Cognition xxx (2018) xxx-xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com

Original Articles

Your visual system provides all the information you need to make moral judgments about generic visual events

Julian De Freitas , George A. Alvarez

Harvard University, USA

ARTICLE INFO

Article history:
Received 29 November 2017
Received in revised form 21 May 2018
Accepted 22 May 2018
Available online xxx

Keywords:
Moral judgment
Perceived causality
Visual illusions

ABSTRACT

To what extent are people's moral judgments susceptible to subtle factors of which they are unaware? Here we show that we can change people's moral judgments outside of their awareness by subtly biasing perceived causality. Specifically, we used subtle visual manipulations to create visual illusions of causality in morally relevant scenarios, and this systematically changed people's moral judgments. After demonstrating the basic effect using simple displays involving an ambiguous car collision that ends up injuring a person (E1), we show that the effect is sensitive on the millisecond timescale to manipulations of task-irrelevant factors that are known to affect perceived causality, including the duration (E2a) and asynchrony (E2b) of specific task-irrelevant contextual factors in the display. We then conceptually replicate the effect using a different paradigm (E3a), and also show that we can eliminate the effect by interfering with motion processing (E3b). Finally, we show that the effect generalizes across different kinds of moral judgments (E3c). Combined, these studies show that obligatory, abstract inferences made by the visual system influence moral judgments.

© 2018.

1. Introduction

We have less control of our moral judgments than we might think. Converging evidence shows that priming, highlighting, or framing one factor over another can influence moral judgments (Gu, Zhong, & Page-Gould, 2013; Haidt, Koller, & Dias, 1993; Petrinovich & O'Neill, 1996; Wheatley & Haidt, 2005). Furthermore, scientists have also exploited the dynamics of eye gaze while subjects are making a moral decision to bias subjects toward making one particular moral decision over another (Pärnamets et al., 2015).

Here we hypothesized that, in addition to being susceptible to such *behavioral* manipulations, moral judgments should also be susceptible to quirks of how the *visual system automatically interprets the world*. Specifically, we predicted and found that certain visual illusions, when present within a moral context, will distort the perception of causal relations in those moral contexts, leading people to make different moral judgments than they otherwise would. Furthermore, we used manipulations that had this effect without observers ever being aware that their moral judgments were being changed. This does not mean that moral judgments are not also influenced by non-visual factors (e.g., knowing that a person is a criminal). Yet insofar as our subtle, unrecognized visual manipulations changed moral judgments, this suggests that moral judgments about these visual scenes are based on causal information that is read out from the visual system.

In the introduction that follows, we explain why we have chosen to use visual illusions as well as what is known about visual illusions of causality per se. Next, we discuss how causal perception might be linked to cognitive processing, focusing on moral judgment as a case study.

2. Visual Illusions, and the distinction between perception and cognition

Visual illusions are perceived images or events that differ from objective reality. The best illusions exploit knowledge of how the visual system works in order to make people see features or events that are not an accurate reflection of the world, thereby illustrating the nature of the visual inferences that underlie perception.

By 'visual illusion' we will refer specifically to 'toy' displays that have been deliberately designed to demonstrate that the visual system has made an inference that goes beyond the literal features of the objects or their retinal projections (the experimenters know this, of course, because they created the stimuli). For instance, a perceiver can be tricked into seeing movement in a still image (waterfall illusion; Crane, 1988), or seeing two objects as having different brightness when they are in fact identical (shadow illusion; Adelson, 1999). Yet although these are toy displays, it is likely that the mechanisms uncovered in the illusion also operate on naturalistic stimuli. Indeed, it is likely that many visual illusions expose mental algorithms that capitalize on stable relationships in the statistics of the visual input

Corresponding author at: Department of Psychology, Harvard University, William James Hall 964, 33 Kirkland Street, Cambridge, MA 02138, USA.
Email address: defreitas@g.harvard.edu (J. De Freitas)

https://doi.org/10.1016/j.cognition.2018.05.017
0010-0277/ © 2018.

UNCORRECTED PROOF

present in our environments (Olshausen & Field, 1996; Purves, Monson, Sundararajan, & Wojtach, 2014; Turk-Browne, Jungé, & Scholl, 2005). It is for this reason that visual illusions are not necessarily problematic biases that fall short of how we would want a perfect observer to see the world (Felin, Koenderink, & Krueger, 2017; Rogers, 2014).

Although most famous visual illusions entail a distortion of literal, low-level features such as edges, color, and orientation, some visual illusions entail higher-level inferences about hidden variables such as identity, animacy, and causality (for a review, see Scholl & Tremoulet, 2000). These phenomena are intriguing because they suggest that the visual system also has something to say about features of the world that are typically considered to be squarely within the domain of higher-level cognition.

Here, methods from visual psychophysics, and visual illusions in particular, can be used to determine whether the given phenomenon is truly perceptual or just cognitive. This is because, unlike purely cognitive phenomena, visual illusions show a number of features that are distinctive of visual processing: (1) they are cognitively impenetrable, meaning that knowing it's an illusion doesn't alter what you see, suggesting that the process that gives rise to the percept is encapsulated from other processes, (2) the phenomena occur very fast, i.e., almost instantaneously upon viewing the displays, (3) they are largely stimulus driven, such that objectively small manipulations to the displays can lead the percepts to disappear, (4) they are categorical, i.e., there is only one specific percept, or only a limited set of qualitatively distinct percepts, e.g., bistable stimuli like the necker cube (Necker, 1832), and (5) implicit manipulations give rise to these illusions, such that observers are often unaware that they are experiencing the illusion or what manipulations gave rise to them. We do not mean to say that observers are *unaware of the stimuli at all*, only that they are unaware of how the arrangement of stimuli influences their perception. Indeed, experiencing visual illusions typically requires the observer to perceive and perhaps even attend to the items.

Thus, if a high-level inference meets these various criteria, then we can generally conclude that it is as much of a visual representation as visual inferences like brightness, color, and depth perception.

2.1. Causal perception

One such high-level visual inference is the perception of causality (Michotte, 1946, 1963). The Belgian experimental psychologist Albert Michotte was the first to notice the following: if one object moves and stops next to a second object, and then that second object moves away within a certain temporal window, people cannot help but see the first object *cause* the second to move, even though this causal information is not present in the stimuli themselves nor in their retinal projections (Michotte, 1946, 1963). He noted that this must be a causal illusion, because the events are objectively described as a sequence of objects at different locations at different times, without any need to refer to whether the interaction between them was causal or non-causal (indeed, one only needs the spatiotemporal coordinates to program such events on a computer).

For a while it was debated whether recognizing causality in Michotte's experiments might instead be computed by higher-level cognition, since observers in these experiments are free to reason about the stimuli or may feel pressure to respond in a particular way, e.g., in order to please the experimenter. Yet later work in visual psychophysics, using more subtle and indirect measures, has marshaled strong evidence that at least a subset of the phenomena originally studied by Michotte is indeed perceptual; here is a brief summary of this evidence:

Illusions of causality emerge as early as six-months, before language emerges (Leslie, 1982; Leslie & Keeble, 1987), and even in non-human primates (Matsuno & Tomonaga, 2017), showing that these effects cannot be due to response bias; causal illusions warp other perceived properties of the stimuli, including their extent of spatial overlap (Scholl & Nakayama, 2004); causal illusions are induced by hallmark perceptual manipulations such as grouping manipulations and events that occur post-dictively within a fixed temporal window (Choi & Scholl, 2004; Choi & Scholl, 2006; Scholl & Nakayama, 2002); they interact with other perceptual processes like apparent motion and the perception of space and time (Buehner & Humphreys, 2010; Cravo, Claessens, & Baldo, 2009; Kim, Feldman, & Singh, 2013); they preferentially break into awareness despite continuous flash suppression (Moors, Wagemans, & de-Wit, 2017); they correlate with activity in brain area V5, which is located high in the visual processing hierarchy (Blakemore et al., 2001); they induce retinotopic adaptation (Rolfs, Dambacher, & Cavanagh, 2013), i.e., if you show a number of causal interactions on a specific location of the retina, then subsequently presented interactions look less causal if you present them at that same location on the retina but not at other retinal locations; and factors that influence judgments of causality have no detectable effect on perceived causality (Schlottmann & Shanks, 1992).

One of the most compelling demonstrations of causal perception — both from methodological and phenomenlogical standpoints — is the 'causal capture' illusion (Scholl & Nakayama, 2002). Observers see displays wherein two objects interact in a non-causal manner, because the first object overlaps completely with the second object, before that second object then moves away. At a sufficiently fast speed, this non-causal interaction begins to look ambiguous; that is, it can be seen in one of three ways: although some observers still see (1) the true *overlap* event, others see (2) a *passing* event, in which the first object moves underneath the second object and continues right passed it (suggesting that somehow it magically morphed into the second object), or (3) a *causal launch*, wherein the first object seems to cause the other to move.

Critically, vision scientists can then employ subtle tricks in order to make such an ambiguous event *consistently look causal*, even though of course it still is not. Specifically, when a causal looking event is shown in the periphery, observers now report seeing a causal interaction in the main overlap event as well — a pure illusion of causality.[1] This illusion can also be elicited by showing just a single object that moves in time with one of the objects in the main overlap event (Choi & Scholl, 2004).

Although a number of mechanisms may be at play in these effects, a general explanation for why they occur may be that the visual system has developed a 'coincidence avoidance' heuristic, whereby stimulating a causal receptor at just the right time may lead the receptor to misattribute irrelevant information to the ambiguous event, making it appear causal even though it wasn't (Scholl & Nakayama, 2002). Although such a heuristic may perhaps seem overly sophisticated for a perceptual system to employ, we note that coincidence avoidance heuristics are embodied in various phenomena that are unequivocally visual, including amodal completion (Kanizsa, 1979; Van Lier & Wagemans, 1999), the tunnel effect (Michotte, Thinès, & Crabbé, 1964), apparent motion (Anstis, 1980; Wertheimer, 1912), illusory conjunctions (Treisman & Schmidt, 1982), auditory-induced bouncing (Sekuler, Sekuler, & Lau, 1997), and others (for a review, see Flombaum, Scholl, & Santos, 2009).

[1] Indeed, these events are typically accompanied by the phenomenon of an "oomph", even though the objects in these displays never actually collide.

UNCORRECTED PROOF

2.2. Linking causal perception to cognitive processing

It is possible that perceived causality is computed only for the purpose of interpreting visual input, and does not influence higher-level cognition. Even within the visual system, this sort of dissociation has been proposed in so-called "blind sight patients." These individuals have damage to their visual stream, leading them to report that they are completely blind, and yet they are able to accurately reach and adjust their grasp to the shape of objects (Weiskrantz, 1986) — all the while insisting that they see nothing. A similar effect is observed within healthy individuals, who will report a perceptual size illusion, and yet accurately adjust their fingers to the correct size when picking up the mis-perceived objects (Rossetti, 1998). In a sense, these phenomena illustrate that perceptual output can exist without directly influencing other aspects of cognition for which that information seems directly relevant — even visually guided action! Thus, it is possible that perceptual representations, including perceived causality, exist only for the purpose of "seeing", e.g., the visual system can make more accurate predictions, and therefore interpret the next view more efficiently, if it has an accurate causal model, and that is all perceived causality is used for.

Alternatively, it is possible that higher-level cognition, such as moral decision making, relies on the output of the perceptual system, such that causality-dependent judgments would be influenced by causal illusions. Under this view, the function of vision extends beyond "just seeing" to uncovering the causal and social structure of the world, helping us to make fast and reliable judgments about more abstract domains such as social cognition.[2]

In a limited sense, one could argue that such a link has already been established insofar as people in causal perception studies are often asked to make physical judgments about the displays, e.g., about the amount of force, contact, or causality (e.g., Mayrhofer & Waldmann, 2014; White, 2012). Yet since such judgments are specific to the perceived physical properties of the stimuli, they do not provide evidence that the effects of causal perception extend into abstract, non-perceptual domains of cognition. Although it would be reasonable to expect that these effects exist, existing studies have simply not yet isolated the role of causal perception per se on abstract cognition. In the absence of such studies, it remains possible that the existing causal perception effects are constrained to literal judgments about physical parameters of simple stimuli, and that these effects will be 'washed out' by the richer knowledge and reasoning brought into play when reasoning about abstract cognitive judgments involving real-world images.

Therefore, in order to begin understanding the link between causal perception and the rest of abstract cognition, we need to study an instance of cognition that differs from visual perception in all the obvious ways, i.e., it is highly abstract, domain-general, slow (rather than fast), and deliberative. We then need to use the methods of psychophysics in order to show that various properties of the causal percept per se predict changes in this chosen cognitive domain.

[2] By "abstract", we mean that the outputs are very different from the literal sensory input; or, to put it in more precise computational terms, the outputs are the product of non-linear transformations to the sensory input (Yamins et al., 2014). Even the visual system is involved in extracting increasingly abstract information from the sensory input (e.g., whether the input contains edges, bounded contours, or object identities). In this sense of abstraction, non-perceptual cognitive systems are traditionally characterized as computing even more abstract information such as causality, intentionality, harm, as well as judgments such as moral judgment, which can also accept other kinds of abstract representations as inputs.

2.3. Moral judgment as a case study

For this purpose, we turn to moral judgment. Not only is moral judgment abstract — a judgment of the extent to which a person has done something blameworthy or wrong — but it can also depend on information that itself is abstract, such as whether an event was caused by an agent, whether the event entailed harm, and whether the agent was causally linked to that harm.

In addition to itself being abstract and depending upon abstract inputs, moral judgment depends on specific combinations of these abstract inputs. To illustrate, someone who initiates a causal interaction that does not result in a harmful outcome is typically less morally blameworthy than someone whose action does lead to a harmful outcome, and similarly, someone is typically less morally blameworthy for a harmful outcome that is not causally related to something that she did than for an outcome that is causally related to her actions. This sort of processing is thus different in many respects from visual processing, which is highly stimulus-driven and occurs obligatorily when light touches the retina.

Moral judgments are also highly domain general, depending on outputs from a host of other cognitive factors, including emotion (e.g., disgust and arousal; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Strohminger & Kumar, 2018), cognitive heuristics and biases (De Freitas & Johnson, 2015; Gu et al., 2013; Haidt et al., 1993; Petrinovich & O'Neill, 1996; Wheatley & Haidt, 2005), inferences about mental states (e.g., Patil, Calò, Fornasier, Cushman, & Silani, 2017; Young & Saxe, 2009), and a person's subjective values, e.g., purity, patriotism, attitude toward other groups, and cultural upbringing (De Freitas & Cikara, 2017; Haidt, 2012; Haidt et al., 1993). Moreover, moral values in turn inform a variety of other mental concepts, such as the notion of a self, and mental state concepts such as happiness and weakness of will (e.g., De Freitas, Cikara, Grossmann, & Schlegel, 2017; Phillips, De Freitas, Mott, Gruber, & Knobe, 2017).

Despite these general differences between visual perception and moral judgment, they share a potential link insofar as causal perception entails an inference that may be relevant to moral judgment. Indeed, almost all models of moral judgment implicate causal inferences (Alicke, 2000; Cushman, 2008; Knobe, 2009; Malle, Guglielmo, & Monroe, 2014; Newman, De Freitas, & Knobe, 2015; Noothigattu et al., 2017; Samland & Waldmann, 2016; Shaver, 1985; Weiner, 1995). Yet, surprisingly, almost all models of moral cognition do not suggest a role for causal perception, instead concerning themselves with cognitive processing that takes place after perception has already occurred. Where models of visual processing are occasionally mentioned, it is typically with the aim of pointing out that these perceptual models are too limited to explain moral judgment (Kim et al., 2018) or other aspects of intuitive psychology more generally (Lake, Ullman, Tenenbaum, & Gershman, 2017). Relatedly, almost all studies of moral judgment are conducted using short verbal vignettes, many involving philosophical conundrums such as the trolley dilemma (Foot, 1978). But, of course, in everyday life we also make moral judgments about events that we directly perceive.

In the handful of moral judgment studies where visual stimuli have been employed (Caruso, Burns, & Converse, 2016; Iliev, Sachdeva, & Medin, 2012; Nagel & Waldmann, 2012), these studies have not (by design) tried to isolate visual processing per se, but they have studied how people reason about such displays. Since observers in these tasks are asked to attend carefully to changes in *task-relevant* events, which are also often manipulated within-subjects, these studies do not bear on whether causal perception per se drives moral

UNCORRECTED PROOF

judgments. To address this gap, the kinds of psychophysical methods reviewed above are necessary.

We believe that considering the potential link between causal perception and moral judgment has a number of promising implications that fall directly out of the different cognitive architectures of these systems. We discuss these implications more deeply in the general discussion. For now, suffice it so say that before such exciting possibilities can even be entertained, we first have to determine whether the link even exists. This is the purpose of the current studies.

2.4. The current studies

Here we employ the causal capture paradigm to test whether a visual illusion of causality in a moral context can systematically change people's moral judgments.

We then test whether these effects are tuned to parametric features of the stimuli that are known to affect, and extinguish, the illusion of causality — including the exact millisecond duration and synchrony of subtle contextual information. Our prediction is that these manipulations should affect moral judgments, suggesting that these judgments, although abstract, reflect a readout of perceptual inferences of causality that have already been made by the visual system.

Finally, to help rule out the alternative possibility that these effects can be explained away by patterns of reasoning that occur beyond the visual system, we also conceptually replicate this effect using a subtler version of the illusion that addresses potential alternative accounts of these findings.

3. Experiment 1: illusory moral wrongs

Observers saw one of four visual events (presented between-subjects), all of which involved an interaction between a red and blue car, immediately followed by the blue car apparently driving into a man and knocking him over (see Fig. 1a–d). In the *launching* condition the red car stopped adjacent to the blue car just before the blue car moved (Fig. 1a). These kinds of temporally contingent events are overwhelmingly seen as causal, whereby the first object causes the movement of the second (Michotte, 1946, 1963). In the other three conditions the red car overlapped completely with the blue car before the blue car moved. This overlap event tends to be seen ambiguously as either causal (red car caused blue car to move) or non-causal (the red car spontaneously transformed into the blue car, or the blue car just happened to move at the same instant that the red car moved in front of it) (Michotte, 1946, 1963).

Of these three overlap conditions, two were accompanied by a task-irrelevant contextual event in which two small circles interacted in the bottom periphery. These events were task-irrelevant since they were not required for the moral evaluation that observers were later asked to make about the driver of the red car. In the *launching context* condition (Fig. 1b), the first circle stopped adjacent to the second circle before the second circle moved. Such causal-looking contextual events are known to elicit higher causal perception for a main, task-relevant event, even when the contextual events are completely task irrelevant (Scholl & Nakayama, 2002). In the *spatiotemporally offset context* condition both circles were spatially offset to the bottom *left* of the screen (rather than immediately below the cars), and there was also a delay between when the first object stopped next to the second object and when the second object moved; Fig. 1c). This last condition has little influence on causal perception, since it is not spatiotemporally aligned with the main event (Scholl & Nakayama, 2002), and thus we expected it to have no systematic affect on blame judgments.

3.1. Observers and testing environment

Observers for all experiments were recruited from the United States via Amazon's Mechanical Turk (for discussion of this pool's reliability see Crump, McDonnell, & Gureckis, 2013). The link to the study specified that observers should have normal or corrected-to-normal visual acuity, and an automatized screening procedure prevented observers from taking part in more than one experiment. After agreeing to participate, all participants were redirected to a web server where platform-independent stimulus presentation and data collection were completed by custom software run in observers' web browsers, written using a combination of html, CSS, and javascript. All observers saw only a single condition in a between-subjects design, and the screening procedure prevented participants from taking part repeatedly in more than one experiment. For all experiments, we planned to run exactly half the number of trials per condition as Scholl and Nakayama (2002), 140 trials, since (1) our a priori experience of the illusions convinced us that this would provide sufficient power to find the effects, and (2) we wanted to save on trials, since, unlike previous studies, we planned to run each trial type completely between subjects in order to prevent across-trial effects. Thus, given

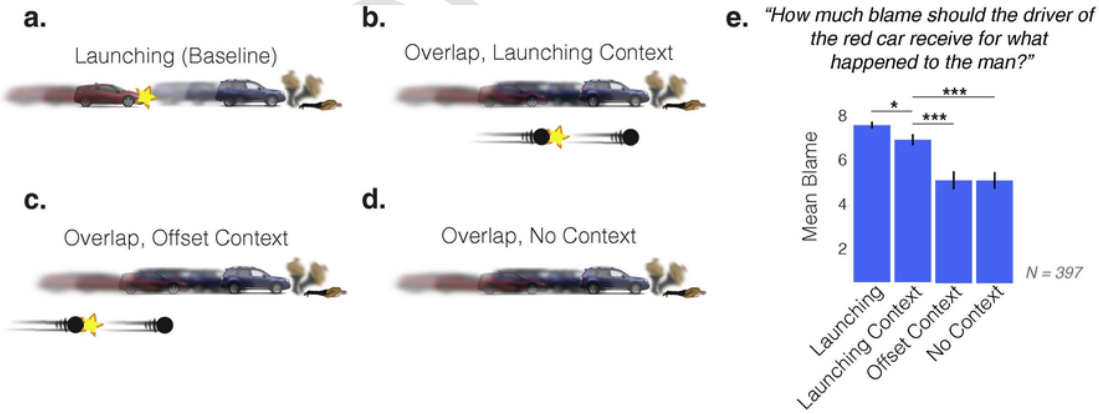

Fig. 1. Stimuli (a–d) from Experiment 1: Causal launch in main event (a), and overlap in main event with (b) launching in two contextual circles, (c) spatiotemporally offset launching in contextual circles, or (d) no contextual circles. (e) Depicts mean blame judgments for the launching condition and three overlap conditions. Error bars depict standard error of the mean. The illustrations show yellow stars to indicate perceived collisions, though yellow stars were not shown in the actual displays. The dynamics are simplified for illustration; see the full demos for exact dynamics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

PREPRINT

that every observer saw only one trial, we ran 140 observers per condition.

Because the display loaded within observers' own web browsers, viewing distance and screen resolutions could vary widely, and so we report dimensions of the stimuli using pixel (px) values and positions of the stimuli as percentage values from the left and top borders (X%, Y%) of a gray (red [R]: 221, green [G]: 221, blue [B]: 221) task window (800 px × 400 px), within which all displays took place.

To ensure precise timing of stimuli, we animated the stimuli using the velocity.js Javascript package (http://velocityjs.org/), which is optimized for this purpose. We also timed the duration of the full display in each observer's browser, and excluded observers whose display lasted either 25 ms longer or shorter than the full, intended duration for that condition. We conservatively chose 25 ms, since it is half as long as the shortest duration that we manipulate in any experiment (50 ms, in Experiment 2a), such that any remaining variability in stimulus duration cannot explain differences between conditions.

Complete code, analyses, and demos for all experiments can be found at the following link: https://github.com/juliandefreitas/moral_capture.

Participants. 562 naïve observers (140 per condition) participated for a small payment, and 72 were excluded for incorrectly answering a comprehension question at the end of the experiment (see exact wording for all exclusion questions under 'design', below), or for failing the stimulus timing check mentioned above. This yielded a final sample of 490 observers (M_{age} = 36, 49% female). We also excluded participants who failed to see the critical event of the blue car hitting the man (see wording below). As 81% of observers said that they saw the blue car hit the man, subsequent analyses were limited to this subset (N = 397). Previous research found that excluding participants who failed comprehension checks can reduce noise in responses due to inattention in online studies (Thomas & Clifford, 2017), and our exclusion percentages were within the range of published exclusions for data collected on Amazon's mechanical Turk (e.g., De Freitas, DeScioli, Nemirow, Massenkoff, & Pinker, 2017; Thomas, De Freitas, DeScioli, & Pinker, 2016).

Stimuli. In order to study these effects in a controlled manner, we intended for the stimuli to have the same spatiotemporal characteristics as the original causal capture stimuli (Scholl & Nakayama, 2002), while also being morally relevant.

In order for the events to be morally relevant, we showed people images, found using Google image search, of a red car (130px × 73.13px; 1%, 26%), blue car (130px × 52px; 32%, 30%), and a man (32px × 93.14; 77%, 22%). People instantly recognize images and understand the moral implications of interactions between depicted agents, e.g., a car hitting a man over. Given the spatiotemporal parameters of the stimuli, these interactions do not look 100% natural, although only 1 participant out of 397 mentioned any complaints about the quality of the stimuli (see the coded free-form descriptions in our online data repository).[3] Furthermore, we expected that this decrement in realism would not detract from people's understanding of the core morally relevant details of the display: there are agents, an interaction occurs, an agent is harmed. Another option would have been to use fully naturalistic videos, although it is more challenging to isolate the causal illusion for study under these conditions.

[3] The first author and one hypothesis-blind coder coded the sentences for any comments suggesting that observers thought the displays were not an accurate reflection of reality; both coders were condition-blind. Inter-coder reliability for the full set of sentences was perfect (Cohen's kappa value of 1.00).

Yet another option would have been to ask people to make moral judgments about simple shapes (e.g., see Iliev et al., 2012; Nagel & Waldmann, 2012), since even simple shapes can be used elicit impression of animacy in certain displays (Heider & Simmel, 1944), and so should be seen as morally relevant. We chose not to use such stimuli because we wanted our manipulated, task-irrelevant contextual events — which were comprised of two simple circles — to seem unrelated to the main, task-relevant event —which were comprised of images of cars. If the main event was made up of circles too, then the task-irrelevant circles would have drawn attention, since feature-based attention spreads to similar-looking items (Rossi & Paradiso, 1995). Furthermore, we expected that observers would be more interested in attending to the task-relevant cars than to the task-irrelevant circles, making them less likely to notice what happened with the circles.

Unlike in the original causal capture stimuli, the overlapping objects we used were not completely identical, since we thought it would be more realistic to show an interaction between two different cars. We expected that this wouldn't be a problem for eliciting the illusion, however, since spatiotemporal features trump surface features in most temporal illusions (Flombaum et al., 2009). That said, the stimuli did look somewhat less causally ambiguous upon overlap, so we increased spatiotemporal uncertainty by presenting the images at a fifth of their true opacity. Critically, our main planned experimental contrasts look at how moral judgments change when we alter features of the task-*irrelevant* events, not this main overlap event, which we hold constant across conditions. Therefore, any differences between conditions cannot be explained by features of the task-relevant event. Since these illusions are inherently phenomenological, we used our own subjective experience of the demos to decide on final parameters such as stimulus size and speed.

Design. The red car, blue car and the man were initially present at rest within the task window. Above this task window observers read the instructions, "Please press play and pay VERY CLOSE ATTENTION. You will only be able to play the video once". These instructions were the same for all experiments, and were deliberately minimal in order to avoid biasing observers to pay attention to specific objects or events. Upon clicking 'play', the instructions disappeared, and 2 s later observers saw one of four displays (presented between-subjects).

(A) In the *launching* condition, the red car translated 140px rightwards at 416.67 px/s, stopping adjacent to the blue car. As soon as the red car stopped, the blue car translated 250px rightwards at the same speed, stopping adjacent to the man. As soon as the blue car stopped, the man fell down; this was conveyed by simultaneously (over 115 ms) rotating him 90° clockwise, and translating him 30px downwards and 90px rightwards. The three overlap conditions were similar to this launching event, except that the red car translated 250px (the same distance as the blue car), overlapping completely with the blue car. (B) In the *overlap with no context condition*, this is all that happened. (C) *In the overlap with launching context* condition, observers simultaneously saw two black circles (50px × 48.92px) — one aligned with the leftmost edge of the red car (1%, 51%) and the other aligned closer to the center of the blue car (37%, 51%) — translate in synchrony with the cars. The reason that the second circle was aligned closer to the center of the right car was so that the first circle would stop adjacent to it, thereby depicting a launching interaction. (D) In the *overlap with spatiotemporally offset context* condition, the circles were still located beneath the cars, but were both shifted to the left of the screen and positioned closer to each other (10%, 51% and 20%, 51%), so that each of the circles translated only 45 pixels. Furthermore, the first circle only began

UNDER PEER REVIEW

translating when the blue car began translating, and the second circle only began translating when the patient began to fall. The net effect of these parameters was to create a spatially offset launching context that had a delay before the second circle moved.

After 1 s, all the objects disappeared and were replaced by a series of questions, each presented on a separate page: a scaled rating of blame for the first driver (*How much blame should the driver of the red car receive for what happened to the man? 1 = None at all, 9 = Severe blame)*; a forced-choice between which driver was more blameworthy (*Who deserves more blame for what happened to the man: the driver of the red car, or the driver of the blue car? A. The driver of the red car; B. The driver of the blue car);* a question regarding which car observers saw hit the man (*Which car directly hit the man: the red car, or the blue car?*); and a request to describe, in a provided textbox, why they made their specific moral judgment (*Please describe why you gave the specific blame judgment that you did.*). The questions were accompanied by a miniature (300px × 152.56px) screenshot of the start of the display, positioned in the top left corner of the task window, so that observers knew to whom in the display the questions referred.

Observers then answered a comprehension question (Which, if any, of the following events occurred in the scenario that you read? A. Two men jumped up and down, B. A car broke into tiny pieces, C. A man punched another man, D. All of the above, E. None of the above), said whether they had participated in a similar experiment previously (Have you ever completed a HIT containing a similar scenario, perhaps involving the same sorts of questions? A = Yes, B = No), and completed two demographic questions (gender and age).

4. Results & discussion

Descriptions. We were curious whether people spontaneously described the interactions in causal terms when asked to explain their moral judgments. To this end, the first author and one hypothesis-blind coder coded the sentences. Both coders were condition-blind, and coded the sentences as causal whenever the language suggested that the red car caused the blue car to move. After checking that inter-coder reliability was high on the first 10 sentences, they coded the rest of the sentences. Inter-coder reliability for the full set of sentences was high: coders selected the same category for 375 of the 397 sentences, giving a Cohen's kappa value of .87. Looking at the subset of sentences on which coders agreed, 70% of sentences were rated causal. Given the different conditions employed, this is a reasonable proportion of causal sentences, and suggests that the causal illusions worked. Furthermore, moral judgments were significantly higher for sentences coded causal ($M = 8.02$. $SD = 1.25$) than sentences not coded causal ($M = 2.72$. $SD = 2.83$; $t(373) = 25.17$, $p < .0001$, $d = 2.85$).

Likert Scale Question. We found that moral judgments were influenced by the irrelevant contextual events (Fig. 1e; $F(3, 393) = 19.60$, $p < .0001$, $\eta p^2 = 0.13$). The overall pattern was predicted by the known effects of these contextual events on the illusion of causality (Scholl & Nakayama, 2002). Moral blame judgments for causal launch events (Fig. 1a; $M = 7.65$, $SD = 1.73$) were higher than for ambiguous overlap events accompanied by a contextual causal launch event (Fig. 1b; $M = 7.01$, $SD = 2.57$; $t(230) = 2.26$, $p = .025$, $d = 0.30$). It is impressive, however, that they were still fairly similar in magnitude. Furthermore, these overlap events with causal contexts elicited higher moral blame than overlap events without a context (Fig. 1d; $M = 5.21$, $SD = 3.59$; $t(197) = 4.11$, $p = .0001$, $d = 0.59$), or overlap events with contexts that were offset in location and time (Fig. 1c; $M = 5.22$. $SD = 3.47$; $t(180) = 4.00$, $p < .0001$, $d = 0.60$). The fact that the effect did not occur for the spatiotemporally offset contexts confirms that the effect is not driven by the presence of just any contextual event, and is consistent with existing knowledge that such offset events do not enhance the causal illusion (Scholl & Nakayama, 2002).

Forced-Choice Question. This pattern of results replicated when the same observers then made a forced-choice of which of the two drivers was most morally blameworthy for hitting the man. Mirroring the pattern of scaled blame ratings, 85% of observers thought the driver of the red car was more blameworthy in the launching condition, which in this case did not even differ significantly from the percentage (78%) in the overlap event with a launching context, $\chi^2(1, N = 232) = 1.82$, $p = .177$, = .09, which in turn was higher than the percentage for the identical overlap events without a context, (51%), $\chi^2(1, N = 199) = 14.99$, $p < .001$, = .27, or the spatiotemporally offset context (51%), $\chi^2(1, N = 182) = 12.71$, $p < .001$, = .26.

Therefore, across conditions we find that the visual illusion of causality determined moral judgments, consistent with the hypothesis that this obligatory high-level visual inference provided the abstract information required to make the moral judgments.

5. Experiment 2: extinguishing the effect with millisecond manipulations of task-irrelevant factors

If, as these results suggest, this manipulation of moral judgments is due to the illusion of causality in these moral events, then moral judgments should also be more subtly affected by cues that are known to alter the strength of the causal illusion (Scholl & Nakayama, 2002), such as millisecond manipulations of the duration of the contextual events (Experiment 2a, Fig. 2) and their timing relative to the main event (Experiment 2b, Fig. 3).

5.1. Experiment 2a: effects of context duration

Participants. 562 naïve observers (140 per condition × 4 conditions) participated for a small payment, and 52 were excluded for incorrectly answering a comprehension question at the end of the experiment, or for failing the stimulus timing check. This yielded a final sample of 510 observers ($M_{age} = 39$, 51% female). As 74% of observers said that they saw the blue car hit the man, subsequent analyses were limited to this subset (N = 379). The mean scaled blame ratings of the first driver are shown in Fig. 2.

Design. The design was identical to the overlap main event with launching context condition of Experiment 1, except for the following factors. In addition to this condition, some observers saw one of three other conditions (presented between-subjects), in which the context event was shown for only 500 ms, 100 ms, or 50 ms (rather than 1200 ms, as in the normal launching context condition). All of these durations were temporally centered around the launch interaction, which was always presented.

5.2. Results & discussion

Likert Scale Question. We found that the effect of the launching context decreased when the context was presented for shorter durations (Fig. 2; $F(3, 375) = 12.02$, $p < .0001$, $\eta p^2 = 0.09$). Planned comparisons confirmed that the normal full duration launching context differed significantly from only the two shortest duration conditions: 1200 ms *vs*. 500 ms, $t(203) = 0.89$, $p = .376$, $d = 0.12$; 1200 ms *vs*. 100 ms, $t(198) = 3.18$, $p = .002$, $d = 0.45$; 1200 ms *vs*. 50 ms, $t(188) = 5.45$, $p < .0001$, $d = 0.80$.

UNCORRECTED PROOF

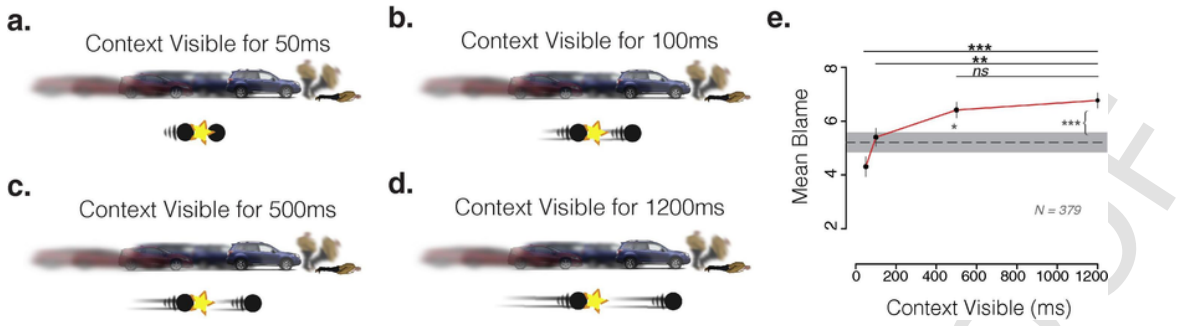

Fig. 2. Stimuli from Experiment 2a, depicting overlap events with contextual launching events that lasted for different durations (a–d). (e) Depicts comparison of these duration conditions to the no context baseline (Fig. 1d). Dotted line depicts this no context baseline. Error bars depict standard error of the mean. The illustrations show yellow stars to indicate perceived collisions, though yellow stars were not shown in the actual displays. The dynamics are simplified for illustration; see the full demos for the exact dynamics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

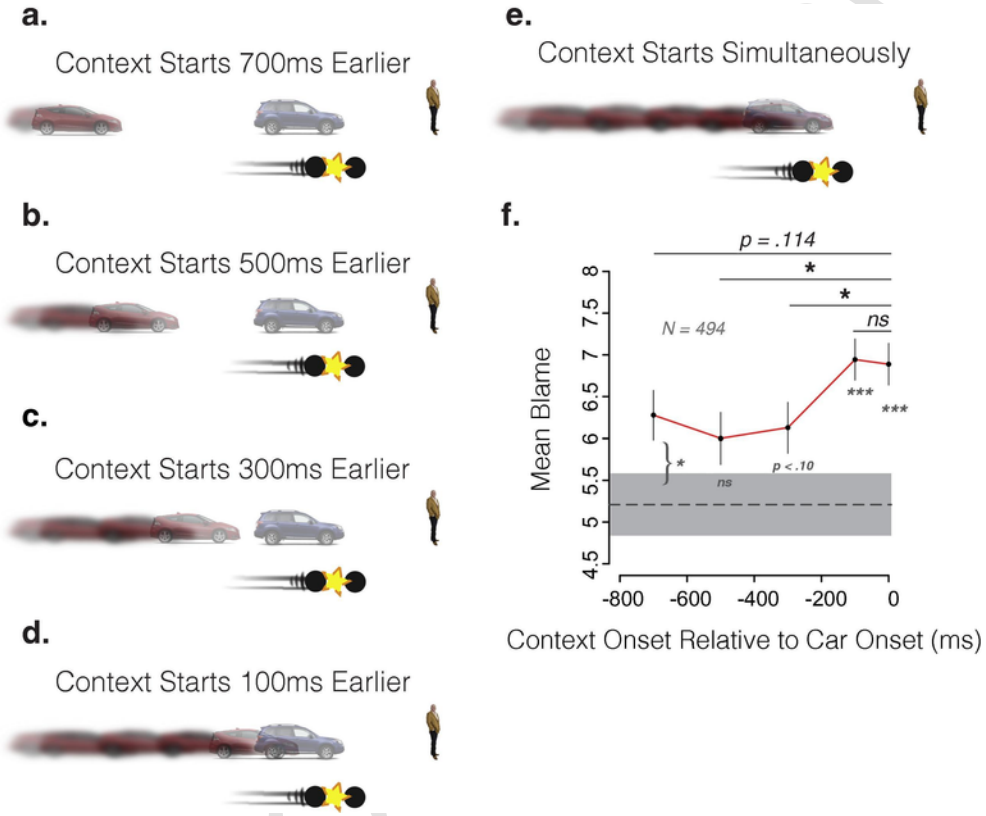

Fig. 3. Stimuli and results from Experiment 2b. Examples show the position of the red car at the point of collision of the contextual objects (a–e). (f) Depicts comparison of these conditions to the no context baseline (Fig. 1d). Dotted line depicts this no context baseline. Error bars depict standard error of the mean. The illustrations show yellow stars to indicate perceived collisions, though yellow stars were not shown in the actual displays. The dynamics are simplified for illustration; see the full demos for exact dynamics.

Further planned comparisons between each duration and the context absent baseline of Experiment 1 (Fig. 1d) revealed that only the longer durations induced significantly higher moral blame judgments than this baseline: *absent vs. 1200* ms, $t(197) = 3.39$, $p < .001$, $d = 0.48$; *absent vs. 500* ms, $t(186) = 2.56$, $p = .011$, $d = 0.37$; *absent vs. 100* ms, $t(181) = 0.39$, $p = .700$, $d = 0.06$; *absent vs. 50* ms, $t(171) = -1.71$, $p = .088$, $d = 0.26$.

Forced-Choice Question. We found a similar pattern for the forced-choice between which driver was most morally blameworthy. The full duration contextual event (78%) only differed from the two shortest durations: 1200 ms *vs.* 500 ms (72%), $\chi^2(1, N = 205) = 0.59$, $p = .444$, = .05; 1200 ms *vs.* 100 ms (57%), $\chi^2(1, N = 200) = 9.36$, $p = .002$, = .22; 1200 ms *vs.* 50 ms (44%), $\chi^2(1, N = 190) = 21.55$, $p < .0001$, = .34.

Furthermore, only the longer durations induced significantly higher moral blame judgments than the context absent baseline (51%) from Experiment 1: *absent vs. 1200* ms (78%), $\chi^2(1, N = 199) = 14.99$, $p < .001$, = .27; *absent vs. 500* ms (72%), $\chi^2(1, N = 188) = 8.39$, $p = .004$, = .21; *absent vs. 100* ms (57%), $\chi^2(1, N = 183) = 0.44$, $p = .508$, = .05; *absent vs. 50* ms (44%), $\chi^2(1, N = 173) = 0.52$, $p = .470$, = .05.

5.3. Experiment 2b: effects of temporal asynchrony

Participants. 699 naïve observers (140 per condition × 5 conditions) participated for a small payment, and 64 were excluded for incorrectly answering a comprehension question at the end of the experiment, or for failing the stimulus timing check. This yielded a final sample of 635 observers (M_{age} = 36, 50% female). As 78% of observers said that they saw the blue car hit the man, subsequent analyses were limited to this subset (N = 494).

Design. The design was identical to the overlap with launching context condition of Experiment 1, except for the following factors. Aside from this condition, some observers saw one of four other conditions (presented between-subjects), in which the context event began 100 ms, 300 ms, 500 ms, or 700 ms *before* the red car began translating.

5.4. Results & discussion

Likert Scale Question. We found that the effect of the launching context decreased when the context started several milliseconds before the main event (Fig. 3, $F(4, 489) = 2.55$, $p = .039$, $\eta p^2 = 0.02$). Planned comparisons confirmed that the temporally synchronous launching context tended to differ significantly from only the most asynchronous conditions: *0* ms *vs. −100* ms, $t(220) = 0.16$, $p = .875$, $d = 0.02$; *0* ms *vs. −300* ms, $t(208) = 1.96$, $p = .051$, $d = 0.27$; *0* ms *vs. −500* ms, $t(202) = 2.26$, $p = .025$, $d = 0.32$; *0* ms *vs. −700* ms, $t(204) = 1.59$, $p = .114$, $d = 0.22$.

Further planned comparisons between each condition and the context absent baseline of Experiment 1 revealed that only the most synchronous conditions tended to differ significantly from the context absent baseline: *absent vs. 0* ms, $t(205) = 3.85$, $p < .001$, $d = 0.54$; *absent vs. −100* ms, $t(195) = 3.96$, $p < .001$, $d = 0.57$; *absent vs. −300* ms, $t(183) = 1.91$, $p = .058$, $d = 0.28$; *absent vs. −500* ms, $t(177) = 1.62$, $p = .107$, $d = 0.24$; *absent vs. −700* ms, $t(179) = 2.23$, $p = .027$, $d = 0.33$.

Forced-Choice Question. Comparing each asynchronous condition to the synchronous condition (78%) revealed a similar trend of forced-choices between who was most blameworthy: *0* ms *vs. −100* ms (77%), $\chi^2(1, N = 222) = 0.00$, $p = 1.00$, = .00; *0* ms *vs. −300* ms (67%), $\chi^2(1, N = 210) = 2.42$, $p = .120$, = .11; *0* ms *vs. −500* ms (66%), $\chi^2(1, N = 204) = 2.86$, $p = .091$, = .12; *0* ms *vs. −700* ms (68%), $\chi^2(1, N = 206) = 2.02$, $p = .156$, = .10.

We also found a consistent pattern when comparing the forced-choices for each of these conditions to the context absent baseline (51%) of Experiment 1: *absent vs. 0* ms (78%), $\chi^2(1, N = 207) = 15.36$, $p < .001$, = .27; *absent vs. −100* ms (77%), $\chi^2(1, N = 197) = 14.31$, $p < .001$, = .27; *absent vs. −300* ms (67%), $\chi^2(1, N = 185) = 4.52$, $p = .033$, = .16; *absent vs. −500* ms (66%), $\chi^2(1, N = 179) = 3.73$, $p = .054$, = .14; *absent vs. −700* ms (68%), $\chi^2(1, N = 181) = 4.87$, $p = .027$, = .16.

6. Experiment 3: conceptual replication with a subtler manipulation

So far, we have operated under the assumption that observers were unaware of the manipulations in Experiment 2, for a couple of reasons: these manipulations consisted of subtle, task-irrelevant events; each observer only saw the display once; and the different manipulations were tested between-subjects. Furthermore, after each participant gave their blame ratings in Experiments 1–2, we asked them to explain why they made their particular moral blame judgments (*Please describe why you gave the specific blame judgment that you did.*). We then checked whether any of these sentences referred to "ball", "black", "circle" or "context", and also checked each sentence manually for any mention of the context event in any form whatsoever. Only a *single* observer (N = 1270) mentioned the contextual event at all, and it was clear that this observer did not reason differently because of this information (*The blue and red car were almost "racing" and the blue car went ahead and knocked the man down…I'm not sure what the two black balls mean though.*; a complete list of descriptions is included in our online data repository).

We consider it highly unlikely, therefore, that observers knew that the task-irrelevant events were affecting their judgments about the task-relevant events. To do this, they would have to (i) explicitly notice the contextual event was causal, and then (ii) reason that, for some reason, the main event must be causal too, even though the task-irrelevant events are divorced in both appearance and space from the main events.

Nevertheless, in order to remove the logical possibility of such a concern altogether, we next used a contextual object that was not itself causal-looking, so that a theoretical observer who noticed this contextual event could not make the similarity inference outlined above. In fact, this time the contextual event was not even a separate, co-occurring event, but was embedded within the scene itself.

Another theoretical possibility is that the contextual stimuli drew attention away from the task-relevant stimuli, somehow explaining the effects. For this to be true, this would have to happen to different extents for the different conditions. Although this seems implausible, Experiment 3 also addresses this concern to some extent, since the contextual stimuli are no longer spatially disconnected from the main events, but are part of the main events.

6.1. Experiment 3a: grouping

To achieve these ends, we leveraged another finding about causal illusions: ambiguously causal events look causal when one of the objects in the interaction is accompanied by a single, moving contextual object, e.g., one moving circle (Choi & Scholl, 2004). We employed this manipulation within our morally relevant displays, while making the contextual event look like part of the scene itself — a pile of logs resting on top of the second car (Fig. 4).

Again, notice that all that literally happens in such a display is that the first car completely overlaps with a second car, which in this case also happens to have logs resting on top of it. Yet we knew that the mere presence of these logs moving together with the second car would induce the illusion that the red car actually hit into the blue car, and so we predicted that moral judgments would be manipulated by this illusion.

We also knew from previous work (Choi & Scholl, 2004) that the number of grouped objects (e.g., one, two, three) does not alter the strength of the illusion, yet we could imagine a number of reasons why people might reason about this factor differently[4]. Therefore, we also varied the number of objects (all between-subjects), in order to further test whether even this subtler property of the illusion is reflected into moral judgments.

Participants. 837 naïve observers (140 per condition × 6 conditions) participated for a small payment, and 115 were excluded for in-

[4] For instance, different objects visibly resting on top of a car might make the negative outcome seem more or less unavoidable (conditioned on the first car hitting the second car), or lead to different inferences about recklessness (i.e., since the driver of the red car in the many logs condition should have clearly seen the other car, crashing into the blue car indicates reckless behavior that makes that driver even more morally blameworthy for the negative outcome). We expected that if such effects exist, they might be squashed by the visual illusion.

UNCORRECTED PROOF

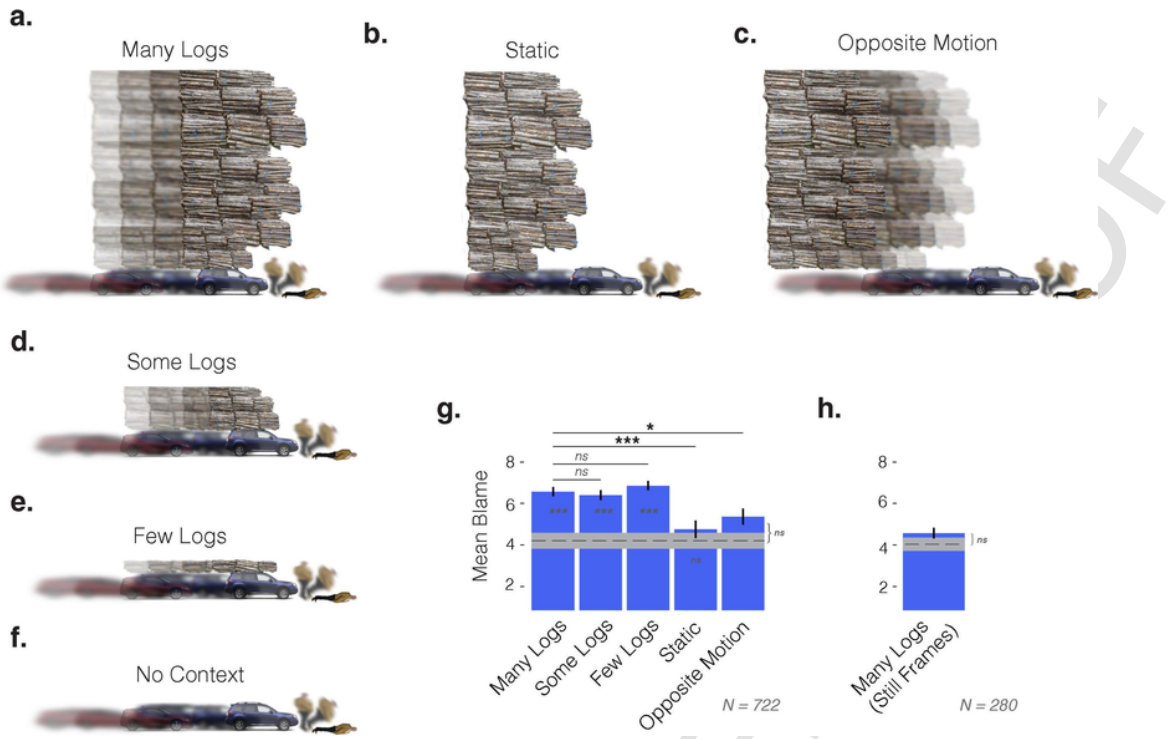

Fig. 4. Stimuli from Experiments 3a-b (a–f). (g) Depicts comparison of these conditions to the no context baseline (Fig. 1d), and (h) depicts this comparison for a still frame version of (a). Dotted line depicts the no context baseline. Error bars depict standard error of the mean. The dynamics are simplified for illustration; see the full demos for exact dynamics.

correctly answering a comprehension question at the end of the experiment, or for failing the stimulus timing check. Since the current paradigm was different from that used in Experiments 1–2, we did not exclude participants for saying they had participated in a similar experiment before. This yielded a final sample of 722 observers (M_{age} = 36, 41% female). As 72% of observers said that they saw the blue car hit the man, subsequent analyses were limited to this subset (N = 521).

Stimuli. These stimuli were meant to employ the same spatiotemporal features as the original causal grouping studies (Choi & Scholl, 2004), while still being morally relevant. Given that our interactions involved real images, we also tried to ensure that the logs wouldn't draw attention as a special manipulation of the experiment. Therefore, whereas the original causal grouping studies employed grouping objects that all looked the same as the objects in the main event, our grouping objects (logs) were meant to look like they could meaningfully fit into the scene, i.e., they were strapped to the top of a car. As such, our grouping objects were also made to look like they were making contact with one of the objects in the main event, unlike the original causal grouping studies, where the grouping objects were physically disconnected from the main event.

We were curious about what kinds of things people would say about the logs, so we asked a hypothesis and condition-blind coder to code the descriptions people provided when asked to explain their moral judgments. Out of 521 sentences, 50 mentioned the logs, and only 2 of these sentences suggested that people were not sure about exactly what they saw or thought that the display looked surprising/unnatural. Almost all observers who mentioned the logs did so only to say that they thought it was an unsafe load for a car to carry, e.g., "The red car started the whole thing. The blue car had a pretty unsafe load on top of their car, but they didn't cause an accident." Full descriptions, together with the codes, are provided in our online data repository.

Design. The same objects from Experiments 1–2 were now positioned near the bottom of the task window: red car (1%, 76%), blue car (32%, 80%), man (77%, 72%). In the context absent condition we simply played the display at this new location. In other conditions, the blue car appeared to have a pile of wooden logs (found using Google image search) tied to its roof. In the *few objects* condition, a clipped version of this log image (130px × 86.59px) was presented above the blue car (28%, 62%). In the *some objects* condition, the entire image was presented at that same location; and in the *many objects* condition we again presented this full image, in addition to three larger copies of the same image that appeared to be stacked on top (1: 190px × 126.56px; 28%, 44%; 2: 200px × 133.22px; 28%, 24%; 3: 210px × 139.89px; 28%, 5%). In all three conditions, the contextual objects translated in complete synchrony with the blue car.

The *static* and *opposite motion* conditions were similar to the *many objects* condition, except that the contextual objects either remained stationary or translated the same amount *leftwards* respectively.

6.2. Results & discussion

Likert Scale Question. We first ran a linear regression with just context motion type (same direction, static, opposite, or absent) as a factor. Relative to the context absent baseline, the same direction and opposite conditions elicited higher moral blame judgments ($t = 7.07$, $p < .0001$ and $t = 2.34$, $p = .020$ respectively), but not so the static condition ($t = 1.05$, $p = .294$). We conducted simple comparisons on the least-squares means using the lsmeans package in R (Length, 2016). As predicted, we found that *same direction* ($M = 6.60$, $SE = 0.16$) differed from each of the other motion types: *same direction vs. absent* ($M = 4.17$, $SE = 0.31$), $t(517) = 7.07$, $p < .0001$, $d = 0.44$); *same direction vs. opposite* ($M = 5.34$, $SE = 0.40$), $t(517) = 2.96$, $p = .017$, $d = 0.18$); and *same direction vs. static* ($M = 4.73$, $SE = 0.44$), $t(517) =$

4.05, $p < .001$, $d = 0.25$). This suggests that the effects were driven by the causal illusion, but not by the mere presence of the contextual object.

We also ran a similar model, but with the number of grouped contextual objects (*many, some, or few*) nested within the *same direction* motion type. As predicted, because the illusion is the same for different numbers of contextual objects (Choi & Scholl, 2004), this more complicated model did not outperform the model with just motion type as a factor, likelihood ratio test, $\chi^2(7, N = 521) = 2.86$, $p = .240$.

Finally, for the simpler model with just motion type as a factor we also compared least squares means of the context absent baseline to each of the opposite and static conditions. The absent baseline condition did not differ significantly from the static condition (*absent vs. static:* $t(517) = 1.05$, $p = .720$, $d = 0.07$), nor the opposite motion condition (*absent vs. opposite motion:* $t(517) = 2.34$, $p = .091$, $d = 0.14$).

Forced-Choice Question. As for forced-choice blame ratings, we first ran an analysis with just motion type as a factor, using logistic regression. Relative to the baseline context absent condition, the same direction and opposite conditions elicited higher moral blame judgments ($z = 5.87$, $p < .0001$ and $z = 2.23$, $p = .026$ respectively), but not so the static condition ($z = 1.31$, $p = .190$). We conducted simple comparisons on the least-squares means. We found that the same direction condition ($M = 0.81$, $SE = 0.12$) differed from both the absent and static conditions: *same direction vs. absent* ($M = 0.68$, $SE = 0.22$): $z = 5.87$, $p < .0001$, and *same direction vs. static* ($M = 0.18$, $SE = 0.30$): $z = 3.06$, $p = 0.012$; and differed marginally from the opposite condition, *same direction vs. opposite* ($M = 0.11$, $SE = 0.28$): $z = 2.33$, $p = .091$).

We also ran a model with both motion type and the number of grouped objects nested within the same direction motion type. A likelihood ratio test showed that this model did not outperform the model with just motion type as a factor, $\chi^2(6, N = 521) = 3.47 p = .176$.

Finally, for the simpler model with just motion type as a factor we also compared least squares means of the context absent baseline to each of the opposite and static conditions. In line with the scaled results and previous work on causal perception, the context absent baseline did not differ from the static condition (*absent vs. static:* $z = 1.31$, $p = 0.556$), nor the opposite motion condition (*absent vs. opposite:* $z = 2.23$, $p = 0.117$).

The percentages of people in each context condition who said that the driver of the red car was more morally blameworthy were as follows: context absent baseline (34%); same direction with many (71%), some (63%), and few grouped objects (74%); static (45%); and opposite motion (53%).

6.3. Experiment 3b: eliminating motion cues

Since the illusion in Experiment 3a requires smooth, continuous motion, this means that if we depict the same events using still frames then we can eliminate the illusion while preserving any information that one would typically need to make a moral judgment. Therefore, if people provide the same pattern of moral judgments for still frames, then this would suggest that Experiment 3a's results were driven by explicit reasoning rather than by a visual illusion. We predicted that we would not find this result.

Participants. 283 naïve observers (140 per condition × 2 conditions) participated for a small payment, and 3 were excluded for incorrectly answering a comprehension question at the end of the experiment. Since our stimuli were dramatically prolonged (in order to interfere with motion processing), we did not record exact millisecond stimulus timing nor exclude subjects based on exact stimulus timing. This yielded a final sample of 280 observers ($M_{age} = 35$, 45% female). As 94% of observers said that they saw the blue car hit the man, subsequent analyses were limited to this subset (N = 264).

Design. We presented five still frames taken from Experiment 3a's context absent condition and same direction condition (the one that used many grouped objects). The still frames were taken from when the red and blue cars were at the start of the display (red car: x = 1%; blue car: x = 32%), beginning their overlap (25%; 32%), completely overlapped (32%; 32%), ending their overlap (32%; 39%), and at the end of the display (32%, 63%). Although we did not actually show the moment of contact between the blue car and the man, we expected that it would be very obvious what happened, since the final frame depicts the man lying down next to the blue car. Indeed, when asked to explain why they made their particular moral judgment, only 1% of participants expressed any uncertainty about what happened to the man, and 63% of participants explicitly indicated that they thought a car hit a man (this is of course an underestimate, since not all participants needed to state this explicitly in order to explain their particular moral judgment; see the coded free-form descriptions in our online data repository). Furthermore, we also replicate all findings when we analyze only the 37% of participants who did not explicitly indicate that a car hit the man (see Appendix).

Each still frame was presented for a set duration of 2000 ms (except for the last frame, which lasted 3000 ms), interleaved with 1000 ms blank displays, in order to eliminate any apparent motion effects.

6.4. Results & discussion

Likert Scale Question. Unlike Experiment 3a's same direction condition, which used smooth motion, moral blame judgments for the still frames same direction condition ($M = 4.61$, $SD = 3.07$) were indistinguishable from the still frames context absent condition ($M = 4.09$, $SD = 3.06$), $t(262) = 1.38$, $p = .170$, $d = 0.17$), as well as from Experiment 3a's context absent condition that involved motion ($M = 4.17$, $SD = 3.31$), $t(224) = 1.01$, $p = .313$, $d = 0.14$ (Fig. 4h). Moreover, we observed a significant experiment (smooth motion *vs.* still frames) × condition (same direction *vs.* absent) interaction, $F(1, 462) = 10.52$, $p = .001$, $\eta p^2 = 0.02$), indicating that the effect of grouping context was greater for smooth motion displays than still frame displays. Thus, we again conclude that moral judgments were driven by the motion-based visual illusion, not by explicit reasoning.

Forced-Choice Question. We found a similar pattern for the forced-choice of which driver was most blameworthy. A logistic regression found a significant experiment (smooth motion *vs.* still frames) × condition (same direction *vs.* absent) interaction ($\beta = 1.26$, $SE = 0.40$, $p = .002$). The still frames same direction condition (36%) elicited statistically indistinguishable moral blame than the still frames absent baseline (30%), $\chi^2(1, N = 264) = 1.00$, $p = .317$, $= .06$, and smooth motion context absent condition from Experiment 3a (34%), $\chi^2(1, N = 226) = 0.08$, $p = .775$, $= .02$.

6.5. Experiment 3c: Generalizing to judgments of moral wrongness

Moral blame is a prototypical moral judgment, with many models of moral judgment dedicated to explaining it (Alicke, 2000; Alicke, Mandel, Hilton, Gerstenberg, & Lagnado, 2015; Malle et al., 2014; Shaver, 1985; Weiner, 1995). Yet in some circumstances it can be difficult to tease apart judgments of causality from judgments of blame (Knobe, 2009; Samland & Waldmann, 2016). This may be because judgments of blame primarily track the negativity of an outcome caused by the agent (Cushman, 2008). In contrast, some other moral judgments may primarily track other factors. For instance,

UNCORRECTED PROOF

judgments of moral wrongness and permissibility primarily track the negativity of the agent's intentions (Cushman, 2008). Thus, we also wanted to know whether the current effects would generalize to a different kind of moral judgment such as wrongness.

Participants. 281 naïve observers (140 per condition × 2 conditions) participated for a small payment, and 38 were excluded for incorrectly answering a comprehension question at the end of the experiment, or for failing the stimulus timing check. This yielded a final sample of 243 observers (M_{age} = 38, 53% female). As 67% of observers said that they saw the blue car hit the man, subsequent analyses were limited to this subset (N = 163).

Design. We used the same design as Experiments 3a, except we only ran the no context and grouping (with few logs) conditions. We also asked for judgments of wrongness rather than blame: "How wrong was the driver of the red car's behavior?" (1 = Not at all, 9 = Very much); "Whose behavior was more wrong: the driver of the red car, or the driver of the blue car?" (1 = The driver of the red car, 2 = The driver of the blue car).

6.6. Results & discussion

Likert Scale Question. As predicted, the grouping condition elicited higher moral wrongness judgments than the no context condition $t(1\,6\,1) = 4.91$, $p < .0001$, $d = 0.78$).

Forced Choice Question. We found a similar trend of forced-choices when comparing the two conditions: *no context* (28%) *vs*. *grouping* (51%), $\chi^2(1, N = 163) = 7.44$, $p = .006$, = .21.

7. General discussion

How does causal perception interface with the rest of cognition? Here we isolated the effect of causal perception on moral judgment, finding that task-irrelevant manipulations known to alter causal perception also change people's moral judgments (Experiment 1). Furthermore, observers' explanations for their moral judgments suggest that they were unaware that these manipulations induced illusory causal impressions of the main events, suggesting that their moral judgments relied on mandatory, implicit causal perception. We were also able to extinguish this effect by employing subtle, millisecond changes to the task-irrelevant event's duration and temporal asynchrony with the main event (Experiment 2), suggesting that the effects were highly stimulus driven in the way that visual illusions are. In order to address potential alternative accounts that observers were somehow reasoning about the task-irrelevant events or were distracted by them, we then conceptually replicated these effects using a subtle grouping manipulation that was embedded within the main event itself (Experiment 3a). We found that the effect of this manipulation on moral judgments obeyed various known properties of this causal illusion. We then showed that we could also eliminate this effect by interrupting the spatiotemporally continuous motion paths needed for perceptual grouping (Experiment 3b). Finally, we generalized the effect to another kind of moral judgment, moral wrongness (Experiment 3c).

7.1. An effect of causal perception on moral Judgment?

Our best interpretation of this collection of results is that moral judgments were driven by *perceived* causality. This is not to say that observers did not also engage in deliberative reasoning about the stimuli. But to the extent that their moral judgments were predicted by factors inherent to the causal illusion, we conclude that causal perception per se may explain the main shifts found in people's moral judgments. A number of features of these results lead us to this conclusion:

1. We used stimuli very similar to the previous stimuli that have been shown to elicit a number of other automatic perceptual effects that people are unaware of.
2. The manipulations we used were task-irrelevant: they were spatially separated and physically distinct (two circles, rather than pictures of real-world objects), and observers were never asked anything about them.
3. Each observer saw only a single trial of the display in a fully between subjects design, and they were only asked a single question about the display *after* the display vanished. This design gave subjects less of a chance to potentially figure out how the illusion was generated.
4. Only a single observer out of >1000 in Experiments 1–2 mentioned the task-irrelevant event in any form whatsoever, suggesting that many observers didn't even think about the circles. This does not mean that they did not see these task-irrelevant events at all. But the fact that the task-irrelevant event was almost never mentioned suggests that people were not trying to figure out what they were 'for'. In Experiment 1, we also coded the descriptions that observers provided when asked to explain their moral judgments. 70% of observers used language consistent with them seeing one car cause the other to move, and those who used causal language tended to provide more severe moral blame judgments than those who did not, suggesting that moral judgments were driven by the illusion itself — rather than by their reasoning about the task-irrelevant events.
5. Like many visual illusions, we found that the effects of these illusions on moral judgment could be destroyed by employing subtle millisecond adjustments to parameters of the task-irrelevant events, such as their duration and synchrony with the main event. We think it is highly unlikely — especially given the single trial, between-subject design employed — that subjects would have said that they gave a particular moral judgment because the task-irrelevant event was out of sync with the main event. Rather, as suggested by their descriptions, their judgments fell out of their (sometimes illusory) perception of the main event.
6. In Experiment 3 we used a manipulation that was even subtler, because the task-irrelevant event was embedded within the display itself. It seems particularly unlikely that someone would give a higher moral judgment in the grouping condition because logs were resting on top of the second car. Instead, it's more likely that the logs influenced perceived causality, and that perceived causality influenced moral judgments.

7.2. How to build a moral human (or Machine)

Previous work has either not studied how causal perception interfaces with the rest of cognition (in the case of the causal perception literature), or has hardly considered how people make moral judgments in the visual domain (in the case of the moral psychology literature). In the small subset of studies where experimenters have studied moral judgments about visual events (Caruso et al., 2016; Iliev et al., 2012; Nagel & Waldmann, 2012), the experiments have not employed the kinds of implicit measures that isolate underlying mechanisms of causal perception — nor have the authors claimed to do so. Therefore, to our knowledge the current studies are the first to demonstrate that causal perception has downstream effects on moral judgment, a domain of cognition that is in many respects the very opposite of automatic perception.

UNCORRECTED PROOF

This link makes much sense when one considers the fact that certain high-level inferences, such as of causality, provide exactly the sort of information that certain domains of abstract cognition, such as moral judgment, tend to rely on. Yet, the current studies provide the first direct demonstration of this link between causal perception and moral judgment. As such, we can now begin to seriously entertain the implications of this connection.

At the most general level, a clear implication of these findings is that causal inferences trafficked in by the visual system contribute to moral judgments. Indeed, given that causal perception applies to any physical domain, not just social interactions, causal perception may be the main way that our moral systems get causal information about physical interactions between agents in the visual world. This is not to say that we don't also reason about the visual realm; but in many cases this may not even be necessary, if the visual system has already made a fast and accurate inference. Thus, the reliance of moral judgment on causal perception appears to present a good example of how the mind efficiently uses and reuses representations from multiple systems — in this case, causal perception is sufficiently abstract not only to tell us about the causal structure of physical interactions in the visual world, but also to inform our abstract moral judgments about causal social interactions in the visual world.

The suggestion that causal perception informs moral judgments is not just an abstract metaphysical point, but also has direct bearing on how we might reverse engineer moral judgments about visual events, for instance, when taking an artificial intelligence approach. Some prominent papers have suggested that biologically inspired models of the visual system, such as hierarchical convolutional neural networks (HCNNs), are effectively too limited to solve aspects of social cognition like moral judgment (Kim et al., 2018; Lake et al., 2017). However, the current results suggest that high-level visual inferences may serve non-perceptual systems like moral judgment, and so we should expect HCNNs to do so too. This idea is consistent with other recent work (De Freitas et al., in preparation) in which we found that a HCNN trained only to do object recognition is able to reliably learn to predict when harm occurs as well as who the agent in a moral interaction is, with only minimal training (a linear transform from the processed visual features to the learned label). Therefore, if the object recognition system provides a basis for higher-level areas of the visual system to get information about a moral event such as whether there was harm, who the agent was, and whether he/she was causally related to the harm, then these outputs can be combined to produce a moral judgment. Indeed, when the information about harm and agency was combined in this other work, the AI results were a close match to human moral judgments about generic visual events (De Freitas et al., in preparation). A similarly promising avenue for future work may be the potential interaction between intentionality and moral judgment, given that the visual system automatically computes representations of intentionality (for a review, see Scholl & Gao, 2013), and intentionality is relevant for moral judgment (e.g., Patil et al., 2017; Young & Saxe, 2009).

One common reason that models based on the visual system are criticized as being too limited for learning aspects of social psychology is that they purportedly do not contain general 'concepts' or 'modules' (Lake et al., 2017). This is thought to be the reason why, for instance, a human baby learns to do a visual task quicker than most deep learning programs (e.g., Mnih et al., 2015). It is thought that instead of learning by leveraging statistical information, one has to pre-specify more general symbolic rules that make up the grammar of our intuitive physics and psychology. Yet even if the mind does contain so-called modules, a complete explanation of cognition must also specify how these modules are learned in the first place. In the present case, the effects are highly stimulus driven, suggesting that causal perception has been learned and tuned to the statistics of the environment, i.e., only kinematics that are reliably associated with causation in the real, complex world drive automatic causal perception (Flombaum et al., 2009; Purves et al., 2014; Scholl & Nakayama, 2002). In fact, such statistical tuning may be the *best* way to tune any perceptual 'module', since this learning process can incorporate multiple statistical contingencies, whereas a theoretically-specified module should miss the full set of contingencies. Therefore, we hope that the current results will help invigorate modeling approaches that synthesize statistical models of how a module is *acquired* with symbolic models of how the module is *deployed*. Indeed, since statistical models like HCNNs are built on a biologically plausible foundation of neural networks, we should first ask how far their utility can extend *before* we settle on symbolic models of computation with little or no biological basis. Unless this approach is taken, we suspect that theoretical models will not generalize beyond toy examples with smaller sets of statistical contingencies.

7.3. Moral Intuitions, fast and slow

We have been suggesting that one reason that certain causal percepts are so entrenched in the mind is that they have presumably been tuned to the statistical regularities of causal events in the environment. As a result, certain spatiotemporal features reflexively give rise to certain percepts, forcing upon us a specific causal interpretation of the world. Here we note that this manner of drawing causal inferences is distinct from how much of cognitive causal inferences are made: in general, cognitive inferences of causality tend to be more graded (rather than categorical). For instance, when talking about the cause of a given event (e.g., why our flowers died) we can talk about one event being more the cause than another (e.g., the gardener neglected to water them, the weather was bad, the flower didn't take well to our soil, Obama didn't water the flowers etc.), and to the extent that we view something as the *the* cause it's because we view it as the *best* cause (Halpern & Hitchcock, 2014). Thus, in the face of ambiguity, cognition is free to ponder this ambiguity and draw graded conclusions, whereas vision commits to its best interpretation. Furthermore, although some causal judgments are based on quick heuristics, many are more slow and deliberative/analytic (rather than fast and automatic). For instance, some causal judgments (e.g., whether it's *always* true that if you water a plant well, then the plant will stay green) are based on reasoning that takes place in multiple consecutive stages, where people first make quick conclusions, and then revise these conclusions by searching for situations that contradict their initial conclusion (Verschueren, Schaeken, & d'Ydewalle, 2005). Further, we can often inspect at least some of the steps that led to our cognitive inferences, whereas we do not have introspective access to the mechanisms that give rise to our perceptual phenomena. In sum, causal perception and causal cognition seem to arise from broadly distinct computational architectures (Kahneman, 2011). As such, we may expect that these differences have consequences for moral judgment.

For one, we should expect people to generally have more conviction in moral judgments drawn from causal percepts than those drawn from more overt reasoning. Since causal percepts should not feel like multi-step inferences but like a mirror reflection of the world, people should be less inclined to question these inferences or the moral judgments that follow from them. As one of the observers in our experiments explained when asked to justify his moral judgment, "I saw it happen". Furthermore, it is well known that the more fluently/effortlessly a belief comes to mind, the more conviction one has in that be-

UNCORRECTED PROOF

lief (for reviews, see Clore, 1992; Jacoby, Kelley, & Dywan, 1989; Whittlesea, Jacoby, & Girard, 1990). Thus, immediate causal percepts may have similar fluency effects on moral judgment. Finally, a causal percept, being more immediate and potentially more salient, may also give rise to stronger emotional reactions, which are known to guide moral judgments in some causal contexts (Greene et al., 2001).

These possibilities would have implications for any legal situation where visual footage is used to determine a person's moral responsibility for harmful interactions. These days, many harmful interactions are caught on surveillance cameras, smart-phones, and "on-officer" cameras. These videos are often thought to meet the best standards of objective evidence, although one empirically motivated article by the New York Times (Williams, Thomas, Jacoby, & Cave, 2016) found that the same event can be construed in very different moral terms depending on the camera angle used to film it. For instance, the view from an on-officer camera may suggest that a suspect pushed the officer over, whereas a side angle shot reveals that no actual contact was made; the cop just tripped. The current experiments suggest one explanation for such differences. In cases of perceptual ambiguity, so long as certain spatiotemporal features are satisfied, viewers should reliably perceive causality. If they do, then they might be less inclined to look for additional evidence. The cautionary note is that even ambiguous videos may lead to strong conclusions. Thus, the mere impression of causality should not be taken as sufficient proof, and more reliable signs of causality— such as actual contact made, or weapons drawn — should be sought. Also, when available, multiple recordings of the event should be consulted, regardless of how compelling any single recording may appear.

In this vein, causal perception may also be exploited to manipulate others. For instance, this may explain why soccer and basketball players often manage to convince referees that they have been fouled, when in fact they have not. These players are presumably adept at exploiting spatiotemporal coincidence, falling at just the right time, such as when another player brushes by. Crucially, the whole reason that referees may fall for such acts is that their visual systems have already come to the strong conclusion that a causal interaction occurred, and so they should feel relatively confident about calling the foul. Echoing the stimulus-driven nature of our effects, playing with the spatiotemporal factors of these sports interactions — as when one replays the video at a slightly slower speed — is often sufficient to extinguish the illusion, often to comic effect. We may conclude that referees' visual systems put them at a disadvantage when judging these interactions in real time, but slowing down the interaction would reveal more objective features, such as contact and the exact delay between contact and falling. Therefore, referees should consult such objective evidence before making their decisions (as is already done cricket), unless, of course, fake fouls (and fooled referees) are part of the entertainment spectacle.

Future work should also pit causal perception and cognition against each, in order to determine how they are reconciled to make moral judgments: Under what circumstances does one modality get weighed more heavily than the other toward moral judgments? In a similar vein, such work can present information from multiple domains, e.g., written, auditory, and visual. A weakness of the current studies is that they only studied moral judgments within the visual domain, which may have led us to overestimate the effects of causal perception on moral judgment more generally. Thus, one question is how much causal perception continues to exert an influence within more complicated domains. For example, future work could study these effects within the context of real-world videos of morally relevant interactions, e.g., while ambient sounds and conversations are simultaneously occurring. In contrast, here we have taken the first-step approach of distilling the effect in a very simple moral scenario, in order to determine whether it even exists.

7.4. Trust and origins of moral intuitions

More generally, the findings may be epistemically challenging, because they may shake our belief that we have full control over our moral intuitions — after all, we often have moral disagreements with others because we have different values that we take to be our own (e.g., De Freitas, Tobia, Newman, & Knobe, 2017). This work contributes to a line of studies suggesting that this view is not always true (Gu et al., 2013; Haidt et al., 1993; Petrinovich & O'Neill, 1996; Pärnamets et al., 2015; Wheatley & Haidt, 2005). Of course, this is the same take-home message of almost every visual illusion. Yet almost all illusions have been limited to isolated demonstrations that are not connected in any way to abstract judgments like moral judgment, which also happens to be a uniquely human capability with everyday significance and social consequence. Luckily, in most cases an automatic causal inference should serve our moral judgments well. Though as we have noted above, it is also likely that mistaken conclusions of moral blame may extend beyond the laboratory.

We also speculate that causal percepts may not only inform our moral intuitions about visual events, but may also indirectly underlie many of the causal intuitions that we have in the absence of visual input, as when deciding how to solve the kinds of moral dilemmas that have inspired the most popular models of moral cognition (e.g., Alicke, 2000; Alicke, 2015; Cushman, 2008; Knobe, 2009; Malle et al., 2014; Noothigattu et al., 2017; Samland & Waldmann, 2016; Shaver, 1985; Weiner, 1995). Solving such moral questions requires engaging in causal imagery and tacit reasoning which may bring online the relevant representations and processes that are associated with causal percepts, as well as memories of such causal percepts (e.g., see Kosslyn, Ganis, & Thompson, 2001). Because these processes are primitive and operate flexibly, they may serve as an inescapable lense through which we not only view but also think about the world.

Providing some support for this view, Amit and Greene (2012) found that individuals with greater visual imagery ability were more likely to provide severe deontological moral judgments in the trolley dilemma (Foot, 1978), i.e., they were more likely to say that it is wrong to push someone on to a train track in order stop an oncoming train from killing five other people further along on the track (footbridge condition), than it is to turn a switch which diverts the train onto a separate track where it kills one person (switch condition). The authors suggest that the reason for this effect may be that participants were more likely to simulate the harm in the footbridge condition than in the switch condition. We suggest that a reason for this is specifically that the footbridge condition involves direct harm, which fulfills the standard spatiotemporal criteria for causal perception, whereas the switch condition involves indirect harm that does not satisfy these criteria. If this is the case, then we should expect people to show signs of reasoning based on this intuitive difference. We found such evidence in a separate study, which asked people to make moral judgments about these cases, and then explain these judgments. Independent coders then coded these sentences for their use of direct (lexical) *vs*. indirect (periphrastic) verb constructions (De Freitas et al., 2017). Sure enough, people's language suggested that they were encoding footbridge events in more direct, causal terms than they were switch events.

UNCORRECTED PROOF

7.5. Causal illusions that drive moral judgments, not moral illusions

When we discuss our results with others, we are sometimes asked whether we are claiming that the visual system perceives moral content per se. We are not. Instead, we are claiming that causal perception per se can drive moral judgments. For our results to be an instance of *moral perception*, we would have to show that moral content per se was inferred by the visual system, yet our results only require that causal content was inferred.

Although some authors have claimed that 'moral perception' per se does actually occur (Gantman & Van Bavel, 2015), others have empirically revealed alternative explanations for these results that suggest that this idea may still need evidence (Firestone & Scholl, 2015; Firestone & Scholl, 2016). Thus, it would be premature for us to conclude that the current results reflect moral perception without first defining what we take to be an instance of moral perception, and then isolating this hypothesized process with the kinds of psychophysical methods that can be used to conclusively arbitrate between perception and cognition.

8. Conclusion

We find that moral judgments can be determined by causal illusions embedded in visual stimuli depicting morally relevant interactions. To our knowledge, these studies provide the first demonstration that causal perception drives moral judgment, leaving a number of implications for current models of moral cognition, AI approaches to social psychology, and the use of visual evidence in any moral domain. This is not to say that the current effects were not mediated by explicit inferences, nor that there are not cases in which moral judgments will also need to factor in additional information, such as the mental states of the agents, and whether they are violating values that are held by the observer. Yet to the extent that the patterns of moral judgment originated in different causal percepts (or a lack thereof), our results do suggest that moral judgments are not only a matter of decision and behavior, but also a matter of visual inference.

10. Uncited reference

De Freitas (2018).

Acknowledgments

We thank Steven Pinker, Patrick Cavanagh, Roger Strong, Bria Long, and Johan Wagemans for helpful comments, and Patcharaporn Thammathorn for coding the sentences.

Appendix A. Replication of Experiment 3b analysis with the subset of participants who did not mention a car hitting the man

Likert Scale Question. Unlike Experiment 3a's same direction condition, which used smooth motion, moral blame judgments for the still frames same direction condition (M=4.71, SD=3.00) were indistinguishable from the still frames context absent condition (M=4.16, SD=2.89), $t(95) = 0.90$, p=.370, d=0.19), as well as from Experiment 3a's context absent condition that involved motion (M=4.17, SD=3.31), $t(146) = 1.01$, p=.314, d=0.17. Moreover, we observed a significant experiment (smooth motion *vs*. still frames) × condition (same direction *vs*. absent) interaction, $F(1, 295) = 5.85$, p=.016, $\eta p^2 = 0.02$), indicating that the effect of grouping context was greater for smooth motion displays than still frame displays. Thus, even for the subset of participants who did not mention a car hitting a man, moral judgments appeared to be driven by the motion-based visual illusion, not by explicit reasoning.

Forced-Choice Question. We found a similar pattern for the forced-choice of which driver was most blameworthy. A logistic regression found a significant experiment (smooth motion *vs*. still frames) × condition (same direction *vs*. absent) interaction (β = 1.44, SE=0.56, p=.010). The still frames same direction condition (29%) elicited statistically indistinguishable moral blame than the still frames absent baseline (26%), $\chi^2(1, N=97) = 0.00$, p=.971, = .00, and smooth motion context absent condition from Experiment 3a (34%), $\chi^2(1, N=148) = 0.20$, p=.656, = .04.

References

Adelson, E.H., 1999. Lightness perception and lightness illusions. In: Gazzaniga, M. (Ed.), The cognitive neurosciences. MIT Press, Cambridge, pp. 339–351.
Alicke, M.D., 2000. Culpable control and the psychology of blame. Psychological Bulletin 126 (4), 556–574.
Alicke, M.D., Mandel, D.R., Hilton, D., Gerstenberg, T., Lagnado, D.A., 2015. Causal conceptions in social explanation and moral evaluation: A historical tour. Perspectives on Psychological Science 10 (6), 790–812.
Amit, E., Greene, J.D., 2012. You see, the ends don't justify the means: Visual imagery and moral judgment. Psychological Science 23 (8), 861–868.
Anstis, S.M., 1980. The perception of apparent movement. Philosophical Transactions of the Royal Society of London B 290 (1038), 153–168.
Blakemore, S.J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A.N., … Decety, J., 2001. How the brain perceives causality: An event-related fMRI study. NeuroReport 12 (17), 3741–3746.
Buehner, M., Humphreys, G., 2010. Causal contraction: Spatial binding in the perception of collision events. Psychological Science 21 (1), 44–48.
Caruso, E.M., Burns, Z.C., Converse, B.A., 2016. Slow motion increases perceived intent. Proceedings of the National Academy of Sciences 113 (33), 9250–9255.
Choi, H., Scholl, B.J., 2004. Effects of grouping and attention on the perception of causality. Perception & Psychophysics 66 (6), 926–942.
Choi, H., Scholl, B.J., 2006. Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. Perception 35 (3), 385–399.
Clore, G.L., 1992. Cognitive phenomenology: Feelings and the construction of judgment. In: Martin, L.L., Tesser, A. (Eds.), The construction of social judgments. Lawrence Erlbaum Associates Inc., Hillsdale, NJ, pp. 133–163.
Crane, T., 1988. The waterfall illusion. Analysis 48 (3), 142–147.
Cravo, A.M., Claessens, P.M., Baldo, M.V., 2009. Voluntary action and causality in temporal binding. Experimental Brain Research 199 (1), 95–99.
Crump, M., McDonnell, J., Gureckis, T., 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS ONE 8 (3), e57410.
Cushman, F., 2008. Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. Cognition 108 (2), 353–380.
De Freitas, J., & Cikara, M. (2017). Deep down my enemy is good: Thinking about the true self reduces intergroup bias. *Journal of Experimental Social Psychology*.
De Freitas, J., Cikara, M., Grossmann, I., Schlegel, R., 2017. Origins of the belief in morally good true selves. Trends in Cognitive Sciences 74, 307–316.
De Freitas, J., DeScioli, P., Nemirow, J., Massenkoff, M., Pinker, S., 2017. Kill or die: Moral judgment alters linguistic coding of causality. Journal of Experimental Psychology: Learning, Memory, and Cognition 43 (8), 1173–1182.
De Freitas, J., Hafri, A., Yamins, D. L. K., & Alvarez, G. A. (in preparation). Learning to recognize objects provides category-orthogonal features for social inference and moral judgment. in preparation.
De Freitas, J., & Johnson, S. G. B. (2015). Behaviorist thinking in judgments of wrongness, punishment, and blame. In D.C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, *Proceedings of the 37th annual conference of the cognitive science society*, Jennings CD, Maglio PP (University of Texas at Austin, Austin, TX) (pp. 524–529).
De Freitas, J., Tobia, K., Newman, J.E., Knobe, J., 2017. Normative judgments and individual essence. Cognitive Science 41 (S3), 382–402.
Felin, T., Koenderink, J., Krueger, J.I., 2017. Rationality, perception, and the all-seeing eye. Psychonomic Bulletin & Review 24 (4), 1040–1059.
Firestone, C., Scholl, B.J., 2015. Enhanced visual awareness for morality and pajamas? Perception vs. memory in 'top-down'effects. Cognition 136, 409–416.
Firestone, C., Scholl, B.J., 2016. 'Moral perception' reflects neither morality nor perception. Trends in Cognitive Sciences 20 (2), 75.
Flombaum, J.I., Scholl, B.J., Santos, L.R., 2009. Spatiotemporal priority as a fundamental principle of object persistence. In: Hood, B., Santos, L. (Eds.), The origins of object knowledge. Oxford Univ. Press, London, pp. 135–164.
Foot, P., 1978. Virtues and vices and other essays in moral philosophy. University of California Press, Berkeley, CA.

UNCORRECTED PROOF

Gantman, A.P., Van Bavel, J.J., 2015. Moral perception. Trends in Cognitive Sciences 19 (11), 631–633.
Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293 (5537), 2105–2108.
Gu, J., Zhong, C.-B., Page-Gould, E., 2013. Listen to your heart: When false somatic feedback shapes moral behavior. Journal of Experimental Psychology: General 142 (2), 307–312.
Haidt, J., 2012. The righteous mind: Why good people are divided by politics and religion. Paragon, New York, NY.
Haidt, J., Koller, S.H., Dias, M.G., 1993. Affect, culture, and morality, or is it wrong to eat your dog?. Journal of Personality and Social Psychology 65 (4), 613–628.
Halpern, J.Y., Hitchcock, C., 2014. Graded causation and defaults. The British Journal for the Philosophy of Science 66 (2), 413–457.
Heider, F., Simmel, M., 1944. An experimental study of apparent behavior. The American Journal of Psychology 57 (2), 243–259.
Jacoby, L.L., Kelley, C.M., Dywan, J., 1989. Memory attributions. In: Roediger, H.L., Craik, F.I.M. (Eds.), Varieties of memory and consciousness: Essays in honour of endel tulving. Lawrence Erlbaum Associates Inc, Hillsdale, NJ, pp. 391–422.
Kahneman, D., 2011. Thinking, fast and slow. Macmillan.
Kanizsa, G., 1979. Organization in vision. Praeger, New York.
Kim, S.H., Feldman, J., Singh, M., 2013. Perceived causality can alter the perceived trajectory of apparent motion. Psychological Science 24 (4), 575–582.
Kim, R., Kleiman-Weiner, M., Abeliuk, A., Awad, E., Dsouza, S., Tenenbaum, J., & Rahwan, I. (2018). A Computational model of commonsense moral decision making. Also Available at: arXiv:1801.04346.
Knobe, J., 2009. Folk judgments of causation. Studies In History and Philosophy of Science Part A 40 (2), 238–242.
Kosslyn, S.M., Ganis, G., Thompson, W.L., 2001. Neural foundations of imagery. Nature Reviews Neuroscience 2 (9), 635–642.
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*.
Leslie, A.M., 1982. The perception of causality in infants. Perception 11 (2), 173–186.
Leslie, A.M., Keeble, S., 1987. Do six-month-old infants perceive causality?. Cognition 25 (3), 265–288.
Iliev, R.I., Sachdeva, S., Medin, D.L., 2012. Moral kinematics: The role of physical factors in moral judgments. Memory & Cognition 40 (8), 1387–1401.
Malle, B.F., Guglielmo, S., Monroe, A.E., 2014. A theory of blame. Psychological Inquiry 25 (2), 147–186.
Matsuno, T., Tomonaga, M., 2017. Causal capture effects in chimpanzees (Pan troglodytes). Cognition 158, 153–164.
Mayrhofer, R., Waldmann, M.R., 2014. Indicators of causal agency in physical interactions: The role of the prior context. Cognition 132 (3), 485–490.
Michotte, A., 1946. La perception de la causalité [the perception of causality]. Études de Psychologie, Louvain.
Michotte, A. (1963). The perception of causality. (T.R. Miles & E. Miles, Trans.). London: Methuen. (English translation of Michotte, 1954).
Michotte, A., Thinès, G., & Crabbé, G. (1964). Les complements amodaux des structures perceptives. In Studia Psychologica. Louvain: Publications Universitaires. Reprinted and translated as Michotte, A., Thinès, G., & Crabbé, G. (1991). Amodal completion of perceptual structures. In G. Thines, A. Costall, & G. Butterworth (Eds.) Michotte's Experimental Phenomenology of Perception (pp. 140–167). Hillsdale, NJ: Erlbaum.
Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., … Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533.
Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster than non-causal events. *Peer Journal 5*:e2932; http://doi.org/10.7717/peerj.2932.
Nagel, J., & Waldmann, M. R. Force dynamics as a basis for moral intuitions. In N. Miyake, D. Peebles, R.P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society*, (Austin, TX) (pp. 785–790).
Necker, L.A., 1832. Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. London and Edinburgh Philosophical Magazine and Journal of Science 1 (5), 329–337.
Newman, G.E., De Freitas, J., Knobe, J., 2015. Beliefs about the true self explain asymmetries based on moral judgment. Cognitive Science 39 (1), 96–125.
Noothigattu, R., Gaikwad, S. N. S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P., & Procaccia, A. D. (2017). A voting-based system for ethical decision making. Also Available at: arXiv:1709.06692.
Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381 (6583), 607–609.
Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M.J., Richardson, D.C., 2015. Biasing moral decisions by exploiting the dynamics of eye gaze. Proceedings of the National Academy of Sciences 112 (13), 4170–4175.
Patil, I., Calò, M., Fornasier, F., Cushman, F., Silani, G., 2017. The behavioral and neural basis of empathic blame. Scientifc Reports 7 (1), 5200.
Petrinovich, L., O'Neill, P., 1996. Influence of wording and framing effects on moral intuitions. Ethology and Sociobiology 17 (3), 145–171.
Phillips, J., De Freitas, J., Mott, C., Gruber, J., Knobe, J., 2017. True happiness: The role of morality in the concept of happiness. Journal of Experimental Psychology: General 146 (2), 165–181.
Purves, D., Monson, B.B., Sundararajan, J., Wojtach, W.T., 2014. How biological vision succeeds in the physical world. Proceedings of the National Academy of Sciences 111 (13), 4750–4755.
Rogers, B., 2014. Delusions about illusions. Perception 43 (9), 840–845.
Rolfs, M., Dambacher, M., Cavanagh, P., 2013. Visual adaptation of the perception of causality. Current Biology 23 (3), 250–254.
Rossetti, Y., 1998. Implicit short-lived motor representations of space in brain damaged and healthy subjects. Consciousness and Cognition 7 (3), 520–558.
Rossi, A.F., Paradiso, M.A., 1995. Feature-specific effects of selective visual attention. Vision Research 35 (5), 621–634.
Samland, J., Waldmann, M.R., 2016. How prescriptive norms influence causal inferences. Cognition 156, 164–176.
Schlottmann, A., Shanks, D., 1992. Evidence for a distinction between judged and perceived causality. Quarterly Journal of Experimental Psychology 44 (2), 321–342.
Scholl, B. J. & Gao, T. (2013) Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M.D. Rutherford, V.A. Kuhlmeier (Eds.), Social perception: Detection and interpretation of animacy, agency, and intention, (pp. 197–230). MIT Press.
Scholl, B.J., Nakayama, K., 2002. Causal capture: Contextual effects on the perception of collision events. Psychological Science 13 (6), 493–498.
Scholl, B.J., Nakayama, K., 2004. Illusory causal crescents: Misperceived spatial relations due to perceived causality. Perception 33 (4), 455–470.
Scholl, B.J., Tremoulet, P.D., 2000. Perceptual causality and animacy. Trends in Cognitive Sciences 4 (8), 299–309.
Sekuler, R., Sekuler, A.B., Lau, R., 1997. Sound alters visual motion perception. Nature 385 (6614), 308.
Shaver, K.G., 1985. The attribution of blame: Causality, responsibility, and blameworthiness. Springer-Verlag, New York.
Strohminger, N., & Kumar, V. (2018). The Moral Psychology of Disgust.
Thomas, K.A., Clifford, S., 2017. Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. Computers in Human Behavior 77, 184–197.
Thomas, K.A., De Freitas, J., DeScioli, P., Pinker, S., 2016. Recursive mentalizing and common knowledge in the bystander effect. Journal of Experimental Psychology: General 145 (5), 621–629.
Treisman, A., Schmidt, H., 1982. Illusory conjunctions in the perception of objects. Cognitive Psychology 14 (1), 107–141.
Turk-Browne, N.B., Jungé, J.A., Scholl, B.J., 2005. The automaticity of visual statistical learning. Journal of Experimental Psychology: General 134 (4), 552–564.
Van Lier, R., Wagemans, J., 1999. From images to objects: Global and local completions of self-occluded parts. Journal of Experimental Psychology: Human Perception & Performance 25 (6), 1721–1741.
Verschueren, N., Schaeken, W., d'Ydewalle, G., 2005. A dual-process specification of causal conditional reasoning. Thinking & Reasoning 11 (3), 239–278.
Weiner, B., 1995. Judgments of responsibility: A foundation for a theory of social conduct. The Guilford Press, New York.
Weiskrantz, L., 1986. Blindsight: A case study and implications. Clarendon Press, Oxford, England.
Wertheimer, M. (1912/1961). Experimentelle Studien über das Sehen von Bewegung. Zeitschrift für Psychologie, 61,161–265. Reprinted and translated as Wertheimer, M. (1961). Experimental studies on the seeing of motion. In T. Shipley (Ed. and Trans.) Classics in Psychology (pp. 1032–1089). New York: Philosophical Library.
Wheatley, T., Haidt, J., 2005. Hypnotic disgust makes moral judgments more severe. Psychological Science 16 (10), 780–784.
White, P.A., 2012. Visual impressions of causality: Effects of manipulating the direction of the target object's motion in a collision event. Visual Cognition 20 (2), 121–142.
Whittlesea, B.W.A., Jacoby, L.L., Girard, K., 1990. Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. Journal of Memory and Language 29 (6), 716–732.
Williams, T., Thomas, J., Jacoby, S., Cave, D. (2016). Police body cameras: What do you see? Accessed at https://www.nytimes.com/interactive/2016/04/01/us/police-bodycam-video.html.
Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences 111 (23), 8619–8624.
Young, L., Saxe, R., 2009. An FMRI investigation of spontaneous mental state inference for moral judgment. Journal of Cognitive Neuroscience 21 (7), 1396–1405.

Data References

De Freitas, J. Moral_capture. Github, 2018. Link: https://github.com/juliandefreitas/moral_capture.

UNCORRECTED PROOF