi.org/10.1068/p230253

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*(5). http://doi.org/10.1037/h0093599

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., & Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive Brain Research*, *24*(1), 41–47. http://doi.org/10.1016/j.cogbrainres.2004.12.001

Gruber, H. E., Fink, C. D., & Damm, V. (1957). Effects of experience on perception of causality. *Journal of Experimental Psychology*, *53*(2), 89–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13406186

Guski, R., & Troje, N. F. (2003). Audiovisual phenomenal causality. *Perception & Psychophysics*, *65*(5), 789–800. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/12956586
Hamrick, J. B., Battaglia, P. W., & Tenenbaum, J. B. (2011). Internal physics models guide probabilistic judgments about object dynamics. In L. Carlson, C. Holscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society Austin, TX: Cognitive Science Society* (pp. 1545–1550). Austin, TX: Cognitive Science Society. Retrieved from http://palm.mindmodeling.org/cogsci2011/papers/0350/paper0350.pdf
Hubbard, T. L. (2013). Phenomenal Causality I: Varieties and Variables. *Axiomathes*, *23*(1), 1–42. http://doi.org/10.1007/s10516-012-9198-8
Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.
Kanizsa, G., & Vicario, G. (1968). The perception of intentional reaction. In G. Kanizsa & G. Vicario (Eds.), Experimental Research on Perception (pp. 71–126). Trieste: University of Trieste.
Kant, I. (1781). *Critique of pure reason.* London: MacMillan.
Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and Constraints in Causal Perception. *Psychological Science*, 95679761771993. http://doi.org/10.1177/0956797617719930
Lake, B. M., Salakhutdinov, R., Gross, J., & Tenenbaum, J. B. (2011). One shot learning of simple visual concepts. In *Proceedings of the Cognitive Science Society* (Vol. 33). Retrieved from http://escholarship.org/uc/item/4ht821jx
Leslie, A. M. (1986). Getting Development off the Ground: Modularity and the infant's perception of causality. In *Theory Building in Developmental Psychology* (pp. 405–437). Elsevier Science Publishers B.V.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*, 265–288. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010027787800069
Leslie, A. M. (1988). The necessity of illusion. In L. Weiskrantz (Ed.), Thought without language (pp. 185–210). New York: Clarendon Press/Oxford University Press. Retrieved from https://ruccs.rutgers.edu/images/personal-alan-leslie/publications/Leslie 1988.pdf
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. http://doi.org/10.1111/desc.12018
McCloskey, M. (1983). Intuitive physics. *Scientific American, 284*(4), 122–130. Retrieved from http://www.iub.edu/~koertge/H205SciReas/McCloskey_IntuitivePhysics.pdf
Michotte, A. (1963). *The Perception of Causality*. London: Methuen & Co Ltd.
Moors, P., Wagemans, J., & De-Wit, L. (2017). Causal events enter awareness faster than non-causal events. *Peerj*, *5*(e2932). http://doi.org/10.7717/peerj.2932
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. http://doi.org/10.1016/j.cogpsych.2008.02.003
Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology, 30(6)*, 869–879. Retrieved from http://psycnet.apa.org/journals/dev/30/6/869/
Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193–207. http://doi.org/10.1016/0885-2014(90)90026-P
Powesland, P. F. (1959). The effect of practice upon the perception of causality. *Canadian Journal of Psychology*, *13*(3), 155–168. Retrieved from http://psycnet.apa.org/journals/cep/13/3/155/

Rips, L. J. (2011). Causation From Perception. *Perspectives on Psychological Science*, *6*(1), 77–97. http://doi.org/10.1177/1745691610393525
Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology*, *23*(3), 250–254. http://doi.org/10.1016/j.cub.2012.12.017
Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., & Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain. *Neuropsychology*, *19*(5), 591–602. http://doi.org/10.1037/0894-4105.19.5.591
Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), 1–77. http://doi.org/10.1037/a0031912
Schlottmann, A. (1999). Seeing it happen and knowing how it works: how children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology*, *35*(1), 303–317. https://doi.org/10.1037/0012-1649.35.1.303
Schlottmann, A., Allen, D., Linderoth, C., & Hesketh, S. (2002). Perceptual causality in children. *Child Development*, *73*(6), 1656–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12487485
Schlottmann, A., & Anderson, N. H. (1993). An information integration approach to phenomenal causality. *Memory & Cognition*, *21*(6), 785–801. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8289656
Schlottmann, A., Cole, K., Watts, R., & White, M. (2013). Domain-specific perceptual causality in children depends on the spatio-temporal configuration, not motion onset. *Frontiers in Psychology, 4, 365*. http://doi.org/10.3389/fpsyg.2013.00365
Schlottmann, A., Ray, E. D., Demetriou, N., & Mitchell, A. (2006). Perceived physical and social causality in animated motions: spontaneous reports and ratings. *Acta Psychologica*, *123*(1–2), 112–143. http://doi.org/10.1016/j.actpsy.2006.05.006
Schlottmann, A., Ray, E. D., & Surian, L. (2012). Emerging perception of causality in action-and-reaction sequences from 4 to 6months of age: Is it domain-specific? *Journal of Experimental Child Psychology*, *112*(2), 208–230. http://doi.org/10.1016/j.jecp.2011.10.011
Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *The Quarterly Journal of Experimental Psychology*, *44*(2), 321–342. Retrieved from http://www.tandfonline.com/doi/full/10.1080/02724989243000055
Schlottmann, A., & Surian, L. (1999). Do 9-month-olds perceive causation-at-a-distance? *Perception*, *28*(9), 1105–1113. http://doi.org/10.1068/p2767
Schlottmann, A., Surian, L., & Ray, E. D. (2009). Causal perception of action-and-reaction sequences in 8- to 10-month-olds. *Journal of Experimental Child Psychology, 103(1)*, 87–107. https://doi.org/10.1016/J.JECP.2008.09.003
Scholl, B. J., & Nakayama, K. (2002). Causal capture: contextual effects on the perception of collision events. *Psychological Science*, *13*(6), 493–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12430831
Scholl, B. J., & Nakayama, K. (2004). Illusory causal crescents: misperceived spatial relations due to perceived causality. *Perception*, *33*(4), 455–469. http://doi.org/10.1068/p5172
Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. http://doi.org/10.1016/S1364-6613(00)01506-0
Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: the effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*, 759-775. https://doi.org/10.1037/a0025140
Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgement. In G. H.

Bower (Ed.), *Psychology of learning and motivation-advances in research and theory Volume 21* (pp. 229 – 261). San Diego, CA: Academic Press.
Straube, B., & Chatterjee, A. (2010). Space and time in perceptual causality. *Frontiers in Human Neuroscience, 4,* 1–10. http://doi.org/10.3389/fnhum.2010.00028
Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-Based Causal Inference. In S. Becker, S. Thrun, & K. Obermaye (Eds.), *Advances in neural information processing systems (Vol. 15)* (pp. 35–42). Cambridge: MIT Press. Retrieved from http://papers.nips.cc/paper/2332-theory-based-causal-inference.pdf
Tinbergen, N., & Perdeck, A. C. (1950). On the Stimulus Situation Releasing the Begging Response in the Newly Hatched Herring Gull Chick (Larus Argentatus Argentatus Pont.). *Behaviour (Vol. 3)*. Retrieved from https://www.jstor.org/stable/pdf/4532715.pdf?refreqid=excelsior%3Aa51de3efc3bcc5d055e95b9d3e236c6c
Wagemans, J., van Lier, R., & Scholl, B. J. (2006). Introduction to Michotte's heritage in perception and cognition research. *Acta Psychologica*, *123*(1–2), 1–19. http://doi.org/10.1016/j.actpsy.2006.06.003
Waldmann, M. R. (Ed.). (2017). *The Oxford handbook of causal reasoning*. New York: Oxford University Press.
Weir, S. (1978). The perception of motion: Michotte revisited. *Perception*, *7*(3), 247–260. http://doi.org/10.1068/p070247
White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 710–727. http://doi.org/10.1037/0278-7393.29.4.710
White, P. A. (2006). The role of activity in visual impressions of causality. *Acta Psychologica*, *123*(1–2), 166–85. http://doi.org/10.1016/j.actpsy.2006.05.002
White, P. A. (2009). Property transmission: an explanatory account of the role of similarity information in causal inference. *Psychological Bulletin, 135(5)*, 774–793. http://doi.org/10.1037/a0016970
White, P. A. (2012). Visual impressions of causality: Effects of manipulating the direction of the target object's motion in a collision event. *Visual Cognition, 20(2)*, 121–142. http://doi.org/10.1080/13506285.2011.653418
White, P. A. (2014). Perceived causality and perceived force: Same or different? *Visual Cognition*, *22*(5), 672–703. http://doi.org/10.1080/13506285.2014.911234
White, P. A. (2014). Singular clues to causality and their use in human causal judgment. *Cognitive Science, 38(1)*, 38–75. https://doi.org/10.1111/cogs.12075
White, P. A. (2016). Visual impressions of generative transmission. *Visual Cognition*, *6285*(March), 1–37. http://doi.org/10.1080/13506285.2016.1149533
White, P. A. (2017). Perceptual impressions of causality are affected by common fate. *Psychological Research*, pp. 1–13. http://doi.org/10.1007/s00426-017-0853-y
White, P. A., & Milne, A. (1999). Impressions of enforced disintegration and bursting in the visual perception of collision events. *Journal of Experimental Psychology: General*, *128*(4), 499–516. Retrieved from http://psycnet.apa.org/journals/xge/128/4/499/
Woods, A. J., Lehet, M., & Chatterjee, A. (2012). Context modulates the contribution of time and space in causal inference. *Frontiers in Psychology*, *3*(October), 371. http://doi.org/10.3389/fpsyg.2012.00371
Yela, M. (1952). Phenomenal causation at a distance. *Quarterly Journal of Experimental Psychology*, *4*(4), 139–154. http://doi.org/10.1080/17470215208416612
Yilmaz, M., & Meister, M. (2013). Rapid Innate Defensive Responses of Mice to Looming Visual Stimuli. *Current Biology*, *23*(20), 2011–2015. http://doi.org/10.1016/J.CUB.2013.08.015

Young, M. E., & Falmier, O. (2008). Color change as a causal agent revisited. *American Journal of Psychology*, 121(1), 129–156. http://doi.org/10.2307/20445447

Appendix

The table below summarizes the available data regarding causal judgements in Michottean-like collisions that feature a temporal delay between the first object stopping and the second object starting to move, a positive or negative (pass) spatial gap between them, or an angle of reflection between the first object's incoming direction and the second object's outgoing direction.

Authors	Year	Condition	Results
Sanborn et al.	2013	Gap – 4mm	Probability of reporting causality close to 0
Fugelsang et al	2005	Gap -12mm	10.4% of participants reported a causal impression
Michotte	1963 (exp.31)	Gap-20mm	No impression of causality
Yela	1952	Gap-50mm	44% of participants reported a causal impression
Yela	1952	Gap-90mm	28% of participants reported a causal impression
Schlottmann & Anderson	1993	Gap-2.1mm	Mean causal rating about 150/300
Schlottmann et al	2006	Gap-30mm	Mean rating was 0.083 (0 was non causal and 1 was causal impression, -1 was an impression of social causality)
Straube & Chatterjee	2010	Delay-164.67ms (mean)	Sequences judged as non-causal
Fugelsang et al	2005	Delay-170ms	4.2% of participants reported a causal impression
Yela	1952	Delay-167ms	30% of participants reported a causal impression
Sanborn et al	2013	Delay-250ms	Probability of reporting causality close to 0
Schlottmann & Anderson	1993	Delay-170ms	Mean causal rating about 150/300
Schlottmann et al	2006	Delay-1300ms	Mean rating was -0.42 (0 was non causal and 1 was causal impression, -1 was an impression of social causality)

White	2014	Delay-120ms	Mean causal rating 45%
Michotte	1963	Delay-200ms or more	No causal impression
White	2012	Angle-20^{o}	Mean causal rating 60%
Michotte	1963	Angle-25^{o}	Impression considerably weakened
Straube & Chatterjee	2010	Angle-31.53^{o} (mean)	Sequences judged as non-causal
White	2012	Angle-40^{o}	Mean causal rating 49%
White	2012	Angle-60^{o}	Mean causal rating 37%
White	2012	Angle-80^{o}	Mean causal rating 29%
Michotte	1963	Angle-90^{o}	Causal impression disappears
Beasley	1968	Angle-90 ^{o}	45% of participants reported a causal impression
Scholl & Nakayama	2002	Pass	10.7% of trials were perceived as causal launches