Realism of Visual, Auditory, and Haptic Cues in Phenomenal Causality

Elyse D. Z. Chase
Dept. of Mechanical Engineering
Stanford University
Stanford, CA, USA
elysec@stanford.edu

Tobias Gerstenberg
Dept. of Psychology
Stanford University
Stanford, CA, USA
gerstenberg@stanford.edu

Sean Follmer
Dept. of Mechanical Engineering
Stanford University
Stanford, CA, USA
sfollmer@stanford.edu

***Abstract*—Interacting in real environments, such as manipulating objects, involves multisensory information. However, little is known about how multisensory cue characteristics help us determine what has occurred in a scene, including whether two events were causally linked. In virtual environments, the number of sensory modalities present and levels of realism often vary. In this work, we explore what role multisensory information and physical realism play in people's causal perception. So far, haptic cues have rarely been studied in causal perception. Here, we combined visual, auditory, and haptic cues in a psychophysical study in which participants were asked to judge whether one billiard ball caused another to move. We manipulated the temporal delay between cause and effect events, and the physical realism of each cue. While temporal delays generally decreased causal judgments, the number of multisensory cues and their physical realism increased causal judgments. We highlight the implications of this work for building immersive environments.**

***Index Terms*—realism, multisensory, causality, causal perception, vision, audio, vibration, haptics**

I. INTRODUCTION

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 happened, as is feeling the reaction force and impact-based vibration. Additionally, haptic information plays a critical role in how we learn about the world. For example, children often bang objects against one another to observe the effect [1], [2]. As we seek to make interactive, multi-modal virtual environments, we must consider how different modalities influence causal perception – the impression that one event is the result of the occurrence of another [3].

Think of a ball that collides with a second ball, causing it to move (see Fig. 1A). What makes you see this event as causal? The perceived realism of sensory information matters. For example, hearing "buzz" may be less indicative than hearing "crash" upon seeing a collision. Researchers have shown that people use realism in their criteria for making causal judgments about events [4], [5], but see also [6]. Here, we investigate how multisensory information that is either physically realistic or not affects causal perception. Specifically, we explore whether having either more realistic, or just more, signals increases the likelihood that people perceive a collision as causal.

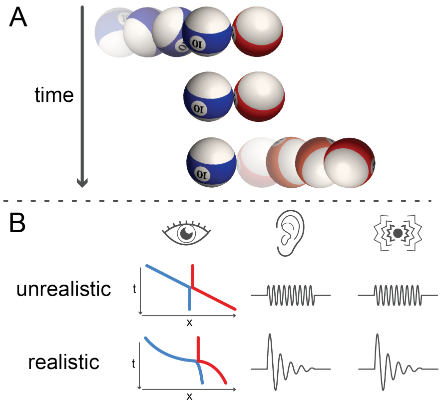

Fig. 1. (A) An overview of the study in which the blue ball starts rolling towards the red ball, makes contact with the red ball, stops, and then the red ball rolls away (B) Conditions for realism of sensory cues

Little work on causal cognition to date, has explored the role of haptic feedback; [7] found that participants detected a haptic signal more quickly when viewing causal events versus non-causal events (see also [8]). Prior work on the physical realism has explored how delays and mismatch between haptic and other sensory information affect the extent to which participants feel immersed in Virtual Reality (VR) environments [9]. 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.

To our knowledge, no work so far has integrated these perspectives and studied what role physical realism plays in multisensory causal perception. Here, we consider vision, audio, and vibrotactile haptics with both realistic and unrealistic cues to determine how these factors affect people's causal perception. We answer the following questions:

- Are people more likely to judge an event as causal when they have evidence from multiple sensory sources?
- Does it matter whether information from different senses is physically realistic?

To study these questions, we use an extension of Michotte's launching paradigm [3] in which a *launcher* touches a stationary object, after which the launcher stops and the stationary object begins to move (see Fig 1A). In our study, two simulated

Supported by Stanford Human-Centered Artificial Intelligence seed grant.

billiard balls roll along a table while additional auditory and haptic feedback occurs upon collision of the objects. After watching the animation, participants rate the probability that the two objects were causally linked. Our results show that participants' causal judgments decrease the longer the temporal delay is between when the first ball stops and the second ball starts to move. We also find that causal judgments increase when multiple sensory signals are present, and that judgments are highest when all of the signals are physically realistic. We discuss the implications of these findings for the design of multisensory virtual environments.

II. RELATED WORK

A. Causal Perception

Humans naturally perceive causal relationships. This was first formally studied by Michotte, in which a *launcher* hits a stationary object with some amount of delay and people treat this as a casual scenario, deemed *phenomenal causality* [3]. In these studies, many different visual animations were explored, and later work has extended it – but generally has remained in the purely visual domain [10]. However, in our everyday lives, we perceive the causal events through many senses. We see, hear, and feel what happens.

Some have considered multisensory cues in causal perception by adding audio cues at different points during the visual contact of two circles [11]. Additional cues, either auditory or visual (the object changes color momentarily), increased causal judgments beyond visual information alone.

Little research so far has explored what role haptics plays in causal perception. One project explored if haptic feedback primed user responses when viewing causal scenarios, and found that it did have an effect on response times – with reduced response times for trials that had causal scenarios [7]. More recent work has shown that kinesthetic force feedback applied in unison with the motion of the second object moving away increases people's perception of causality [8]. However, no work has examined how multisensory information, beyond vision and audio, affects our causal perception.

B. Realism

Within the causal perception literature, work has shown that people are using realism as part of their criteria for making causal judgments [4], [5]. More recent work has demonstrated that people sometimes judge physically unrealistic event sequences as causal, such as when, for example, one ball stops at a distance and the other ball moves at the same time the first one stopped [6]. However, when they later see a more canonical causal collision event, they now judge these other events as less causal. This suggests that participants are initially uncertain about what experimenters mean when asking them to judge causality, and that being presented with canonical collision events leads them to interpret the question to be about physically realistic causation. This highlights the importance of context and the flexibility of people's definitions of causality.

People may be using kinematic information to match stored representations about forces and causality [1], [2], [6]. This could motivate the idea that more realistic cues would lead to higher ratings of causal perception.

Researchers have also studied how visual rendering and physical dynamics affect people's judgments about collision events. 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 rendering (balls that rotated and undergo elastic collision) resulted in lower causal ratings compared to a 2D rendering, but adding physical dynamics (rotation and inelastic collision) to a realistic 3D rendering increased causal ratings above all conditions [12]. Work in haptics has considered the realism of cues and discovered an uncanny valley. Participants used a VR system with vibrotactile stimuli that provide spatial haptic effects via two controllers [9]. Participants' rated sense of presence in VR 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 further understand how it affects causal perception.

C. Haptic Delay

Much work in haptics has focused on delay, asking directly about the perception of visual and haptic signals being synchronous or asynchronous. Delay is important for creating realistic scenarios and is one of the main variables controlled for in our study. Humans are quite acute at determining the delay between a visual collision and feeling feedback through a joystick, on the order of 45 ms [13]. Later work found similar results for tapping in VR; people judged visual and haptic information as synchronous if haptic feedback occurred no more than 50 ms after the event. However, when haptic feedback preceded visual contact the delay needed to be less than 15 ms [14]. Several works have extended these concepts to determine the effects that haptic delay could have upon the perceived stiffness of objects [15]–[17]. As delay is common in many computer and VR systems, this is an important measure for haptic designers to consider. In this paper, we take a different approach, focusing on the perception of causality rather than the perception of asynchrony.

III. METHOD

A. Parameter Choice

To understand the effects of more realistic cues on causal perception, we selected two different cues for each sensory modality: vision, audition, and touch[1]. Here, we use realistic to mean physical, impact-based dynamics for rigid objects. We discuss how we implemented each cue in turn.

1) Vision: The unrealistic visual sequence was designed based upon an early causal study [3], in which both objects move at a constant velocity and undergo a perfectly elastic collision. While the way in which each object rolls looks realistic, the collision itself doesn't look realistic. Upon contact

---

[1] github.com/ShapeLab/realism_multisensory_causality

Fig. 2. A participant seated at the study setup with the haptic device in front of a computer screen. Users grasp the handle of the device, which provides vibrotactile feedback, while receiving audio cues via headphones.

the first object immediately stops rolling (without sliding), and the second one immediately starts rolling.

The realistic dynamics were modeled as a sphere rolling with drag. Upon contact, an inelastic collision occurs after which both balls slip until they again begin to roll.

In both conditions[2], the objects have consistent initial positions which ensures that contact between the two balls will occur at the center of the computer screen.

2) Audio: The realistic, impact-based audio cue[3] was edited to be of a set duration and amplitude. The unrealistic audio cue was a non-decaying sinusoid of reduced amplitude, in order to match the overall intensity of the realistic cue.

3) Vibration: We used work on contact realism for event-based haptics [18], [19] in order to design a vibration that felt like a collision between two billiard balls. This vibration 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.

For the unrealistic vibration, a sinusoid without decay was used and the frequency was set to match the unrealistic audio cue (240 Hz). The amplitude was decreased to normalize perceived intensity between cues. All audio and vibration cues were held at a constant length of 150 ms.

B. Device

Our system consists of a haptic device with a handle that contains a voice coil motor (VCM: Dayton Audio DAEX19CT-4, Amplifier: TPA3116D2) and micro-controller (Teensy LC) as well as a computer screen, headphones, and keyboard (Fig. 2). The study was built using CHAI3D [20], which affords the fast refresh rates necessary for haptic interactions. The monitor (Dell P2715Q 27", 3840 x 2160) displayed the scene, while the haptic device was secured to the table. Users listened to white noise during the study – which was used to mask unintentional sounds from the hardware – in addition to hearing the programmed audio signals.

IV. HYPOTHESES

To address our questions about the role of multisensory cues and realism on causal perception, we preregistered five hypotheses[4]. The first hypothesis tests whether our work replicates prior results in causal perception, which have shown that temporal delay effects causal judgments [8], [11].

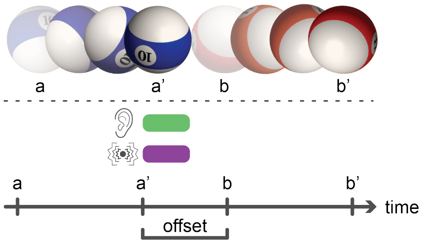

Fig. 3. Timeline highlighting key study events – (a) first ball starts motion; (a') first ball stops motion while audio, vibration, and offset begin; (b) second ball starts motion; (b') second ball stops motion

(H1) *Participants' causal judgments decrease as the temporal offset increases.*

We predict that people are more likely to judge an event as causal with evidence from multiple sensory sources:

(H2) *Causal judgments are higher in multisensory conditions (containing either audio, vibration, or both) compared to vision alone.*

Relating to our second question – does it matter whether information from different senses is physically realistic? – we consider the following three hypotheses. Additional sensory cues provide synchronous temporal information, but in some cases (realistic cues) the information is more diagnostic of what has occurred in the scene. Thus,

(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.*

Related work has shown a possible uncanny valley in haptics [9] as well as for phenomenal causality [12], therefore:

(H5) *Causal judgments are lower in conditions with mismatched cues (realistic visuals with unrealistic sensory signals) compared to the realistic vision-only condition.*

V. STUDY

A total of 22 right-handed participants (*age*: M = 25, SD = 4; *sex*: 9 female, 13 male) completed the study in accordance with our IRB and were compensated $20.

A. Experimental Setup

1) Design: The study 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 [21]). Temporal offsets of duration greater than 100 ms are not something

---

[2] For equations of motion and videos, see the Supplementary Material.

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

[4] doi.org/10.17605/OSF.IO/CYD8J

we experience in the world, and thus disrupt impressions of causality, as well as revealing thresholds for causal perception. Audio and vibration cues were played when the balls contacted (point a' in Fig. 3). All conditions were repeated 4 times for a total of 360 trials. Trial order was randomized within each repetition to ensure people saw different conditions comparably throughout the study.

2) Procedure: Participants were seated at a table with a keyboard, mouse, computer screen, and rigidly mounted haptic device (Fig. 2). Noise-canceling headphones played white noise (to remove any distraction from external noise or sound made due to vibrations) in addition to audio cues from the study. During each trial, participants held onto the handle of the device using a precision grip.

Before beginning, participants were guided through two sets of practice trials that served to both introduce them to the system input and provide a range of sensory information enabling people to build their definition of causality. We first presented the smallest and largest visual delays across both visual conditions, to help task adaption and to mitigate possible order effects [6]. The presentation of visual conditions was counterbalanced between participants. Second, participants were exposed to the range of sensory conditions, through 8 randomized trials (all combinations of vibration and audio, excluding no vibration/no audio as this was in the first practice). Afterwards, participants began the study.

At the beginning of each trial, two billiard balls would initialize at set positions. Then, the blue ball would move towards the red ball and make contact. Upon contact, the two balls would stop for a set amount of time (temporal delay) at which point the red ball would move away (Fig. 3). In the dynamic case, the blue ball would also move after this delay. After the trial, this question (modified from [11]) appeared:

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

Participants were asked 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 would appear on the scale, and participants would confirm their selection to continue to the next trial. Three breaks were provided at regular intervals.

At the end, participants completed a short survey about strategies and general comments. Most participants finished the study in less than 60 minutes, but were allotted 90 minutes.

VI. CONFIRMATORY ANALYSIS

Generalized linear mixed effects models with fixed effects of temporal offset, number of signals, realism, condition, and/or match-mismatch were fit to the data. All models have random effects of the participant as well as the other predictors. Temporal offset and number of signals are treated as continuous variables, while the rest are coded as factors. To test our preregistered hypotheses, we fit models (one for each hypothesis), where j is the participants' causal judgments:

M1: $\text{j} \sim 1 + \text{offset} + (1 + \text{offset}|\text{participant})$

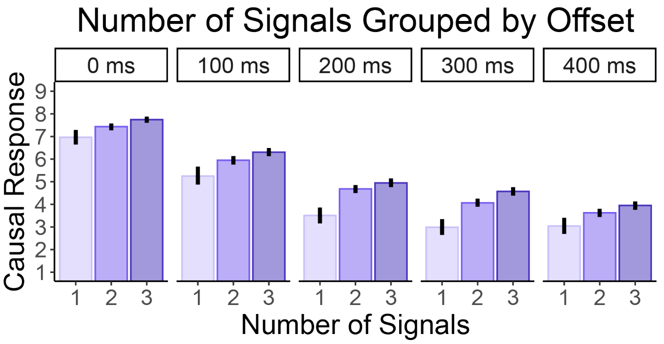

Fig. 4. Means and 95% bootstrapped confidence intervals (CI) shown for causal response across participants.

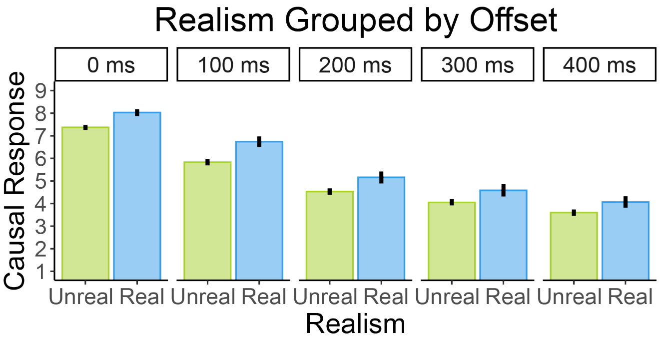

Fig. 5. Means and 95% CI shown for causal response across participants.

M2: $\text{j} \sim 1 + \text{signals} * \text{offset} + (1 + \text{signals} * \text{offset}|\text{participant})$

M3: $\text{j} \sim 1 + \text{realism} * \text{offset} + (1 + \text{realism} * \text{offset}|\text{participant})$

M4: $\text{j} \sim 1 + \text{condition} + (1 + \text{condition}|\text{participant})$

M5: $\text{j} \sim 1 + \text{mismatch} + (1 + \text{mismatch}|\text{participant})$

These models were fit using Bayesian data analysis with the *brms* package in R [22]. To analyze which model best predicts participants' causal judgements, we used approximate leave-one-out cross-validation via the *loo* function. For contrasts, we use the *emmeans* package and function [23]. Across all analyses, we use the inference criterion that the 95% credible interval excludes zero. We chose to use Bayesian data analysis, as compared to a frequentist approach, because it is more robust when fitting complex models.

A. H1: Temporal Offset

To see whether there was an effect of offset, consider M1, which predicts causal judgments with a fixed effect of temporal offset and random effects of participant and offset. Here, temporal offset is treated as a continuous variable.

Overall *causal judgments decreased with temporal offset* ($\beta = -0.01$, 95% Credible Interval $[-0.01, -0.01]$), meaning that with greater delay, causal judgments are predicted to decrease. While not explicitly shown in a figure, the general trend can be seen in results for (H2) & (H3) (Fig. 4 - 5).

B. H2: Number of Signals

We hypothesized that the number of signals provided is important for people's causal judgments. Specifically, we hypothesize that with more sensory information people will attribute higher levels of causality to the event. To test this, M2 predicts causal judgments with fixed and random effects of temporal offset and number of signals, as well as their interaction, and an additional random effect of participant.

M2, which includes number of signals, was found to better predict the data compared to M1. There was a *positive relationship between causal judgments and number of signals* ($\beta = 0.44$, 95% CrI [0.29, 0.57]) as shown in Fig 4.

C. H3: Cue Realism

To determine the effect of cue realism, we built M3 which predicts causal judgments with fixed and random effects of temporal offset, realism, and their interaction as well as a random effect of participant. While temporal offset is treated as a continuous variable, realism is a factor. Conditions that contain any unrealistic cues, are coded as "unreal."

M3, which accounts for realism, better predicts the data than M1. Additionally, *real signals, compared to unreal, increase causal judgments* ($\beta = 0.82$, 95% CrI [0.42, 1.25]) (Fig. 5).

D. H4: Condition

We predicted that one condition (realistic vision, audio, and vibration) would have larger causal ratings than all other conditions. To determine whether this was the case, we built M4, which predicts causal judgments with fixed and random effects of condition as well as an additional random effect of participant. The condition is set as a factor.

To test this prediction, we evaluated contrasts on the levels of condition, specifically comparing one condition to the rest. The causal judgments in the *realistic, multisensory condition were greater than the other conditions* ($\beta = 1.18$, 95% CrI [0.74, 1.65]). This result can be seen in Fig. 6 where the blue bar furthest to the right represents that condition.

E. H5: Mismatch

This final hypothesis considers a possible uncanny valley, in which we consider only the realistic vision cases. Within that, specifically we want to compare the mismatched cases (unrealistic signals or a mix of realistic and unrealistic signals) to purely vision. Thus, we took only these portions of the data and established a factor for each of these groups[5]. M5 predicts causal judgments with a fixed effect of mismatch and random effects of participant and mismatch.

Similar to (H4), we evaluated contrasts on the levels of mismatch. Causal judgments were higher in the mismatched conditions than in the realistic vision only condition ($\beta = 0.67$, 95% CrI [0.28, 1.08]), which is not what we predicted. We consider possibilities for this discrepancy in the Discussion.

VII. Exploratory Analysis

In the hypotheses, we made no distinct claims about how vision, audio, and vibration cues would differ in the effects they had on causal judgments, as it was unclear what might happen given the limited prior work. From observing the data, we considered two models to predict causal judgments – one more basic and a second with an interaction effect. Both models have fixed effects of audio, vibration, and vision as well as random effects of those variables and participant. Audio, vibration, and vision are all treated as factors.

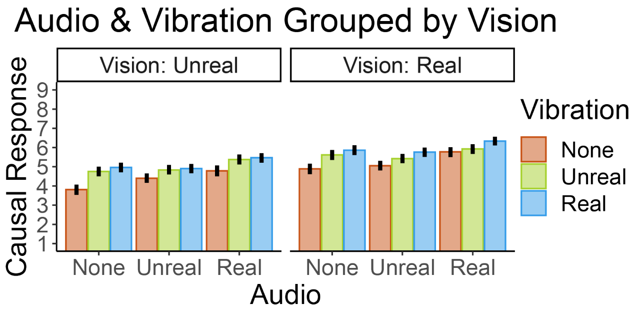

Fig. 6. Means and 95% CI shown for causal response across participants.

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

M7: $\text{j} \sim 1 + \text{audio} + \text{vibration} * \text{vision} + (1 + \text{audio} + \text{vibration} * \text{vision}|\text{participant})$

With the same methods we used in our previous analysis, we compared these models. We found that M7 is a better predictor of the data, which indicates that the interaction effect between vision and vibration matters.

We computed contrasts between audio cues as well as the interaction between vibration and vision. The causal ratings of realistic audio cues are greater than they are in the absence of audio ($\beta = 0.63$, 95% CrI [0.27, 0.98]) and also greater than they are with the inclusion of unrealistic audio ($\beta = 0.55$, 95% CrI [0.21, 0.92]). Within the unrealistic vision condition, unrealistic vibration is higher rated than no vibration ($\beta = 0.66$, 95% CrI [0.36, 0.92]), and realistic vibration is higher rated than unrealistic vibration ($\beta = 0.78$, 95% CrI [0.48, 1.06]). Within realistic vision, there is again a difference between no vibration and unrealistic vibration, the former being higher rated than the latter ($\beta = 0.415$, 95% CrI [0.05, 0.78]) as well as between unrealistic and realistic vibration, the latter being higher rated than the former ($\beta = 0.75$, 95% CrI [0.42, 1.10]).

While the difference between the vibrations does not meet our inference criteria, the estimate between no and unrealistic vibration is smaller in the realistic vision condition (red and green bars across panels of Fig. 6).

VIII. Discussion

We aimed to understand the role that physical realism plays in multisensory causal perception. First we considered whether people's causal judgments increase with multiple sensory signals, and we find that they did (Fig. 4). Beyond the increase with one additional signal (as shown in prior work [8], [11]), here the increase extended beyond just two signals. Additionally, regarding the role of physically realistic information, we found that the information provided within cues is important, as more realistic signals were viewed as more causally linked (Fig. 5), which aligns with results based on only vision [12]. So, not only

---

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

does more temporally contiguous sensory information lead to higher levels of causal perception but so does providing more realistic cues. This finding is further highlighted by results for (H4), in which the condition with all three realistic signals was the one with the strongest perception of the events being causally linked. This could be due to realistic cues aligning with people's intuitive understanding of physics [24] and internalized representations of how collisions look, sound, and feel.

We did not find an uncanny valley effect (H5). This could be due to a variety of reasons, such as the specific vibration/audio cues selected or the timing (all cues led to an increase in the number of co-occurring events as the collision and cues occur simultaneously across all conditions, which others have theorized plays a key part in how people make these judgments [11]). As mismatched conditions (a mix of realistic and unrealistic sensory cues) had larger causal judgments compared to realistic vision alone, this could imply that regardless of the quality of feedback, postdictive understanding of the action overrides visual-based understanding. Further investigation is necessary to explore these effects in multisensory settings.

From the exploratory analysis, we found that audio, vibration, vision, and the interaction between vibration and vision were all factors in predicting causal judgments. Prior work has shown that the addition of audio to visual scenes increased causal judgments [11], which we saw 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. 6). Similarly, both types of vibration had an effect regardless of vision, however the "boost" provided by adding unrealistic vibration decreased when realistic visuals were used (Fig. 6, compare the green and blue columns across both panel). This increase from vibrotactile haptics is similar to what was found previously with kinesthetic haptics [8].

IX. LIMITATIONS & FUTURE WORK

We chose to synchronize the audio and vibration with the stopping of the first ball (the moment of collision). However, there are other locations that still coincide with causal cues, such as when the second ball starts moving or during the delay. While previous work tested locations of audio feedback, there were no conclusive results on the role of location [11]. Additionally, there are options that do not align with these suggestions, such as before or after all events have occurred. It is unclear how cue locations would affect the results.

The study design has several limitations. First, likert scales are prone to subject bias. Future work could avoid this by using a 2-alternative forced choice question about causality or by asking about other objects in the world that would be affected by the presence (or absence) of a causal linkage. Second, we programmatically created the realistic condition. Future work could use real-world footage of objects colliding to provide a more realistic comparison, and possibly find an uncanny valley effect. However, this would reduce the ability to explore realism that extends beyond the binary (realistic vs. unrealistic) considered in this work. Finally, we introduced a spatial mismatch between where people saw and felt the collision. By running similar studies in VR, we could avoid any undesired effects from spatial mismatch.

Additionally there are limitations to our experimental paradigm. First, we only tested one scenario (collision) and the results may not extend outside of an experimental setting. Even though this study paradigm is common in psychology, its simplicity could lead to high ratings of causality. Future work should focus on testing similar questions in situ, with instrumented objects in order to observe people's interactions directly. Additionally, more information on how and what cues people utilize from the environment could aid development of a model for how humans form causal judgments. Some work has begun to create models that predict and explain causal judgments [25], [26], but none have considered the role of haptic information or realism. Second, this paradigm focuses on people's retrospective judgments, which may vary from results that consider moment-to-moment perception of causality.

This experiment tested vibrotactile haptic information, so we cannot comment on how kinesthetic cues would interact with realism and audio. Future work should consider this additional sensory mode, as it could provide highly diagnostic information about a collision (or other causal event) and has been shown to have an effect on causal perception [8].

Finally, research has shown that perception is different in active compared to passive touch. Vogels found that sensitivity for asynchrony was higher in passive compared to active touch [13]. Other work has explored active touch and agency, a person's feeling of control over their actions [27], finding that haptic information significantly increases the users sense of agency [28], [29]. Although we only considered passive touch, other settings would allow people to actively explore and naturally experience multisensory causal interactions.

X. CONCLUSION

We ran a psychophysical study to explore the effect of realism and multisensory cues on causal perception. We found that people are more likely to judge an event as causal when there is evidence from multiple sensory sources, which extends beyond what was found in prior work that only explored two signals. Additionally, the type of information, realistic or unrealistic, has an effect on causal judgments, with the largest perception of causal linkage given to the condition with realistic cues across all senses (vision, audio, and vibration). From exploratory analysis, we found that both types of audio and vibration cues increase causal response, but the degree of increase is affected by the realism of vision.

This is the first work to combine three sensory cues in the study of causal perception. Realism and multisensory integration are essential to understanding how people interpret their environment, as we experience the world across our senses. These findings will help designers of future systems determine how and what cues to utilize to achieve desired perception of interactions, with a focus on the benefits of utilizing realism across visual, audio, and haptic information.

ACKNOWLEDGMENT

Thanks to Dr. Paul Mitiguy for helping solve the dynamics.

REFERENCES

[1] P. A. White, “Causal processing: Origins and development.” *Psychological bulletin*, vol. 104, no. 1, p. 36, 1988.

[2] ——, “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*, vol. 138, no. 4, p. 589, 2012.

[3] A. Michotte, *The perception of causality*. Basic Books, 1946/1963.

[4] J. F. Kominsky, B. Strickland, A. E. Wertz, C. Elsner, K. Wynn, and F. C. Keil, “Categories and constraints in causal perception,” *Psychological Science*, vol. 28, no. 11, pp. 1649–1662, 2017.

[5] J. F. Kominsky and B. J. Scholl, “Retinotopic adaptation reveals distinct categories of causal perception,” *Cognition*, vol. 203, p. 104339, 2020.

[6] C. Bechlivanidis, A. Schlottmann, and D. A. Lagnado, “Causation without realism.” *Journal of Experimental Psychology: General*, vol. 148, no. 5, p. 785, 2019.

[7] P. Wolff and J. Shepard, “Causation, touch, and the perception of force,” in *Psychology of learning and motivation*. Elsevier, 2013, vol. 58, pp. 167–202.

[8] E. D. Chase, P. Wolff, T. Gerstenberg, and S. Follmer, “A causal feeling: How kinesthetic haptics affects causal perception,” in *2021 IEEE World Haptics Conference (WHC)*. IEEE, 2021, pp. 347–347.

[9] C. C. Berger, M. Gonzalez-Franco, E. Ofek, and K. Hinckley, “The uncanny valley of haptics,” *Science Robotics*, vol. 3, no. 17, p. eaar7010, 2018.

[10] B. J. Scholl and P. D. Tremoulet, “Perceptual causality and animacy,” *Trends in Cognitive Sciences*, vol. 4, no. 8, pp. 299–309, 2000.

[11] R. Guski and N. F. Troje, “Audiovisual phenomenal causality,” *Perception & psychophysics*, vol. 65, no. 5, pp. 789–800, 2003.

[12] K. Meding, S. A. Bruijns, B. Schölkopf, P. Berens, and F. A. Wichmann, “Phenomenal causality and sensory realism,” *i-Perception*, vol. 11, no. 3, p. 2041669520927038, 2020.

[13] I. M. Vogels, “Detection of temporal delays in visual-haptic interfaces,” *Human Factors*, vol. 46, no. 1, pp. 118–134, 2004.

[14] M. Di Luca and A. Mahnan, “Perceptual limits of visual-haptic simultaneity in virtual reality interactions,” in *2019 IEEE World Haptics Conference (WHC)*. IEEE, 2019, pp. 67–72.

[15] H. Ohnishi and K. Mochizuki, “Effect of delay of feedback force on perception of elastic force: a psychophysical approach,” *IEICE transactions on communications*, vol. 90, no. 1, pp. 12–20, 2007.

[16] A. Pressman, L. J. Welty, A. Karniel, and F. A. Mussa-Ivaldi, “Perception of delayed stiffness,” *The International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1191–1203, 2007.

[17] B. Knorlein, M. Di Luca, and M. Harders, “Influence of visual and haptic delays on stiffness perception in augmented reality,” in *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2009, pp. 49–52.

[18] A. M. Okamura, M. R. Cutkosky, and J. T. Dennerlein, “Reality-based models for vibration feedback in virtual environments,” *IEEE/ASME transactions on mechatronics*, vol. 6, no. 3, pp. 245–252, 2001.

[19] K. J. Kuchenbecker, J. Fiene, and G. Niemeyer, “Improving contact realism through event-based haptic feedback,” *IEEE transactions on visualization and computer graphics*, vol. 12, no. 2, pp. 219–230, 2006.

[20] F. Conti, F. Barbagli, R. Balaniuk, M. Halg, C. Lu, D. Morris, L. Sentis, J. Warren, O. Khatib, and K. Salisbury, “The chai libraries,” in *Proceedings of Eurohaptics 2003*, Dublin, Ireland, 2003, pp. 496–500.

[21] L. A. Jones and H. Z. Tan, “Application of psychophysical techniques to haptic research,” *IEEE transactions on haptics*, vol. 6, no. 3, pp. 268–284, 2012.

[22] P.-C. Bürkner, “brms: An R package for Bayesian multilevel models using Stan,” *Journal of Statistical Software*, vol. 80, no. 1, pp. 1–28, 2017.

[23] R. Lenth, H. Singmann, J. Love, P. Buerkner, and M. Herve, “Package ‘emmeans’,” 2019.

[24] J. R. Kubricht, K. J. Holyoak, and H. Lu, “Intuitive physics: Current research and controversies,” *Trends in cognitive sciences*, vol. 21, no. 10, pp. 749–759, 2017.

[25] K. P. Körding, U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, and L. Shams, “Causal inference in multisensory perception,” *PLoS one*, vol. 2, no. 9, p. e943, 2007.

[26] J. F. Magnotti, W. J. Ma, and M. S. Beauchamp, “Causal inference of asynchronous audiovisual speech,” *Frontiers in psychology*, vol. 4, p. 798, 2013.

[27] J. W. Moore, “What is the sense of agency and why does it matter?” *Frontiers in psychology*, vol. 7, p. 1272, 2016.

[28] J. Bergstrom-Lehtovirta, D. Coyle, J. Knibbe, and K. Hornbæk, “I really did that: Sense of agency with touchpad, keyboard, and on-skin interaction,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–8.

[29] S. Kasahara, J. Nishida, and P. Lopes, “Preemptive action: Accelerating human reaction using electrical muscle stimulation without compromising agency,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–15.