Running head: MORAL KINEMATICS

Iliev, R., Sachdeva, S., & Medin, D. (2012) Moral kinematics: The role of physical factors in moral judgments. *Memory and Cognition*. Online. DOI 10.3758/s13421-012-0217-1

Moral Kinematics: the Role of Physical Factors on Moral Judgments

Rumen I Iliev*, Sonya Sachdeva and Douglas L Medin

Northwestern University

*Corresponding Author: r-iliev@northwestern.edu

Department of Psychology

Northwestern University

2029 Sheridan Road - 219 Swift Hall

Evanston, IL 60208-2710

Abstract

Harmful events often have a strong physical component, e.g. car accidents, plane crashes, fist fights, and military interventions. Yet there is very little systematic work on the degree to which physical factors influence our moral judgments about harm. Since physical factors are related to our perception of causality, they should also influence our subsequent moral judgments. In three experiments we tested this prediction, focusing in particular on the role of motion and contact. In Experiment 1 we used abstract video stimuli and found that intervening on a harmful object was judged less bad than intervening directly on the victim, and that setting an object in motion was judged worse than redirecting an already moving object. Experiment 2 showed that participants were sensitive not only to the presence or absence of motion and contact, but also to the magnitudes and frequencies associated with them. Experiment 3 extended the findings from Experiment 1 to verbally presented moral dilemmas. These results suggest that domain-general processes play a larger role in moral cognition than what is currently assumed.

Keywords: moral judgments, causality, motion, agency, physical factors

In this paper we ask a simple empirical question: Do people take physical factors into account when making moral judgments? At first glance, our ability to distinguish good from bad seems more or less independent from the way we perceive the physical aspects of our surroundings. Moral principles, such as "do no harm," are abstract and largely decontextualized, and many would argue that they have transcendental or sacred aspects embedded in them that go beyond the physical plane. Yet our cognitive system is attuned to the physical aspects of reality (Michotte, 1946; Heider, 1958; Wolff, 2008; White, 2006, 2009; Saxe & Carey, 2006), and we suggest that moral cognition will similarly be sensitive to physical properties. More specifically, we focus on motion and contact, which have been shown to be important factors in judgments of causality and agency, and thus presumably relevant to attributing moral blame.

The rest of this paper is organized as follows: first we review some relevant findings linking causality to moral judgments; then we briefly describe some of the physical properties that have been associated with causal reasoning; next we present three empirical studies exploring the role of motion and contact on moral judgments; and finally we turn to a discussion on the broader implications of the experimental results.

Causality in moral judgments

Among the different cognitive processes that are relevant to moral judgments, our ability to perceive meaningful connections between events, to make causal inferences and to assign causal roles seems particularly important (Heider, 1958; Shaver, 1985; Darley & Schultz, 1990; Waldmann & Dieterich, 2007; Sloman, Fernbach, & Ewing, 2009; Baron & Ritov, 2009, Pizarro, Uhlmann, & Bloom, 2003; Driver, 2008; Lagnado & Channon, 2008). Nevertheless, it is not only the presence or absence of a causal link between actor and outcome that guides our moral judgments. To a large degree moral judgments hinge on the particular properties of the link. Different actors bringing equivalent harm through different causal paths are often judged differently.

Three causality-based moral factors are particularly important for the current work:

Directness of the harm and locus of intervention

Royzman and Baron (2002) made the distinction between direct and indirect harm. Pushing a person who falls as a result is perceived as worse than pushing a fence on which the person is leaning causing the person to fall, even though the result is the same. In the first case, the falling of the person is an inherent part of the pushing; in the second, the inherent result concerns the fence, which in turn influences the person (see also Anscombe, 1963 and Davidson, 1980). Recent work by Paharia, Kassam, Greene and Bazerman (2009) extends these findings to cases of an indirect agency, even when the actors and victims are collective entities, such as corporations and the public.

Waldmann and Dieterich (2007) proposed an alternative distinction between victim-based and harm-based interventions. In a series of studies they found that actors who bring about a negative outcome through harm-based intervention (e.g., throwing a bomb on a person) are blamed less than actors bringing the same harm through victim-based intervention (e.g., throwing the person on the bomb).

Action-omission

Another important causal factor which has been linked to moral judgments is the distinction between action and omission (Ritov & Baron, 1990). An actor who harms a victim through his own action is judged worse than an actor who simply fails to prevent the harm. In one scenario Spranca, Minsk and Baron (1991) presented participants with a story about tennis player who intends to eliminate the favorite from the competition by making him eat something that he is allergic to. In the action version, the player orders food that will make the favorite sick, while in the omission case he intentionally does not prevent the competitor from accidentally ordering the same food. The player was judged more harshly in the first scenario than in the second one. Similarly, in the case of vaccination it has been found that people are more reluctant to cause harm through action (vaccinating), than through omission (not vaccinating), even when the odds favor acting (Ritov & Baron, 1990; Baron & Ritov, 2004, 2009)

Personal force

Recently Greene et al. (2009) suggested that another relevant causal factor that influences moral judgments is the amount of personal force involved in the event. For example, harm can be caused entirely by the contractions of the muscle of an actor, or by using other existing forces (see also Unger, 1996, p.101). Although the authors gave the example of pushing someone off a bridge as harm brought about through personal force, perhaps a better example (which does not include gravity) would be strangling someone. Contrast this with flipping a switch that redirects an already moving trolley toward a victim on the track. Greene et al found that the larger the use personal force in producing harm, the larger the blame assigned to the actor.

The physical aspect of causality and its implications to moral judgments

One common feature to all three factors described above is that contrast causal structures where the role of the actor is more salient than in the other. The clearest example is the action-omission distinction, where an actor causing harm through action is more easily seen as the cause for the harm, compared with the same harm brought about through omission, where attention may be directed elsewhere

(Spranca et al., 1991; see also Wolff, Barbey, & Hausknecht, 2010). A similar diffusion of causal responsibility is observed in the case of indirect harm: “when harm is indirect, other causes of the harm, aside from a decision maker’s choice, are salient” (Royzman & Baron, 2002). Finally in the personal force case the presence of other types of forces reduces the relative causal responsibility of the actor (see also work on causal discounting: Kelley, 1972).

Since varying the salience of the agentic role of an actor is important for moral judgments, a relevant question is wheter manipulating causal roles in terms of physical causality alone will affect moral judgments. Notice that even though the factors described above have strong physical components, they are still inherently semantic distinctions. Take for example the action-omission distinction, where we can readily assume that action is inherently associated with more motion or more muscular activity. Although such an assumption will be right in the majority of the cases, the mapping is not necessarily one-to-one. Bennett (1980), for example, points to the following exception: a person having to force his body to stay still in order to allow the dust in a room to fall on the ground and close an electric circuit, which, in turn, leads to some outcome. Stillness in this case can be considered as an action, and, as such, it implies a stronger agentic role of the actor, despite the fact that the actor does not generate or transfer a causal quantity[1]. From this perspective, understanding the role of physical factors in moral judgments is closely related, but not equivalent to, previous work on omission bias and directness.

The two physical factors most relevant to our studies are motion and contact. In physicalist approaches to causality, a causal interaction includes an agent-object that possesses some causal quantity and transfers it to a patient-object, ultimately changing its state. A prototypical example is a moving billiard ball, hitting a static billiard ball and stopping after the contact, while the second ball starts moving. Here motion is the causal quantity which is transferred via physical contact from agent to patient. In simple cases like this one, the presence of motion itself is enough for the assignment of agent or patient role to an object. For example, when participants are asked to describe such an event they will typically say “One ball set the other one into motion”, but very few will say “One ball stopped the other”, suggesting that they naturally associate motion and agency[2] (White, 2007).

In order to study the role of motion and contact on moral judgments we designed a simple setting which included three interacting objects - Actor, Victim and Harm - where the Actor's behavior leads to the Victim contacting the Harm, and

[1] Some will argue that “forcing yourself” to stand still is highly agentic, and Wolff (2008) suggests cases where intention is treated similarly to a physical cause, but these parallels are beyond our scope here.

[2] Notice that this agent-patient distinction is psychological, and does not reflect physical interactions in terms of Newtonian mechanics, where the forces applied by moving and static objects during collision have exactly the same magnitude.

subsequently dying. The first hypothesis is that people will be sensitive to motion patterns when assigning blame to an actor. If motion is assigned causal properties, then two predictions follow, one related to the motion of the actor and the other to the motion of the other two objects. The first is that when outcomes are the same, a moving actor will be blamed more then a static actor. This straightforward prediction is closely related to the action/omission distinction and has been already confirmed by other researchers (Spranca et al., 1991). The second prediction is that if there is preexisting motion not associated with the Actor, the Actor will be blamed less. For example, if an Actor harms a static Victim by setting him in motion, he will be blamed more than an Actor who redirects an already moving Victim. In other words, when there is a causal quantity that exists independently from the Actor and contributes to the outcome, the Actor's blameworthiness will be diminished[3].

Another hypothesis is that observers will be sensitive to the contact between Actor and Victim. An Actor intervening directly on the Victim will be blamed more than an Actor who intervenes on the Harm, which then contacts the Victim. This prediction is closely related to the directness factor (Royzman & Baron, 2002) and to Waldmann and Dieterich's (2007) locus of intervention, both of which predict that intervention on the victim is worse than intervention on the harm. It also agrees with the results of Cushman, Young, and Hauser (2006) and Cushman and Young

[3] Unger (1996) made similar prediction linking “protophysics” to “pseudoethics”, but he did not provide any theoretical justification or empirical testing of his intuitions.

(2011) who found that contact between Actor and Victim affects moral judgments. However, this prediction disagrees with data by Greene et al. (2009) who found that contact between Actor and Victim does not influence moral judgments. We suggest that the pattern the Greene and colleagues observed may be due to the way contact was operationalized: they compared a case where an Actor pushes a Victim off a bridge using his hands to an Actor who uses a pole to do the same. In this situation a pole might not be seen as a mediating object, in the same way as a shoe is not necessarily seen as a causal mediator when a person kicks something. In our experiments we present more nuanced tests of the role of contact between Actor and Victim.

A third hypothesis concerns the quantitative aspect of physical factors. Typically, when researchers study the role of different factors on moral judgments, they use binary factors, where the action is either present or not, or the harm is either direct or not, or the causal connection is present or not, harm intended or not, outcome achieved or not, etc. When studying the role of physical factors, however, it is natural to extend such binary comparisons to quantitative comparisons, involving magnitudes and frequencies. Thus, as an extension of the first two hypotheses, we predict that subjects will not only be sensitive to the presence or absence of motion and contact, but they will also be similarly sensitive to quantitative differences un motion and contact. “

Overview of experiments

A series of three experiments were conducted to test the hypotheses outlined above. Experiment 1 used video presentations to test the role of motion and physical contact on the perception of wrongness. Experiment 2 used similar stimuli to explore the quantitative aspects of motion and contact, manipulating factors such as distance, acceleration, resistance and duration. Experiment 3 followed up the results of the first study, and extended the findings to verbal trolley-type scenarios.

Experiment 1

The first experiment we provides initial evidence for the role of motion and contact on moral judgments. This study uses abstract video-stimuli, representing dynamic interaction between three objects, resulting in one of the objects being harmed. We evaluate two main hypotheses. The first one deals with contact and directness of intervention.

H1: Contact between Actor and Victim will lead to harsher moral judgments.

The second hypothesis is related to the origin of causal quantity, which in this particular case is motion:

H2: Pre-existing motion not associate with the Actor will reduce moral blame.

However, since we have two other objects, one on which the Actor intervenes, and one on which the Actor does not, we split this general hypothesis into two, more specific ones:

H2.1: If the intervened on object is static, the Actor will be blamed more, compared to when the intervened on object is moving.

H2.2: If the non-intervened on object is static, the Actor will be blamed more, compared to when the non-intervened on object is moving.

Method

Participants. Twenty-two Northwestern University undergraduate students participated as a partial requirement for an entry level psychology class.

Stimuli. A short movie introduced life on a hypothetical planet, inhabited by *Cylinders* and *Cones*. In addition, there were non-animate objects called *Fireballs*, which were harmless for the Cylinders but deadly for the Cones. The introduction was followed by a presentation of episodes where the action of a Cylinder resulted in a Cone touching a Fireball, and the subsequent death of the Cone. All video clips were generated in the physics simulator of the Blender software (www.blender.org). For clarity, henceforth we will refer to Cylinders, Cones and Fireballs as Actors, Victims and Harms, respectively.

Design and procedure. The vignettes manipulated 1. whether the Actor intervened on the Harm, or directly on the Victim, 2. whether the intervention modified preexisting motion of an object or set a static object into motion and 3. the dynamic state of the non-intervened object, which was also either static or moving. Crossing these three factors resulted in a 2x2x2 factorial design. One of these conditions is presented at Figure 1, and the introductory movie with the eight clips can be seen at https://depot.northwestern.edu/xythoswfs/webui/_xy-7745679_1-t_jeReppCo.

First the participants saw the introduction and all eight clips. After this familiarity experience, participants were told that they would have to compare the behavior of the Actor in each of the eight clips to the behavior of the Actors in the remaining seven clips, resulting in a total of 28 pairwise comparisons. Then participants saw the two clips from each of the 28 pairs again, and marked their answers on a six-point scale, ranging from the *"The action of the Cylinder in case A was much worse "* to *"The action of the Cylinder in case B was much worse. "* The comparisons were made in two pseudo-random orders. The task was self-paced and took approximately 15 minutes.

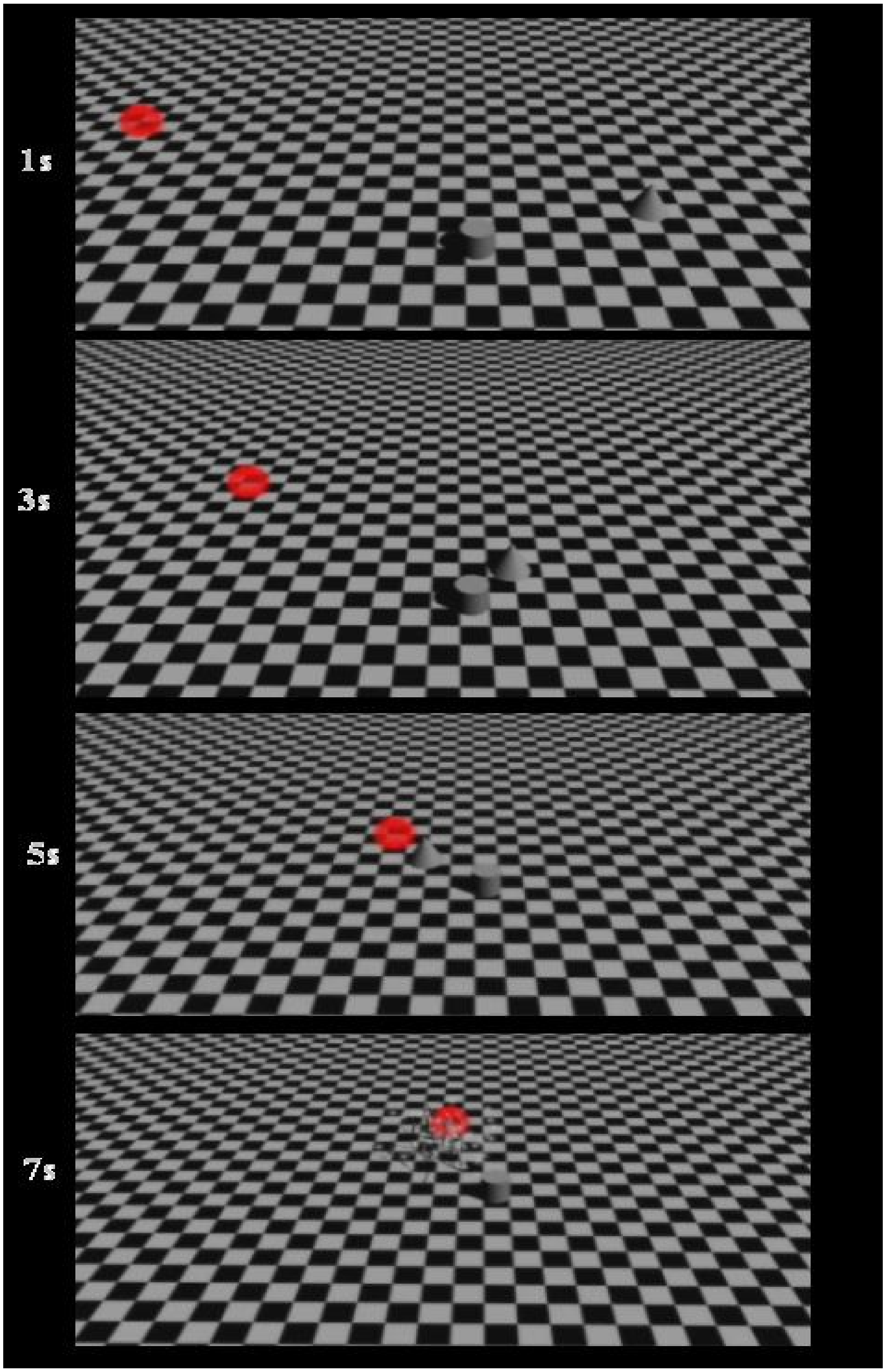


Figure 1. A time-sequence of one of the eight video-clips used as stimuli in Experiment 1. Here the Actor (Cylinder) intervenes directly on a moving Victim

(Cone), pushing it in the way of a moving Harm (Fireball).

Results and Discussion

We dichotomized the data for all 28 comparisons, assigning 1 if the action was judged worse in case A and -1 if the action was judged worse in case B[4]. Next, we computed the mean disapproval for each of the 8 scenarios. For clarification, if a scenario was always judged to be worse relative to the other 7 scenarios, then its mean score (M) would be -1; if it was judged to be always better, then its mean score would be 1, and if it was judged to be better (worse) in half of the cases, then its mean score would be 0. The mean scores and the corresponding proportions of "worse" choices for the eight events are presented in Table 1. Each of the main effects for the three factors we manipulated was analyzed separately.

To analyze the effect of contact, we looked at the pairs which compared the harm-based interventions to victim-based interventions. Since there were 4 harm-based intervention videos and 4 victim-based intervention videos, crossing them resulted in 16 comparisons[5]. In Table 1, this is the comparison between videos 1, 2, 3, 4 versus 5, 6, 7, 8. Across these comparisons, in 57% of the cases participants judged the victim-based intervention to be worse than the harm-based intervention, which was significantly above chance ($M = .14$, $SD = .21$, $t(21) =$

---

[4] If we use the original scale as a continuous variable the results do not change.

[5] For clarity, the other 12 pairs compared videos that had the same level of the factor, i.e. they juxtaposed harm to harm or victim to victim interventions.

3.43, $p<.01$[6]). This result supports our hypothesis that contact between Actor and Victim leads to harsher moral judgments.

The same type of analysis was applied for the role of the dynamic state of the intervened object. Out of the 8 videos, 4 involved intervention on a moving object (videos 1, 2, 5, 6), and 4 on a static one (videos 3, 4, 7, 8), again resulting in a total of 16 pairwise comparisons. In 59% of these, participants judged that intervention on a static object was worse than an intervention on a moving one (M = .18, SD = .22, $t(21)$ = 3.68, $p<.01$). This finding supports hypothesis 2.1, that preexisting motion associated with the intervened on object will lead to less harsh moral judgments. Next we analyzed the role of the dynamic state of the non-intervened object. Looking at the 16 comparisons which juxtapose videos of moving non-intervened object (videos 1, 3, 5, 7) to static non-intervened object (videos 2, 4, 6, 8) no reliable difference was found (M = .02, SD=.14, $t(21)$ = .64, $p>.05$). Thus, hypothesis 2.2 that moving non-intervened object will reduce blame was not supported.

We also tested whether the effect of one factor depended on the levels of the other factors. The design of the experiment does not allow the use of a regular ANOVA because the means for the eight clips are not independent from each other. Instead, we looked to se whether different groupings of these means revealed any noticeable patterns. There appeared to be an interaction-type mediation related to

---

[6] All statistical tests are two-tailed unless otherwise stated.

the role of the non-intervened object. When the non-intervened on object was static, the disapproval for intervening on a static rather than on moving object was much stronger (clip 2 and 6 paired against clips 4 and 8; M_1 =-0.39), compared to when the non-intervened object was moving (clips 1 and 5 versus 3 and 7; M_2 =-0.09). In other words, the effect of the static/moving distinction for the intervened on object was significantly mitigated when the non-intervened on object was moving (η_1^2 =.38, η_2^2 =.04, $t(21)$ =3.70, $p<.05$).

These results provide some initial evidence that the physical properties of a causal interaction can inform moral judgments in forms not limited to directness of the intervention. Instead, judgments are also affected by the dynamic state of the participating objects. Nevertheless, it was not simply the presence of motion that played a role. Given that motion can be a cue to agency and causal responsibility, one might expect that any kind of motion not associated with the Actor will mitigate the Actor’s moral responsibility. However, we found that only motion of the intervened object produced a main effect associated with less blame. We speculate that these results may be linked to what Waldmann and Dieterich (2007) labeled *interventional myopia,* where the information immediately associated with the causal intervention is weighed more than more peripheral information.

Clip	Intervened Object	State of the Intervened object	State of the Non-intervened object	Mean disapproval
1	Victim	moving	moving	-0.08 (54%)
2	Victim	moving	static	0.16 (42%)
3	Victim	static	moving	-0.24 (62%)
4	Victim	static	static	-0.40 (70%)
5	Harm	moving	moving	0.18 (41%)
6	Harm	moving	static	0.36 (32%)
7	Harm	static	moving	0.06 (47%)
8	Harm	static	static	-0.04 (52%)

Table 1. The design of Experiment 1, where three binary factors are crossed, in a 2x2x2 design. The last column shows the mean disapproval score, where more negative numbers indicate stronger disapproval for the action. The percentages next to the means reflect the proportion of choices where a scenario was judged “worse” than the remaining seven.

Experiment 2

The first experiment revealed that moral judgments are sensitive to the presence or absence of motion for the intervened on object, and to the presence or absence of contact between Actor and Victim. Taking into account the nature of physical quantities, a follow up question is whether participants will be similarly sensitive to physical factors when the differences are in magnitudes or frequencies. If motion can be seen as a causal quantity, then “more” motion generated by the actor should lead to harsher moral judgments. Similarly, since contact implies transfer of some causal quantity, "more" contact between Actor and Victim should suggest a stronger causal connection, again leading to harsher moral judgments. Thus our general prediction is:

H3: Higher magnitudes and frequencies associated with motion and contact will lead to harsher moral judgments.

This hypothesis can be further divided into five more detailed predictions:

H3.1 <u>Magnitude of force</u>. Larger force used by the Actor should to lead to harsher judgments. Larger force can be indicated by more acceleration or by the presence of an obstacle or source of resistance (such as the slope of the surface).

H3.2 <u>Physical contact</u>. Outcomes which include physical contact will lead to harsher judgments.

H3.3 <u>Distance traveled by the Actor before the contact</u>. Longer distances traveled are predicted to lead to harsher judgments. Here motion (or work) is used as an indicator of the causal quantity generated by the Actor, so a longer distance traveled is associated with more motion, thus larger causal quantity.

H3.4 <u>Duration of contact between Actor and Victim</u>. Longer durations of contact are predicted to lead to harsher judgments. The rationale for this prediction is twofold. First, if participants are sensitive both to physical contact and to magnitudes of physical factors, they should be sensitive to the duration of physical contact. Second, longer physical contact may imply more resistance encountered, thus larger force applied.

H3.5 Frequency of contacts. A greater number of contacts should be associated with harsher judgments. The rationale is the same as in the last prediction. Alternatively, if contact is treated as “boundary” of an act, more contacts can be also interpreted as more acts, leading to the same prediction.

Given that we are exploring much broader set of physical factors, the current experiment does not follow the factorial design of the previous one, but rather relies on paired comparisons which differ on a single factor (whenever possible) to examine this broader range of physical characteristics.

Method

Participants. Sixteen Northwestern University undergraduate students participated for a course credit.

Design and procedure. The participants saw the same introductory movie from Experiment 1. After the introduction they saw 15 pairs[7] of video-clips, where the behavior of the Actor leads to the death of the Victim. Most pairs differed on a single factor. For example, in the first video-clip in Comparison 1 the participants saw an Actor who traveled a short distance before pushing the Victim,

[7] Another four pairs were also presented, but they are not of interest for the current work

while in the second video-clip from the same comparison the Actor traveled a longer distance. Comparison 6 juxtaposed an Actor who traveled on flat surface to one who traveled uphill. The fifteen comparisons and the factors manipulated are represented in Table 2, and the actual stimuli can be seen at https://depot.northwestern.edu/xythoswfs/webui/_xy-7745556_1-t_m0AgbcWl . After seeing each comparison the participants were asked to judge in which case the behavior of the Actor was worse, using the same 6-point scale as in Experiment 1. The comparisons were presented in 4 pseudo-random orders.

Results and Discussion

For each pair, choices predicted from a physicalist framework were coded as 1, while the alternative choices were coded as 0. First we looked at the overall effect of physical factors. Collapsing across all 15 comparisons, 80% of the choices were in the predicted direction, which was significantly higher than chance ($t(14) = 8.89$, $p<.01$). We interpret this as an overall support for Hypothesis 3. Next, we looked at each comparison separately. Twelve of the comparisons achieved statistical significance by a sign test, two were not statistically reliable but in the predicted direction (67% and 69% of the cases), and one comparison showed no trend (50%). The means for each comparison are reported in Table 2.

Now we will consider each of the predictions separately. The magnitude of force (H 3.1) was varied in comparisons 3, 4, 5, 6, 7 and 8. Averaging across these probes, in 76% of the cases, participants judged the Actor using larger force as worse ($t(15) = 5.72$, $p<.01$). Physical contact between Actor and Victim (H 3.2) was varied in comparisons 10, and 11, and in 72% of the cases participants judged contact to be worse ($t(15)=2.41$, $p<.05$). The distance traveled by the Actor (H 3.3) was varied in comparisons, 1, 2, 3 and 8, and in 86% of the cases participants judged the longer distance to be worse ($t(15) = 7.91$, $p<.01$). Analyzing duration of contact (H 3.4) in comparisons 13,14 and 15, revealed that for 92% of the answers longer duration led to harsher judgments ($t(15)=11.19$, $p<.01$). Lastly, number of contacts (H 3.5) was varied in only comparison 12, and 75% of participants judged two contacts to be worse than one ($p <.05$, one-tailed sign test). In summary all five predictions were supported by the data.

Although this experiment was mainly designed to explore the role of separate physical properties rather than to estimate the relative importance of different factors by juxtaposing them, there was one scenario which directly compared the relative role of two factors. Comparison 9 presented one clip where an Actor moves away from the oncoming Victim, resulting in the Victim crashing into the Harm, and a second clip where a Victim bounces off a static Actor, going into the Harm. If participants are more sensitive to physical contact than they are to motion, they should judge the second Actor to be worse than the first one, since in the first case there is no physical contact. If, on the other hand, motion is more important than physical contact, then the opposite pattern would be expected and, indeed, that is what we observed. Eighty one percent of subjects judged the motion without contact to be worse than no motion with contact ($p < .05$, sign test). This finding suggests that physical contact is only a way to trace physical quantities rather than a major indicator of moral responsibility. In this case, when no physical energy originated from the actor, the role of physical contact was diminished. Nevertheless, physical contact was taken into account even in the absence of any movement on the part of the actor. In Comparison 10 a static Actor was either contacted or not by a moving Victim. In this case, when no motion was present, 75% of the participants judged contact to be worse than no contact ($p < .05$, one-tailed sign test). This pattern suggests that both motion and contact are taken into account, but when they conflict motion may be taken as the more important cue. Of course motion and contact are not categorical variables and the relative importance of the two likely depends on how strongly they are manipulated.

This study provides further support for the hypothesis that physical factors are consistently used as a basis for forming moral judgments. They also demonstrate that the difference in factors does not have to be qualitative or categorical, but can be quantitative, based on magnitudes and frequencies. Whenever the Actor traveled further or faster, pushed the Victim for a longer distance, overcame obstacles such as gravity and resistance, participants judged the action to be worse.

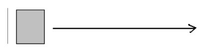
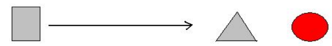
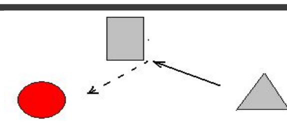
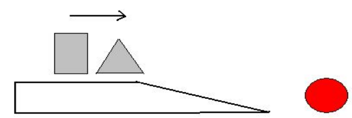
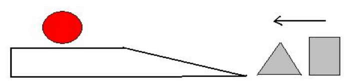
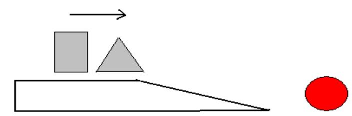
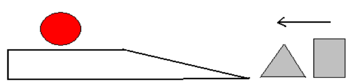
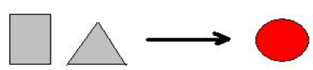
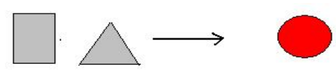
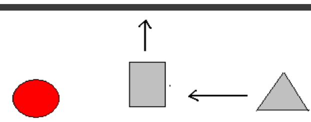
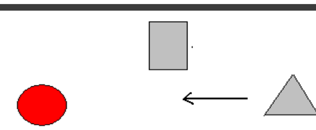
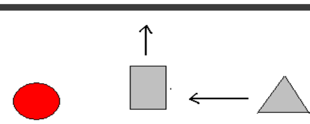
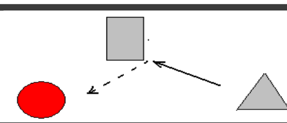
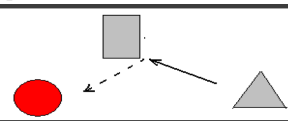
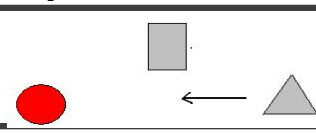
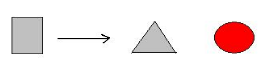
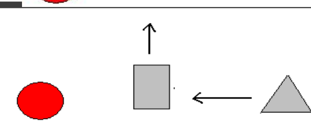
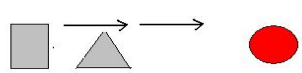
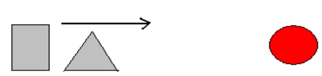
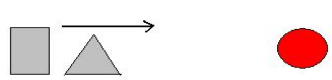
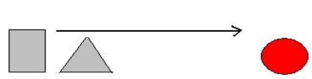
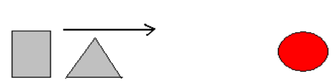
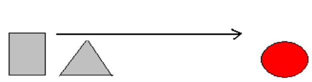
	Factor	Implementation	Clip A (predicted "worse")	Clip B	Agreement
1	Distance	Long distance versus short distance travelled by the Cylinder before the contact			0.82*
2	Distance	Long distance versus medium distance travelled by the Cylinder before the contact			0.88*
3	Force/Distance	Cylinder is moving versus static, physical contact in both cases			0.75*
4	Force	Cylinder moves the Cone downhill versus on a flat surface			0.5
5	Force	Cylinder moves the Cone downhill versus uphill			0.67
6	Force	Cylinder moves the Cone uphill versus on a flat surface			0.84*
7	Force	Cone travels faster versus slower after the contact			0.75*
8	Force/Distance	Cylinder is moving versus staying static. In both cases the Cone hits the Fireball without the Cylinder touching either of them.			1*
9**	Force vs Contact	Cylinder is static and the Cone bounces into it (physical contact) versus the Cylinder moves out of the way of the Cone (no physical contact)			0.82*
10	Contact	A static Cylinder is either contacted or not by a Cone			0.75*
11	Contact	Moving Cylinder contacts vs does not contact a Cone			0.69
12	Frequency	Two contacts versus one contact			0.75*
13	Duration	Cylinder pushes versus touches the Cone			1*
14	Duration	Cylinder pushes the Cone for a long versus medium period of time			0.82*
15	Duration	Cylinder pushes the Cone for a long versus short period of time			0.94*

Table 2. Summary of the comparisons in Experiment 2. The graphical representations indicate the motion patterns of the participating objects, juxtaposing the two clips in each pair. In the “Clip A” column are the events that a physicalist framework predicts will be judged “worse” than the ones in the “Clip B” column. In all events the final outcome was contact between the Harm and the Victim. The numbers in the rightmost column, labeled “Agreement”, indicate the proportion of participants making choices consistent with a physicalist framework. Notice that if participants were choosing randomly the proportion would be .50, while if they were consistently choosing the opposite answer the agreement with a physicalist framework would be less than .50.

*The mean is different from chance (.50) at $p<.05$

**Comparison 9 involves two opposing physical factors, where the directionality of the comparison favors motion over physical contact.

Experiment 3

The first experiment demonstrated that when motion and contact are presented visually, observers use them as a cue to inform their moral judgments. These findings, however, may be limited to abstract visual presentations and to events where harm is the only outcome. The current experiment is designed to test whether the role of motion and contact can be generalized further to less abstract scenarios. Compared to Experiment 1, there are two important changes in the current study: first, we present information in verbal form describing humans in real-life situations; and second, we employ a dilemma structure, where the intervention of the actor brings about both a positive and a negative outcome.

The scenarios were variations of the trolley problem, where the Actor intervenes either on the Victim or on the oncoming Harm. Notice that in the original version of the problem, flipping the switch is also an intervention on a moving object, whereas in the footbridge case the intervention is on a static object, so part of the asymmetry in our intuitions might be due to the dynamic state of the intervened object. To test this hypothesis, in the current experiment we manipulate the dynamic state of the victim, expecting that intervention on a moving victim will be judged less harshly than intervention on a static victim (H 2.1). Based on the results from Experiment 1, we do not expect the dynamic state of the victim to play a role when the intervention is on the harm (H 2.2). Finally, we also expect to replicate H 1, finding again that intervention on the Victim is judged worse than intervention on the Harm.

Method

Participants. One hundred and twenty Northwestern University undergraduate students participated in the experiment as a partial fulfillment for a class requirement.

Stimuli. Four different situations described a person in a position of responsibility who was in a trolley-type moral dilemma, where the only way to save five of his subordinates is to sacrifice one of them. The four actors were: a zoo manager, a biohazard lab manager, a railway station manager and an army officer. Each scenario had four different versions wherein two factors were manipulated: the first was whether the Actor intervenes on a Harm or on a Victim, and the second whether the dynamic state of the Victim was moving or static. The scenarios are given in Appendix A.

Design and Procedure. Each participant received four scenarios in total, one of each content type (army, zoo, train, biohazard) and one of each of the four experimental conditions (harm intervention – static victim, harm intervention - moving victim, victim intervention – static victim, victim intervention – moving victim). After the participants read the scenarios they were asked to go back and draw a schematic representation to help them visualize each of the four situations. After finishing the drawings the participants were asked to compare the decision in each of the scenarios to the other three, resulting in six total pairwise comparisons. Participants indicated their answers on the same 6-point scale used in Experiments 1 and 2. The comparisons were made in 8 different pseudo-random orders.

Note that this design allows us to assess both the dynamic factors manipulated and the effects of content domain.

Results and Discussion

As in Experiment 1, the scenario that was judged better in each pairwise comparison received a score of 1, while the one judged worse received a score of -1, i.e. an average score of 0 means that a scenario was judged better in half of the cases. The mean scores for each comparison are presented in Table 3. First, we tested whether intervention on the Victim was judged worse than invention on the Harm. In Table 3 this is a comparison of scenarios 3 and 4 versus scenarios 1 and 2. Collapsing across different contexts, decisions which involved direct intervention on the victim were judged more inappropriate compared to the decisions which intervened on the harm ($M=-.22$, $t(119)=6.08$, $p<.05$). Next, we tested the general version of Hypotheses 2; the prediction that preexisting motion in general, regardless of whether it is associated with the intervened or the non-intervened object, will lead to less harsh moral judgments. This is the comparison between scenarios 1 and 3, versus 2 and 4 in Table 3. There was no statistically reliable effect of the overall moving/static victim distinction ($M=-.06$, $t(119) = 1.69$, $p>.05$). However, looking only at the moving/static comparison when the intervention was directly on the Victim (comparing scenarios 3 versus 4), showed a reliable effect. Sixty-one percent of participants judged the intervention on a static victim to be worse than intervention on a moving victim (M=-.22, t(119)=2.42, p<.05). The moving/static victim distinction did not seem to play a role when the intervention was on the harm (M=0, t(119)= 0, p>.05). In other words, as in Experiment 1, H 2.1 was supported, while H 2.2 was not. [8]

		Scenario	1	2	3	4
Harm Intervention	Moving Victim	1	.	0 (50%)	0.22 (39%)*	0.46 (27%)*
	Static Victim	2	0 (50%)	.	0.32 (34%)*	0.36 (32%)*
Victim Intervention	Moving Victim	3	-0.22 (61%)*	-0.32 (66%)*	.	0.22 (39%)*
	Static Victim	4	-0.46 (73%)*	-0.36 (68%)*	-0.22 (61%)*	.

Table 3. A summary of the six pairwise comparisons in Experiment 3. Each cell shows the mean disapproval score of the scenarios in the rows when compared to the scenarios in columns, where more negative scores indicate higher disapproval. For clarity this number is also converted in percentages that show how often a scenario in the rows was judged worse than a scenario in the columns. The asterisk indicates that the preference was statistically different from chance at p<.05.

In addition to the effects of the physical factors, there were also strong

[8] We also analyzed the schematic representations the participants drew after reading the scenarios, but we did not find any patterns reflecting the experimental manipulation, or their subsequent judgments.

content effects. For example, intervention by an army officer was judged less bad in 86% of all comparisons in which it participated, regardless the particular causal structure the scenario described ($M=-.72$, $t(119)= 14.47$, $p<.05$). In fact, when the army scenario was part of the classical “switch-footbridge” comparison (with both victims being static), we did not even observe the typical trolley effect ($t(35) = 1.35$, $p=.19$), which shows that participants were more influenced by content than by the victim/harm or static/moving distinctions. The content effects we observed here are important, but not surprising. Moral norms and expectations are often dependent on the particular social role, regardless the abstract causal structure of the situation. Haidt and Baron (1996), for example, found that the effect of the action-omission distinction is dependent on the particular social role of the hypothetical Actor. Going back to our scenarios, an army officer sacrificing soldiers is less unexpected than a lab manager sacrificing research assistants.

Overall, the results replicate those from Experiment 1, where both the directness of the intervention (H 1) and the dynamic state of the intervened object (H 2.1) matter. The current study also showed that the effects of motion were not limited to seeing moving objects, but were also present when motion was just mentioned in a scenario. And last, these effects of causal path were accompanied by strong content effects related to the social role of the Actor.

General Discussion

The ways in which we understand the social and the physical worlds are very different (Anderson, 1972) and we become sensitive to this difference very early on in our development (Leslie, 1994; Wynn, 2008). Yet, the physical aspects of cognition are not limited to understanding simple interactions between inanimate objects, but instead influence how we process a much broader set of information, including events from the social world (see diSessa, 2000). In the present studies we demonstrate that physical quantities also influence our moral judgments. In Experiment 1 we used video-clips with a minimal context, varying the directness of the intervention and the dynamic state of the participating objects. In agreement with a number of previous findings, intervening directly on the victim was judged worse than intervention on the harm. In addition, intervention on a moving object was judged less bad than intervention on a static object. Experiment 2 explored a broader range of physical factors, again finding a pattern consistent with a physicalist framework. For example, participants in this study were sensitive both to the properties of the physical contact between Actor and Victim, and to the amount of energy that the Actor expended. Experiment 3 extended these findings to verbal presentations and to bivalent outcomes that combined positive and negative consequences.

The finding that the dynamic state of the objects and the physical properties of the contact influence subsequent moral judgments has two important implications. On a general level, it is directly related to the affect-cognition debate in the field (Haidt, 2007, 2010; Narvaez, 2010), emphasizing the importance of understanding the cognitive processes involved in moral judgments. A common criticism to affect-based approaches is that they cannot make ad-hoc predictions about the particular properties of the information that leads to stronger affective response. From this perspective, theories that focus on the cognitive processes involved in moral judgments will be a necessary part of any broader theory about morality (see Clore & Ortony, 2000; 2008 on cognition and emotion).

Further, our results suggest a much larger role for domain-general causal principles than is currently accepted in the field of moral cognition (for a discussion see Haidt & Joseph, 2007). For example, Hauser (2006a, 2006b) and Mikhail (2007, 2009) argue that, similar to language, domain-specific rules which are part of a universal[9] moral grammar are the primary determinants of our moral reasoning. Among these rules, for instance, is the *doctrine of double effect*, which, on one of its readings, allows harm to happen as a side effect of otherwise good action, but not as a means for achieving a better end. A pro-life doctor might perform a life-saving surgery on pregnant women that will result in the death of

[9] For empirical tests of cultural universality see Hauser, Cushman, Young, Jin, & Mikhail (2007) and Banerjee, Hubner & Hauser (2011)

the fetus, but would not abort the fetus even if this will save the mother's life too. Notice that this rule is actually a detailed analysis of the causal structure of the action, but unlike the other causal principles discussed here, it is focused solely on morally-relevant situations, where both positive and negative outcomes result from the same intervention, and where a situation can be parsed into means, ends and side-effects. Although the experiments presented here were not designed as a direct argument against domain-specific causal principles in morality, the results demonstrate a surprising role of domain-general causal factors (see also Rai and Holyoak, 2010). When combined with previous findings on action/omission, the role of directness and physical contact, and personal force, our results suggest a large and consistent role of causal inferences in moral judgments (for a recent discussion on domain-specificity and domain-generality of moral principles see Young & Dungan, 2012).

Before closing, we need to address two limitations of the current work. The first one is related to the role of mental states in moral judgments. Here we have established connections between physical factors and morality, but we have not explored in depth the specific mechanisms behind this link. We know that patterns of physical motion are not only linked to transfer of causal quantities, but they can also imply animacy, agency, intentionality (Scholl & Tremoulet, 2000; Gao, Newman, & Scholl, 2009) or even emotional states and social relations (Heider & Simmel, 1944; Barrett, et al., 2005, Bloom & Veres, 1999). From this perspective, one possibility is that physical factors are used only as cues to infer mental states, which subsequently are used to form moral judgments. A full answer to this question is beyond the scope of this paper, yet we believe that inferred differences in mental states will not be enough to explain our main findings. For example, in Experiment 3 we explicitly described the intention of the Actor, yet judgments were still sensitive both to contact and to the dynamic state of the intervened object. Further support for the idea that the role of causal cues goes beyond inferences about mental states comes from previous work on omission bias by Spranca et al. (1991). They found that although the properties of the causal path between Actor and Victim often imply different intentionality, controlling for intentionality does not eliminate the link between the causal path and moral judgments.

Also, we have largely ignored the role of normality, or surprisingness associated with an action, which has been proposed as another major factor for causal inferences (Kelley, 1972; Hart & Honore, 1985; Kahneman & Miller, 1986; Hilton & Slugosky, 1986). Witnessing a hammer hitting the crystal of a wristwatch, which breaks as a result, easily implies an agentic causal role for the hammer. When presented with the additional information that the event takes place in a testing facility in a wristwatch factory, however, we no longer see the hammer as the main cause, but rather attribute the outcome to the weakness of the crystal (Einhorn & Hogarth, 1986). In the second case the physical cues to causality are overshadowed by the knowledge activated when thinking about the testbed context. The strong content effects observed in Experiment 2 further illustrate the role of expectations and norms when distinguishing between otherwise causally-similar events.

Conclusion. One needs to keep in mind that moral cognition is not simpler than cognition itself. Hauser (2006b) pointed out that one of the central questions that moral psychology faces is to find how the moral mind is different from, and how it is similar to, the non-moral mind: “...we not only wish to uncover those processes that clearly support our moral judgments, but in addition, identify principles or mechanisms that are selectively involved in generating moral judgments.” In this paper we have demonstrated one surprising way in which moral cognition is similar to rather than different from non-moral cognition. We take these results as a demonstration that cognitive theories of domain-general causality can be a useful tool when charting the boundaries of the moral mind.

References

Anderson, P.W. (1972). *Science*, New Series, Vol. 177, No. 4047.

Anscombe, G. (1963). *Intention*, 2, Blackwell, Oxford.

Banerjee, K., B. Huebner, & M. D. Hauser (2011). Intuitive moral judgments are robust across demographic variation in gender, education, politics, and religion: A large-scale web-based study. *Journal of Cognition and Culture, 37, 151–187.*

Baron, J. & Ritov, I. (2004). Omission bias, individual differences, and normality. Organizational Behavior and Human Decision Processes, 94, 74-85.

Baron, J. & Ritov. I. (2009). Protected values and omission bias as deontological judgments. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral Judgment and decision making*, Vol. 50 in B. H. Ross (series editor), *The Psychology of Learning and Motivation*, pp. 133–167. San Diego, CA: Academic Press.

Barrett, H.C., Todd, P.M., Miller, G.F., & Blythe, P. (2005). Accurate judgments of intention from motion alone: A cross-cultural study. *Evolution and Human Behavior*, 26, 313-331.

Bloom, P. & Veres, C. (1999). The perceived intentionality of groups. *Cognition* 71, B1–B9

Clore, G. L. & Ortony, A. (2000). Cognition in emotion: Always, sometimes, or never? In L. Nadel, R. Lane & G. L. Ahern (Eds). *The Cognitive neuroscience of emotion*. New York: Oxford University

Press.

Clore, G. L., & Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In M. Lewis, J.M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of Emotions* (3rd ed., pp. 628-642). New York: Guilford Press.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108(2): 353-380.

Cushman, F, Young L., & Hauser M. (2006). The role of reasoning in moral judgments: Testing three principles of harm *Psychological Science* 17(12): 1082-1089.

Darley, J. M., & Schultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, 41, 525-556.

Davidson, D. (1980). *Essays on Actions and Events*, Oxford University Press, Oxford.

diSessa, A. A. (2000). Does the mind know the difference between the physical and social worlds? In L. Nucci, G. Saxe, and E. Turiel (Eds.), *Culture, Development and Knowledge* (pp. 141-166). Mahwah, NJ: Lawrence Erlbaum Associates. 141-166). Mahwah, NJ: Lawrence Erlbaum Associates.

Driver, J. (2008). Attributions of causation and moral responsibility.

in W. Sinnott-Armstrong (Ed.) *Moral psychology (Vol: 2):, The cognitive science of morality: Intuition and diversity* (pp.423-439). Cambridge, MA: MIT Press.

Einhorn, H. & Hogarth, R. (1986). Judging probable cause. *Psychological Bulletin*, 99 (1): 3-19.

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, 59 (2).

Greene, J.D., Cushman, F.A., Stewart. L.E., Lowenberg, K., Nystrom, L.E., and Cohen, J.D. (2009) Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364-371.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316,

Haidt, J. (2010). Moral psychology must not be based on faith and hope: Commentary on Narvaez. *Perspectives on Psychological Science*, *5.*

Haidt, J., & Joseph, C. (2007). The moral mind: How 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, and S. Stich (Eds.) *The Innate Mind, Vol. 3.* New York: Oxford, pp. 367-391.

Hart, H. & Honoré, T. (1985) *Causation in the Law*, 2nd ed., Oxford:

Clarendon, 1985.

Hauser, M. (2006a). Moral minds. New York: Harper Collins Publishers.

Hauser, M. (2006b). The liver and the moral organ. Social Cognitive and Affective Neuroscience. 1(3):214-220.

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R. & Mikhail, J. (2007) A dissociation between moral judgments and justifications. Mind & Language. 22(1):1-21.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: John Wiley & Sons.

Heider, F. & Simmel, S. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.

Kahneman, D. & Miller D. (1986). Norm Theory: Comparing reality to its alternatives. *Psychological Review*, 80, 136-153.

Kelley, H. (1972). *Causal schemata and the attribution process.* New York: General Learning Press.

Lagnado, D. & Channon, S. (2008). Judgments of cause and blame: The influence of intentionality and foreseeability, *Cognition* 108, 754–770.

Leslie, A. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture.* New York: Cambridge University Press, 119-148.

Malle, B. F. & Knobe, J. (1997). The Folk Concept of Intentionality. *Journal of*

Experimental Social Psychology, 33, 101-121.

Michotte, A. (1946/1963). *La Perception de la Causalite*. Louvain: Institut Superieur de Philosophie, 1946. [English translation by T. Miles & E. Miles, The perception of causality, Basic Books, 1963.]

Mikhail, J. (2007). Universal Moral Grammar: Theory, Evidence, and the Future. Trends in Cognitive Sciences, 11 (4), 143-152.

Mikhail, J. (2009). Moral Grammar and intuitive jurisprudence: A formal model of unconscious moral and legal knowledge. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of Learning and Motivation*, Vol. 50: *Moral Judgment and Decision Making*. San Diego, CA: Academic Press.

Narvaez, D. (2010). Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science, 5*(2), 163-181.

Paharia, N., Kassam, K.S., Greene, J.D., & Bazerman, M.H. (2009). Dirty work, clean hands: the moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes*, 109, 134-141.

Pizarro, D.A., Uhlmann, E., & Bloom, P. (*2003*). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39

Rai, T. S. and Holyoak, K.J. (2010). Moral Principles or Consumer Preferences?

Alternative Framings of the Trolley Problem. *Cognitive Science, 34,* 311–321

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, 3, 263–277.

Royzman, E. B. & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15, 165-184.

Saxe, R. & Carey, S. (2006). The perception of Causality in Infancy *Acta Psychologica,* 123, 144 – 165.

Scholl, B. J., & Tremoulet, P. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4, 299 - 309.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. NY: Springer-Verlag.

Sloman, S. A., Fernbach, P. M. & Ewing, S. (2009). Causal models: The representational infrastructure for moral judgement. In D. Bartels, C. W. Bauman, L. J. Skitka, & D. Medin (Eds.) *Moral judgment and decision making: The psychology of learning and* *motivation* (Vol 50). San Diego, CA: Elsevier.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27(1), 76–105.

Unger, P. (1996*). Living high and letting die: Our illusion of innocence*. Oxford: Oxford University Press.

Waldmann, M. & Dieterich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, 18 (3), 247-253.

White, P. (2006). The causal asymmetry. *Psychological Review*, 113, 132 - 147.

White, P. (2009). Perception of forces exerted by objects in collision events. *Psychological Review*, 116, 580 - 601.

White, P. (2007). Impressions of force in visual perception of collision events: A test of the causal asymmetry hypothesis. *Psychonomic Bulletin & Review*, 14, 647 - 652.

Wolff, P. (2008). Dynamics and the perception of causal events. In T. Shipley & J. Zacks (Eds.), Understanding events: How humans see, represent, and act on events (pp. 555 - 587). Oxford University Press.

Wolff, P., Barbey, A. K., & Hausknecht, M. (2010). For want of a nail: How absences cause events. *Journal of Experimental Psychology: General*.

Wynn, K. (2008). Some innate foundations of social and moral cognition. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind: Foundations and the Future*. Oxford: Oxford University Press.

Young, L., Dungan, J. (2012). Where in the Brain is morality? Everywhere and maybe nowhere. *Social Neuroscience,* 7(1), 1-10.