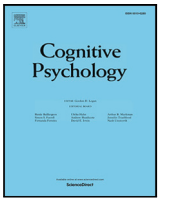
Cognitive Psychology 165 (2026) 101814

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych



Seeing, hearing, and feeling causation

Elyse D.Z. Chase [a,*], Kevin A. Smith [b], Sean Follmer [c], Tobias Gerstenberg [d]

[a] Department of Mechanical Engineering, Rice University, United States
[b] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States
[c] Department of Mechanical Engineering, Stanford University, United States
[d] Department of Psychology, Stanford University, United States

ARTICLE INFO

Keywords:
Causal perception
Causal inference
Multimodal integration
Haptics
Touch

ABSTRACT

How do people decide whether one event caused another? While previous research has focused on visual and auditory cues in causal perception, the role of touch remains underexplored. Here, we investigate how haptic feedback contributes to causal judgments across three psychophysical experiments. In Experiment 1, we introduced force-based haptic feedback to a visual launching paradigm and found that haptic information increased causal judgments compared to visual feedback alone. Experiment 2 combined vision, audio, force-based haptics, and vibrotactile haptics, revealing that additional sensory cues increase causal judgments with diminishing returns — the largest benefit comes from adding a second modality. In Experiment 3, we examined how both the number and physical realism of multisensory cues affect causal perception, finding that both factors boosted causal judgments. We present a Bayesian multimodal inference model that captures human judgments by integrating visual, auditory, and haptic information based on their relative timing, uncertainty, and realism. Taken together, these experiments show that haptic information contributes to causal judgments by shaping the multisensory evidence observers use when deciding whether one event physically caused another, including how realistic and physically coherent the event appears. More broadly, we find that temporal alignment and cross-modal coherence are key constraints for how multisensory evidence shapes causal judgments, with implications for virtual reality, robotics, and human–computer interaction systems.

1. Introduction

How do we determine whether one event caused another? This fundamental question underlies our ability to navigate and understand the physical world, from recognizing that a collision caused a ball to move to inferring that pressing a button activated a machine. While this process seems effortless, it involves sophisticated information integration from multiple sensory modalities operating under uncertainty. We rely on various senses to perceive and interpret our surroundings — we see, hear, smell, taste, and feel the world around us. Haptic feedback is particularly important for successfully interacting with the world. Indeed, some have argued that experiencing forces through touch grounds infants' understanding of causality (White, 1988, 1990; Wolff, 2007; Wolff & Thorstad, 2017). Children learn about the causal structure of the world by hitting objects into each other and examining how they move, combining haptic feedback with visual and auditory information (Baillargeon et al., 1995; Spelke, 2013; White, 2012).

Despite the fundamental role of touch in causal understanding, this sensory modality remains severely understudied in causal perception research. This gap has significant theoretical and practical implications. Theoretically, understanding multisensory causal

* Corresponding author.
E-mail address: ec100@rice.edu (E.D.Z. Chase).

https://doi.org/10.1016/j.cogpsych.2026.101814
Received 22 August 2025; Received in revised form 2 May 2026; Accepted 5 May 2026

Available online 17 June 2026

perception is crucial for developing complete models of human cognition. Practically, as we increasingly interact with virtual and augmented reality systems, robots, and other technologies that provide haptic feedback, we need to understand how to design these systems to support accurate causal perception.

Even though haptic experiences may be fundamental to the development of causal knowledge, little work has examined how haptic feedback affects causal perception and inference. Most research has focused on the role of visual evidence in causal perception (Bechlivanidis & Lagnado, 2013, 2016; Bechlivanidis et al., 2019; Choi & Scholl, 2004, 2006; Michotte, 1946/1963), while some studies have explored the role of auditory evidence as well (Agrawal & Schachner, 2022; Gerstenberg et al., 2018; Guski & Troje, 2003; Outa et al., 2022; Sekuler et al., 1997). Arguably, this imbalance says less about the importance of the different senses for causal inference and more about the fact that studying vision and sound empirically is easier than studying haptic feedback — computers come with screens and speakers but not with haptic sensors.

A notable exception is Wolff and Shepard (2013), who explored whether perceiving causation makes people more sensitive to feeling a force. In their experiment, participants viewed causal and non-causal scenarios (such as a marble rolling on a kitchen counter and impacting another marble or moving across the table at the same rate without impacting another marble). They felt a force through a haptic controller at the end of the clip. Participants' task was to indicate when they felt the force. Causal perception affected response times: participants responded more quickly to the haptic feedback in causal scenarios than in non-causal ones. In control conditions, participants received visual or auditory stimuli (rather than a force). The response times to an audio cue or a visual flash did not differ based on whether they had viewed a causal or non-causal scenario. Wolff and Shepard argued that seeing causation primes the feeling of force. In this study, we examine how force and other sensory information affect causal perception, rather than how causal perception affects the feeling of a force.

Outside the lab, designers of video games, VR systems, and consumer devices routinely exploit haptic feedback to make virtual events feel causal, using controllers that rumble when a collision occurs and force-feedback steering wheels to make virtual events feel real (Israr et al., 2012; Israr & Poupyrev, 2011). There is also substantial academic literature on haptic rendering, delay (Di Luca & Mahnan, 2019; Vogels, 2004), and realism (Shin & Choi, 2018; Srinivasan et al., 1996), including work on contact-event vibration models (Okamura et al., 2002; Park & Choi, 2016), event-based haptic feedback (Kuchenbecker et al., 2006), and multimodal haptic rendering (Park et al., 2019). However, much of that literature has focused on perceived realism, synchrony, or interaction quality more generally, rather than on the more specific question of how realism in haptic cues contributes to judgments of physical causality. Our goal here is not to advance haptic rendering itself, but to use haptic cues as controlled sensory signals for studying multisensory integration in causal perception.

What do we know about how people combine information from different senses to make causal inferences? (Ernst & Banks, 2002) show that people optimally integrate visual and haptic information when judging how tall an object is. Participants selected the taller of two objects under visual, haptic, and visual-haptic conditions with varying levels of visual uncertainty. Participants benefited from access to both modalities, particularly when the visual signal was highly uncertain. People's inferences were captured well by a Bayesian account that weights each source of evidence based on its associated uncertainty. Körding et al. (2007) looked at how people combine visual and auditory evidence in a spatial location task. Participants had to locate the source of an auditory or visual signal based on both seeing a light and hearing a sound. When participants had to locate a sound, their judgments were strongly affected by the light source. In contrast, when participants had to locate the light, the sound source had little effect on their judgments. To capture these judgments, Körding et al. (2007) developed a hierarchical Bayesian model that infers whether the two sources of evidence stem from the same common cause or are independent from one another. This model captures that people's inferences about the spatial location are affected more strongly by the evidence they have less uncertainty about. To figure out where something came from, visual evidence is more diagnostic than auditory evidence.

We approach multisensory causal perception through the lens of Bayesian inference, which provides a principled framework for understanding how humans combine uncertain evidence from multiple sources. In this framework, people maintain beliefs about whether events are causally related and update these beliefs based on sensory evidence, weighing each source according to its reliability and relevance. This approach is motivated by evidence that humans integrate multisensory information in approximately optimal ways (Ernst & Banks, 2002; Körding et al., 2007). However, prior work has largely focused on inferences based on evidence from only two modalities (e.g., vision and haptics, or vision and sound), generally finding that multimodal inferences are more accurate than unimodal ones. Critical questions therefore remain: Are three or four sources better than two? Do people integrate information across more than two modalities in the same way? And how does the realism of sensory cues shape this process?

We build on this foundation to explore how people make causal inferences based on evidence across up to four modalities: vision, audition, kinesthetic haptics, and vibrotactile haptics. We conducted three psychophysical experiments that systematically manipulated the available information sources and the realism of each signal. We also developed and tested a Bayesian computational model that integrates information from all these sources, accounting for their relative timing, uncertainty, and physical realism. Our work makes three key contributions:

1. **Haptic Evidence Affects Causal Judgments**: We test whether haptic information changes how people judge causality in a launching paradigm. Experiment 1 shows that kinesthetic feedback systematically shifts causal judgments relative to vision alone, establishing touch as an important cue for judging causation.
2. **Constraints on Multisensory Integration**: We examine when additional cues help. Experiments 2 and 3 show that the key contrast lies between unimodal and multisensory evidence, and that the contribution of added cues depends on their temporal alignment and coherence with the event.

3. **Role of Realism & Computational Explanation**: We explore whether physically realistic cues strengthen causal judgments, and we introduce a Bayesian model that explains these effects in terms of uncertainty, temporal information, and realism-sensitive weighting of evidence.

Our results show that people use multisensory evidence – including haptic evidence – when judging whether one object caused another to move. Additional cues increase causal judgments most strongly when moving from a single modality to multisensory modalities, and this benefit is shaped by temporal alignment and how coherent the cues are with a physically plausible collision.

The remainder of this paper is organized as follows: We begin by reviewing the relevant literature on causal perception, with particular attention to the role of visual and auditory feedback, while highlighting the gap in research on kinesthetic and vibrotactile feedback. We also examine prior work on the role of physical realism in causal perception. We then present three experiments. Experiment 1 tests the effects of kinesthetic haptic feedback on causal perception. Experiment 2 augments the setup of Experiment 1 with vibration and audio to explore up to four sensory signals at once (vision, audio, kinesthetic haptics, and vibrotactile haptics). Experiment 3 focuses on realism and explores the combination of visual, auditory, and vibrotactile cues. We show how a Bayesian model of causal perception based on multimodal evidence can account for the experimental results. In the General Discussion, we highlight key findings, limitations, and directions for future research.

1.1. Causal perception

Michotte (1946/1963) established the foundation for studying causal perception through the canonical "launching" paradigm. In these displays, one disc contacts another, stops moving, and the second disc begins to move. Systematic manipulation of temporal dynamics (delays between contact and movement) and spatial dynamics (gaps between objects) revealed that both factors critically influence causal perception (Cravo et al., 2009; Scholl & Tremoulet, 2000; White & Milne, 1997).

Importantly, causal perception is flexible and context-dependent. Work by Bechlivanidis et al. (2019) demonstrated that people sometimes judge physically unrealistic event sequences as causal, including those with temporal delays or spatial gaps between objects. However, after participants view a more canonical causal collision event, they subsequently judge these same unrealistic events as less causal. This suggests that participants initially adopt a broader view of causation that includes physically unrealistic causes. After exposure to physically realistic events, they adopt a narrower view of physical causation. These results highlight the importance of context and the flexibility of people's definitions of causality across varying levels of realism.

1.2. Visual and auditory integration in causal perception

The integration of visual and auditory information significantly enhances causal perception. Sekuler et al. (1997) demonstrated that sound can disambiguate visually ambiguous stimuli: without sound, discs appear to pass through each other, but with sound, they appear to bounce. This principle extends to causal inference more broadly—(Gerstenberg et al., 2018) showed that people integrate visual and auditory cues to infer past events (see also Schachner & Kim, 2018; Wu et al., 2024), while other work shows these modalities help determine *when* and *where* events occurred (Körding et al., 2007; Shams & Beierholm, 2010). Guski and Troje (2003) provided the most direct evidence for audiovisual integration in causal perception. Adding audio cues during launching events increased causal judgments. Critically, the specific type of additional signal mattered less than having multiple co-occurring events — even color changes produced similar benefits. This suggests that multisensory causal perception operates on a principle of convergent evidence: multiple simultaneous signals increase the probability that a single causal event occurred.

1.3. Kinesthetic and vibrotactile integration in causal perception

While haptic information is essential to how we interact with the world, it has not been studied as much as other senses. Haptic information is more difficult to manipulate experimentally. Nonetheless, many researchers consider force and touch-based information integral to how people learn about the world (White, 1988, 2012). For example, Wolff (2007) and Wolff and Thorstad (2017) argue that our perception of causality is grounded in our perception of forces. Many creators of haptic devices strive to build systems that emulate the feeling of interacting with the real world. The goal is to present vibrational cues that feel like an impact and forces that emulate real-world contact forces. Okamura et al. (2002) evaluated vibration to improve contact realism and determined that reality-based vibration models enhance the perceived realism of virtual contacts. Shin and Choi (2018) considered how force and vibration could be used independently to recreate realistic textures. They find that force feedback delivers more realistic texture cues than vibration. However, neither is sufficient to mirror the feedback humans experience in real interactions. Additionally, visual information can drastically affect haptic perception and higher-level perceptual judgments, such as how stiff an object is Chase and O'Malley (2024), Chase et al. (2025) and Srinivasan et al. (1996).

1.4. Realism in causal perception

Physical realism matters for how people judge causation (Bechlivanidis et al., 2019; Kominsky & Scholl, 2020; Kominsky et al., 2017). In the real world, hearing a "buzz" would be less indicative that a collision happened than hearing a "crash". When engaging with their environment, people observe kinematic (motion) information about how objects interact. This includes details like relative timing, trajectories, and changes in velocity. People then compare these observations against their previous experiences with forces and causal patterns to make judgments about cause, force magnitude, and agency (Bechlivanidis et al., 2019; White, 1988, 2009, 2012). If we compare current events to prior interactions, the more realistic the cues are, the more likely they will align with that prior experience. White (2012) argues that people use a combination of perceived forces (from our mechanoreceptors) and a model of action (from our motor control system). He argued that humans use stored information to make future predictions and to run mental simulations for how objects interact.

Researchers have also studied how visual rendering and physical dynamics affect people's judgments about collision events. In Meding et al. (2020), participants experienced three conditions: 2D rendering with constant velocity motion and elastic collision, 3D rendering with constant velocity rolling and elastic collision, and 3D rendering with rolling and inelastic collision. They found that a more realistic 3D rendering (balls that rotated and underwent elastic collisions) yielded lower causal ratings than a 2D rendering. However, adding physical dynamics (rotation and inelastic collision) to a realistic 3D rendering increased causal ratings compared to the other conditions. Thus, people use the realism of cues to make causal judgments beyond just the temporal and spatial information within a cue. Berger et al. (2018) considered the interaction of realism in visual and haptic cues — discovering an uncanny valley. Participants used a VR system with vibrotactile stimuli delivered via two controllers to produce spatial haptic effects. Participants rated their sense of presence in VR, which decreased as the realism of the haptic cues exceeded that of the visual feedback. Overall, this demonstrates the importance of considering the realism of cues in multisensory environments and motivates us to explore further how they affect causal perception.

1.5. Haptic rendering

Within haptics research, a substantial body of work has examined how contact events should be rendered to feel realistic. Reality-based vibration models improve the perceived realism of virtual contacts, and event-based approaches leveraging physics-based cues to provide impact transients at contact further enhance contact realism (Kuchenbecker et al., 2006; Okamura et al., 2002; Park & Choi, 2016). Other work has compared force and vibration as distinct channels or combined them into multimodal feedback for conveying contact properties, showing that they contribute differently to perceived realism (Park et al., 2019; Shin & Choi, 2018). The literature has also shown that people are highly sensitive to visual-haptic delay, highlighting the importance of temporal alignment for coherent multisensory interactions (Di Luca & Mahnan, 2019; Vogels, 2004). This prior work is directly relevant to our stimulus design, but its primary focus has generally been realism, synchrony, or interaction quality. By contrast, our goal is to use haptic cues as experimentally controlled signals to ask how touch contributes to judgments of physical causality based on multisensory evidence. In this sense, our work is complementary to the haptics literature: rather than proposing a novel technique, we test how kinesthetic and vibrotactile signals alter observers' judgments of whether one event caused another.

2. Overview of the experimental paradigm

Across experiments, our goal was to investigate short, multisensory contact events rather than full, naturalistic collisions. Accordingly, the stimuli were designed to isolate the temporally localized sensory evidence that accompanies contact – when one object appears to transfer motion to another – rather than to reproduce the full mechanics of impact, deformation, or rebound. Cue durations, therefore, emphasized the contact-related phase of the event, and amplitudes were tuned for clear but comfortable perception across modalities.

All our experiments investigate how people combine multiple sources of information to judge whether one event caused another. These sources include four modalities: vision, audition, vibration, and force (kinesthetic feedback). During all trials, participants viewed two objects on a computer screen, either trains (Fig. 1A, Experiments 1 and 2) or billiard balls (Fig. 1B, Experiment 3). At the beginning of each trial in Experiments 1 and 2, the blue train moves toward the red train. The red train begins to move at some point after that. After having seen the animation, we asked participants to select one of the following two statements:

(1) *The blue train caused the red train to move.*
(2) *The red train moved by itself.*

We manipulated the temporal offset between when the first train stopped and when the second train started moving. In the ideal *causal* case, the trains make momentary contact, with the first train stopping and the second train starting motion simultaneously. A *positive temporal offset* introduces a delay in the red train's movement, creating the appearance that the trains stick together temporarily. A *negative temporal offset* creates a visual gap where the second train begins moving before the first train stops.

Participants received haptic feedback in the form of a lateral force that temporally matched the second train's motion. The lateral, force-based cue was intended as a controlled proxy for the reaction force one might feel through a held object or interface during contact. Rather than a momentary pulse, the cue was sustained over the second object's movement so that participants experienced a clear kinesthetic correlate of the motion. For example, the short sustained pull or driven motion one might feel through a controller

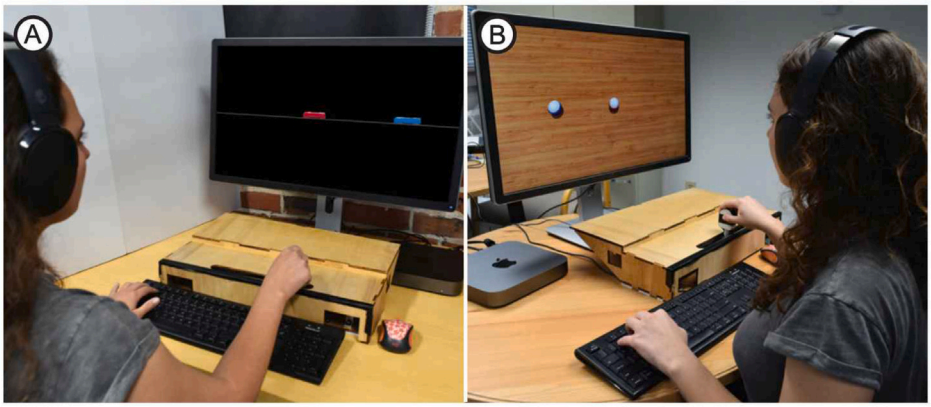


Fig. 1. A participant seated at the experimental setup with the haptic device in front of a computer screen. Participants grasp the device's handle, which provides kinesthetic and vibrotactile feedback while receiving audio cues via headphones. (A) Screen display for Experiments 1 & 2. (B) Screen display for Experiment 3.

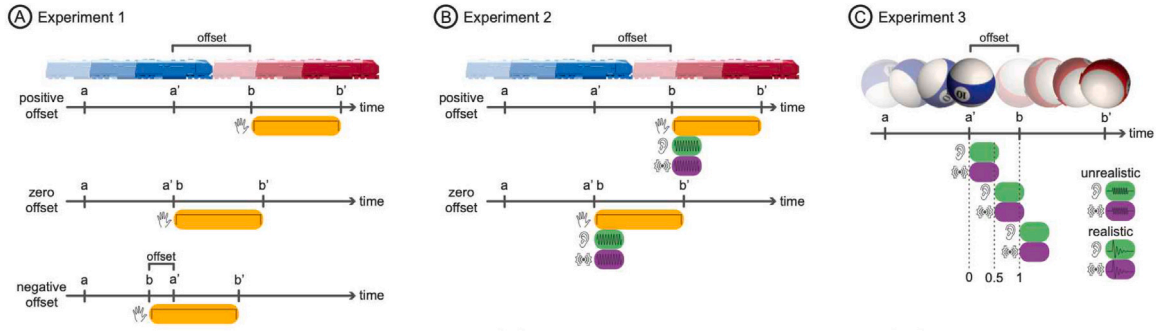


Fig. 2. Overview of experiments: (a) first object starts motion, (a') first object stops motion, (b) second object starts motion, (b') second object stops motion, and (offset) is the difference between a' and b. When additional sensory information is present, all cues begin between a' and b. (A) Experiment 1: Only kinesthetic haptic feedback was added at b and played for the second train's motion length. (B) Experiment 2: Audio and vibration cues were played in addition to the kinesthetic cue at b. Additionally, the length of each signal was reduced, as was the time of travel for each train. (C) Experiment 3: In Part A, audio and vibration cues were played at a' when present. In Part B, the audio and vibration cues were shifted to begin at three locations: a', b, and halfway between a' and b.

or handle during a game, or when someone taps or briefly tugs on an object as you hold onto it. We chose trains as stimuli because their engines and brakes make it plausible for them to start or stop at any moment. Participants could also experience vibrotactile feedback, similar to the momentary impact of real-world contact. Together, these haptic cues allow participants to draw on prior experience with short, hand-coupled contact events, extending previous work by incorporating two related but distinct types of haptic feedback: kinesthetic (force-based) and vibrotactile.

Our experiments test how multiple sources of sensory information affect causal perception. In Experiment 1, we used temporal offsets to compare vision and kinesthetic haptic feedback with vision alone. In Experiment 2, participants viewed events with the addition of audio, kinesthetic, or vibrotactile cues. In both of these experiments, the cues were temporally aligned with the movement of the red train (see Fig. 2A and B). In Experiment 3, we manipulated the timing of the sensory cues. Their onset was aligned with the first billiard ball stopping, with the second one moving, or halfway between these two times (Fig. 2C). We synchronized non-visual cues to event timings most consistent with the percept being represented. In Experiments 1–2, haptic cues were synchronized with the second object's motion. We were interested in avoiding additional timing cues for the stopping of the first object and instead having the user feel forces and move with the second object during the interaction. In Experiment 3, where visual collisions were physically realistic, participants strongly expected cues at the moment of contact, consistent with pilot testing and prior work on event-based haptics.

Testing haptic devices requires in-person studies with custom hardware and thus usually has fewer participants than online studies. Additionally, participant fatigue is a concern for longer studies. Therefore, we limited the number of trials each participant experienced and provided ample breaks to ensure participants could complete the tasks attentively. All materials, data, and code are available at our GitHubrepository.

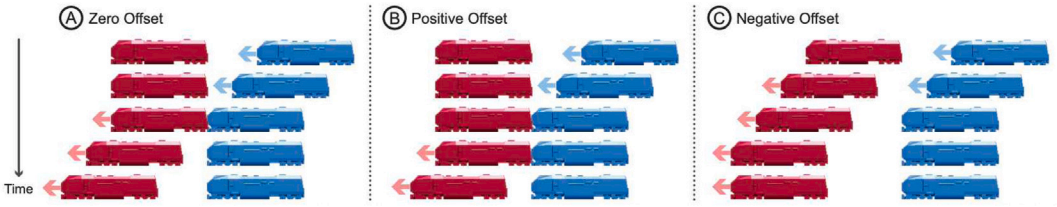


Fig. 3. Demonstration of the three different scenarios described for two trains. (A) The causal case is when the trains make contact momentarily, and the second moves away. (B) In the delayed case (positive temporal offset), the trains stick together for some time before the second train moves away. (C) In the spatial gap case (negative temporal offset), the second train moves away before the first train contacts it.

2.1. Hardware & setup

Across all experiments, participants interacted with the same system – a custom one-degree-of-freedom (1-DoF) haptic device, computer screen, keyboard, and headphones (Fig. 1). The haptic device was constrained to move in only one direction with a constant force of 3 Newtons (which matched the motion of the trains on screen) via a linear capstan drive and impedance-based control (Chase & Follmer, 2019). Additionally, we modified the handle to include a voice coil motor that provided controlled vibrotactile feedback. In Experiment 2, vibrotactile cues were 250 ms in duration, while in Experiment 3 they were 150 ms — to more closely match the length of signals we experience daily. We implemented the experimental designs using CHAI3D (Conti et al., 2005), which affords fast refresh rates necessary for haptic interactions. The monitor (Dell P2715Q 27″, 3840 × 2160) displayed the scene while the haptic device was secured to the table in front of the monitor. Participants listened to white noise during the experiment to mask unintentional sounds produced by the hardware. In Experiment 1, participants wore earplugs while white noise was played aloud in the room. For Experiments 2 and 3, headphones played white noise in addition to the audio signals. This change was due to the inclusion of audio signals in Experiments 2 and 3 that were not present in Experiment 1.

The kinesthetic device is a 1-DoF handle with closed-loop force control capable of delivering precise lateral contact-like forces. Participants held the handle with a light precision grip and were instructed that the device would move/apply force independently during each trial. We used a constant-force cue, not because it closely matches the physics of a moving object, but because it provides a reproducible, controlled kinesthetic signal across participants with different arm impedance and grip stiffness. Thus, using constant force rather than constant velocity ensures that all participants experience the same sensation during each trial, regardless of how tightly or loosely they hold the device. In this sense, the force cue should be understood as a controlled approximation of the force required to set the second object in motion after contact, rather than a literal replay of an object's trajectory. The closest real-world analogs are hand-coupled forces transmitted through a held interface or other rigid intermediary. To help visualize the device and kinesthetic feedback, we provide example trial videos for experiments in our code repository.

In our studies, we chose to synchronize the haptic feedback with the movement of the second object. In principle, we could also have synced the feedback with the first object, or used two devices: one synced with the first object (held by the left hand) and one with the second object (held by the right hand). Our model (discussed below) predicts no difference in the haptic signals associated with the first and second objects. However, we know from prior research that there is an important asymmetry in how people perceive launching events. They generally perceive the first object as having launched the second, rather than perceiving the second object as having stopped the first (Mayrhofer & Waldmann, 2016; White, 2009). Thus, participants' judgments may differ depending on whether their haptic feedback is associated with the "agent" or the "patient" in the causal interaction. We therefore chose to render the haptic cue on the side of the effect (the second object's movement), because that is the part of the event whose causal status participants were judging. This choice kept the task focused on whether a controlled kinesthetic signal synchronized with the second object's motion could serve as evidence that the second object's motion was caused. By contrast, associating the force cue only with the first object would emphasize the initiating action rather than the effect that participants were evaluating. Using both objects would introduce a more complex bimanual mapping, in which each hand could correspond to a different part of the event, thereby changing the question being tested. This is not the only ecologically valid mapping, of course. One could imagine completing the same task, but instead being in sync with the first train rather than the second.

To help visualize the device and kinesthetic feedback, we provide example videos in our code repository.

3. Experiments

3.1. Experiment 1: Vision & kinesthetic haptics

To explore how sensory information affects the perception of causal events, we investigated kinesthetic haptic feedback. This sensory mode can provide diagnostic information for collisions and has been understudied. This experiment was based on Michotte's seminal launching paradigm, presenting both the positive offset (delay Fig. 3B) and negative offset (spatial gap Fig. 3C) with vision alone and then vision & kinesthetic haptics.

Since few previous works have explored haptic feedback in phenomenal causality, we drew on studies examining the role of auditory feedback. In their experiments on whether audio enhances causal impressions, Guski and Troje (2003) found that it did

so when it occurred within a synchronous temporal window with the objects' visual interactions. Intuitively, with each additional co-occurring event (i.e., the first object stopping motion, the second object starting motion, the sound playing), the chance of all events occuring simultaneously if the events are not causally related becomes increasingly unlikely. From this, we hypothesized that additional haptic feedback would make people more likely to judge that an event was causal compared to the condition where they only saw (but did not feel) what happened.

We pre-registered four hypotheses.[1] Three consider positive offset (delay) conditions:

(H1) *The probability of viewing an event as causal will decrease as the positive temporal offset increases.*
(H2) *The probability of viewing an event as causal for positive temporal offsets is greater under the 'vision & haptics' condition than the 'vision only' condition.*
(H3) *The probability of viewing an event as causal decreases more strongly as the positive offset increases for the 'vision only' condition compared to the 'vision & haptics' condition.*

In the negative temporal offset condition (spatial gap), the haptic feedback will add an additional cue indicating that the second train began motion before making contact. Thus, we have one hypothesis for negative offset conditions as follows:

(H4) *The difference in the probability of viewing an event as causal when comparing the 'vision & haptics' condition with the 'vision only' condition will be lower for negative offsets compared to positive offsets.*

3.1.1. Design & measures

Experiment 1 follows a within-subjects, repeated measures design with two factors: temporal offset and sensory modality ('vision only' vs. 'vision & haptics'). We blocked the experiment by sensory modality, using the method of constant stimuli (Jones & Tan, 2012) to estimate the temporal threshold at which people perceived the events to be causally connected. Each offset was repeated twelve times per block in a randomized order. We counterbalanced the order of blocks across participants.

Temporal offsets create either a delay (positive offset) or a visual gap (negative offset) (Fig. 1A). We used thirteen temporal offsets with equal separation between each group of offsets (Jones & Tan, 2012), including six positive offsets (100, 200, 300, 400, 500, 600 ms), six negative offsets (−84, −70, −56, −42, −28, −14 ms), and zero offset (0 ms). These offsets were selected based on pilot studies and prior work on launching paradigms (e.g., Bechlivanidis et al., 2019; Guski & Troje, 2003). The two sets of offsets were created to include causal judgments that spanned from clearly causal on one end to clearly non-causal on the other. The ranges (0 to −84 ms and 0 to 600 ms) span conditions, in both directions, that were almost always judged as causal, to conditions that were almost always judged as non-causal in pilot experiments. This limited range provides sufficient intermediate points to estimate the psychometric function while ensuring the experiment would not last too long. Haptic feedback, when present, was delivered to the right hand via a 1-DoF device that provides a constant force (3 N) synchronized with the motion of the red train (Fig. 1A).

3.1.2. Participants

A total of 18 right-handed participants (8 Female, 10 Male; age $\mu = 25.4$, $\sigma = 1.8$) were recruited and compensated for participating.

3.1.3. Confirmatory analysis

First, we considered hypotheses that used only the positive offset (delay) trials. We fit generalized linear mixed-effects models with a binomial linking function to the data using the `lme4` package in R (Bates et al., 2015). All models included a combination of fixed effects of offset and condition, and a random effect of participant to predict causal judgments. In the model notation below and throughout the paper, we use * to indicate that both the simple effects and the interaction between effects are included.

M1: causal judgment $\sim 1 + \text{offset} * \text{condition} + (1 \mid \text{participant})$
M2: causal judgment $\sim 1 + \text{offset} + \text{condition} + (1 \mid \text{participant})$
M3: causal judgment $\sim 1 + \text{offset} + (1 \mid \text{participant})$
M4: causal judgment $\sim 1 + \text{condition} + (1 \mid \text{participant})$

We used likelihood-ratio tests to assess our different hypotheses. For within-subjects paired t-tests, effect sizes are reported as Cohen's d_z, which standardizes the mean paired difference by the standard deviation of the difference scores.

H1: Temporal offset. As predicted, the probability of a causal response decreased as the positive temporal offset increased (see Fig. 4). M2 (AIC = 1854.5) provides a better fit to the data than M1 (AIC = 3641.9, $\chi^2(1) = 1813.6$, $p < 0.0001$). Within M2, there is a significant effect of offset ($\beta = -2.92$, $SE = 0.11$, $z = -26.06$, $p < 0.0001$)

[1] https://osf.io/58un9.

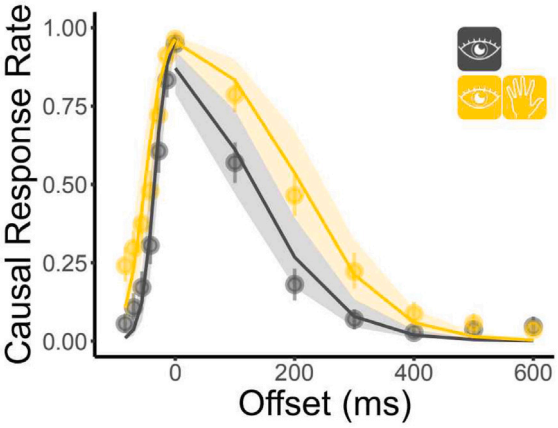


Fig. 4. Experiment 1 results. Causal judgments (0 = non-causal, 1 = causal) as a function of temporal offset. Points show participant means with 95% bootstrapped confidence intervals; lines show model predictions (M2). Gray: 'vision only'; yellow: 'vision + haptics'. Haptic feedback increased causal judgments across all offsets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

H2: Condition. As predicted, the probability of a causal response was higher in the 'vision & haptics' condition than in the 'vision only' condition. M2 (AIC = 1854.5) provides a better fit to the data than M3 (AIC = 1936.3, $\chi^2(1) = 89.9$, $p < 0.0001$). Within M2, the 'vision & haptics' condition has significantly greater causal response rates than 'vision alone' ($\beta = 1.15$, $SE = 0.13$, $z = 9.15$, $p < 0.0001$).

H3: Interaction. In contrast to what we predicted, the probability of a causal response did not decrease more strongly with temporal offset for the 'vision only' condition compared to the 'vision & haptics' condition. M1 (AIC = 1860.5) did not provide a better fit to the data than M2 (AIC = 1854.4, $\chi^2(1) = 1.9$, $p = 0.17$).

H4: Negative vs. Positive offsets. To analyze whether the differences in causal responses between conditions were greater for positive compared to negative offsets, we fit psychometric curves with lapses using the `quickpsy` package in R. These thresholds represent the detection threshold of 0.5, indicating a 50% causal judgment (an even split between causal and non-causal responses). We calculated four threshold values for each participant across two conditions (with and without haptics) and two offset groups (negative and positive). We conducted a one-tailed paired t-test comparing the absolute difference between positive and negative offset thresholds for 'vision & haptics' minus 'vision only'. As predicted, we found larger differences for positive compared to negative offsets ($t(17) = 3.83$, $p < 0.001$, $d_z = 0.90$).

3.1.4. Exploratory analysis

In our preregistration, we made no predictions about what participants' causal responses would look like for negative temporal offsets. In the negative offset condition, the second train began to move before the first train stopped motion (Fig. 3). Because in these situations, the second train starts moving before the first train stops (Fig. 3), one could imagine that the haptic feedback (which was locked to the second train) would provide an additional cue for the lack of contact between the trains. Alternatively, haptic information may provide more evidence that contact must have occurred (based on prior interactions and understanding of forces in collision events), thereby downplaying the visual information that suggests otherwise.

We constructed models with the same structure as M1-4 above, but fit only on the negative offset trials. First, we considered the role of offset by comparing M2 to M4. We found that M2 is a better predictor of the data, indicating that offset is a significant predictor of causal judgment. Second, we compared M2 and M3 to determine whether the conditions differed. M2 was found to be a better model. Thus, condition was a significant predictor of causal judgment. Third, we examined an interaction effect between offset and condition by using M1 and M2. M1 was a better predictor of the data, indicating an interaction effect between condition and offset in this subset. As Fig. 4 shows, the data between both conditions are quite similar at offset zero and then begin to separate as the offset decreases, which accounts for the interaction effect.

3.1.5. Qualitative results

When asked what strategies they used to complete the task, several participants mentioned considering two main factors for judging causality: contact and timing. One participant noted that the haptic feedback could alter their visual perception, stating, "the onset of the movement of the handle seemed to overwrite the visual cues". This comment is in line with the finding that kinesthetic haptic cues increased causal responses compared to vision alone (Fig. 4). Interestingly, one participant mentioned that they used expected timing to judge causality in the haptic condition. They stated "when I thought the red train would move, I would move my hand and if the haptic device moved earlier/later than my hand then I would know the timing of the collision was off".

3.1.6. Discussion

Our results show that kinesthetic haptic feedback can alter people's perception of causal events. Participants were more likely to say that an ambiguous event was causal when they received visual and haptic feedback compared to visual feedback only (Fig. 4). This was the case for both negative temporal offsets (where the second object moves before the first one stops) and positive offsets (where the second object moves only after some delay). This increase in causal judgments in the multimodal condition aligns with prior work showing a similar effect when audio cues were presented together with visual information (Guski & Troje, 2003).

Some participants reported using expectations about *when* the second object should move to make their judgment. We interpret this strategy as part of their causal inference rather than abstracting the task into one that merely relies on predicting event timing. Importantly, these observers did not rely exclusively on temporal information. They also wrote that it mattered whether the trains made contact and how plausible the collision looked overall. Indeed, as we will see in Experiment 3, participants' causal judgments are affected by the realism of the different signals, and not just by their timing alone. Together, these findings suggest that participants genuinely reported perceived causality, with temporal information among several important cues.

3.2. Experiment 2: Vision, audio, kinesthetic haptics, & vibrotactile haptics

Having established that haptic feedback increases causal judgments in Experiment 1, we added additional sensory signals in Experiment 2. Here, we focused on situations with positive temporal offsets.[2] No work to date has explored people's causal judgments in situations involving more than two sensory signals. In Experiment 2, participants experienced all combinations of vision, audio, kinesthetic haptics, and vibrotactile haptics for a range of conditions: one unimodal, three bimodal, three trimodal, and one quadrimodal. In this experiment (and in Experiment 3), we focused on situations with either no temporal offset or a positive one. Experiment 1 showed that the effect of an additional sensory signal was stronger for positive offsets compared to negative ones. This experiment was preregistered[3].

Based on the results from Experiment 1 and prior work, we predicted that:

(H1) *The probability with which participants select the causal statement decreases as the temporal offset increases.*

Prior work has only explored bimodal conditions. We present two competing hypotheses for what will occur as signals increase. Some prior work in rhythmic synchronization has found an upper limit to sensory integration when combining visual, audio, and tactile cues (Johnson et al., 2020). So, additional sensory signals (beyond bimodal) may not further increase causal judgment. Alternatively, with each additional sensory cue, participants have more information suggesting that a collision has occurred (as it would be increasingly unlikely for multiple independent sensory signals to happen simultaneously). Accordingly, causal judgments should increase with each additional sensory cue. From these two positions, we present two competing hypotheses:

(H2.1) *The probability with which participants select the causal statement is greater in the multisensory conditions (bimodal, trimodal, quadrimodal) compared to the 'vision only' (unimodal condition).*
(H2.2) *The probability with which participants select the causal statement increases with the number of signals.*

3.2.1. Design & measures

Experiment 2 was a within-subjects, repeated measures design with two factors: temporal offset and sensory modality. We had one unimodal condition (vision), three bimodal conditions (vision & audition, vision & vibrotactile haptics, vision & kinesthetic haptics), three trimodal conditions (vision, audition, & kinesthetic haptics; vision, audition, & vibrotactile haptics; vision, kinesthetic, & vibrotactile haptics), and one quadrimodal condition (vision, audition, kinesthetic, & vibrotactile haptics). We treated sensory modality as a blocking factor and then randomized offsets within each block. Each offset was repeated six times per block. We randomized the block order between participants. We used seven temporal offsets, equally separated in time (0, 100, 200, 300, 400, 500, 600 ms). We used seven offsets in 100 ms steps to sample the psychometric function identified by Experiment 1 and pilot testing, while keeping the total number of trials feasible given the eight sensory modality conditions.

We synchronized all additional sensory modality cues to begin with the motion of the second-moving red train. The audio cue was delivered via noise-canceling headphones and consisted of a 200 ms sine wave at 240 Hz. We adjusted the vibrotactile and kinesthetic cues to ensure people perceived the signals as having similar intensities. Notably, we adjusted the length of the visual cues (250 ms of motion each), reducing the time that the trains move to fit the sensory cues presented (Fig. 2) – thus, all sensory cues were 250 ms in length. The kinesthetic force feedback was kept at a constant 3 N force applied to the right hand, matching the red train's timing and direction. We set a 250 ms duration for visual, audio, vibration, and kinesthetic cues to capture the brief sensory transient associated with short contact events such as a mug being set onto a table, a laptop lid closing, a drawer reaching its stop, or keys landing on a surface. Our aim was not to simulate the full dynamics of such events, but to preserve their temporally localized impact signature across modalities. To avoid dominance by a single modality, intensities were matched across modalities through pilot adjustments, ensuring that no single cue dominated the percept or produced discomfort, consistent with recommendations for haptic rendering (Jones & Tan, 2012). We note that longer-impact events (e.g., automotive collisions) involve additional phases and were not the ecological target of our stimuli.

[2] In the spatial-gap conditions, participants produced many competing causal narratives (e.g., inferring that the red train "had" to move out of the way of the blue one when there was a visible gap), the design doubled the required trials relative to positive temporal offsets, and haptic effects were stronger only for positive offsets.

[3] https://osf.io/jmz8q

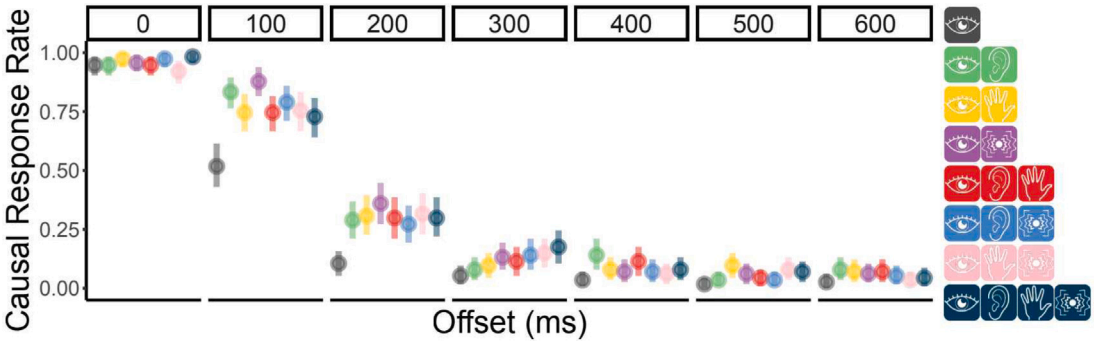


Fig. 5. Experiment 2 results: the x-axis shows the temporal offset in milliseconds (subdivided by offset for comparisons between conditions), the y-axis shows average causal response where 1 indicates *causal* and 0 indicates *non-causal*, circles indicate participant data mean, and vertical error bars mark 95% bootstrapped confidence interval, lines are statistical model predictions, ribbons are confidence intervals on the statistical models; all eight conditions are shown in different colors. Participants were more likely to judge events as causal in multimodal than unimodal conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2.2. Participants

Nineteen right-handed participants (8 Female, 8 Male, 3 Non-Binary; age $\mu = 28.0$, $\sigma = 10.8$) were recruited and compensated for participating. None of them participated in Experiment 1.

3.2.3. Confirmatory analysis

To test our hypotheses, we fit the following three generalized linear mixed-effects models, assuming a binomial linking function:

M1: causal judgment $\sim 1 + \text{offset} + (1 + \text{offset}|\text{participant})$
M2: causal judgment $\sim 1 + \text{condition} * \text{offset} + (1 + \text{condition} * \text{offset}|\text{participant})$
M3: cause judgment $\sim 1 + \text{signals} * \text{offset} + (1 + \text{signals} * \text{offset}|\text{participant})$

All models have a main and random effect of offset and participant. M2 has an additional main and random effect of condition (coded as a binary variable with 0 = *vision only* and 1 = *multisensory*) and an interaction between condition and offset. M3 has a similar additional main and random effect but with signals (coded as a numeric variable, where 1 = one signal, 2 = two signals, etc.) and the interaction between signals and offset.

These models were fit using Bayesian data analysis with the *brms* package in R. We used approximate leave-one-out cross-validation via the *loo* function to analyze which model best predicts participants' causal judgments. We used the *emmeans* package and function for contrasts. Across all analyses, we used the inference criterion that the 95% credible interval excludes zero.[4]

H1: Temporal offset. To test H1, we predicted that the effect of offset would be negative and that the 95% credible interval would exclude 0. As predicted, causal judgments decreased the longer the temporal delay between when the first train stopped and the second train started ($\beta = -0.02, 95\%$ Credible Interval (CrI)$[-0.03, -0.02]$; see Fig. 5).

H2: Multisensory condition vs. Number of signals. To test H2, we compared M2 and M3 using approximate cross-validation to determine which one better explained the data.

M2 (which only encodes whether there were one or several signals) captured the data better than M3 (which encodes the number of signals; $elpd_D = -20.8, \text{SE}(elpd_D) = 9.3$). This result is in line with H2.1 and not with H2.2. Fig. 6 shows the results either grouped by unimodal vs. multimodal (Fig. 6A), or by the number of signals (Fig. 6B).

3.2.4. Qualitative results

When asked how they decided which statement to select, and what thoughts they had about the combination of the different signals (audio, kinesthetic haptic, and vibrotactile), participants preferred different sensory cues and differed in their perception of how those cues related to the trains' contact. P3 said "the more sensations that the trials had, the easier it was for me to determine the correct statement. However, having too much sensation sometimes clouded my rationality as I was hyperfocused on making sure I felt/ heard everything. I also would say 'hit' in my head everytime the two trains made contact and would determine the statement by how long it took the red train to move after I said it". Whereas, P16 said "I think that the kinesthetic was the most convincing in terms of differentiating train motion. I noticed that I was gripping the handle and anticipating the motion for the kinesthetic

[4] We used Bayesian analysis here because evaluating H2 requires comparing two models (M2 and M3) that are not nested (i.e. they feature a different set of predictors and it is not the case that one model is a reduced version of the other.).

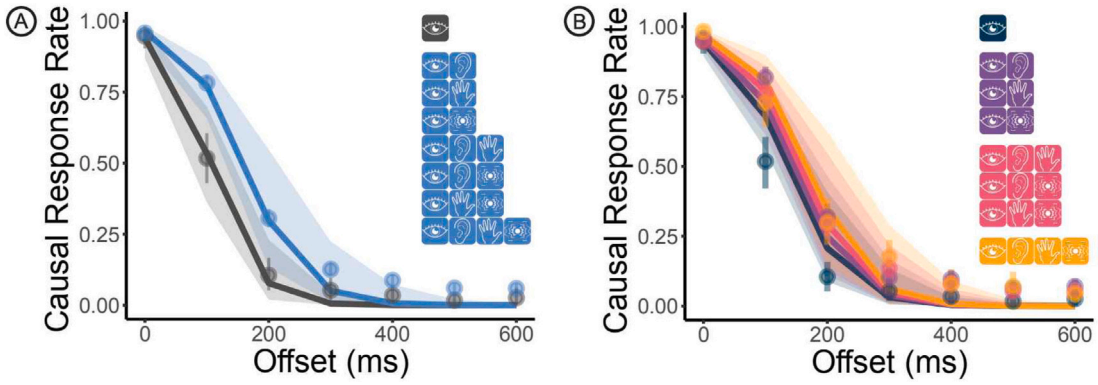


Fig. 6. Experiment 2 results grouped: *x*-axis shows the temporal offset in milliseconds, *y*-axis shows average causal response where 1 indicates *causal* and 0 indicates *non-causal*, circles indicate participant data mean, and vertical error bars mark 95% bootstrapped confidence interval, lines are statistical model predictions, ribbons are confidence intervals on the statistical models; (A) M2 results grouped into a vision and then multisensory — meaning vision with at least one and at most three more sensory sources; (B) M3 results grouped by the number of signals present. (Both) A, which assumes that people differentiate between one versus several signals, works better than B, which assumes that people's causal judgments are sensitive to the number of signals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

feedback. In essence, I saw the collision, and already started to try to move my hand to the left for the red train. However, I noticed that sometimes my hand would move before the stick, and thats how I determined when the red train started on its own". The feedback suggests that people used different strategies to interpret information from the sensory sources.

3.2.5. Discussion

This experiment examined how the number and type of sensory signals affected people's causal judgments. Consistent with Experiment 1, participants were more likely to judge ambiguous events as causal in multimodal than unimodal conditions. Additionally, we found that a model that assumes people differentiate between one and several signals works better than one that assumes people's causal judgments are sensitive to the number of signals. Put differently, it mattered mostly whether only a single signal or more than one signal was present, but it did not matter as much whether two or more signals were available. We interpret this not as evidence for a precise law governing the effect of cue number, but as evidence showing that the largest gain comes from moving from a unimodal to a multisensory event representation (see also Johnson et al., 2020), with smaller additional benefits from additional cues. Moreover, the effect of the different types of sensory signals was very similar (see Fig. 5). For example, whether the additional signal was auditory or haptic did not matter much.

4. Experiment 3: Realism

In the physical world, we use multiple senses to better understand what events occur around us. Additional sensory information helps us infer what happened. For example, hearing a loud "crash" is a strong cue that a collision occurred, as is feeling the reaction force and impact-based vibration. Work within virtual reality has shown that including multisensory cues increases people's sense of presence and memory of the environment (Dinh et al., 1999). Additionally, haptic information plays a critical role in how we learn. Also, haptic information is important for how children learn about forces, for example, by hitting objects against one another and observing what happens (White, 1988, 2012). As we seek to create interactive, multimodal virtual environments, we must consider how different modalities influence causal perception — the impression that one event is the result of the occurrence of another (Michotte, 1946/1963).

In Experiments 1 and 2, we considered collisions between two trains, providing temporally matched sensory information. Participants heard a "buzz" when the trains made contact and felt a strong, constant vibration. However, if you think of a ball that collides with a second ball, causing it to move (see Fig. 7A), what makes you see this event as causal? The perceived realism of sensory information matters. For example, hearing "buzz" (as provided in Experiment 2) may be less indicative than hearing "crash" on seeing a collision. Researchers have shown that people use realism in their criteria for making causal judgments about events (Kominsky & Scholl, 2020; Kominsky et al., 2017, but see also Bechlivanidis et al., 2019). And haptics research has likewise shown that the perceived realism of contact depends strongly on the structure and timing of the rendered cues (Di Luca & Mahnan, 2019; Okamura et al., 2002; Vogels, 2004).

Here, we investigate how multisensory information that is either physically realistic or not affects causal perception. Prior work on physical realism has explored the effect of delays and mismatches between haptic and other sensory information on participants' feelings of immersion in Virtual Reality (VR) environments (Berger et al., 2018). They find that if haptic feedback is not rendered in concert with vision and audio, the impression of realism decreases, producing a haptic uncanny valley.

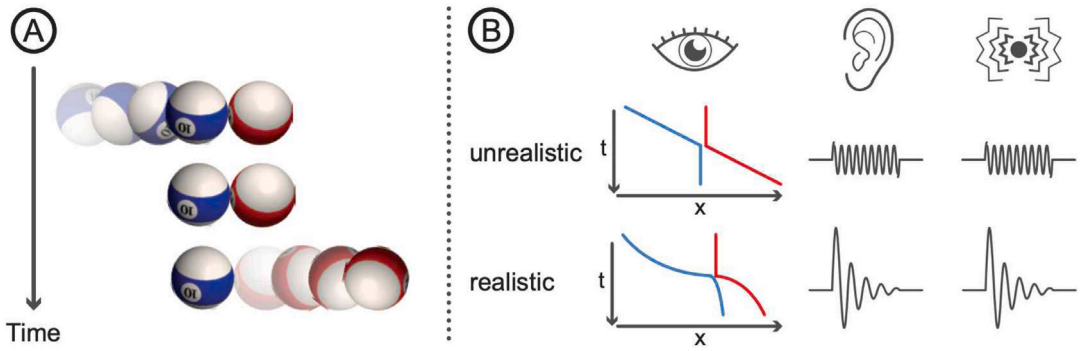


Fig. 7. (A) An overview of the experiment in which the blue ball starts rolling toward the red ball, makes contact with the red ball, stops, and then the red ball rolls away (B) Conditions for the realism of sensory cues. The visual cue is either unrealistic (constant velocity with instantaneous starting and stopping) or realistic (encounters drag as it rolls and undergoes an inelastic collision). The unrealistic audio and vibrotactile cues are sinusoids of a fixed length that sound and feel like a generic vibration (such as a phone notification). The realistic audio is a contact cue for billiards. The realistic vibrotactile cue is a sinusoid with exponential decay to mimic a real contact.

To our knowledge, no work has studied how physical realism affects multisensory causal perception. Here, we consider vision, audio, and vibrotactile haptics with realistic and unrealistic cues. We ask the following questions: First, are people more likely to judge an event as causal when they have evidence from multiple sensory sources? Specifically, does this extend beyond our previous results when the signals are realistic, or is the difference still only between vision alone and with any additional sensory signals? Second, does it matter whether information from different senses is physically realistic?

We use the same paradigm as in Experiments 1–2. However, instead of trains, two simulated billiard balls roll along a table, and additional auditory and haptic feedback occurs on their collision (Fig. 7B). The use of billiard balls, rather than trains, was to simulate rigid-body dynamics and provide strong visual cues between the physically realistic dynamics and constant velocity motion (as the objects roll and slide, which would be less visible to participants if presented as the wheels of a train). After watching the animation, participants rated the probability that the two objects were causally linked.

To address our questions about the role of multisensory cues and realism on causal perception, we preregistered[5] the following five hypotheses:

(H1) *Participants' causal judgments decrease as the temporal offset increases.*
(H2) *Causal judgments are higher in multisensory conditions (containing either audio, vibration, or both) compared to vision alone.*
(H3) *Causal judgments are higher in conditions that have only realistic signals compared to conditions with at least one unrealistic source of information.*
(H4) *Causal judgments are highest in the condition that is realistic and that contains all three sensory cues.*
(H5) *Causal judgments are lower in conditions with mismatched cues (realistic visuals with unrealistic sensory signals) compared to the realistic vision-only condition.*

Our predictions for H3–H5 were based on work showing a possible uncanny valley in haptics (Berger et al., 2018) and for phenomenal causality (Meding et al., 2020).

There is also a plausible alternative: increasing the number of cues that are inconsistent (temporally misaligned) or uniformative (their content is not realistic) could decrease perceived causality.

4.1. Design & measures

To understand the effects of more realistic cues on causal perception, we selected two cues for each sensory modality: vision, audition, and touch. One is a physically realistic cue, and one is not. The non-realistic cues are similar to the cues used in Experiment 2. The realistic cues mirror the impact-based dynamics for rigid objects in real life. We discuss how we implemented each cue in turn. The same hardware and software were used as in Experiments 1 and 2, with the new cue implementation (Fig. 1B).

[5] https://osf.io/cyd8j

4.1.1. Vision

The display depicted two spheres on a uniform surface. At the beginning of each trial, one sphere (Ball A) rolled smoothly toward a stationary sphere (Ball B). On contact, Ball A slowed, and Ball B began to move, suggesting momentum transfer. Immediately after contact, both balls briefly slid against the surface before resuming rolling. As they rolled, both gradually decelerated. Ball A came to rest sooner, while Ball B continued rolling for a short time afterward. In some conditions, the animation included a brief temporal pause at the exact moment of contact, during which both balls appeared to freeze before continuing their motions.

In both conditions, the objects share the same initial positions to ensure that they make contact at the center of the computer screen. The first object moves a set distance, and the second object moves for a predetermined amount of time. The shadows for the balls were rendered using two thin discs that move with the same dynamics (realistic and unrealistic).

The unrealistic visual cue shows both objects moving at a constant velocity and undergoing a perfectly elastic collision. While the way each object rolls looks realistic, its collision does not. On contact, the first object suddenly stops rolling (without sliding), and the second one immediately starts rolling.

4.1.2. Audio

The realistic cue matched the impulse that would occur when two balls collide in the real world. We edited it to have a fixed duration and amplitude.[6] The unrealistic audio cue was a non-decaying sinusoid of reduced amplitude to match the overall intensity of the realistic cue.

4.1.3. Vibration

The realistic cue was based on work in contact realism for event-based haptics (Okamura et al., 2002). We designed a vibration that felt like a collision between two billiard balls. It takes the form of an exponentially decaying sinusoidal model,

$$Ae^{-Bt}sin(2\pi\omega t), \tag{1}$$

where $A = 0.5$, $B = 40$, $\omega = 90$, and $t = 0 : 150$ ms.

The unrealistic cue used a sinusoid without decay, and we set the frequency to match the unrealistic audio cue (240 Hz). The amplitude was reduced to normalize the perceived intensity between cues. We held all audio and vibration cues at a constant length of 150 ms. The parameter values (A, B, omega) and a 150 ms duration were selected based on Okamura et al. (2002)'s work and pilot tuning to evoke a single, crisp impact feeling, while maintaining a similar overall intensity to the audio cue. Intensity was evaluated both analytically and empirically. Analytically, we calculated the signal's energy by considering the amount of activation it undergoes over time. Empirically, we pilot tested with humans to determine whether they perceived the two signals as having similar intensities.

4.2. Participants

A total of 22 right-handed participants (*age*: M = 25, SD = 4; *sex*: 9 female, 13 male) completed the experiment in accordance with our IRB and were compensated for their participation. None had participated in any of the previous experiments.

4.3. Part A: Varying temporal delay

Similar to previous experiments, this experiment is a within-subjects, repeated measures design with four factors: vision (2 levels: unrealistic, realistic), audio (3 levels: none, unrealistic, realistic), vibration (3 levels: none, unrealistic, realistic), and temporal offset (5 levels: 0, 100, 200, 300, 400 ms, method of constant stimuli Jones & Tan, 2012). Fewer temporal offsets were used to reduce the total number of trials, as this experiment has more conditions than the previous ones. Audio and vibration cues were played when the balls contacted (point a' in Fig. 2) – this is different from the time at which cues were played in Experiments 1 and 2. In these experiments, sensory cues were played when the second object began moving (rather than at the moment of contact). Here, we chose to change this as pilot studies indicated that people found it confusing to hear a realistic collision sound ("clack") when the second object moved away compared to when two objects made contact. All conditions were repeated four times for a total of 360 trials. We used a pseudo-randomized trial order in each repetition to ensure that people saw different conditions comparably throughout the experiment.

Participants were seated at a table with a keyboard, mouse, computer screen, and rigidly mounted haptic device (Fig. 1B). Noise-canceling headphones played white noise (to reduce distractions from external noise or vibrations) and audio cues from the experiment. Participants held onto the device's handle during each trial using a precision grip.

Before beginning, we guided participants through two sets of practice trials that introduced them to the system's input and provided a range of sensory information. We first presented the smallest and largest visual delays across both visual conditions to help task adaptation and mitigate potential order effects (Bechlivanidis et al., 2019). We counterbalanced the presentation of visual conditions between participants. Second, participants were exposed to the full range of sensory conditions through eight randomized trials (all combinations of vibration and audio, excluding no vibration/no audio, as this was in the first practice). Afterward, participants began the experiment.

[6] freesound.org/people/Za-Games/sounds/539854/.

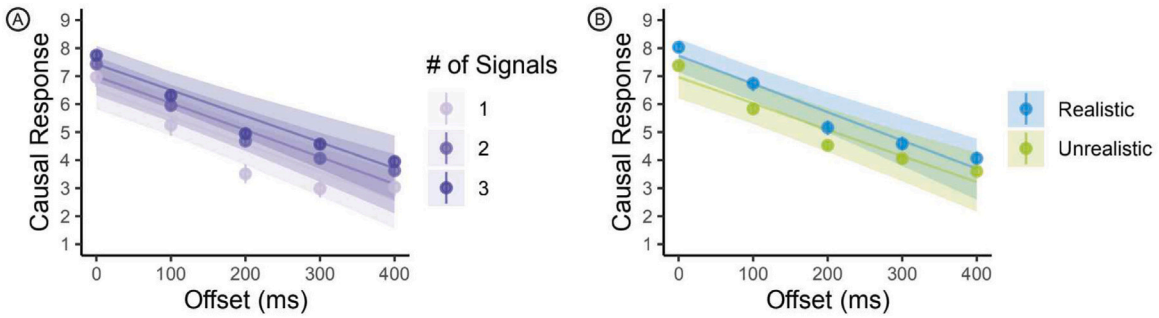


Fig. 8. Means and 95% bootstrapped confidence intervals (CI) shown for causal response across participants. Lines are statistical model predictions, and ribbons are confidence intervals on the statistical models. (A) Results by number of signals with predictions based on M2. (B) Results by realism of the cues with predictions based on M3. (Both) A shows that causal judgments increased with more sensory signals. B shows that realistic signals led to higher causal ratings than non-realistic ones. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Similar to the previous experiments, two billiard balls initialize at the beginning of each trial at set positions. Then, the blue ball moves toward the red ball and makes contact. On contact, the two balls stop for a set amount of time (temporal delay), after which the red ball moves away (Fig. 2). If the visual cue is realistic, then the blue ball also moves slightly, given the coefficient of restitution. After the trial, the following question (modified from Guski & Troje, 2003) appeared:

How probable is it that the blue object caused the movement of the red object?

We asked participants to select a number between one and nine that best described what happened, where one means *not at all probable* and nine means *very probable*. Afterward, a visual indicator appeared on the scale, and participants confirmed their selection to continue to the subsequent trial. Participants were able to take a break after every 90 trials.

4.4. Confirmatory analysis

Generalized linear mixed-effects models with fixed effects of temporal offset, number of signals, realism, condition, and match-mismatch were fit to the data. All models have random effects for participant and the other predictors. The temporal offset and number of signals are continuous variables, while the rest are coded as factors.

To test our preregistered hypotheses, we fit the following models to participants' causal judgments:

M1: causal judgment $\sim 1 + \text{offset} + (1 + \text{offset}|\text{participant})$
M2: causal judgment $\sim 1 + \text{signals} * \text{offset} + (1 + \text{signals} * \text{offset}|\text{participant})$
M3: causal judgment $\sim 1 + \text{realism} * \text{offset} + (1 + \text{realism} * \text{offset}|\text{participant})$
M4: causal judgment $\sim 1 + \text{condition} + (1 + \text{condition}|\text{participant})$
M5: causal judgment $\sim 1 + \text{mismatch} + (1 + \text{mismatch}|\text{participant})$

Like in Experiment 2, we used Bayesian data analysis and approximate leave-one-out cross-validation for model comparison.

H1: Temporal offset. Causal judgments decreased with temporal offset ($\beta = -0.01, 95\%$ CrI$[-0.01, -0.01]$).

H2: Number of signals. Participants gave higher causal judgments when there were more rather than fewer signals (Fig. 8A). M2 captured the data better than M1 ($elpd_{loo} = 203.5$, $SE = 23.0$). In M2, there was a positive relationship between the number of signals and participants' causal judgments ($\beta = 0.43, 95\%$ CrI$[0.29, 0.58]$). Just like in Experiment 2, there was a larger difference between one signal and two signals, compared to the difference between two signals and three.

H3: Cue realism. Participants' causal judgments were affected by the realism of the cues (see Fig. 8B). M3 captured the data better than M1 ($elpd_{loo} = 280.2$, $SE = 29.5$). Within M3, there was a positive effect of signal realism on causal judgments ($\beta = 0.77, 95\%$ CrI$[0.34, 1.21]$).

H4: Condition. We predicted that the condition in which all signals were realistic would receive greater ratings than the other conditions. We computed a contrast based on M4 that compares this condition with the rest. As predicted, the causal judgments in the realistic, multisensory condition were greater than in the other conditions ($\beta = 1.19, 95\%$ CrI$[0.73, 1.65]$). This result can be seen in Fig. 9, where the blue bar furthest to the right represents that condition.

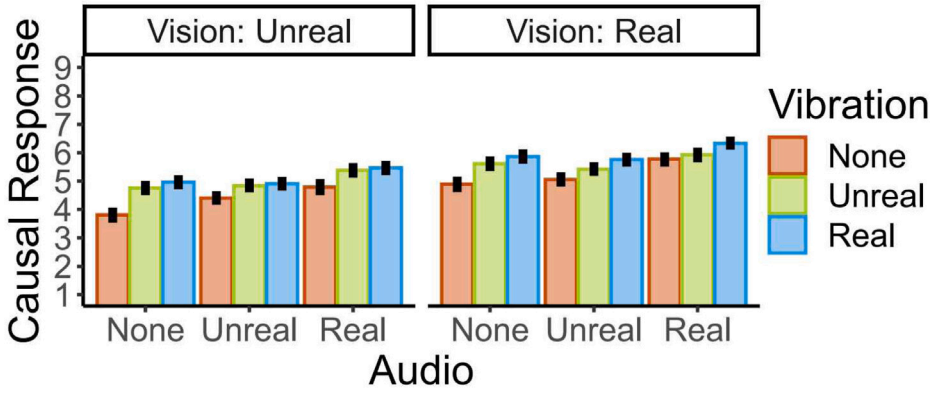


Fig. 9. Means and 95% bootstrapped confidence intervals (CI) shown for causal response across participants. Visual realism has a stronger influence on causal judgments than the realism of the other signals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

H5: Mismatch. This final hypothesis considers a possible uncanny valley, in which mixing realistic and unrealistic cues might lead to lower causal ratings than in a situation that features only the visual cue. Thus, we took only these portions of the data and established a factor for each of these groups.[7]

We computed a contrast comparing all the mismatched conditions with the vision-only condition. Unlike what we predicted, causal judgments were higher in the mismatched conditions than in the vision-only condition ($\beta = 0.67, 95\%$ CrI$[0.24, 1.07]$).

4.5. Exploratory analysis

We did not hypothesize whether the cues would differ in how strongly they affected causal judgments. To explore whether they do, we ran a generalized linear mixed-effects model with fixed effects of audio, vibration, and vision. Vision, audio, and vibration are coded as factors. The model has random effects for participant and the other predictors.

M6: causal judgment $\sim 1 + \text{audio} + \text{vibration} + \text{vision} + (1 + \text{audio} + \text{vibration} + \text{vision}|\text{participant})$

In M6, the baseline condition (unrealistic vision, no audio, no vibration) had a population-level predicted response of 4.16. This is the fixed-effect intercept of the mixed-effects model, indicating the average causal rating score in the baseline condition before adding other predictors. The realistic vision has the largest coefficient ($\beta = 0.82$), followed by realistic vibration ($\beta = 0.77$), and realistic audio ($\beta = 0.61$). Unrealistic vibration also had a positive influence ($\beta = 0.54$), while unrealistic audio did not have a clear impact ($\beta = 0.08$).

4.6. Part B: Varying signal location

The second part of the experiment examined whether the timing of the sensory signals mattered for people's causal judgments (Fig. 2). This part of the experiment had three factors: condition (2 Levels: realistic vision, audio, and vibration; unrealistic vision, audio, and vibration), location (3 Levels: with the first object; with the second object; halfway between), and temporal offset (5 Levels: 0, 100, 200, 300, 400 ms). We randomized all conditions within one repetition (and repeated for each of the three remaining repetitions). Thus, each condition was experienced four times, for 120 total trials.

After completing the set of trials from Part A, participants took a break before starting with the trials in Part B. The instructions for this part were largely the same as for Part A. Participants were not told the signal's location would vary. We gave participants a break midway through this part of the experiment.

At the end, participants completed a short survey with questions about strategies and comments on the experiment. Most participants finished the whole experiment (Part A and B) in less than 60 min, but were given up to 90 min.

4.7. Exploratory analysis

Previous work has shown that placing additional sensory cues between the first object stopping and the second object starting motion does not consistently affect causal rating (Guski & Troje, 2003). However, with more realistic cues, the effect of temporal cue placement may be stronger. Fig. 10 shows that participants' causal judgments decrease as the cues move from being co-located with the first ball stopping toward the second ball moving as the temporal offset increases.

[7] This is a deviation from our preregistered hypothesis (which grouped all conditions, not just realistic vision) due to an oversight on what was stated.

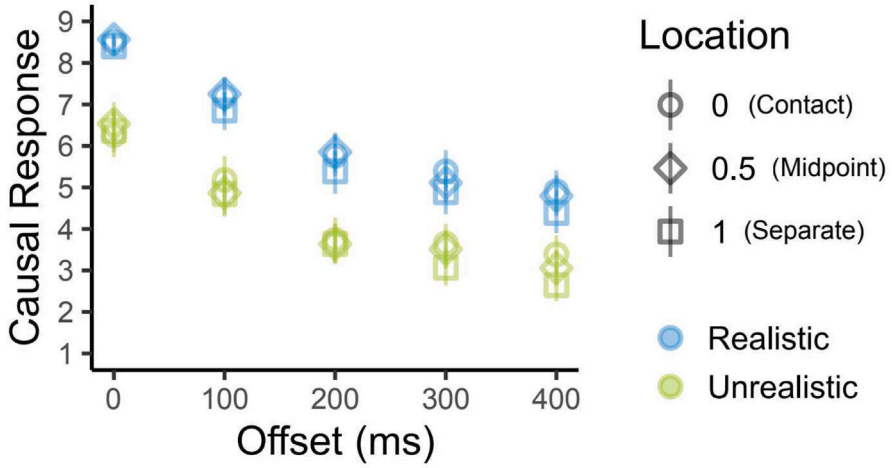


Fig. 10. Means and 95% bootstrapped confidence intervals (CI) shown for causal response across participants. The marker shape corresponds to different relative locations within the contact (0, 0.5, 1) where additional sensory cues were played. The marker color corresponds to whether the sensory cues are all realistic or all unrealistic. Participants' causal judgments were slightly lower for signals that occurred later in the window (1) than earlier (0). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.8. Qualitative results

We asked participants how the combination of signals (vision, audio, and vibration) affected their response, whether the type of signal (realistic or non-realistic) affected their response, and what their (least) favorite condition was. Participants commented on the specific types of cues. P2 said that even when "there was a delay, vibration somehow filled that gap to create a sense of causality", echoing a prior interpretation of audiovisual phenomenal causality cues (Guski & Troje, 2003).

Realism strongly shaped perception. P19 wrote "I think the realistic signals may have made me think the ball movements were more realistic than they actually were sometimes … It seemed like the non-realistic signals affected me less". This aligns with the effect of cue realism (Fig. 9). Visual information was "critical" for P22, yet if "the vision was slightly off and the audio and vibration were both correct it almost felt realistic", indicating non-visual cues also supported causal judgments.

Several participants noted differences in vibration duration and intensity. P18 commented "too much vibration made the scenario much less convincing. A vibration with weak strength and a similar length to the audio made this most convincing". They added "For audio cues it was clearly ranked for me- any other sound than the actual billiard ball impact was not convincing, and at times contributed to the feeling that the ball was stopping or pausing for longer periods of time". P13 commented "I gave higher scores when the combinations made sense, so when there were realistic audio and vibration signals". These remarks suggest priors about plausible cue strength and congruence.

Participants highlighted their enjoyment of particular cue combinations. P9 noted a "big dopamine hit with the 0-delay, realistic sound effect, and the shot ball following up behind the hit ball". P11 commented "when the red ball initially slips, it feels satisfying". Such reactions may reflect prior multisensory experiences (e.g., video games pairing realistic audio with generic haptics).

Matching vs. mismatching realism emerged as a final group. P7 wrote "When the realistic signal (e.g., billiard ball sound, accurate vibration) corre[s]ponded with a well-timed impact and movement, the greater my score was. In other words, the more realistic, the better I perceived the movement". P8 explained "When different conditions were combined, it was harder to give a response since some of the signals matched what I would expect to happen while the other signals did not match at all. When this happened, I gave a score in the middle. When there was a combination of realistic signals, it was easier to give a higher probability to these trials, and similarly for non-realistic signals". Mismatched relative realism tended to reduce scores compared to internally congruent (all realistic or all unrealistic) sets.

4.9. Discussion

What role does the physical realism of sensory signals play in whether people judge that an event was causal? Realistic signals led to higher causal ratings than non-realistic signals (Fig. 8B). So, not only do more temporally contiguous sensory signals lead to higher causal judgments, but so does providing more realistic cues. This could be due to realistic signals aligning with people's intuitive understanding of physics (Kubricht et al., 2017) and internalized representations of how collisions look, sound, and feel. In other words, realism matters here not only because it changes how well the different cues seem to fit together as having resulted from a plausible collision, but also because it changes how strongly the scene supports a causal interpretation. In this sense, realism

is not simply an alternative judgment participants might make about the event; it is one of the factors that shape whether the event is interpreted as causal. Like in Experiments 1 and 2, we found that causal judgments increased for more sensory signals (Fig. 8A).

We did not observe an uncanny valley effect under the present mismatch manipulations. Instead, conditions that mixed realistic and unrealistic signals still yielded higher causal judgments than the vision-only condition. We do not believe that realism mismatch is unimportant in general, nor that prior work reporting stronger mismatch costs was mistaken. Rather, our results suggest a narrower point, that some mismatches attenuate the multisensory benefit without eliminating it. One plausible reason for this is that the mismatches in our study were moderate enough that the added cues still provided usable evidence for a collision, even if they were less coherent than fully realistic combinations. More work is needed to better understand when uncanny valley effects occur.

An uncanny valley effect would occur if mismatched cues yielded a drop below both the vision-only baseline and the fully realistic, multisensory condition. We did not observe this pattern. Instead, we found that mismatched cues produced intermediate causal ratings, which were higher than those in vision-only conditions but lower than those in fully realistic multisensory conditions, indicating that the added signals were informative despite the inconsistency. This pattern suggests a tempered multisensory benefit: additional cues help even when not perfectly realistic, but they help less than consistent cues. Thus, the key point is not simply that "more signals are better", but that the multisensory evidence is most effective when the cues are jointly consistent with a single physical event. This interpretation aligns with broader findings on multisensory cue-conflict (Meding et al., 2020). For designers of multisensory systems, this means adding cues can improve causal impressions, but the largest gains should come from cues that are both well-timed and realistic.

These findings suggest that realism matters here not only because it alters the perceptual quality of the scene, but because it changes how strongly the scene supports a causal interpretation. In other words, observers appear to use realism as part of their evidence for deciding whether a collision occurred.

The exploratory analysis found that audio, vibration, vision, and the interaction between vibration and vision were all predictive of causal judgments. Prior work has shown that adding audio to visual scenes increases causal judgments (Guski & Troje, 2003), and we observed this effect across visual conditions. However, the difference between no audio and an unrealistic audio cue was smaller in the realistic vision case (the red columns in Fig. 9). Similarly, both types of vibration affected performance regardless of whether the visual cues were realistic. However, the *boost* provided by adding unrealistic vibration decreased when realistic visuals were used (compare the green and blue columns across both panels in Fig. 9). This increase from vibrotactile haptics is similar to what we found previously in Experiments 1 and 2. Additionally, from Fig. 9 we can see that people care strongly about whether the visual information is realistic, whereas the realism of the other signals matters less.

In Part A of the experiment, we synchronized the audio and vibration with the first ball stopping (the moment of collision). Previous work found no conclusive effect of manipulating the temporal location of audio feedback (Guski & Troje, 2003). Our results from testing three locations (moment of collision, moment of separation, and halfway between) showed that causal judgments were slightly lower for later signals than for earlier signals (Fig. 10). This aligns with our general experience, as we often hear and feel impacts when objects come into contact — rather than when they separate.

5. A Bayesian model of multisensory causal inference

Our experiments reveal several robust behavioral regularities: causal judgments depend on the *relative* timing between cues, increase with the number of signals up to a point, and receive a sensory realism boost when signals look, sound, and feel physically realistic. Descriptive analyses establish these trends, but, by themselves, they do not specify what kind of evidence observers are integrating when they judge that one event caused another. To provide insights into potential underlying cognitive mechanisms, we introduce a Bayesian model that accounts for sensory uncertainty, uses a finite temporal binding window, and is sensitive to physical realism. In this model, realism is not an auxiliary property of the stimuli, but one of the factors that changes how strongly a cue supports a common-cause interpretation, thereby affecting causal judgments. The model's main role is to formalize the claim that observers treat haptic, auditory, and visual signals as evidence about a common collision event, rather than treating any single cue in isolation. More specifically, it allows us to test whether the empirical patterns are better explained by bounded temporal integration than by alternative models that ignore event timing or require close synchrony. In this sense, the model serves as a computational explanation of the behavioral findings (as well as prior audiovisual study findings) rather than as a separate result. The model is intended to capture causal judgments for brief multisensory contact events, where observers use temporally localized visual, auditory, and haptic evidence to infer whether a single collision occurred. It is not intended as a full simulation of extended-object dynamics or as a richly interactive manipulation task.

5.1. Modeling causal perception

We briefly discuss three different classes of computational models for causal perception, each with a unique perspective on how humans process sensory information to infer causality.

5.1.1. Uncertainty-based Bayesian models

Several works have considered uncertainty-based models that use Bayesian cue integration to make causal judgments. Researchers have explored multisensory cue combinations between vision and audition to determine whether perceived cues come from a single causal source or independent sources that are not causally connected (Körding et al., 2007; Shams & Beierholm, 2010). Magnotti et al. (2013) considered an audio-visual task based on the ventriloquism effect, where a mouth moves, and a sound is played that is either synchronous or asynchronous with the mouth movement. The ventriloquism effect is a perceptual illusion in which the perceived source of a sound is influenced by the co-occurring visual stimuli, such that we perceive the sound as having come from the source that visually co-occurs with it. They built and validated a generative model that accounts for whether the sound and visually presented face have a common cause. However, none of the works presented focused on launching paradigms.

5.1.2. Temporal binding & window-based models

During causal events, people perceive both the actions and the outcomes, the key components that make up events. By thinking about the action and its outcome rather than just the event as a whole, you add temporal information. Researchers have argued that people compress the perceived time interval between actions and events when causally linked – a concept known as *temporal binding*. Bechlivanidis and Lagnado (2013, 2016) determined that sometimes causal beliefs are so strong that they can lead to a reordering of the events. Haggard et al. (2002) found that people perceived their actions as occurring later and the outcomes as occurring earlier than they did. This observation can help inform potential generative models that handle multiple pieces of related, but separate, temporal information.

Colonius and Diederich (2004) considered audio-visual events that do not co-occur but are perceived by the brain as a unitary event. They introduce a "temporal window of integration" in which people judge temporal order or simultaneity of events. They found that the temporal window in a reaction time task is wider than in an order judgment task. Later, they extended this work to consider saccadic responses (rapid eye movements that change the point of fixation) to visual-tactile attention tasks and found that these could be predicted with a temporal window model (Diederich & Colonius, 2015). Legaspi and Toyoizumi (2019) developed a Bayesian model that accounts for temporal binding and reveals that reliable sensory cues can make a sense of agency emerge even if the action was unintended.

5.1.3. Physics-based models

Some researchers have taken a more intuitive, physics-based approach for modeling tasks that depend on motion, such as collision or object interaction. They suggest that people may use their own internal physics model to predict and compare actions to outcomes. Sanborn et al. (2013) propose a noisy Newtonian framework as an alternative to more task-specific heuristic modeling for predicting people's responses in estimating the mass of objects in the launching paradigm. This framework focuses on people's intuitive understanding of physics, with some dynamic noise added. Later work by Smith et al. (2019) developed a model that uses inference and probabilistic physical simulation to predict perceptual observations, such as one object disappearing behind another.

In summary, uncertainty-based models highlight how people combine noisy evidence, temporal binding models emphasize how action and outcomes are compressed in time when causally linked, and physics-based models focus on internal simulations of dynamics. Our approach builds most directly on the uncertainty-based models, but incorporates a temporal binding window and treats realism as a prior over causal structures rather than explicitly simulating forces.

In this model, sensory cues are encoded with modality-specific uncertainty, evaluated within a temporal window that determines whether they are likely to share a common cause, and weighted by a realism prior that favors physically realistic events. This model captures the aspects of temporal binding and realism most relevant for the current launching paradigm without requiring a full Newtonian simulation.

5.2. Model architecture

The model assumes that people try to infer whether all observed cues stem from the same collision event ($C = 1$) versus from unrelated events ($C = 0$), conditioned on the observed data (which includes the onset times and (potential) spatial gap: $\mathbf{d}$) and their uncertainty about the data (θ):

$$P(C \mid \mathbf{d}; \theta) \propto P(\mathbf{d} \mid C; \theta)\, P(C), \tag{2}$$

Fig. 11 visualizes the model's inferential process. We assume people consider two alternatives: a causal event or (several) non-causal events. If it is causal, all cues (visual, auditory, vibrotactile, and kinesthetic) are assumed to be generated within a similar temporal window, and any perceptual asynchrony is attributable only to the temporal uncertainty of the cues. If it is non-causal, then the cues may arise at different times. Likewise, for spatial judgments, the model considers whether there is a gap (which is noisily estimated) or no gap, in which case any differences in object location are due solely to visual uncertainty. The likelihood of an observation is a combination of the temporal and spatial evidence.

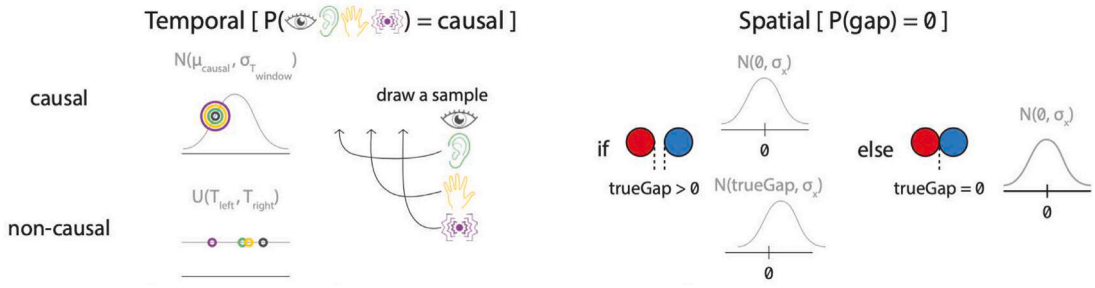


Fig. 11. A schematic of how the model's Monte-Carlo approximation works. Each simulated "run" starts by flipping a fair coin: the upper panel traces the branch in which all sensory cues come from the same collision; the lower panel traces the branch that they do not. Based on that choice, the engine draws plausible arrival times for every cue (visual, auditory, vibrotactile, kinesthetic). Similarly, for the spatial gap at contact, in the causal case the model assumes that there was contact between the two objects — regardless of whether that is true. In the non-causal case, the gap is accurately represented (i.e., the model incorporates whether there is a gap or not). Repeating this draw-and-weight cycle many times lets the model tally how often the causal branch outvotes the non-causal one. This tally yields an empirical estimate of the posterior probability of a causal link.

5.2.1. Key model parameters

The model builds on a small set of psychologically interpretable parameters that capture how an observer encodes when and where events occur, and how cue realism matters:

- σ_{visual}: the temporal uncertainty of visual onsets;
- σ_{sensory}: a shared temporal uncertainty for the three non-visual cues (auditory, vibrotactile, kinesthetic)
- σ_x: spatial uncertainty for the perceived gap between objects at assumed contact
- $\sigma_{T_{window}}$: uncertainty about the width of the temporal binding window within which separate sensory events may be merged into one;
- δ_{real} and δ_{unreal}: realism multipliers that up- or down-weight the odds of a causal interpretation when cues are physically plausible or implausible, respectively.

Collectively, $\theta = \{\sigma_{\text{visual}}, \sigma_{\text{sensory}}, \sigma_x, \sigma_T, \delta_{\text{real}}, \delta_{\text{unreal}}\}$ summarize the observer's internal uncertainty profile and expectations about collisions. Small values of σ_{visual} and σ_{sensory} make the observer sensitive to millisecond-scale asynchronies, producing the steep psychometric functions seen in Experiments 1 and 2. A tighter spatial prior (low σ_x) makes visible gaps significantly reduce causal impressions (Experiment 2). A broader temporal window (high $\sigma_{T_{window}}$) allows more variability in the times that different signals are produced from the same event. Finally, $\delta_{\text{real}} \geq 1 \geq \delta_{\text{unreal}}$ formalizes the intuition that a physically believable click or vibration should count more heavily toward an integrated causal percept than a mismatched one. In other words, a wider temporal window makes close signals more likely to be judged as causal, and so will up-weighting the prior with realistic signals.

Repeating this procedure yields an ensemble of hypothetical percepts that we compare to the actual stimulus; their relative frequency furnishes an importance-sampled estimate of the likelihood term in Eq. (2). Interested readers can consult the pseudocode in Appendix A.1 or the full model implementation on GitHub.

5.2.2. Model–fitting

The model was designed and fit using the WebPPL language, a probabilistic programming language designed for cognitive models (Goodman et al., 2016). The model was implemented in R 3.6.1 using the `webppl` package. For Experiments 1–3 we performed a grid search (Table A.1) to obtain maximum-likelihood estimates of θ. For Experiments 3A–B, where responses were on a 1–9 scale, we first transformed the model's binary causal probability via linear regression (2-fold cross-validation) before computing the root-mean-square error. We also computed Pearson's r as a measure of correlation between the model predictions and averaged human responses for each study condition.

The model was fit to best match its estimate of the causal likelihood of each stimulus with participants' judgments, using Eq. (2). Parameters were fit separately for Experiments 1 & 2, and for Experiment 3, due to stimuli and experimental setup differences. The likelihood $P(\mathbf{d} \mid C; \theta)$ is intractable in closed form because it must integrate over multiple noisy latent cues. We approximate it numerically via Markov Chain Monte Carlo sampling. The realism modifiers enter the likelihood as multiplicative boosts (or penalties) that tilt the evidence toward $C = 1$ when cues look and feel right.

We compare the full model to a lesioned version that removes the temporal window. This contrast reveals whether temporal integration is important for making causal inferences. First, we shrink the window by setting $\sigma_{T_{window}}$ to 1 ms (*no window* model), effectively ignoring the impact of the temporal window. Second, we expand the window to 1000 ms, which effectively acts as an infinite-size window in our setting (∞-*window* model).

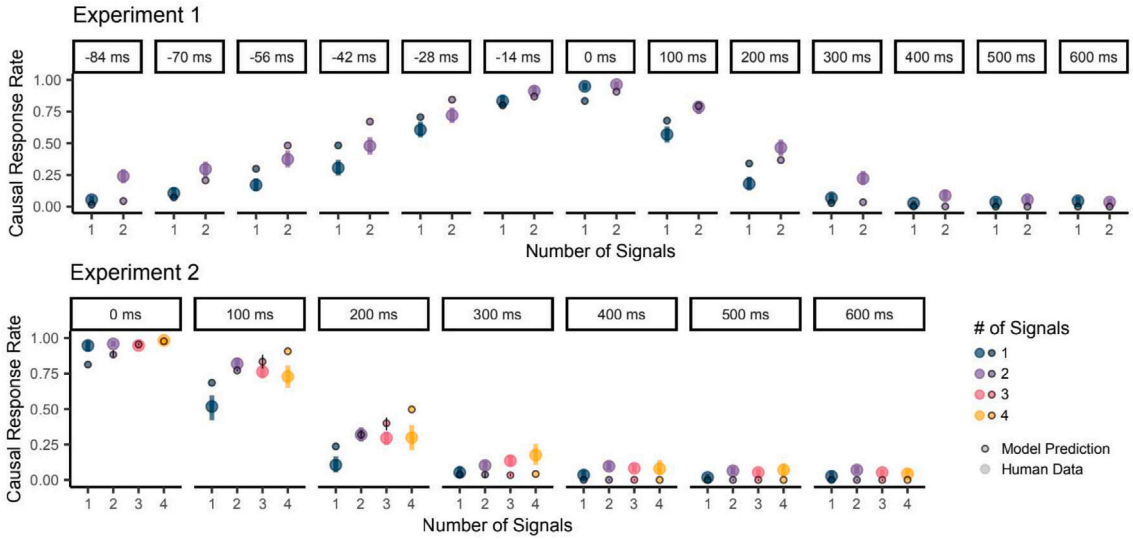


Fig. 12. Experiments 1 & 2 modeling results: *x*-axis shows temporal offset in milliseconds, *y*-axis shows causal response where 1 indicates *causal* and 0 indicates *non-causal*, color indicates the number of sensory signals present, and shape indicates if it is our model's prediction or human experiment data. All circles represent mean causal judgments for the model's predictions or human experiment data, and vertical error bars mark 95% bootstrapped confidence interval. The Bayesian model successfully captures human judgments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.3. Model results: Experiments 1 & 2

We fit the model's free parameters to the results of Experiments 1 and 2, which share similar visual stimuli, with the main difference being the addition of multisensory cues in Experiment 2. Overall performance and parameter fits are presented in Table A.2. The generative model captures the major patterns in human behavior (Fig. 12), including the decline in causal judgments with increasing temporal offset, the benefit conferred by additional sensory modalities, and the diminishing returns with three or more cues. The model does not yield perfect point-by-point predictions, yet it explains the broad structure of the data with a compact parameterization (Pearson's r = 0.97 across Experiments 1-2).

The lesioned variants produced poorer fits: the no-window model predicts nearly identical causal judgments for different numbers of signals, while the ∞-window model underestimates causal judgments at short delays. Overall, these results are consistent with the assumption that human observers rely on a finite temporal window when integrating multisensory evidence. See Appendix A.2 for additional visualizations — specifically Fig. A.1 and Fig. A.2.

5.4. Model results: Experiment 3

Due to the similarity between Experiment 3 A and 3B, we fit the free parameters of our modified model jointly. In these studies, participants reported values between 1 and 9 (compared with Experiments 1 and 2, which used binary responses). Thus, we linearly regress the human data onto the model results using 2-fold cross-validation and use root mean squared error (RMSE) as the evaluation metric. The modified generative model produced results that align with experimental results (Figs. 13–14). Unlike Experiments 1 & 2, here both the model's no-window and ∞-window lesioned versions produced similar RMSE values (Table A.3). Still, these similarities are due to the linear regression performed on the binary model predictions, as the general model trends do not align with the human data. Our results and the best-fit model predictions demonstrate that people's causal judgments increase as the number of signals increases (regardless of their relative realism). In contrast, the no-window lesioned model predicts similar results for unrealistic conditions regardless of the number of signals, and the infinite window lesioned model predicts a decrease in unrealistic signals as the number of signals increases. See Appendix A.2 for additional visualizations — specifically Fig. A.3 and Fig. A.4.

5.5. Applying the model to prior work

5.5.1. Vision and audio

To further validate our model, we consider previous work that has experimentally tested the effect of phenomenal causality with visual and audio signals. Guski and Troje (2003) introduced a *clack* sound and separately a visual *blink* at the midpoint between

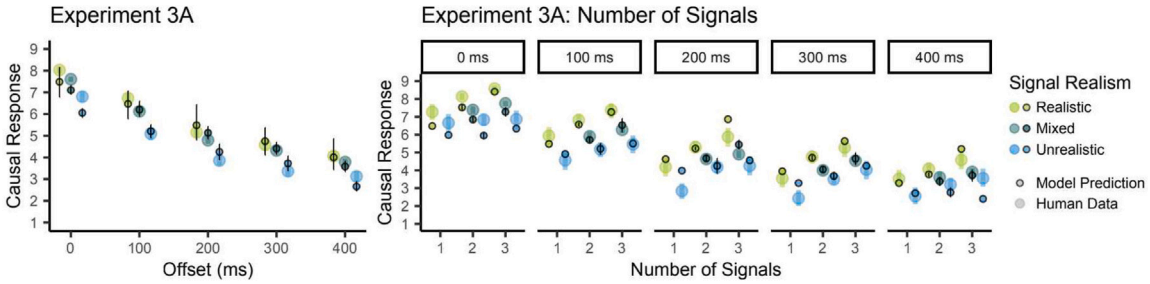


Fig. 13. Experiments 3A modeling results: *x*-axis shows temporal offset in milliseconds, *y*-axis shows causal response where 9 indicates *causal* and 1 indicates *non-causal*, color indicates the relative realism of the sensory signals, and shape indicates whether it is the model's prediction or human data. All circles represent mean causal judgments for the model's predictions or human experiment data, and vertical error bars mark 95% bootstrapped confidence interval. The model captures both realism and temporal offset effects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

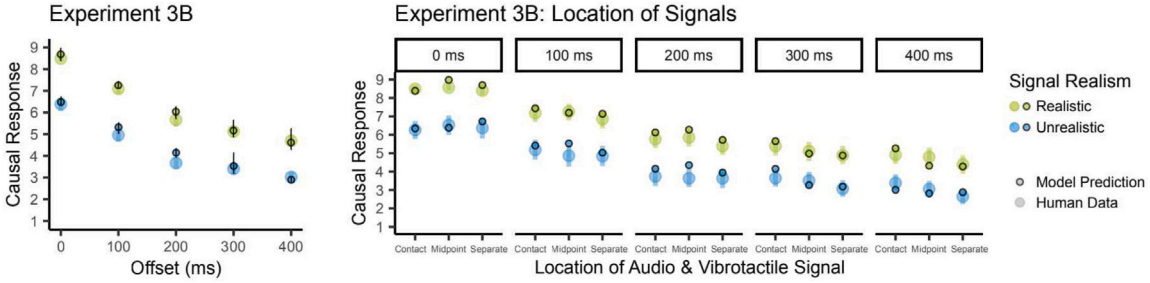


Fig. 14. Experiment 3B modeling results: *x*-axis shows temporal offset in milliseconds, *y*-axis shows causal response where 9 indicates *causal* and 1 indicates *non-causal*, color indicates the relative realism of the sensory signals, and shape indicates if it is our model's prediction or human experiment data. All circles represent mean causal judgments for the model's predictions or human experiment data, and vertical error bars mark 95% bootstrapped confidence interval. The model successfully predicts human judgments across all timing conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

when the objects touched and separated. Wang et al. (2018) added audio as well; however, they placed it at the moment of collision between the two objects. The visual information provided and the temporal sequence of events differed between these two studies, so we fit their model parameters separately. This includes Experiment 1 from Guski and Troje (2003), which has a control condition of 'vision only' and then three additional conditions: blink (50 ms color change), sound (10 ms *clack*), or a blink with a sound, and Experiment 1 from Wang et al. (2018), which explored causal perception in 2-dimensions with a control condition of 'vision only' and then one additional condition with sound (10 ms *bowling ball* sound). The authors' datasets are not readily available online, so we fit the models based on the averages reported in their figures.

Our model captures the main empirical trends reported in Guski and Troje (2003) and Wang et al. (2018) as shown in Fig. 15, with the model parameters stated in Table A.4. However, it does not quite capture a bigger increase in causal ratings from one signal to two signals compared to two signals to three. Future iterations of the model could address this.

5.5.2. Realism

Meding et al. (2020) studied the effects of realistic motion and rendering on the perception of phenomenal causality. They tested three categories: 2D animation with constant velocity, 3D animation with constant velocity, and 3D animation with collision-based dynamics. However, for our parameter fitting, we focus on two similar conditions that differ only in how the objects move (3D animations with constant velocity and collision-based dynamics — the *physical* and *rendered* conditions). Further model modification could include a 2D animation with constant velocity. Still, this case requires a richer understanding of people's priors and expectations about how physics might work in a fake world (2D circles) compared to relating physics to how we normally experience it in our day-to-day lives. We fit our model parameters based on the averages reported in the paper figures, as the raw data were not readily available.

Our model predicts similar trends to those observed in the experimental work (Fig. 16). The hypothesis that realism has differences generalizes to data from this prior research and is well represented by introducing two variables: δ_{real} and δ_{unreal}. This

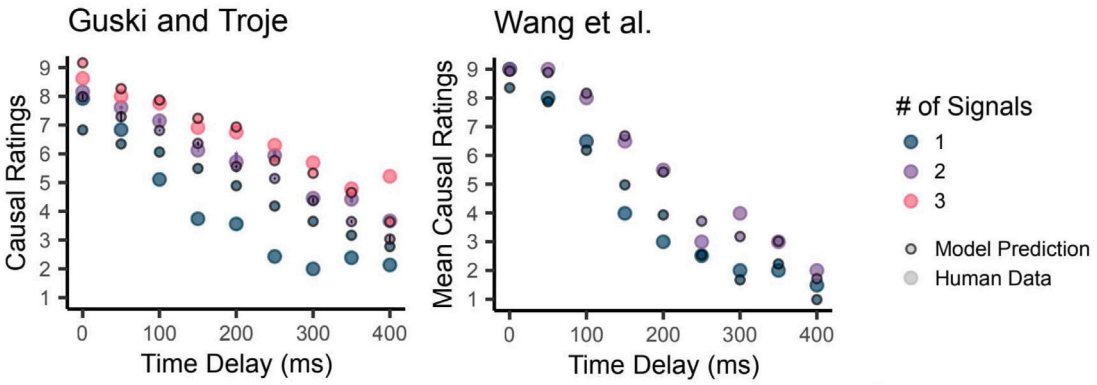


Fig. 15. Related work modeling results: x-axis shows temporal offset in milliseconds, y-axis shows causal response where 9 indicates *causal* and 1 indicates *non-causal*, color indicates the number of sensory signals present (1 = vision, 2 = vision and audio, 3 = vision, audio, and an additional visual cue *blink*), and shape indicates if it is our model's prediction or human experiment data. All circles represent mean causal judgments for the model's predictions or human experiment data, and vertical error bars mark 95% bootstrapped confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

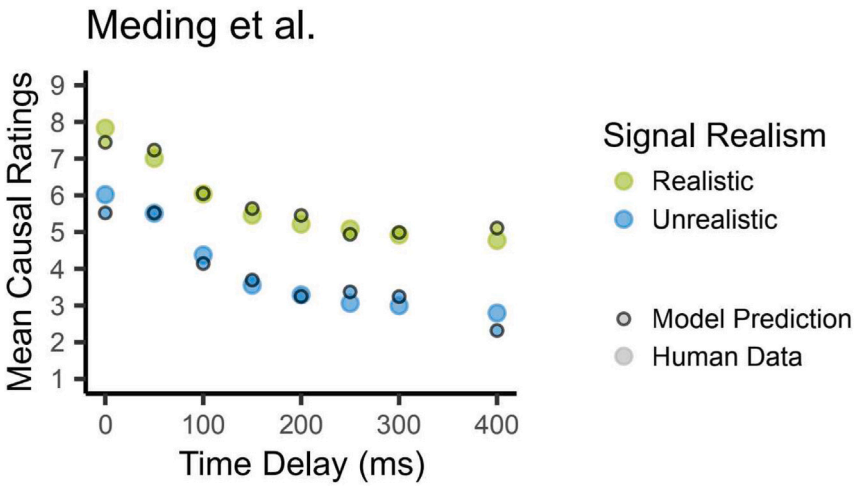


Fig. 16. Related work modeling results: x-axis shows temporal offset in milliseconds, y-axis shows causal response where 9 indicates *causal* and 1 indicates *non-causal*, color indicates the number of sensory signals present (1 = vision, 2 = vision and audio, 3 = vision, audio, and an additional visual cue *blink*), and shape indicates if it is our model's prediction or human experiment data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

gives us insights into how people might process information within a temporal signal (e.g., realism). Specifically, it suggests that it is not necessarily about any difference in causal processing, but rather that more realistic events are a priori more likely to be causal than unreal ones (because those parameters change the prior over $P(C)$). For model parameters, see Table A.4.

6. General discussion

How do people combine evidence from different sensory modalities to determine whether one event caused another? What role, if any, does haptic information play? We investigated these questions using the traditional launching paradigm from causal perception research, but extended it by introducing kinesthetic and vibrotactile haptic cues alongside vision and sound. The central empirical result is that haptic information systematically changes causal judgments in these displays. Across experiments, observers

were more likely to judge events as causal when convergent multisensory evidence was present than when they relied solely on vision, suggesting that touch is part of the evidence people use when deciding whether one object caused another to move.

The remaining results clarify how that multisensory evidence is used. Additional cues were most helpful when temporally aligned with the event and consistent with a physically plausible collision. Realistic cues yielded stronger causal judgments than unrealistic ones, and mismatched cues produced intermediate judgments — they still increased judgments relative to vision alone, but less so than fully realistic multisensory combinations. Overall, causal judgments are shaped by how well multiple sensory signals jointly support a single collision event. In this framework, realism is not separate from causal judgment but is one of the factors observers use when judging causation. Cues that make the event feel more like a plausible collision also make it more likely to be judged causal. This differs from prior haptics work, which often treats realism itself as the outcome of interest (e.g., Kuchenbecker et al., 2006; Okamura et al., 2002; Park et al., 2019; Shin & Choi, 2018). By contrast, here, realism serves as part of the evidence that shapes causal judgments.

While the behavioral patterns might be summarized informally as more and better-timed signals yielding higher judgments of perceived causality, the computational model operationalizes this result by tying causal judgments to noisy inferences over event timing, spatial gaps, and realism-weighted priors. The model shows how a small set of assumptions can account for the effects of haptic evidence, temporal tolerance, and realism across experiments. Comparisons between the full model and lesioned alternatives suggest that observers integrate evidence within a bounded temporal window, rather than simply accumulating cues without regard to their timing. The model, therefore, helps clarify the underlying cognitive mechanisms by which people reach their causal judgments. People appear to treat multiple sensory signals as uncertain evidence about a common cause, with sensory cues contributing more when they are well-timed and physically coherent. Future research can build on this framework to test stronger mismatches between signals, richer haptic signals, and more interactive paradigms.

A potential concern is that participants might have treated our task as about judging time (i.e., whether 'all events happen at the same time') rather than causality (i.e., 'did [the signals] come from the same cause'). However, several aspects of the data suggest otherwise. First, participants were explicitly asked to answer causal questions (e.g., "caused" vs. "moved by itself"). Second, causal ratings varied with cue realism and modality even when timing was held constant (Experiment 3). This raises a related question: were participants judging causality, or were they instead judging realism or cross-modal fit? Our view is that, in this task, this is not a clean either/or distinction. When people judge whether one object caused another to move, they do not rely solely on timing; they also consider whether the event looks, sounds, and feels like a plausible physical interaction. We therefore treat realism not as a competing interpretation of the task, but as one of the inputs to causal judgment itself. In other words, the present study does not ask whether individual cues, by themselves, feel realistic, but instead whether how realistic these cues feel affects how people judge physical causation. Our results suggest that observers integrated temporal information with expectations about contact, force, and realism to decide whether an event was causal. Thus, while our design does not isolate causality from realism in a strict sense, it shows that observers use realism as part of the evidence they consider when deciding whether an event was causal.

To our knowledge, this work is the first to explore multisensory integration in causal perception beyond the traditional focus on vision and sound. Including haptic feedback allows us to explore, for the first time, what happens to people's causal perception when more than two signals are present. We find that participants are more likely to say that an event was causal (Experiments 1 and 2) and that they give higher causal judgments (Experiment 3), the more sources of evidence point toward a causal event (as long as all the events occur within a window of time). We also observe diminishing returns with additional causal signals. The difference between two and one source of information is larger than the difference between three and two sources. Haptic information – just like information from the other modalities – contributes significantly to causal judgments, and its impact is most pronounced when the haptic signals are physically realistic.

Beyond research, haptic signal and device designers for video games and other consumer haptic products routinely augment visual events with tactile and auditory feedback. Although there is an active literature on haptic rendering and realism, many practical design decisions in deployed systems are still guided by domain-specific heuristics and internal testing. Our study complements this work by isolating the effects of timing, sensory modality, and realism on judgments of physical causality in a controlled collision paradigm. Rather than proposing a new rendering method, we provide quantitative guidance about when multisensory cues strengthen causal impressions and when additional signals offer smaller incremental gains. This evidence-based framing offers a principled foundation for future multisensory applications in which conveying a clear sense of physical interaction is important.

6.1. Causal perception vs. Causal inference

Causal perception is an instantaneous, often subconscious experience when sensory events, such as visual, auditory, and haptic cues, align in time and space. Michotte's (1946/1963) psychophysical experiments show that when two objects appear to make contact, and one begins moving afterward, people rely on spatial and temporal cues to perceive causation.

In contrast, causal inference is a more deliberative process (Woodward, 2011) in which causality is determined by analyzing relationships and explicitly accounting for uncertainty. However, causal inference models have been used previously to capture perceptual judgments. For example, Körding et al. (2007) found that a Bayesian model of multisensory cue integration captured participants' judgments about where a visual and auditory signal came from.

Similarly, we model participants' causal judgments in our experiments by using a Bayesian inference model that integrates various sensory cues (visual, auditory, kinesthetic, vibrotactile) and assigns probabilities to different possibilities based on the timing and realism of each signal. This model supports a structured evaluation of causality by probablistically integrating multiple modalities.

We are not sure whether what participants are doing in our task is better called "causal perception" or "causal inference". Our participants' causal judgments were likely shaped by both processes: perception and inference. Inferences about what happened may play a more important role, particularly when multiple signals conflict with one another and need to be resolved (Kruschke & Fragassi, 2019; Shams & Beierholm, 2010). A similar point applies to realism. In physical event judgments, people may rely not only on when cues occur, but also on whether those cues match their expectations for how a collision should look, sound, and feel. We therefore do not treat realism as an alternative to causal judgment, but as one of the factors that may help constitute it in the present task. Even so, our experiments show that haptic and auditory signals systematically shift observers' judgments of which object caused another to move. However, because we did not collect separate realism judgments, we cannot determine from the present data the extent to which these cues shaped causal judgments by altering how realistic or cross-modally coherent the event seemed, versus other aspects of multisensory processing. This distinction is important because causal judgments and realism judgments are related, but not identical. Prior work has shown that observers sometimes judge events as causal that are not fully physically realistic (Bechlivanidis et al., 2019). In our task, we therefore treat realism as one source of evidence that may contribute to causal judgment, rather than as a separate outcome cleanly dissociable from it. Future work could more directly dissociate automatic multisensory integration from higher-level reasoning by asking participants to ignore specific cues or by comparing judgments based on direct perceptual displays with judgments based solely on verbal descriptions of the same event structures (van Buren & Scholl, 2018, 2025). Such designs would help clarify which aspects of multisensory causal judgment are obligatory consequences of perceptual integration and which depend more on explicit reasoning or task demands.

6.2. Limitations and future work

One limitation of our work is that the controlled nature of the experimental setup does not capture the full complexity of real-world environments, where sensory cues are noisier and less consistent. Our controlled setup is crucial for isolating the effects of specific variables, but it limits generalizability. In particular, our haptic cues were designed to represent controlled, temporally localized evidence associated with contact and the resulting motion rather than full naturalistic force trajectories. Thus, we expect the present findings to generalize best to short, discrete contact events – especially for interfaces and in virtual environments where audio-haptic signals mark collisions – and less so to extended collisions, collisions with deformable materials, or fully interactive manipulation tasks. While our participants were passive observers, future work should explore causal perception and inference in interactive settings where participants choose when and how to act (see Bramley et al., 2018). Studies in real-time, interactive environments, where participants actively manipulate objects and receive instant feedback, would help better explore the interplay between action, feedback, and causal judgments.

Our experiments were run on a relatively small number of human participants. However, these participants completed many experimental trials. While causal perception experiments that use visual and auditory information can now be run online (which makes it much easier to recruit large numbers of participants), this is not possible for exploring the role of haptic feedback in causal perception (as people do not have haptic controllers at home).

Future research could address these limitations and expand on our findings in several ways. Experiments with more complex and dynamic environments, that is, not only spatial and temporal information but richer impact-based and texture-based cues, could yield a more nuanced understanding of how realism influences multisensory causal perception. Besides increasing sample size, investigating individual differences, such as prior experience with various sensory modalities, would also yield more robust findings. Finally, including neuroimaging techniques could provide valuable insights into the underlying cognitive mechanisms involved in multisensory integration and a more comprehensive understanding of how humans perceive causality.

7. Conclusion

In this paper, we showed that haptic information is an important cue for judging physical causality. Across three psychophysical studies, extending the classic launching paradigm with kinesthetic and vibrotactile haptic feedback revealed that observers do not rely on vision alone when deciding whether one object caused another to move. Instead, they integrate multisensory evidence, with causal judgments shaped most strongly by whether the cues are temporally aligned and cohere with a physically plausible collision. People's judgments are sensitive to the uncertainty associated with each signal and are well explained by a Bayesian model that infers whether the same common cause generated each piece of evidence. More broadly, the present results suggest that, in judgments of physical events, realism is not merely cosmetic; it serves as evidence that people use when deciding whether one event caused another. In this way, the present work complements prior haptics research on realism by showing that realism contributes not only to whether an interaction feels physically plausible and well-matched across modalities, but also to whether events are perceived as causal. Better understanding the role of haptic feedback in causal perception and inference matters. For example, real-time force feedback may improve performance in robotics and teleoperation by accelerating fault detection. The model we develop takes a first step of laying out how disparate sensory feedback can lead to a unified sense of "what caused what".

Acknowledgments

We want to thank Phillip Wolff for the early discussions on the project. In addition, this work was supported by an NSF GRFP under Grant No. DGE-1656518, a Stanford Graduate Fellowship (SGF), and a Stanford Human-Centered Artificial Intelligence (HAI) Seed Grant. TG was supported by a grant from Cooperative AI. All experiments were approved by Stanford's Institutional Review Board. We presented earlier versions of this work in Chase (2023) and Chase et al. (2023, 2021a, 2021b).

Appendix. Model

A.1. Implementation

For each Markov Chain Monte-Carlo sample $s \in \{1 \dots N\}$:

1. **Draw causal flag.** This is a fair coin flip (50%) modified by the relative realism of the cues.

$$C^{(s)} \sim \text{Bernoulli}(0.5 \times \delta_{real}^{\sum RealCues} \times \delta_{unreal}^{\sum Cues - \sum RealCues}) \tag{3}$$

2. **Draw multimodal cue times.** For each modality m in the set of modalities $\mathcal{M}$ (vision: start and stop of motion, audition, vibration, force), draw

$$t_m^{(s)} \sim \begin{cases} \mathcal{N}(0, \sigma^2_{T_{window}}) & \text{if } C^{(s)}=1 \\ \mathcal{U}(T_{\text{left}}, T_{\text{right}}) & \text{if } C^{(s)}=0 \end{cases} \tag{4}$$

When the sample is causal, one time will be drawn and then used for all modalities, whereas when non-causal, the time is redrawn for each modality.

3. **Calculate the likelihood of the timing of the true cues given the simulated cues.** For each modality, consider the likelihood of the true cue time (t_{true}) given the draw time ($t_m^{(s)}$) and the noise associated with that modality (σ_m^2).

$$p_m^{(s)} \sim t_{true} | \mathcal{N}(t_m^{(s)}, \sigma_m^2) \tag{5}$$

4. **Draw spatial gap.** For each sample, consider the true gap (x_{true}).

$$x^{(s)} = \begin{cases} 0 & \text{if } C^{(s)}=1 \\ x_{true} & \text{if } C^{(s)}=0 \end{cases} \tag{6}$$

5. **Observe the likelihood of the drawn spatial gap to the real gap.**

$$p_x^{(s)} \sim x_{true} | \mathcal{N}(x^{(s)}, \sigma_x^2) \tag{7}$$

The sample-wise likelihood $p_m^{(s)}\ \forall m \in \mathcal{M}$ and $p_x^{(s)}$ of the observed data **d**, given branch $C^{(s)}$, is the product of the temporal and spatial terms.

Aggregating the likelihood across N samples yields a numerical estimate of the posterior probability of the study conditions being causal given the observed data and priors.

Unless stated otherwise, we use $N=10^5$ samples per condition, which produced stable posterior estimates (<0.5% coefficient of variation).

A.2. Additional model results

In this section, we present a single table of grid search details for our work (Table A.1). Then we report three tables with results from Experiments 1&2 (Table A.2), Experiment 3 (Table A.3), and related work (Table A.4). Finally, we include four sets of graphs showing the results from the lesions across all Experiments (see Figs. A.1–A.4).

Table A.1
Grid search across parameters for Experiments 1, 2, and 3. The/symbol indicates that no values were searched over, as this parameter is irrelevant to the dataset.

Parameters	Experiment 1&2	Experiment 3
σ_{visual} [ms]	[50, 60, …, 150]	[100, 150, …, 250]
$\sigma_{sensory}$ [ms]	[150, 160, …, 250]	[250, 300, …, 400]
σ_x [cm]	[0.75, 1.00, …, 2.75]	[0.5, 1.0, …, 2.0]
$\sigma_{T_{window}}$ [ms]	[200, 210, …, 300]	[100, 150, …, 450]
δ_{real}	/	[1.0, 1.1, 1.2]
δ_{unreal}	/	[0.7, 0.8, 0.9, 1.0]

Table A.2
Model fitting results for Experiments 1 and 2. Each row refers to a different model: best fit (best overall, with all parameters free), no-window (lesioned, $\sigma_{T_{window}}$ held at 1), and ∞-window (lesioned, $\sigma_{T_{window}}$ held at 1000). Correlation was reported using Pearson's r, computed from average human data compared to model output for each study condition. Log-Likelihood was calculated for each experiment independently, comparing aggregate human data to model output. All σ's are the fit parameters of the model (T_{window} was controlled for in the lesioned case).

Model	Correlation	Log-Likelihood		σ_{visual}	$\sigma_{sensory}$	σ_x	$\sigma_{T_{window}}$
	Pearson's r	Exp 1	Exp 2	[ms]	[ms]	[cm]	[ms]
Best fit	0.970	−2506	−2292	70	210	1.25	240
No-Window	0.962	−2705	−2352	80	250	1.00	1
∞-window	0.937	−2749	−2499	90	150	2.00	1000

Table A.3
Model fitting results for Experiment 3. Each row refers to a different model: best fit (best overall, with all parameters free), no-window (lesioned, $\sigma_{T_{window}}$ held at 1), and ∞-window (lesioned, $\sigma_{T_{window}}$ held at 1000). Correlation was reported using Pearson's r, computed from average human data compared to model output for each study condition. Root Mean Squared Error (RMSE) was calculated for each experiment independently, comparing aggregate human data to model output. All σ's and δ's are the fit parameters of the model (T_{window} was controlled for in the lesioned case). Due to space constraints, this table uses shortened versions for the σ parameters: σ_v for σ_{visual}, σ_s for $\sigma_{sensory}$, and σ_{T_w} for $\sigma_{T_{window}}$.

Model	Correlation	RMSE		σ_v	σ_s	σ_x	σ_{T_w}	δ_{real}	δ_{unreal}
	Pearson's r	Exp 3A	Exp 3B	[ms]	[ms]	[cm]	[ms]		
Best fit	0.950	2.398	2.291	200	400	2.0	400	1.1	1.0
No-Window	0.921	2.418	2.338	200	400	0.5	1	1.1	0.9
∞-Window	0.951	2.397	2.288	150	350	1.0	1000	1.1	0.9

Table A.4
Model fitting results for related work. Each row refers to a related work: Guski and Troje (2003), Meding et al. (2020) and Wang et al. (2018). Correlation was reported using Pearson's r, computed from average human data compared to model output for each study condition. Root Mean Squared Error (RMSE) was calculated for each experiment independently, comparing aggregate human data to model output. All σ's and δ's are the fit parameters of the model. The / symbol represents that realism was not part of these study conditions, so these values were held at 1, and not searched across.

Related work	Correlation Pearson's r	RMSE	σ_{visual} [ms]	$\sigma_{sensory}$ [ms]	σ_x [cm]	$\sigma_{T_{window}}$ [ms]	δ_{real}	δ_{unreal}
Guski &Troje	0.908	0.77	90	40	0.30	250	/	/
Wang et al.	0.984	0.52	200	250	0.05	250	/	/
Meding et al.	0.982	0.29	350	50	0.04	400	1.0	0.9

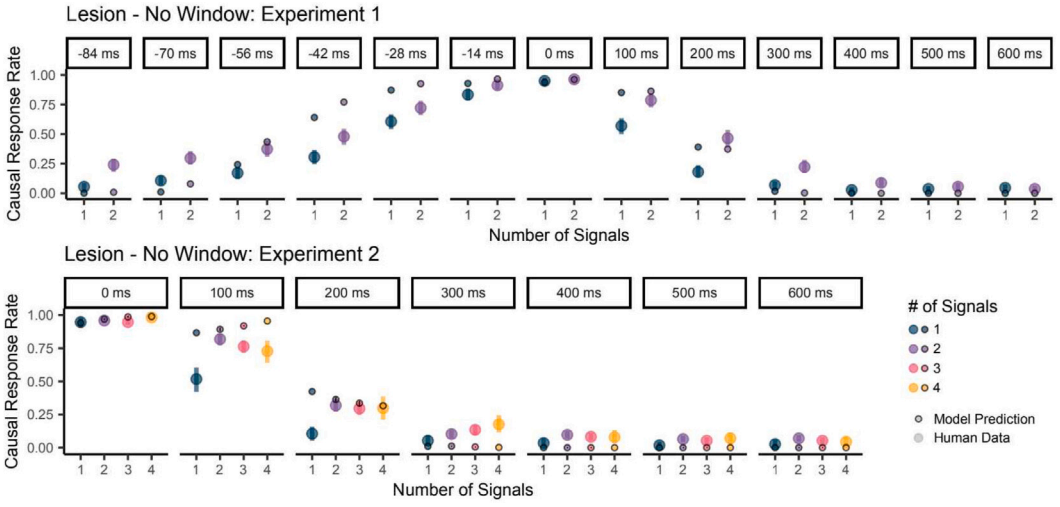


Fig. A.1. Experiments 1 & 2 lesion no window results: *x*-axis shows temporal offset in milliseconds, *y*-axis shows causal response where 1 indicates *causal* and 0 indicates *non-causal*, color indicates the number of sensory signals present, and shape indicates if it is our model's prediction or human experiment data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

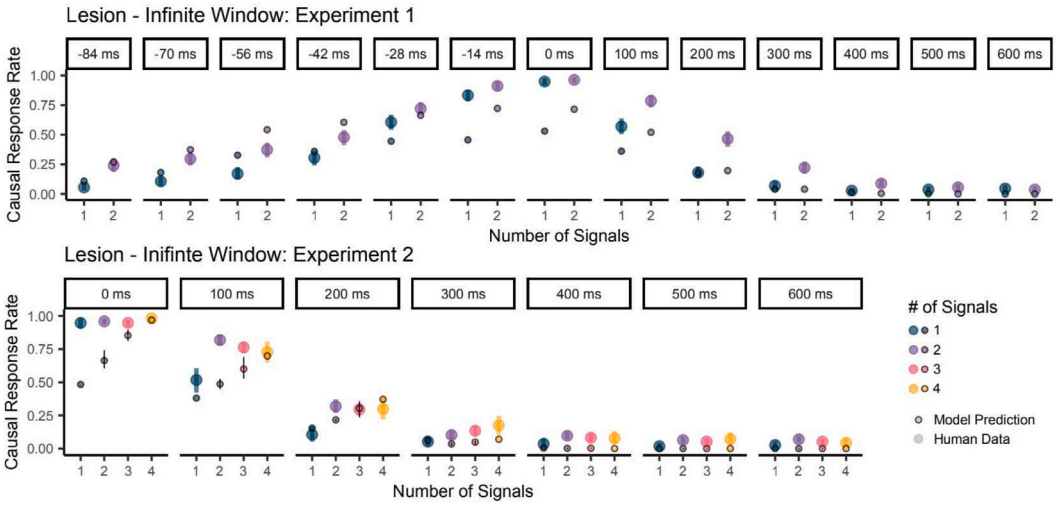


Fig. A.2. Experiments 1 & 2 lesion infinite window results: *x*-axis shows temporal offset in milliseconds, *y*-axis shows causal response where 1 indicates *causal* and 0 indicates *non-causal*, color indicates the number of sensory signals present, and shape indicates if it is our model's prediction or human experiment data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

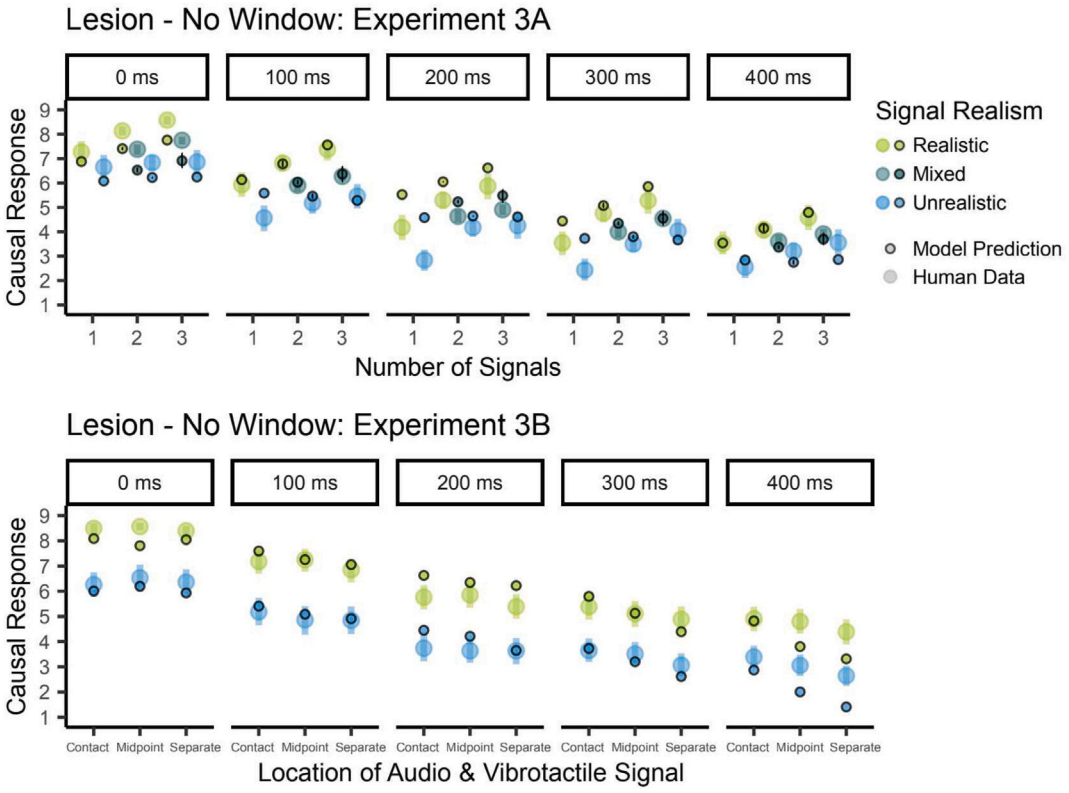


Fig. A.3. Experiments 3 lesion no window results: *x*-axis shows temporal offset in milliseconds, *y*-axis shows causal response where 9 indicates *causal* and 1 indicates *non-causal*, color indicates the relative realism of the sensory signals, and shape indicates if it is our model's prediction or human experiment data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

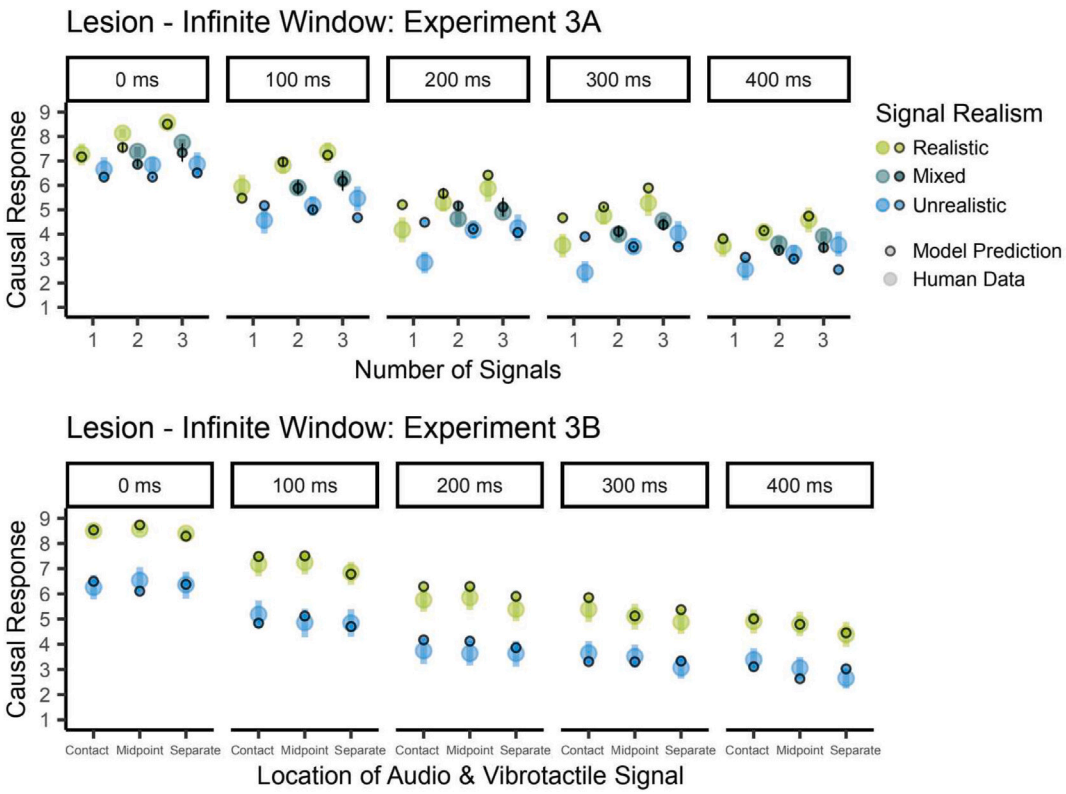


Fig. A.4. Experiment 3 lesion infinite window results: *x*-axis shows temporal offset in milliseconds, *y*-axis shows causal response where 9 indicates *causal* and 1 indicates *non-causal*, color indicates the relative realism of the sensory signals, and shape indicates if it is our model's prediction or human experiment data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Agrawal, T., & Schachner, A. (2022). Hearing water temperature: Characterizing the development of nuanced perception of sound sources. *Developmental Science*, Article e13321.

Baillargeon, R., Kotovsky, L., & Needham, A. (1995). The acquisition of physical knowledge in infancy. In D. Sperber, D. Premack, & J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 79–116). Clarendon Press/Oxford University Press.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Bechlivanidis, C., & Lagnado, D. A. (2013). Does the "Why" tell us the "When"? *Psychological Science*, *24*(8), 1563–1572.

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

Bechlivanidis, C., Schlottmann, A., & Lagnado, D. A. (2019). Causation without realism. *Journal of Experimental Psychology: General*, http://dx.doi.org/10.1037/xge0000602.

Berger, C. C., Gonzalez-Franco, M., Ofek, E., & Hinckley, K. (2018). The uncanny valley of haptics. *Science Robotics*, *3*(17), eaar7010.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., & Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive Psychology*, *105*, 9–38.

Chase, E. D. Z. (2023). *In touch with causation: The role of haptics in multisensory phenomenal causality* [Unpublished doctoral dissertation], Stanford University.

Chase, E. D. Z., & Follmer, S. (2019). Differences in haptic and visual perception of expressive 1dof motion. In *ACM Symposium on Applied Perception 2019* (pp. 1–9).

Chase, E. D. Z., Gerstenberg, T., & Follmer, S. (2023). Realism of visual, auditory, and haptic cues in phenomenal causality. In *2023 IEEE World Haptics Conference* (pp. 306–312). IEEE.

Chase, E. D., & O'Malley, M. K. (2024). The interplay of vision and referred haptic feedback in vr environments. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer.

Chase, E. D., Sullivan, D. H., & O'Malley, M. K. (2025). Hands-on or hands-off? Active touch influences multisensory perception of referred haptics. In *2025 IEEE World Haptics Conference* (pp. 321–331). IEEE.

Chase, E. D. Z., Wolff, P., Gerstenberg, T., & Follmer, S. (2021a). A causal feeling: How kinesthetic haptics affects causal perception. In *2021 IEEE World Haptics Conference*. IEEE, 347–347.

Chase, E. D. Z., Wolff, P., Gerstenberg, T., & Follmer, S. (2021b). In touch with causation: Understanding the impact of kinesthetic haptics on causality. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Choi, H., & Scholl, B. J. (2004). Effects of grouping and attention on the perception of causality. *Perception & Psychophysics*, *66*(6), 926–942. http://dx.doi.org/10.3758/bf03194985.

Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception*, *35*(3), 385–399.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*(6), 1000–1009.
Conti, F., Barbagli, F., Morris, D., & Sewell, C. (2005). Chai 3d: An open-source library for the rapid development of haptic scenes. *IEEE World Haptics*, *38*(1), 21–29.
Cravo, A. M., Claessens, P. M., & Baldo, M. V. (2009). Voluntary action and causality in temporal binding. *Experimental Brain Research*, *199*, 95–99.
Di Luca, M., & Mahnan, A. (2019). Perceptual limits of visual-haptic simultaneity in virtual reality interactions. In *2019 IEEE World Haptics Conference* (pp. 67–72). IEEE.
Diederich, A., & Colonius, H. (2015). The time window of multisensory integration: Relating reaction times and judgments of temporal order. *Psychological Review*, *122*(2), 232.
Dinh, H. Q., Walker, N., Hodges, L. F., Song, C., & Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings IEEE Virtual Reality (cat. no. 99CB36316)* (pp. 222–228). IEEE.
Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.
Gerstenberg, T., Siegel, M., & Tenenbaum, J. (2018). What happened? Reconstructing the past from vision and sound. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
Goodman, N. D., Tenenbaum, J. B., & Contributors, T. P. (2016). *Probabilistic models of cognition* (2nd ed.). http://probmods.org/v2. [Accessed 22 August 2025].
Guski, R., & Troje, N. F. (2003). Audiovisual phenomenal causality. *Perception & Psychophysics*, *65*(5), 789–800.
Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385.
Israr, A., Kim, S.-C., Stec, J., & Poupyrev, I. (2012). Surround haptics: Tactile feedback for immersive gaming experiences. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (pp. 1087–1090).
Israr, A., & Poupyrev, I. (2011). Tactile brush: Drawing on skin with a tactile grid display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2019–2028).
Johnson, V., Hsu, W.-Y., Ostrand, A. E., Gazzaley, A., & Zanto, T. P. (2020). Multimodal sensory integration: Diminishing returns in rhythmic synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(10), 1077.
Jones, L. A., & Tan, H. Z. (2012). Application of psychophysical techniques to haptic research. *IEEE Transactions on Haptics*, *6*(3), 268–284.
Kominsky, J. F., & Scholl, B. J. (2020). Retinotopic adaptation reveals distinct categories of causal perception. *Cognition*, *203*, Article 104339.
Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science*, *28*(11), 1649–1662.
Körding, K., Beierholm, U., Ma, W., Quartz, S., Tenenbaum, J., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, *2*(9), Article e943.
Kruschke, J. K., & Fragassi, M. M. (2019). The perception of causality: Feature binding in interacting objects. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 441–446). Routledge.
Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, *21*(10), 749–759.
Kuchenbecker, K. J., Fiene, J., & Niemeyer, G. (2006). Improving contact realism through event-based haptic feedback. *IEEE Transactions on Visualization and Computer Graphics*, *12*(2), 219–230.
Legaspi, R., & Toyoizumi, T. (2019). A Bayesian psychophysics model of sense of agency. *Nature Communications*, *10*(1), 4250.
Magnotti, J. F., Ma, W. J., & Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. *Frontiers in Psychology*, *4*, 798.
Mayrhofer, R., & Waldmann, M. R. (2016). Causal agency and the perception of force. *Psychonomic Bulletin & Review*, *23*(3), 789–796. http://dx.doi.org/10.3758/s13423-015-0960-y.
Meding, K., Bruijns, S. A., Schölkopf, B., Berens, P., & Wichmann, F. A. (2020). Phenomenal causality and sensory realism. *I-Perception*, *11*(3), Article 2041669520927038.
Michotte, A. (1946/1963). *The perception of causality*. Basic Books.
Okamura, A. M., Cutkosky, M. R., & Dennerlein, J. T. (2002). Reality-based models for vibration feedback in virtual environments. *IEEE/ASME Transactions on Mechatronics*, *6*(3), 245–252.
Outa, J., Zhou, X. J., Gweon, H., & Gerstenberg, T. (2022). Stop, children what's that sound? Multi-modal inference through mental simulation. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual conference of the cognitive science society* (pp. 1359–1366). Cognitive Science Society.
Park, G., & Choi, S. (2016). A physics-based vibrotactile feedback library for collision events. *IEEE Transactions on Haptics*, *10*(3), 325–337.
Park, C., Park, J., Oh, S., & Choi, S. (2019). Realistic haptic rendering of collision effects using multimodal vibrotactile and impact feedback. In *2019 IEEE World Haptics Conference* (pp. 449–454). IEEE.
Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), 411–437.
Schachner, A., & Kim, M. (2018). Entropy, order and agency: The cognitive basis of the link between agents and order. http://dx.doi.org/10.31234/osf.io/kdr9m, PsyArXiv. Retrieved from psyarxiv.com/kdr9m.
Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309.
Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*(6614), 308–308.
Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432.
Shin, S., & Choi, S. (2018). Effects of haptic texture rendering modalities on realism. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (pp. 1–5).
Smith, K., Mei, L., Yao, S., Wu, J., Spelke, E., Tenenbaum, J., & Ullman, T. (2019). Modeling expectation violation in intuitive physics with coarse probabilistic object representations. *Advances in Neural Information Processing Systems*, *32*.
Spelke, E. S. (2013). Where perceiving ends and thinking begins: The apprehension of objects in infancy. In *Perceptual development in infancy* (pp. 209–246). Psychology Press.
Srinivasan, M. A., Beauregard, G. L., & Brock, D. L. (1996). The impact of visual information on the haptic perception of stiffness in virtual environments. *vol. 15281*, In *ASME International Mechanical Engineering Congress and Exposition* (pp. 555–559). American Society of Mechanical Engineers.
van Buren, B., & Scholl, B. J. (2018). *vol. 47*, *Visual illusions as a tool for dissociating seeing from thinking: A reply to Braddick (2018)*. London, England: SAGE Publications Sage UK, No. (10–11).
van Buren, B. F., & Scholl, B. J. (2025). The blindfold test: Helping to decide whether an effect reflects visual processing or higher-level judgment. *Attention, Perception, & Psychophysics*, *87*(2), 445–457.
Vogels, I. M. (2004). Detection of temporal delays in visual-haptic interfaces. *Human Factors*, *46*(1), 118–134.
Wang, D., Kubrlcht, J., Zhu, Y., Lianq, W., Zhu, S.-C., Jiang, C., & Lu, H. (2018). Spatially perturbed collision sounds attenuate perceived causality in 3d launching events. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces* (pp. 259–266). IEEE.
White, P. A. (1988). Causal processing: Origins and development. *Psychological Bulletin*, *104*(1), 36.
White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin*, *108*(1), 3–18. http://dx.doi.org/10.1037/0033-2909.108.1.3.
White, P. A. (2009). Perception of forces exerted by objects in collision events. *Psychological Review*, *116*(3), 580–601.
White, P. A. (2012). The experience of force: The role of haptic experience of forces in visual perception of object motion and interactions, mental simulation, and motion-related judgments. *Psychological Bulletin*, *138*(4), 589–615.

White, P. A., & Milne, A. (1997). Phenomenal causality: Impressions of pulling in the visual perception of objects in motion. *The American Journal of Psychology*, *110*(4), 573.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136*(1), 82–111.

Wolff, P., & Shepard, J. (2013). Causation, touch, and the perception of force. *vol. 58*, In *Psychology of learning and motivation* (pp. 167–202). Elsevier.

Wolff, P., & Thorstad, R. (2017). Force dynamics. In *The oxford handbook of causal reasoning* (pp. 147–168). Oxford University Press New York.

Woodward, J. (2011). Causal perception and causal cognition. In *Perception, causation, and objectivity* (pp. 229–263). Oxford University Press Oxford.

Wu, S. A., Brockbank, E., Cha, H., Fränken, J.-P., Jin, E., Huang, Z., Liu, W., Zhang, R., Wu, J., & Gerstenberg, T. (2024). Whodunnit? Inferring what happened from multimodal evidence. In L. K. Samuelson, S. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Pproceedings of the 46th annual conference of the cognitive science society* (pp. 1809–1816).