# Once and Again: Repeated viewing affects judgments of spontaneity and preparation

Kristin Donnelly[1], William H. Ryan[2], and Leif D. Nelson[2]

[1]University of Chicago Booth School of Business, Chicago, IL 60637

[2]Haas School of Business, University of California, Berkeley, Berkeley, CA 94720

Forthcoming in *Psychological Science*

Author Note

Correspondence concerning this article should be addressed to Kristin Donnelly, Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave #369, Chicago, IL, email: Kristin.Donnelly@chicagobooth.edu

**Abstract**

Reality is fleeting, and any moment can only be experienced once. Re-watching a video, however, allows people to repeatedly observe the exact same moment. We propose that people may fail to fully distinguish between merely *observing* behavior again (through replay) from that behavior being *performed* again in the exact same way. Across an assortment of stimuli that includes auditions, commercials, and potential trial evidence, nine experiments ($N$ = 10,412 adults in the US) demonstrate that re-watching makes a recorded behavior appear more rehearsed and less spontaneous, as if the actor were simply precisely repeating their actions. These findings contribute to an emerging literature showing that incidental video features, like perspective or slow motion, can meaningfully change evaluations. Replay may inadvertently shape judgments in both mundane and consequential contexts. To understand how a video will influence its viewer, one will need to consider not only its content, but also how often it is viewed.

**Statement of Relevance**

Video recordings are a primary source of information, communications, and entertainment, and for all those purposes, they are frequently watched more than once. Replays necessarily diverge from reality, where every experience is somewhat unique. How does the replay experience change the inferences people draw about real life?

We demonstrate that video replay makes people in the videos seem more rehearsed and less spontaneous. Viewers seem to interpret actors in replayed videos as though reenacting the same action again. This bias influences judgments in a broad array of domains, with replays making some actions seem better prepared and others less authentic. These findings are of practical importance for audiences as varied as business managers, lawyers, policymakers, and content creators.

## Research Transparency Statement

### General Disclosures

Conflicts of interest: The authors declare no conflicts of interest. Funding: This research was supported by funding from UC Berkeley Haas School of Business and University of Chicago Booth School of Business. Artificial intelligence: No artificial intelligence assisted technologies were used in this research or the creation of this article. Ethics: This research received approval from UC Berkeley's Committee for Protection of Human Subjects.

### For all studies:

Preregistration: https://researchbox.org/151.

Materials: https://researchbox.org/151.

Data: https://researchbox.org/151.

Analysis Scripts: https://researchbox.org/151.

Human experience is necessarily limited to our direct observations and perception. Video recordings, however, enable people to see and hear completely different observations. Moreover, they transcend the present-bound nature of real-world perception, allowing people to observe a unique moment in time more than once through replay. For those reasons, video recordings are a widespread conduit for information, communication, and entertainment. For example, most Americans get their news from video (Chmielewski, 2018), and every day, more videos are watched on YouTube and Facebook alone than there are people in the world (Donchey, 2021; Aboulsohn, 2020). Contributing to this video consumption, people often watch videos more than once, whether because they are looped on television, shared on social media, or intentionally re-watched because the viewer finds them particularly informative, important, or entertaining (O'Brien, 2021).

Despite providing (repeated) access to moments that the viewer was not present for, videos are merely a close facsimile of real life. As such, viewer inferences may not be accurate, either. Even when aware that a given video clip is a modified representation of reality, viewers may struggle to fully account for their initial, intuitive impressions. Pioneering research on this topic found that viewers failed to fully account for the time distortion introduced by slow motion (Caruso et al., 2016). Rather, viewers inferred that a sequence of events had taken longer when they watched those events in slow motion, instead of regular speed. For example, slow-motion viewing made interpersonal transgressions like illegal football tackles appear more intentional, and the actors, therefore, more culpable. The video's visual perspective may have a similar effect. Officers seemed less involved in police activity when it was viewed from their body cam, which removes them from the scene, compared to a dash cam, which shows them clearly (Turner et al., 2016). In essence, not seeing the actor in a scene reduced her perceived involvement.

These findings suggest that viewers may treat features of the video-watching experience, like replay speed and visual perspective, as though they are features of the actual event (Granot et al., 2018). We build on this foundation with an analogous prediction: Rewatching a recorded event imparts an intuitive impression that the people in the video are repeating their actions or behavior—and repeating them in precisely the same way. In other words, we propose a "replay bias", wherein people intuitively feel that when a video is replayed, the actor is reenacting the exact same actions.

This hypothesis draws from a phenomenon underlying numerous perceptual and cognitive illusions in which the system accurately interprets stimuli in frequently encountered contexts, but then overapplies that interpretation (Gregory, 1968, Segall et al., 1966). In real life, we only see an action repeatedly when it truly occurs more than once. Appropriately, such repetition eliminates the uniqueness and reduces the perceived spontaneity of the actor's behavior, potentially influencing how others judge them. Gershon and Smith (2020), for example, found that comedians delivering the same joke on different occasions seem inauthentic because they falsely present their performance as spontaneous; the same was true for self-repetitions in other contexts (i.e., politics, tour guiding, interviews).

We propose that viewers inappropriately apply this understanding of *actual* repeated behavior to the *apparent* repetition of a single behavior afforded by replay, even when they intellectually know the actions in the video only occurred once. Therefore, just like when behavior in the real world is carried out more than once, a single moment loses its uniqueness and spontaneity when replayed. Of course, exact repetition is physically impossible, but when an actor's repeated behavior nonetheless *appears* identical to their original, it is likely to be highly controlled, practiced, or deliberate. For example, a dancer exactly replicating a series of

movements or a speaker delivering identical monologues must have practiced and prepared extensively. If viewers fail to fully consider that the video captured a unique moment in time, and that the person in a replayed video is not simply repeating the same actions identically, they may infer that those actions are controlled and deliberate, potentially through extensive rehearsal.

The impact of this inference on ancillary judgments should vary with context. In situations where uncontrolled, spontaneous behavior is viewed positively, repeated viewing might lead to negative judgments (e.g., the actor may appear less genuine and more contrived). Conversely, in situations where controlled, planned, or rehearsed behavior is viewed positively, repeated viewing might enhance positive evaluations (e.g., the actor may appear more prepared). We explore this bias and its consequences in nine studies, ruling out alternative accounts that include mere exposure, increased accuracy, and experimenter demand.

**Open Practices Statement**

All studies were preregistered on AsPredicted. The preregistrations, materials, data and code are available at https://researchbox.org/151. The target sample size for each study was specified in advance within our preregistration, as were all exclusion criteria, hypotheses, manipulations, measures, and analysis plans. Generally, we aimed to power studies to detect small-to-medium effect sizes. Studies were approved by the Institutional Review Board. The Supplemental Materials detail additional analyses and exploratory measures.

## Studies 1a – 1c

Studies 1a – 1c test our hypothesis that when re-watching a video, the events in the video feel like they are simply happening again in the exact same way. If so, repeated viewing should make the videotaped behavior seem more rehearsed and less spontaneous.

**Method**

**Participants.** In three separate studies, denoted as Studies 1a, 1b, and 1c, we enlisted a combined total of 870 workers from Prolific Academic. As per our pre-registrations, we aimed for sample sizes of 200, 300, and 350 in Studies 1a, 1b, and 1c, respectively. 223 participated in Study 1a (43% female, 55% male, 2% other/prefer not to answer, $M_{age} = 29$), 303 participated in Study 1b (43% female, 54% male, 3% other/prefer not to answer, $M_{age} = 39$), and 344 participated in Study 1c (55% female, 43% male, 2% other/prefer not to answer, $M_{age} = 34$). Each study employed an English comprehension check and excluded those who failed (5.4%, 1.3%, and 2.0% for Studies 1a, 1b, and 1c, respectively). As an attention check, participants were asked to identify, based on video stills, which video they did not see; we excluded those who answered incorrectly (4%, 3%, and 0.3% for Studies 1a, 1b, and 1c, respectively). These exclusions resulted in a final sample of 204 for Study 1a, 292 for Study 1b, and 336 for Study 1c.

**Materials.** Participants in Study 1a watched homemade audition tapes for the reality TV show Survivor. Each video depicted the contestant introducing themselves. Participants in Study 1b were shown short clips of complicated dance videos from TikTok, a popular video-sharing platform (Rodriguez, 2020). Finally, to examine repeated viewing effects for bad performances, Study 1c used clips of contestants auditioning for the competition singing show American Idol. All contestants delivered performances that received scathing criticism from the judges. Unless noted otherwise, stimuli in all studies were taken from YouTube and trimmed using Kapwing, a free online video editor.

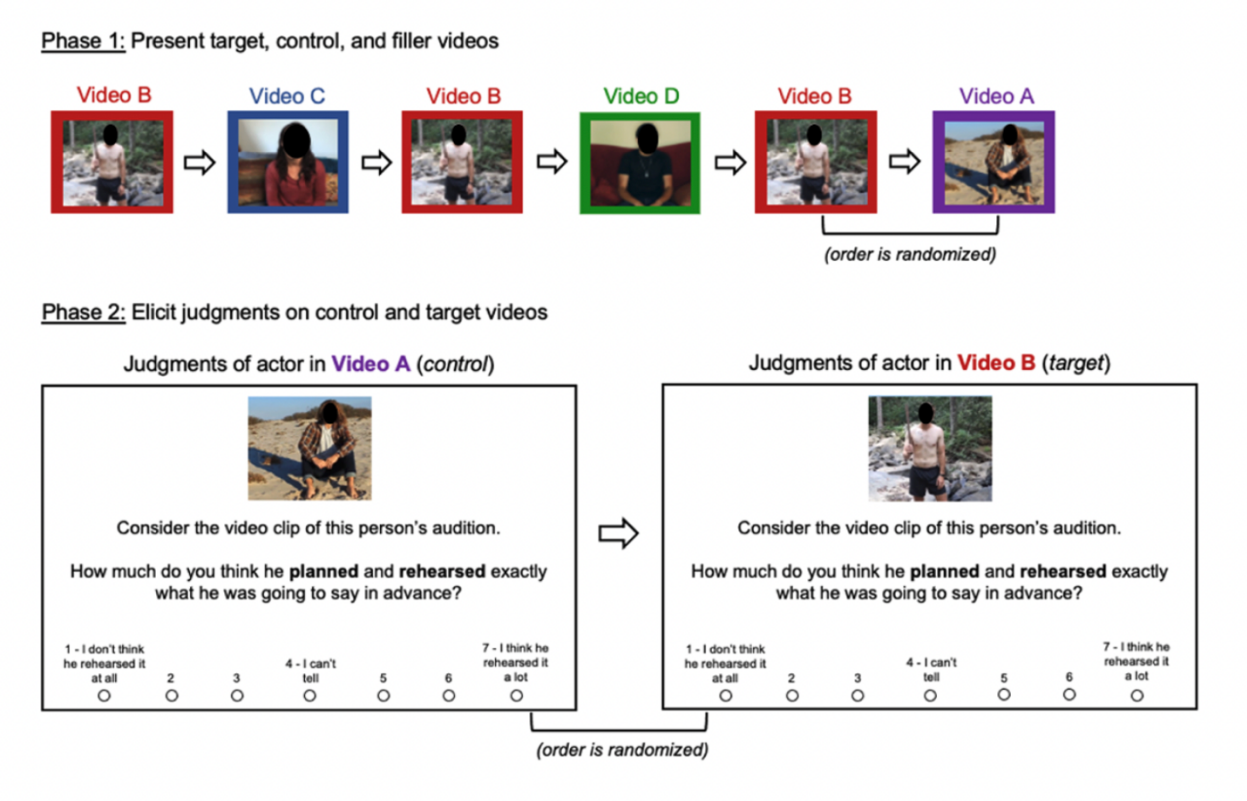
**Procedure.** Studies 1a – 1c use the same basic experimental paradigm (see Figure 1). Participants saw four unique video clips (A through D), each roughly the same duration (between 8-11 seconds for Study 1a, 7-10 seconds for Study 1b, and 3-7 seconds for Study 1c). They watched either video A three times and B once, or B three times and A once, with videos C and

D as fillers preventing back-to-back replay. In Study 1a, we randomized the order of the last two videos such that the viewing sequence was either A-C-A-D or B-C-B-D, followed by either A-B or B-A. Studies 1b-c presented either an A-C-A-D-A-B or B-C-B-D-B-A sequence. Each clip played automatically and automatically advanced to the next.

At the end of the viewing sequence, participants rated videos A and B in random order. In Study 1a, participants evaluated how much the target had planned and rehearsed what he was going to say in advance on a 7-point Likert scale (1 = *I don't think he rehearsed it at all*, 7 = *I think he rehearsed it a lot*). We altered this question slightly for participants in Studies 1b-c to ask about preparation (1 = *I don't think they prepared or practiced at all*, 7 = *I think they prepared and practiced a lot*).

Participants also evaluated targets on a related dimension that we expected would correlate with perceptions of preparation. Specifically, participants in Study 1a rated how much effort the target had put into making the tape (1 = *I don't think he put much effort into making this*, 7 = *I think he put a lot of effort into making this*). Participants in Studies 1b and 1c rated the performer's skill level on the relevant dimension (dancing and singing, respectively; 1 = *I think they are very bad at [dancing/singing],* 7 = *I think they are very good at [dancing/singing]*).

*Figure 1.* Experimental paradigm as illustrated by Study 1a. Faces are redacted.




## Results

Throughout these studies, as well as in several subsequent ones detailed in this paper, we preregistered the same analysis plan: a mixed effects model regressing the dependent measure on how many times the target's clip was played (once vs. three times) and include random effects of participant and video.[1] We do so for each of the primary and secondary dependent measures present in Studies 1a – 1c .

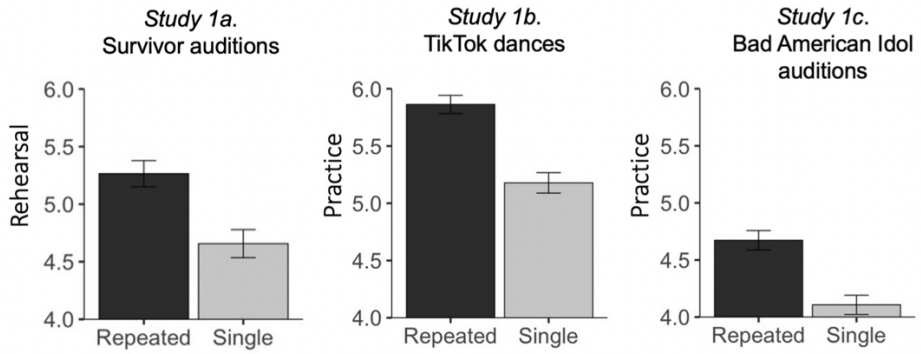
The results provide support for the idea that viewers inappropriately apply their understanding of actual repetition to the replay context. That is, viewers may fail to fully account for the fact that they are merely *observing* the behavior more than once—the actor is not

[1] Including both terms as random effects produced overfitting in Studies 1a, 2, and 4. There, we deviate slightly from the pre-registration and control for the term responsible for the overfitting, following Bar et al. (2013). All statistical conclusions are unchanged. In all studies, the basic effect of interest remains significant under standard linear regressions and paired-sample *t*-tests (see Table S2).

precisely repeating it. This is most obviously reflected in the results for our primary dependent measure of spontaneity versus preparation: When a video was repeated rather than viewed once, the target was rated as having rehearsed more (Study 1a: *M*s = 5.26, 95% CI [5.04, 5.49] vs. 4.66, 95% CI [4.42, 4.9], *SD*s = 1.62 vs. 1.74; $t(404.02) = 4.31$, $p < .001$, $d = 0.36$, 95% CI [0.17, 0.56][2]) and practiced more (Study 1b: *M*s = 5.86, 95% CI [5.71, 6.02] vs. 5.18, 95% CI [5, 5.35], *SD*s = 1.36 vs. 1.52; $t(290.06) = 6.86$, $p < .001$, $d = 0.47$, 95% CI [0.31, 0.64]; Study 1c: *M*s = 4.67, 95% CI [4.51, 4.84] vs. 4.11, 95% CI [3.94, 4.27], *SD*s = 1.56 vs. 1.55; $t(334.00) = 6.79$, $p < .001$, $d = 0.36$, 95% CI [0.21, 0.52]). Results are displayed in Figure 2.

*Figure 2.* Mean rating (7-point scale) for rehearsal (Study 1a) and practice (Studies 1b-1c) as a function of repeated viewing. In separate experiments, participants watched a series of video clips taken from Survivor audition tapes (Study 1a), TikTok dances (Study 1b), and bad American Idol auditions (Study 1c). Two of the videos were "target videos" in which one video was repeated three times and the other was shown once, interspersed with filler videos. For the target that they saw three times and for the target that they saw once, participants evaluated how much the target had rehearsed (Study 1a) and practiced (Studies 1b – 1c) their actions in advance. Error bars reflect ±1 SEM.

[2] There is significant debate about which measures of effect size, if any, are appropriate for mixed effects models. Here, we chose to use Cohen's d as a measure of effect size, which is calculated using a non-mixed effects model, in order to give readers an intuitive idea of effect size (see Table S2 for full reporting of standard linear regressions and paired t-tests).




Our results for the secondary measures of effort (Study 1a) and skill (Studies 1b-c) are also consistent with viewers mistaking replay for true repetition. We selected effort (Study 1a) and skill (Studies 1b-c) because those dimensions are especially likely to produce the tight bodily control that one would require to actually repeat an action in precisely the same way. Indeed, in Study 1a, the target was rated as having put more effort into making the clip when his video was replayed versus shown once ($M$ = 5.00, 95% CI [4.77, 5.22] vs. 4.68, 95% CI [4.44, 4.91], $SD$s = 1.65 vs. 1.73), $t(404.03) = 2.37$, $p = .018$, d = 0.19, 95% CI [-0.01, 0.38]. Repeated viewing also made the target appear significantly more skilled in Study 1b, where participants saw TikTok dance performances ($M$s = 4.87, 95% CI [4.71, 5.03] vs. 4.50, 95% CI [4.34, 4.65], $SD$s = 1.43 vs. 1.33), $t(290.05) = 4.76$, $p < .001$, d = 0.27, 95% CI [0.11, 0.43]. Note that precise repetition may only be straightforwardly related to skill in Study 1b, where targets' performances were, as per participant ratings, somewhat impressive. In Study 1c, targets were terrible singers, and the effect of replay on skill was only marginally significant ($M$s = 2.62, 95% CI [2.48, 2.76] vs. 2.77, 95% CI [2.61, 2.93], $SD$s = 1.31 vs. 1.47), $t(334.03) = 1.78$, $p = .08$, d = 0.11, 95% CI [-0.04, 0.26]. Insofar as viewers confuse repeated viewings with actual repetition, their impressions of skill may have been shaped by two opposing influences—namely, the lack of skill suggested by the poor performance and the skill (i.e., vocal control) required to precisely repeat it.

## Study 2

Studies 1a-c support the hypothesis that viewers mistake the *apparent* repetition from replay for actual repetition, and thus the lack of variability between replays makes the target seem more rehearsed. But it is also possible that repeated viewing simply makes any evaluation more positive. Presumably, replay allows viewers to process content more easily, and the increased fluency might lead to greater liking or more positive evaluations overall. This would also align with the mere exposure effect, whereby repeated exposure increases positive affect (Mrkva & Van Boven, 2020; Zajonc, 1968). To examine this alternative account, Study 2 used stimuli where actors would be judged more *negatively* for rehearsing or controlling their behavior. Specifically, participants saw videos of surprised reactions. Such behavior is inherently spontaneous and involuntary; a person can only have one true first-time reaction to a stimulus. If participants fail to fully distinguish replay from actual repetition, repeated viewing should make these spontaneous reactions seem less authentic.

### Method

**Participants and design.** A total of 501 Prolific Academic workers (48% female, 50% male, 0.2% other/prefer not to answer, 1.2% left the question blank, $M_{age} = 34$) took part in Study 2. Sample size was pre-registered at 500. As in the previous studies, we excluded those who failed the English comprehension check (1.2%) and/or attention check (0.2%). This left a final sample of 494.

**Materials.** Participants saw four unique videos, each 5-11 seconds long, from "unboxing videos", a popular genre of YouTube that blurs the line between entertainment and advertising. These videos typically feature someone unpacking a product that they have supposedly never seen before and expressing surprise and excitement. As before, there were two target videos (A

and B) and two filler videos (C and D). Video A showed a woman opening a Pusheen-themed subscription box[3], and Video B showed a woman opening an Amazon "mystery box" of Christmas lights. Filler videos C and D depicted people unboxing an Apple monitor and a PlayStation.

**Procedure.** Study 2 follows the design of most of the previous studies. Participants saw videos in a sequence of either A-C-A-D or B-C-B-D followed by A-B or B-A. As a small departure from the previous studies, clips did not automatically advance; participants had to press the "next" arrow to proceed, ensuring they attended to each clip. Ratings were provided on a 7-point scale for authenticity (1 = *I think her reaction was completely fake*, 7 = *I think her reaction was completely genuine*).

**Results**

A mixed effects model regressed our dependent measure of authenticity on number of viewings (once vs. thrice), controlling for video and with a random effect of participant. The anticipated effect emerged: When participants saw the unboxing video multiple times (vs. once), they thought that the target's reaction to the product was less genuine (*M*s = 3.95, 95% CI [3.75, 4.15] vs. 4.40, 95% CI [4.22, 4.58], *SD*s = 2.24 vs. 2.04), $t(492.02) = -5.02$, $p < .001$, $d = 0.21$, 95% CI [0.09, 0.34].

These results provide further support for our account and directly challenge the idea that, by improving processing fluency or increasing familiarity via mere exposure (e.g., Montoya et al., 2017), replay simply makes observers' evaluations more positive. In addition, the findings might appear to conflict with the illusory truth effect, whereby a piece of information is perceived to be more true the more times it is repeated or encountered (for a recent review, see

[3] Pusheen is a popular cartoon cat.

Brashier & Marsh, 2020). However, the present work focuses on a very specific, interpersonal evaluative context (i.e., the authenticity of "first-time" reactions). In this particular context, perceptions of spontaneity and veracity may hang together, but more generally, one need not imply the other.

## Study 3

The previous studies involved stimuli with no known ground truth. Thus, replay might have made judgments more accurate. After all, viewers notice new details upon rewatching that might move their perceptions closer to the truth (O'Brien, 2021). Study 3 used stimuli where reactions were indisputably authentic. If replay improves accuracy, it would make these genuine expressions of surprise seem *more* authentic. Our account predicts the opposite.

**Participants.** Participants were 799 Prolific workers (45% female, 52% male, 0.3% other/prefer not to answer, 2.1% left the question blank, $M_{age} = 34$); pre-registered sample size was 800. Study 3 accidentally omitted the attention check used in the previous studies, but retained the English comprehension check and excluded those who failed (3.6%). This resulted in a final sample of 770.

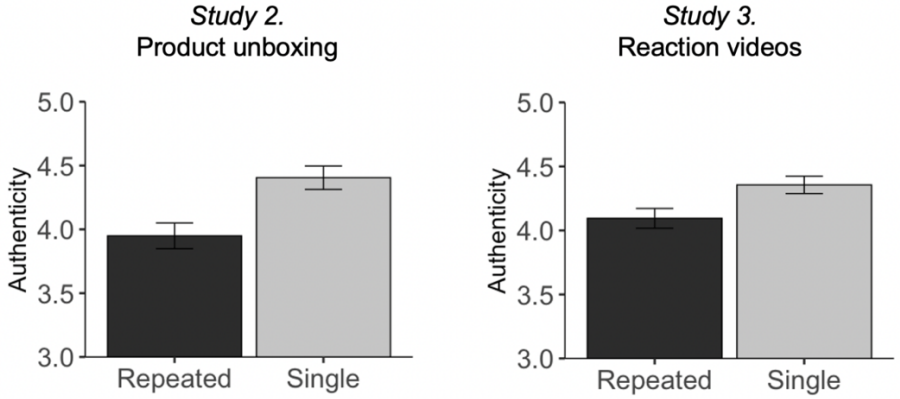
**Materials.** Study 3 replicates Study 2 with stimuli from the Berkeley Reactions to Affective Video Elicitors (BRAVE) database, which contains over 45,000 videos of people reacting to previously unseen emotionally evocative content (Cowen et al., 2021; see also Cowen & Keltner, 2017). We selected videos of surprised reactions. Notably, these reactions should be genuine; the people in the videos were research participants who had no access to the stimuli ahead of time or incentive to fake their reactions.

**Procedure.** The experimental procedure and dependent measure were identical to that of Study 2.

**Results**

A mixed effects model regressed our dependent measure of authenticity on number of viewings (once vs. thrice) and specified random effects of participant and video. The results do not suggest that replay improves accuracy. Once again, the target's surprised reaction was rated as less genuine when watched multiple times ($M$ = 4.09,95% CI [3.94, 4.25], $SD$ = 2.13), versus once ($M$ = 4.36, 95% CI [4.22, 4.49], $SD$ = 1.88), $t$(769.00) = -2.75, $p$ = .006, $d$ = 0.13, 95% CI [0.03, 0.23]. Note that in this study, ratings were closer to the midpoint of the scale when participants saw the target multiple times compared to just once. This pattern of results casts doubt on an account based on evaluative extremity, which might predict that replay simply pulls responses toward scale endpoints (see Mrkva & Van Boven, 2020). Figure 3 displays the results.

*Figure 3.* Mean rating (7-point scale) for authenticity as a function of repeated viewing. Participants evaluated the authenticity of the target's expressions of surprise, a reaction that was elicited by unboxing a product (Study 2) and in a controlled lab setting (Study 3). Error bars reflect ±1 SEM.



## Study 4

Commercials often show people who are purportedly *not* paid actors reacting to a product, ostensibly for the first time; consider Chevrolet's "real people, not actors" campaign and the "Pepsi Challenge". Study 4 examines whether repeated viewing makes these supposedly naïve, would-be customers appear less like "real people" and more like actors.

### Method

**Participants.** Study 4 recruited 1,501 Prolific workers as participants (45% female, 54% male, 1% other/prefer not to answer, $M_{age} = 31$), and 1,417 remained after excluding those who failed the attention (1.7%) and/or English comprehension (4.3%) checks. Our pre-registered sample size was 1,400.

**Materials.** Stimuli were four clips, between 3-10 seconds long, of commercials that featured surprised reactions from ostensibly naïve customers interacting with a product. In a commercial for Chevrolet (Video A), a group of people learn that the brand-anonymized car that they thought was a luxury vehicle is actually a Malibu, an affordable sedan. Similarly, in a commercial for Suave haircare (Video B), a woman discovers that she has been using Suave (a budget-friendly brand) and not a high-end alternative. Filler videos were taken from Chevy Cruise Hatch and "Febreze smell test" commercials.

**Procedure.** We used the same design as the previous studies (see Figure 1) but varied the framing of the dependent measure between-subjects: Participants were randomly assigned to rate the likelihood of the target being either "an actor, not a real person", or "a real person, not an actor" ($N$'s = 771 and 730, respectively). This manipulation aimed to reduce the influence of experimenter demand, whereby participants interpret replay as a signal from the experimenter that the repeated target is different and deserves multiple viewings. If asking about the likelihood

of targets being "actors, not real people" implies that targets are generally "real people", the repeated target might be interpreted as a noteworthy exception, appearing more likely to be an actor. This "replay signaling uniqueness" account would predict the opposite in the "real people, not actors" frame (i.e., replay would make targets appear *less* likely to be actors). Our account predicts that regardless of frame, replay would make the targets seem more likely to be an actor. In line with this, participants used a 101-point slider to answer either "What is the probability that the people in this video are actually actors, not 'real people'?" ("actors" frame; 0 = *They are definitely not actors*, 100 = *They are definitely actors*) or "What is the probability that the people in this video are actually 'real people', not actors"? ("not actors" frame; 0 = *They are definitely actors*, 100 = *They are definitely not actors*).

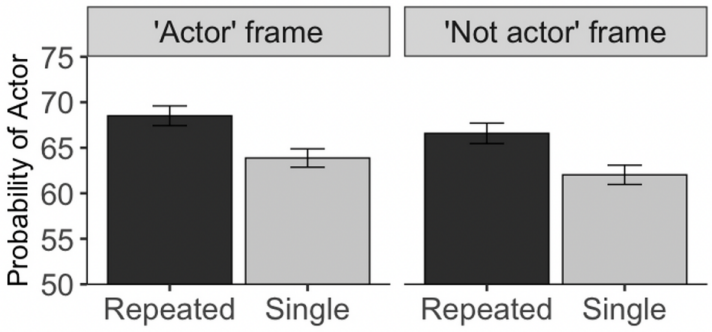
**Results**

First, we subtracted scores in the "not actor" condition from 100 to place responses on the same scale (wherein higher values reflect a higher probability of being an actor). A mixed effects model regressed participants' ratings (i.e. of the target being an actor) onto number of viewings (once vs. thrice, within-subjects), frame ("actor" vs. "non-actor", between subjects), and their interaction, and specified random effects of participant[4] As predicted, when participants saw a commercial thrice rather than once, they thought that the people it featured were more likely to be actors (*M*s = 67.57, 95% CI [66.04, 69.1] vs. 62.9, 95% CI [61.53, 64.4], *SD*s = 29.35 vs. 27.53), $t(1,414) = 4.61$, $p < .001$, $d = 0.16$, 95% CI [0.09, 0.24]. We find no evidence that replay signaled atypicality. The wording of the dependent measure may have shaped participants' general perceptions of targets—they rated targets as marginally more likely to be actors in the "actors" frame compared to the "real people" frame (*M*s = 66.19 vs. 64.3, 95% CI [64.74, 67.65]

[4] All results remain identical in significance and direction if a mixed ANOVA is used instead.

vs. 95% CI [62.79, 65.82], *SD*s = 28.25 vs. 28.83), $t(1,415) = 1.655$, $p = .098$, $d = 0.07$, 95% CI [-0.01, 0.14]. However, the effect of replay did not differ by frame, $t(1,414) = 0.030$, $p = .976$ (see Figure 4). The non-significant interaction is inconsistent with the experimenter demand account proposed earlier, whereby participants infer that a target is replayed because it is "special".

*Figure 4.* Mean rating (from 0 to 100) for the probability of the target being an "actor, not a real person" ("actor" frame) or the probability of the target being a "real person, not an actor" ("not actor" frame) as a function of repeated viewing. Responses in the "not actor" frame were subtracted from 100 to put ratings on the same scale; that is, to reflect the probability of the target being an actor. Error bars reflect ±1 SEM.




## Study 5

Study 5 probes the effect's robustness by removing filler videos and manipulating viewing frequencies. Participants watched a clip from an improvisational ("improv") theater skit and evaluated the extent to which the performance seemed improvised rather than scripted. Critically, the video played anywhere from one to six times in a row; viewing frequency varied between-subjects.

## Method

**Participants.** Three thousand thirty-three participants (50% female, 47% male, 3.3% other/prefer not to answer, $M_{age} = 33$) completed Study 5 through Prolific Academic. Pre-registered sample size was 3,000. This study did not employ an attention check but did include the English comprehension check, excluding those who failed (6.1%) for a final sample of 2,847.

**Materials.** The video was 2 seconds long and showed a man talking animatedly while performing an improv skit onstage.

**Procedure**. Participants were not told that the performance was improvised, but that it came from a "satirical skit at a community theater." After watching the clip anywhere from one to six times in a row, they answered "Do you think that the actor's line was improvised or scripted?" on a 7-point scale (1 = *I think it was scripted*, 7 = *I think it was improvised*). All participants saw the same video clip.

## Results

The results reveal that a single replay affected ratings to the same extent as multiple. A series of planned comparisons found that the actor's line was rated as less improvised (more scripted) for participants who watched the video twice[5] ($N = 488$; $M = 3.79$, 95% CI [3.62, 3.95], $SD = 1.85$) rather than once ($N = 442$; $M = 4.14$, 95% CI [3.96, 4.31], $SD = 1.84$), $t(920) = -2.86$, $p = .004$, $d = 0.19$, 95% CI [0.06, 0.32]. The same was true for those who watched it three ($N = 444$; $M = 3.84$, 95% CI [3.66, 4.02], $SD = 1.96$, $t(875) = -2.29$, $p = .02$, $d = 0.15$, 95% CI [0.02, 0.29]), four ($N = 472$; $M = 3.78$, 95% CI [3.60, 3.95], $SD = 1.95$, $t(904) = -2.85$, $p = .005$, $d = 0.19$, 95% CI [0.06, 0.32]), five ($N = 507$; $M = 3.58$, 95% CI [3.41, 3.74], $SD = 1.91$, $t(933) = -4.53$, $p < .001$, $d = 0.3$, 95% CI [0.17, 0.43]), or six times ($N = 494$; $M = 3.87$, 95% CI [3.70,

[5] The comparison between 1 and 2 viewings was exploratory in our pre-registration.

4.04], $SD = 1.94$, $t(909) = -2.13$, $p = .03$, $d = 0.14$, 95% CI [0.01, 0.27]). In short, any amount of repetition made the video content seem more scripted. In the real world, whether a person performs precisely the same behavior two, three, or five times, that behavior typically appears more controlled and deliberate than when performed just once.

Moreover, an omnibus ANOVA found that the five repetition conditions did not significantly differ from each other, $F(4, 2351) = 1.71$, $p = .145$, suggesting that one repetition may not differ from many, which may not be surprising given this context. Imagine observing an actor *actually* repeat a line word-for-word with identical inflection, facial expression, and body movements. Even one repetition should make the performance seem scripted rather than improvised, and additional repetitions would be unlikely to change this perception.

## Study 6

Thus far, repetition occurs for no apparent reason; participants are merely told that one video "may be shown more than once, so that you can examine it more closely." Such repeated exposure is unnatural and may raise concerns about experimenter demand. In Study 6, repeated exposure is incidental: participants purposely play and replay clips in service of answering various questions.

Study 6 also tests whether repeated listening suffices or if there is something unique about the visual aspect of repeated viewing. We selected audio clips of spontaneous (real) and volitional (fake) laughter. There is some evidence that laughs from the same person sound more similar when they are fake compared to when they are genuine; that is, genuine laughs are more variable (Lavan et al., 2019). If at least to some extent, listeners interpret the repetition from replay as true repetition, the complete lack of variability between replayed laughs may make them sound fake.

## Method

**Participants.** Participants were 1,514 workers from Prolific Academic (54% female, 43% male, 2.2% non-binary/third gender; $M_{age} = 38$); our pre-registration aimed to recruit 1,500. Study 6 specified stringent exclusion criteria in lieu of an attention check. In Study 6, by way of manipulating replay frequency, participants had to indicate whether two laughs were the same or different on three different occasions; the correct answers were "same", "different" "same". We excluded participants who provided a different answer on any of these questions (10.2%) and/or who failed the English comprehension check (3.4%) were excluded, leaving a total sample of 1,321.

**Materials.** Stimuli were 1-second audio clips of laughter selected from unrelated research on laughter authenticity (Bryant & Aktipis, 2014). In that research, spontaneous laughs were recorded from different speakers while they were talking with friends. Volitional laughs were recorded by instructing participants to laugh into a microphone (i.e., "Now, laugh."). All laughs came from female university students.

**Procedure**. Participants answered a series of three questions designed to manipulate their exposure to a given laugh. Each question asked whether two laughs were the same or different, and participants had to press buttons to hear them (i.e., "Is this laugh {button 1} different than this laugh {button 2}?"). All participants compared laugh A to laugh A for the first and third questions and laugh A to laugh C for the second. Next, participants rated the authenticity of either laugh A (repetition condition) or a laugh that they hadn't heard before (laugh B, single condition). Between the three comparison questions and the authenticity rating, participants in the repeated condition heard the target laugh six times, while those in the single condition heard it once. Participants answered the authenticity question on a 7-point scale (1 = *I think it is totally*

*fake*, 7 = *I think it is totally genuine*). This authenticity question was placed on the same page as a question about the gender of the person laughing, which was a filler intended to make our interest in authenticity less obvious.

Importantly, both the repeated laugh and the target laugh were either spontaneous *or* volitional (680 participants heard the spontaneous laugh, 641 heard the volitional). That is, participants rated the authenticity of either a real or fake laugh, and it had either been played earlier (repeated condition, $N = 646$) or preceded by a different laugh that was also real or fake (single condition, $N = 675$). Thus, this study employed a 2(repeated vs. single) x 2(spontaneous laughter vs. volitional) fully between-subjects design.

**Results**

We then examined participants' ratings of the authenticity of the target laugh in a 2 (replay: target played multiple times vs. once) x 2 (laughter: spontaneous vs. volitional) between-subjects ANOVA. Participants reliably distinguished spontaneous laughter from volitional laughter, rating the former to be more genuine (*M*s = 4.45, 95% CI [4.32, 4.59] vs. 3.12, 95% CI [3.00, 3.24], *SD*s = 1.78 vs. 1.49), $F(1,1317) = 218.43$, $p < .001$, $d = 0.81$, 95% CI [0.70, 0.92]. But despite their ability to detect that difference, judgments were also influenced by replay: Participants thought that the laughter was less genuine when they heard it multiple times compared to once (*M*s = 3.68, 95% CI [3.55, 3.82] vs. 3.92, 95% CI [3.79, 4.06], *SD*s = 1.73 vs. 1.81), $F(1,1317) = 7.13$, $p = .008$, $d = 0.14$, 95% CI [0.03, 0.24]. Interestingly, we also observed a significant interaction between laughter type and repetition, $F(1,1317) = 10.31$, $p = .001$. The genuine laughter was rated as less genuine when heard multiple times rather than once (*M*s = 4.19, 95% CI [3.99, 4.38] vs. 4.70, 95% CI [4.52, 4.88], *SD*s = 1.82 vs. 1.70), $t(1317) = -4.10$, $p < .001$, $d = 0.29$, 95% CI [0.14, 0.44], but replay had no effect on the perceived authenticity of

volitional (fake) laughter (repeated: $M = 3.15$, 95% CI [2.99, 3.31], $SD = 1.45$; single: $M = 3.09$, 95% CI [2.92, 3.25], $SD = 1.53$), $t(1317) = 0.49$, $p = .624$, $d = 0.04$, 95% CI [-0.11, 0.2]. That said, a closer examination suggests that the null effect for fake laughter may result from one of the fake laughs showing the opposite pattern (i.e., sounding *more* authentic upon replay). Thus, three out of four laughs were perceived as less authentic when heard repeatedly, and we are unable to identify why one showed the opposite pattern. More broadly, these results indicate that repeated listening produces the same effect as repeated viewing, reducing perceptions of behavior's authenticity when authenticity requires spontaneity. Note that this pattern of results is inconsistent with an alternative account in which repeated exposure to a stimulus increases judgments of its emotional intensity (Mrkva & Van Boven, 2020). Here, and in Studies 2 and 3, replay instead reduced the perception that the behavior reflected a strong, genuine emotion. Further, the design of this study suggests that incidental repetition is sufficient to produce the effect, and replay does not simply enhance accuracy.

## Study 7

Our final study examines replay and authenticity in a different, highly consequential context. Either once or thrice, participants watched a murder suspect reacting to the news of her target's death—a spontaneous emotional reaction that, if rewatched, may appear less genuine, and thus make the suspect appear more guilty.

### Method

**Participants.** Two thousand one hundred ninety-four (46% female, 52% male, 2% other/prefer not to answer; $M_{age} = 39$) were recruited from Prolific Academic. Note that we over-recruited from the pre-registered sample of 2,000 in anticipation of greater exclusions for this study. Namely, as per our pre-registration, subjects were excluded if they indicated that they were

already familiar with the murder case in question (13.4%). We also excluded those who failed to correctly report the number of times they viewed the video (3.6%), or failed the attention (2.0%) and/or English comprehension check (2.7%). This left a final sample of 1,734.

**Materials.** Our stimulus was a 3-second video clip taken from a 2009 episode of "Cops". In the clip, a police officer tells a woman named Dalia Dippolito that her husband had been killed. Dalia reacts with anguish, sobbing and clutching the officer. Notably, her husband was still alive, having avoided an unsuccessful attempt on his life; the police suspected that Dalia was involved so they lied to her to gauge her reaction[6].

**Procedure**. Before watching the clip, we oriented participants to look for either evidence of guilt, or evidence of innocence (*N*'s assigned to each condition were 889 and 845, respectively; "Thus far, the evidence suggests that she is [guilty/innocent], but we want to see whether *you* think that's true."). This framing manipulation helps address whether replay merely magnifies viewers' pre-existing biases; such an account would be supported if replay *reduces* perceived guilt when participants were told to look for evidence of innocence. Participants watched Dalia's devastated reaction either once or three times in a row ($N = 822$ and $N = 912$, respectively). Unlike many of the previous studies, participants were not told about the possibility of replay at the outset; "Here's that video one more time, just so you can get a better sense of it" and "…and one final time" introduced the second and third viewings, respectively.

Participants indicated whether they thought Dalia was guilty or innocent (coded as 1 and 0) and explained their choice in a free-response text box. Importantly, participants had to report how many times they had watched the video either *before* or *after* answering the dependent measure ("intervention" and "no intervention" conditions, respectively; N's = 887 and 847). This

[6] Dalia was eventually charged with attempted murder and is now in prison for her crime. Given widespread media attention at the time, we excluded participants who indicated that they were familiar with Dalia's case (13.4%).

allowed us to explore whether drawing attention to the distinction between replay and true repetition could weaken the effect. Finally, participants indicated whether they had known about this case or seen the video beforehand.

**Results**

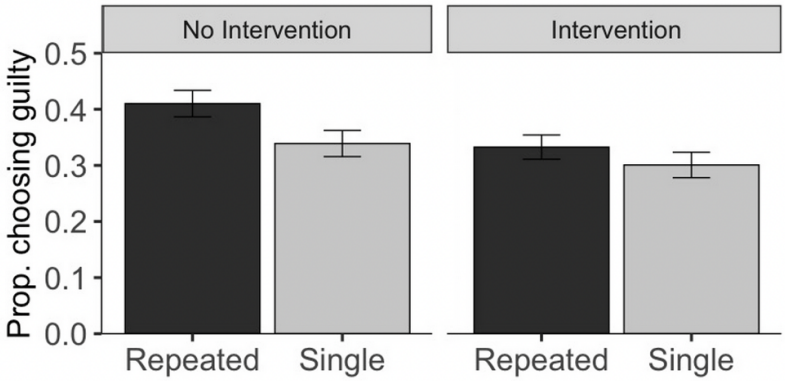
We performed a 2 (number of viewings: one vs. three) x 2 (frame: looking for guilt vs. innocence) x 2 (intervention: yes vs. no) between-subjects logistic regression. The target was more likely to be judged as guilty when participants had been instructed to look for evidence of guilt instead of innocence (respectively, 41%, 95% CI [25%, 31%] vs. 28%, 95% CI [38%, 45%] of participants chose "guilty"), z = -5.98, $p < .001$, OR= 0.73, 95% CI [0.66, 0.81].

More importantly, participants were also more likely to deem the target guilty when they had watched the video three times compared to once (37% vs. 32% chose "guilty"; 95% CI [34%, 40%] vs. 95% CI [29%, 35%] ), $z = 2.16$, $p = .031$, OR= 1.12, 95% CI [1.01, 1.24]. This result echoes our findings in the previous studies, and suggests that this effect also occurs in this highly impactful criminal judgment context.

We had no a priori predictions about the impact of the "intervention"– asking participants to report how many times they had seen the video before making their judgment – and its effects were less clear. The interaction between replay condition and intervention did not achieve significance, $z = -.981$ $p = .327$, although the fully between-subject, binary outcome design of this study may leave us underpowered to detect a full attenuation. Perhaps suggestive of an attenuation, an analysis of the simple effects found that repeated viewing affected judgments in the "no intervention" condition (respectively, 41%, 95% CI [36%, 46%] vs. 34%, 95% CI [29%, 39%] of participants in the repeated and single conditions chose "guilty"), $z = 2.25$, $p = .024$, OR= 1.18, 95% CI [1.02, 1.36]. In contrast, we observed no difference in the "intervention"

condition (33%, 95% CI [29%, 38%] of participants in the repeated and 30%, 95% CI [26%, 35%] of those in the single condition chose "guilty), $z = 0.82$, $p = .412$, OR = 1.06, 95% CI [0.92, 1.23] (see Figure 5).

*Figure 5.* Proportion of respondents who judged the target to be guilty as a function of repeated viewing. Participants in the "no intervention" condition had to indicate the number of times they watched the video at the end of the study, while those in the "intervention" condition did so prior to making their judgments. Error bars reflect ±1 SEM.




A pre-registered exploratory analysis examined whether participants in the repeated condition would, of their own volition, cite preparation as the reason for their choice. After judging the target's guilt, participants were asked to justify their answer, and we coded whether their justification cited preparation (see Supplemental Material). We found that participants were more likely to mention the target seeming prepared in the repeated viewing condition compared to the single (28% of responses vs. 22%), $\chi^2$ (1,734) = 9.2, $p = .002$, $\phi = .07$, 95% CI [.01, .12].

These results demonstrate that replay may impact judgments of criminal culpability, at least in the context studied here. With the growing prevalence of body cameras, surveillance

footage, and cellphone recordings, videos capturing spontaneous emotional reactions are increasingly likely to be key evidence in criminal trials, making any effects of replay on their interpretation consequential (McMillan et al., 2013). Moreover, by reporting how many times they had watched the video, repeated viewers' judgment became identical to that of single viewers, offering a straightforward way to eliminate the effect.

## General Discussion

In the modern world, recording, sharing, and replaying is easier and more common than ever before, underscoring the importance of understanding how these behaviors impact perceiver judgments. The present research examines the consequences of video re-watching for impressions of spontaneity and preparation. We identify a "replay illusion", wherein viewers perceive actors as if they were precisely reenacting their behavior. The lack of variability between apparent repetitions suggests tight control, typically achieved through extensive rehearsal, and diminishes the impression of spontaneity. We document this phenomenon across diverse stimuli, judgments, and experimental designs. Further, we test and eliminate alternate accounts involving experimenter demand, mere exposure, or enhanced viewer accuracy.

Of course, replay may not uniformly *reduce* accuracy. In some contexts, individuals may initially underestimate preparation, with re-watching bringing their judgments closer to the truth. Indeed, re-watching is sometimes necessary for a full understanding, and future research may examine the tradeoffs between replay-induced distortions and the benefits of additional information. But in other situations, replay may impair accuracy, as observed in Study 3, where genuine emotional reactions appeared contrived. In such cases, replay may be particularly damaging because viewers may not actively correct for the distortion, potentially amplifying its

effects. It is also important to note that any bias from replay is especially likely to impact consequential videos, given their higher likelihood of being replayed.

As the mélange of topics explored in the present paper suggests, replay may inadvertently shape a wide range of consequential judgments. A lack of spontaneity means that behavior is more thoughtful, measured, and thorough, positively shaping perceptions of the actor and potentially signaling effort, quality, and skill (e.g., Studies 1a-c). Conversely, replay may transform a public figure's accidental remarks into calculated offenses or diminish the charm of their off-the-cuff humor. Importantly, any effects on public opinion may also impact associated support behaviors, such as voting, purchasing, and shareholding.

These findings also bear relevance to commercial advertising. Advertisers may want to deprioritize repeated exposure to commercials that feature the supposed first-time reactions of "real people". Repeated viewing may undermine the authenticity of such reactions (Study 4), potentially diminishing the effectiveness of ads that feature them.

Many of the most consequential interpersonal judgments are made in a courtroom, a context rife with replay. Study 7 demonstrated how repeated viewing can change viewer's judgments of those involved with a crime by making their actions seem less spontaneous. Before replaying a clip for the jury, lawyers should consider whether changing perceptions of the spontaneity of a defendant's behavior might also change perceptions of his guilt, and the legal system should consider ways to mitigate these effects.

Participants' experience of the stimuli closely matches how much content is consumed, given that it involved watching short clips on personal devices like computers or smartphones. Nonetheless, further research would need to determine how the effects documented in the present paper might ultimately transfer to other contexts, like the courtroom. The generalizability of our

findings may also be limited by geography, as we only recruited participants located in the United States. However, we believe that the basic effect (i.e., failing to fully disentangle replay with true repetition) should persist across cultures, insofar as it results from basic human perception. That said, the specific implications of this perceptual bias may differ across cultural contexts. For instance, in societies that value innate ability above diligence and practice, perceiving actors as having prepared more may lead to less favorable evaluations.

The accessibility of recorded audio and video content has introduced new complexities to social perceptions. Much as altered perspectives and slow motion seems to alter viewer judgments, so too does mere repetition. To understand how a video will influence its viewer, one will need to consider not only whether it is viewed, but whether it is viewed again.

**References**

Aboulhosn, S. (n.d.). *18 Facebook statistics every marketer should know in 2020*. Retrieved January 20, 2021, from https://sproutsocial.com/insights/facebook-stats-for-marketers

Barr, D.J., Levy, R., Scheepers, C., Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. Annual review of psychology, 71, 499-515. doi: 10.1146/annurev-psych-010419-050807.

Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior*, *35*(4), 327–335. https://doi.org/10.1016/j.evolhumbehav.2014.03.003

Caruso, E. M., Burns, Z. C., & Converse, B. A. (2016a). Slow motion increases perceived intent. *Proceedings of the National Academy of Sciences*, *113*(33), 9250–9255. https://doi.org/10.1073/pnas.1603865113

Chmielewski, D. (n.d.). *Majority of Americans Still Get Their News By Turning on the TV, Study Finds*. Retrieved February 22, 2018, from https://deadline.com/2018/12/how-americans-watch-news-study-tv-online-pew-research-center-1202512745/

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38). https://doi.org/10.1073/pnas.1702247114

Cowen, A. S., Prasad, G., Tanaka, M., Kamitani, Y., Kirilyuk, V., Somandepalli, K., Jou, B., Schroff, F., Hartwig, A., Brooks, J. A., & Keltner, D. (2021). *How emotion is*

*experienced and expressed in multiple cultures: A large-scale experiment* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/gbqtc

Donchev, D. (n.d.). *37 Mind Blowing YouTube Facts, Figures, and Statistics*. Retrieved June 2, 2021, from https://fortunelords.com/youtube-statistics/

Gershon, R., & Smith, R. K. (2020). Twice-told tales: Self-repetition decreases observer assessments of performer authenticity. *Journal of Personality and Social Psychology*, *118*(2), 307–324. https://doi.org/10.1037/pspi0000183

Granot, Y., Balcetis, E., Feigenson, N., & Tyler, T. (2018). In the eyes of the law: Perception versus reality in appraisals of video evidence. *Psychology, Public Policy, and Law*, *24*(1), 93.

Gregory, R. L. (1968). Visual Illusions. *Scientific American*, *219*(5), 66–76. https://doi.org/10.1038/scientificamerican1168-66

Grush, J. E. (1976). Attitude formation and mere exposure phenomena: A nonartifactual explanation of empirical findings. *Journal of Personality and Social Psychology*, *33*(3), 281–290. https://doi.org/10.1037/0022-3514.33.3.281

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

Hawkins, S. A., & Hoch, S. J. (1992). Low-Involvement Learning: Memory without Evaluation. *Journal of Consumer Research*, *19*(2), 212. https://doi.org/10.1086/209297

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review*, *26*(1), 90–102. https://doi.org/10.3758/s13423-018-1497-7

McMillan, J. E. R., Glisson, W. B., Bromby, M. (2013). Investigating the Increase in Mobile Phone Evidence in Criminal Activities, *46th Hawaii International Conference on System Sciences*, 4900-4909. doi: 10.1109/HICSS.2013.366

McWhinnie, H. J., Segall, M., Campbell, D., & Herskovits, M. (1970). The Influence of Culture on Visual Perception. *Art Education*, *23*(6), 30. https://doi.org/10.2307/3191488

Montoya, R. M., Horton, R. S., Vevea, J. L., Citkowicz, M., & Lauber, E. A. (2017). A re-examination of the mere exposure effect: The influence of repeated exposure on recognition, familiarity, and liking. *Psychological bulletin*, *143*(5), 459.

Mrkva, K., & Van Boven, L. (2020). Salience theory of mere exposure: Relative exposure increases liking, extremity, and emotional intensity. *Journal of Personality and Social Psychology*, *118*(6), 1118–1145. https://doi.org/10.1037/pspa0000184

O'Brien, E. (2021). A mind stretched: The psychology of repeat consumption. *Consumer Psychology Review*, *4*(1), 42–58. https://doi.org/10.1002/arcp.1062

Rodriguez, S. (n.d.). *TikTok passes Instagram as second most popular social app for U.S. teens*. Retrieved November 18, 2020, from https://www.cnbc.com/2020/10/06/tiktok-passes-instagram-as-second-most-popular-social-app-for-us-teens.html

Turner, B. L., Caruso, E. M., Dilich, M. A., & Roese, N. J. (2019). Body camera footage leads to lower judgments of intent than dash camera footage. *Proceedings of the National Academy of Sciences*, *116*(4), 1201–1206. https://doi.org/10.1073/pnas.1805928116

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

## Supplementary Materials

*More information about recruitment of participants.* Participants were recruited from Prolific Academic. We limited recruitment to those who are fluent in English and reside in the United States.

*Materials.* Unless noted otherwise, all stimuli were taken from YouTube and trimmed using Kapwing, a free online video editor.

*Stimulus sampling*. Any research program should be concerned about how experimental stimuli are chosen, and the ability to generalize any finding will depend in part on that process (Wells & Windschitl, 1999; Yarkoni, 2019). We tried to handle this issue in a few ways. First, across the nine experiments, we use recordings sampled from domains as diverse as choreographed dances, actor auditions, television advertisements, and experimentally induced reactions. Although this is not a random sample, we believe it is extensive enough that stimulus sampling issues are unlikely to drive our results. Second, any given study manipulates whether the participant experiences a stimulus just once or repeatedly, meaning that our critical manipulation is never confounded with stimulus selection. Finally, across the studies we develop and refine our predictions so that variation in stimuli is a meaningful component of our primary predictions and conclusions.

### Exclusion criteria

*English comprehension check.* An English comprehension check was used in every study. The question showed this image:

Participants were given a text entry box to answer the question, “What is the man throwing in this picture?” We considered a “passing” response to be either “basketball” or “basket ball” (any capitalizations allowed). Participants who did not pass this check were excluded (see Table S1 for failure rates broken down by study).

*Attention check.* Studies 1a, 1b, 1c, 2, and 4 included an attention check in which participants had to select which of four images (stills taken from videos) were from a video that they did not

see. Participants who failed to identify the correct video were excluded. Studies 3 and 5 did not include an attention check. Study 6 specified stringent exclusion criteria in lieu of an attention check. In Study 6, by way of manipulating replay frequency, participants had to indicate whether two laughs were the same or different on three different occasions; the correct answers were "same", "different" "same". Participants who provided a different answer on any of these questions were excluded.

Table S1. *Information About Exclusions and Sample Size*

| Study | Original N | Percentage Failing Attention Check | Percentage Failing English Check | Final N |
|---|---|---|---|---|
| 1 | 223 | 4.0% | 5.4% | 204 |
| 1b | 303 | 3.0% | 1.3% | 292 |
| 1c | 344 | 0.3% | 2.0% | 336 |
| 2 | 501 | 0.2% | 1.2% | 494 |
| 3 | 799 | n/a | 3.6% | 770 |
| 4 | 1501 | 1.7% | 4.3% | 1417 |
| 5 | 3033 | n/a | 6.1% | 2847 |
| 6 | 1514 | 10.2% | 3.4% | 1321 |
| 7 | 2194 | 2.0% | 2.7% | 1734* |

*We note that Study 7 also excluded people who failed to report the correct number of viewings (1 in the 1x condition, 3 in the 3x). This was 3.6% of participants. We also excluded participants who indicated that they were familiar with Dalia's case (13.4%).

Additional note: Study 6 had a more substantial exclusion rate based on failing the attention check. This is likely because Study 6 used a much more stringent attention check than did other studies. Importantly, the results from all studies hold without any exclusions (see Table S2).

## Alternate analyses

As a robustness check, in the table below we report results from each study 1) without exclusions, 2) using linear regression, instead of a mixed model, that simply controls for participant and video, and 3) with a basic paired-samples t-test. Note that not all alternate analyses apply to every study. Namely, Studies 5, 6, and 7 are entirely between-subject and thus cannot be examined using a paired-samples t-test. Studies 5 and 6 already use linear regression. Table S2 displays the results from these alternate specifications. Note that for studies that employed them, both primary DVs (e.g., rehearsal, authenticity) and auxiliary DVs (e.g., seriousness) are included.

Table S2. *Alternate Analyses for Each Study's Key Result*

| Exp | DV | Without exclusions | Linear regression | Paired t-test |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 1 | Preparation<br><br>Effort | $t(442.1) = 4.46, p < .001$<br><br>$t(442.1) = 2.66, p = .008$ | $t(404) = 4.31, p < .001$<br><br>$t(404) = 2.38, p = .018$ | $t(203) = 3.38, p < .001$<br><br>$t(203) = 1.95, p = .053$ |
| 1b | Rehearsal<br><br>Skill | $t(298.06) = 7.07, p < .001$<br><br>$t(298.04) = 4.83, p < .001$ | $t(580) = 5.68, p < .001$<br><br>$t(580) = 3.22, p = .001$ | $t(291) = 6.85, p < .001$<br><br>$t(291) = 4.77, p < .001$ |
| 1c | Rehearsal<br><br>Skill | $t(342.00) = 6.92, p < .001$<br><br>$t(342.03) = 1.83, p = .07$ | $t(668) = 4.71, p < .001$<br><br>$t(668) = 1.44, p = .15$ | $t(335) = 6.66, p < .001$<br><br>$t(335) = 1.75, p = .08$ |
| 2 | Authenticity | $t(499.02) = -4.99, p < .001$ | $t(984) = -4.86, p < .001$ | $t(493) = -3.10, p = .002$ |
| 3 | Authenticity | $t(797) = -2.61, p = .009$ | $t(1536) = -2.56, p = .011$ | $t(769) = -2.73, p = .006$ |
| 4 | Probability of target being an actor | $F(1,1498) = 19.77, p < .001$ | $F(1,2828) = 18.52, p < .001$ | $t(1416) = 4.64, p < .001$ |
| 5 | Probability that actor's line is improvised, not scripted | 1x vs. 2x: $t(990) = -2.80, p = .005$<br><br>1x vs. 3x: $t(937) = -2.17, p = .031$<br><br>1x vs. 4x: $t(977) = -1.90, p = .058$<br><br>1x vs. 5x: $t(992) = -4.39, p < .001$<br><br>1x vs. 6x: $t(971) = -2.12, p = .034$ | n/a | n/a |
| 6 | Authenticity | $F(1,1509) = 9.32, p = .002$ | n/a | n/a |
| 7 | Guilt | $z = 2.29, p = .022$ | $z = 2.17, p = .030$ | n/a |

**ANOVA Analysis for Study 4**

Here we report results for Study 4 when they are analyzed using a mixed ANOVA, instead of a mixed effects linear regression. We find results which are the same in both direction and statistical significance. As in the main text, we subtracted scores in the "not actor" condition from 100 to place responses on the same scale (wherein higher values reflect a higher probability

of being an actor). We then examined participants' ratings (i.e., of the target being an actor) in a 2 (replay: target video shown once vs. thrice; within subjects) x 2 (frame: "actor" vs. "not actor"; between subjects) mixed ANOVA. Participants rated targets as marginally more likely to be actors in the "actors" frame compared to the "real people" frame (*M*s = 66.19 vs. 64.31, *SD*s = 28.25 vs. 28.83), $F(1,1415) = 2.74$, $p = .098$. When participants saw a commercial thrice rather than once, they thought that the people it featured were more likely to be actors (*M*s = 67.57 vs. 62.97, *SD*s = 29.35 vs. 27.53), $F(1,1414) = 21.21$, $p < .001$, and this effect did not differ by dependent measure frame, $F(1,1414) = 0.001$, $p = .976$.

**Content Analysis for Study 7**

Five research assistants read through and coded the free responses of participants for instances where the actions of the woman in the video were described as showing preparation or planning. The raters all coded 75 responses in common, which were used to assess their inter-rater reliability with Fleiss's Kappa, a measurement of inter-rater reliability controlling for chance which can be used when there are more than two raters (Fleiss 1971). We found a $\kappa = .49$, which indicates good agreement between the raters.

This was done so that we could see how the mentions of preparation varied by condition – under our account, we expected that mentions of preparation would be more frequent when the video was viewed multiple times (as opposed to once). Indeed, this is what we found. Mentions of preparation were more frequent in the replay condition (28% of responses) than in the single viewing condition (22% of responses, $\chi^2(1,734) = 9.2$, $p = .002$). This finding was confirmed by a logistic regression regressing repetition condition and intervention condition onto a binary indicator of whether the participant's free response mentioned preparation. This regression found that being in the replay condition increased the chances that a response would mention preparation (OR = 1.61, 95%CI [1.29, 1.92]), while all other terms were not significant (all *p*'s > .05).

The table below shows the total number of participants within each condition-intervention cell by the judgment they made (guilty or innocent), as well as the total number of times they mentioned preparation.

Table S3. *Mentions of Preparation as a Function of Verdict, Viewing Condition, and Intervention, with Exclusions*

| Condition | Intervention | Participant Judgment | Total N | Total Mentions Preparation | % |
|---|---|---|---|---|---|
| 1x | No | Guilty | 140 | 84 | 57% |
| 1x | No | Innocent | 273 | 9 | 3% |
| 1x | Yes | Guilty | 123 | 77 | 57% |
| 1x | Yes | Innocent | 286 | 7 | 3% |
| 3x | No | Guilty | 178 | 131 | 69% |
| 3x | No | Innocent | 256 | 7 | 3% |
| 3x | Yes | Guilty | 159 | 106 | 63% |
| 3x | Yes | Innocent | 319 | 10 | 3% |

These results held without exclusions, as mentions of preparation are still more frequent in the replay condition (29% of responses) than in the single viewing condition (22% of responses, $\chi^2$ (2131) = 3.60, $p < .001$), and this result was again confirmed by a logistic regression regressing repetition condition, intervention condition, and their interaction onto a binary indicator of whether the participant's free response mentioned preparation. This regression found that being in the replay condition increased the chances that a response would mention preparation (OR = 1.63, 95%CI [1.36, 1.91]), while all other terms were not significant (all $p$'s > .05).

The table below recreates Table S3, but without exclusions.

Table S4. *Mentions of Preparation as a Function of Verdict, Viewing Condition, and Intervention, without Exclusions*

| Condition | Intervention | Participant Judgment | Total N | Total Mentions Preparation | % |
|---|---|---|---|---|---|
| 1x | No | Guilty | 209 | 117 | 57% |
| 1x | No | Innocent | 336 | 11 | 3% |
| 1x | Yes | Guilty | 201 | 114 | 57% |
| 1x | Yes | Innocent | 336 | 9 | 3% |
| 3x | No | Guilty | 258 | 179 | 69% |
| 3x | No | Innocent | 297 | 9 | 3% |
| 3x | Yes | Guilty | 224 | 142 | 63% |
| 3x | Yes | Innocent | 363 | 12 | 3% |

**Supplement References**

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378–382. https://doi.org/10.1037/h0031619

Wells, G. L., Windschitl, P.D. (1999). Stimulus sampling and social psychological experimentation. *Pers. Soc. Psychol. Bull.* **25**, 1115-1125. https://doi.org/10.1177%2F01461672992512005

Yarkoni, T. (2019). The generalizability crisis, *PsyArXiv*. https://doi. org/10.31234/osf. io/jqw35