# Contrasting Guilty Minds:
# Exposure to Contrast Concepts Narrows Conceptions of Acting Knowingly and Recklessly

Christian Mott[1] & Larisa Heiphetz Solomon[2]

[1] Department of Psychology, Columbia University. 1190 Amsterdam Ave., New York, NY 10027, United States. E-mail: christian.mott@columbia.edu. Phone: 212-853-1406. (Corresponding author)

[2] Department of Psychology, Columbia University. 1190 Amsterdam Ave., New York, NY 10027, United States. E-mail: lah2201@columbia.edu. Phone: 212-854-1348.

**Mott, C. J., & Solomon, L. H. (2026). Contrasting guilty minds: Exposure to contrast concepts narrows conceptions of acting knowingly and recklessly. *Journal of Experimental Psychology: Applied*, *32*, 192-211. doi: 10.1037/xap0000547**

Author Note: This work was supported, in part, by the John Templeton Foundation [grant numbers 61080 and 61808], the NSF [CAREER grant 2044360], and Columbia University (all awards to LHS). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of any of these organizations. The authors wish to thank Joshua Knobe, participants at the Yale Experimental Jurisprudence Workshop, members of the audience at the 2022 SPP/ESPP joint meeting, members of the Yale Experimental Philosophy Lab, and members of the Columbia Social and Moral Cognition Lab for providing feedback. Data and analysis code for all studies are available at the following link: https://osf.io/mn32g/?view_only=399413ebc7284ab88cea1ddd3a70a9cd.

**Abstract**

When one person harms another, the way lay jurors describe the perpetrator's mental state – whether they acted “knowingly” or “recklessly” – can significantly affect their culpability under United States criminal law. Five studies conducted in Fall 2017 show that the meanings of these crucial legal terms can shift depending on whether the jury instructions mention an alternative mental state. In Studies 1-3, lay participants, acting as mock jurors, were less likely to say an agent caused a harm “knowingly” when they could instead describe the person as acting “recklessly” – a less severe but still culpable state of mind. This pattern emerged whether or not participants received legal definitions of these terms. In Study 4, mock jurors were less likely to say an agent acted “knowingly” when “recklessly” appeared in the jury instructions as a contrast, even when they did not have any way to attribute the contrast term to the agent. In Study 5, mock jurors were also less likely to say an agent acted “recklessly” when the possibility of acting “negligently” appeared in the jury instructions. These studies provide evidence that “knowingly” and “recklessly” are *contrast sensitive* – that is, their meanings can shift based on what contrast concept is salient.

**Keywords**: blame; experimental jurisprudence; law and psychology; moral psychology; social cognition

**Public Significance Statement**: The present research suggests that jurors' interpretations of key legal terms like "knowingly" and "recklessly" can vary based on which alternative mental states appear in the jury instructions. These findings have important implications for criminal trials in the United States, where jurors must determine a defendant's mental state to decide guilt. This inconsistency could potentially lead to uneven application of justice across similar cases. These results highlight an urgent need to clarify which meaning of these mental state terms jurors should use when making life-altering decisions in U.S. criminal cases.

**Contrasting Guilty Minds:**
**Exposure to Contrast Concepts Narrows Conceptions of Acting Knowingly and Recklessly**

Two campers in a dry forest built a campfire that ignited a forest fire, causing millions of dollars of property damage (United States Attorney's Office for the Central District of California, 2014). A go-cart race driver squeezing through a tight space between another car and the side of the track clipped the car, causing an accident that killed the driver (*Commonwealth v. Sostilio*, 1949). A person fleeing from the police in a stolen car weaved across a road's centerline and collided headfirst with an oncoming car, killing the driver and a passenger (*Commonwealth v. Subenko*, 2001). In all these cases, an agent or agents (the campers, the drivers) caused a harm (the property damage, the death(s)). How much punishment they deserved for causing that harm depended, to a substantial degree, on what was going on in their minds when they performed the harmful behavior – at least, according to United States criminal law.[1] The agents would be most culpable if they acted with the goal of bringing about the harm – that is, if they caused the harm *intentionally*. The actors would still be highly culpable, but less so, if they perceived the harm as certain but had some other goal in mind – that is, if they caused the harm *knowingly*. The actors would be less culpable still if they pursued another goal and thought the harm was substantially likely rather than certain – that is, if they caused the harm *recklessly*. The actors would be even less culpable if they genuinely but unreasonably did not believe they would cause a harm – that is, if they caused the harm *negligently* (American Law Institute (ALI), 1962, § 2.02; Shen et al.,

[1]The United States contains more than 52 different bodies of criminal law, one for each jurisdiction within it that has the power to enact such laws. When discussing United States criminal law in this paper, we refer either to federal criminal law or to the Model Penal Code, a model set of criminal laws drafted by the American Law Institute in 1962. Around two-thirds of states have adopted at least parts of the code into their statutory law (Robinson & Dubber, 2007), and the United States Supreme Court often cites the Model Penal Code when determining the mental state requirements of federal law (e.g., *Elonis v. United States*, 2015).

2011). In this paper, we investigate whether the meanings of these mental state terms are *contrast sensitive* – that is, whether terms like "knowingly" and "recklessly" have different meanings to laypeople depending on what other mental state terms are salient as contrasts.

In criminal cases in the United States, lay jurors generally must make a yes-or-no judgment about whether the event described by the evidence falls into whichever one of these four culpable mental states is an element of the charged crime (Beatty & Fondacaro, 2018; Macleod, 2016). Determining how laypeople understand and apply these concepts therefore has substantial practical importance (Sommers, 2021). However, prior research suggests that lay jurors may have difficulty understanding the mental state concepts used in the criminal law (Devine, 2012; Shen et al., 2011).

Understanding laypeople's mental state concepts also has substantial theoretical importance. There is an extensive literature in both philosophy and psychology investigating the concept of acting intentionally (e.g., Austin, 1966; Cova et al., 2012; Cova, 2016; Knobe, 2003; Mele & Moser, 1994; Moore, 1997; Quillien & German, 2021; Tobia, 2024), and an emerging literature on the concepts of acting knowingly (e.g., Ginther et al., 2014; Nadelhoffer, 2006; Shen et al., 2011), acting recklessly (e.g., Flick et al., 2024; Kneer & Skoczeń, 2023), and acting negligently (e.g., Nobes & Martin, 2022). All this attention is well-deserved: Judgments about mental states play an essential role in social life, including in judgments of blame and punishment (Cushman, 2008; Cushman et al., 2013; Flick et al., 2024; Young & Saxe, 2008, 2011; for reviews, see Cushman, 2015; Heiphetz & Young, 2014).

There are often at least two different ways to think about mental state concepts, as there are for many other concepts (Austin, 1962). First, one can define these concepts positively by attempting to determine what features of an agent-outcome pair make it appropriate to say that

the agent brought about the outcome "intentionally," "knowingly," "recklessly," or "negligently" (e.g., Flick et al., 2024; Kneer & Skoczeń, 2023; Malle & Knobe, 1997; Mele & Moser, 1994; Moore 2010; Nadelhoffer, 2006). This is the most common approach in philosophical and psychological investigations into these concepts.

Second, one can define these concepts by contrasting them with other mental states. Most commonly, the contrast mental state is an antonym (Austin, 1966; Hart, 1967). The legal philosopher H.L.A. Hart (1967, p. 121) provides a useful example for "intentionally" and "knowingly," considering a case where the action of an agent (named Barrett) caused a foreseen side effect:

> [In cases] where the harmful consequences are merely a foreseen but unwanted byproduct of action we would not say that the agent did what he did *un*intentionally: it would be most misleading to say that Barrett unintentionally killed the victims, if he realized that they would be killed . . . because that would suggest that [he] had not even foreseen the outcome, and had brought it about accidentally or through some mistake. In fact, it is impossible to squeeze such cases into the dichotomy of intentionally and unintentionally, and we must turn to some other expression like 'knowingly' to characterize them.

In other words, in Hart's view "knowingly" means something different from both "intentionally" and "unintentionally." Likewise, "intentionally" means something different from both "knowingly" and "unintentionally."

An implication of Hart's analysis of "intentionally" is that the negation of "intentionally" ("not intentionally" or "non-intentionally") encompasses at least two different ways of acting: acting "unintentionally" and acting "knowingly." In contexts where only one of these two ways of acting non-intentionally is salient, it is plausible that yes-or-no judgments about whether an agent brought about an outcome "intentionally" may differ. If that is true, then the meaning of "intentionally" would "shift[] depending on what it is contrasted with" (Gardner, 2012, p. 11). As noted above, we call terms with this feature "contrast sensitive."

Cova and colleagues (2012) have argued that "intentionally" is contrast sensitive and that fact explains a wide range of empirical findings on laypeople's intentionality attributions. On their view, "intentionally" has at least three different meanings, depending on what conceptual contrast the context makes relevant – i.e., depending on whether it is contrasted with (1) "carelessly," (2) "unwillingly," or (3) "accidentally" (p. 388). Although people may use one of these contrasts by default, features of a scenario – including what attitudes and actions are normative in the scenario – may make a different contrast salient (Cova, 2016).

The proposal that "intentionally" is contrast sensitive raises the same question about other culpable mental states, like "knowingly" and "recklessly." This paper investigates whether those two terms are also contrast sensitive. Such a finding would have substantial practical importance, as in United States criminal law the distinctions between "knowingly" and "recklessly" or "recklessly" and "negligently" often significantly affect both culpability (i.e., degree or class of crime) and punishment (American Law Institute (ALI), 1985). This paper presents four studies showing that "knowingly" is contrast sensitive and one study showing the same for "recklessly."

## Is "Knowingly" Contrast Sensitive?

Even if "intentionally" is contrast sensitive (Cova et al., 2012), one might think that other culpable mental states, like "knowingly" and "recklessly," are not. On this view, "intentionally" has some special feature that other mental state terms lack, and it is this feature that makes it contrast sensitive. We consider three features that appear most closely related to contrast sensitivity and conclude that all apply to "knowingly" as well.

First, the underlying concepts of acting intentionally are complex, involving (at the very least) both motivational and cognitive components (e.g., belief and desire; Cova et al., 2012; Davis, 1984; Jara-Ettinger, 2019; Jara-Ettinger et al., 2016; Sinhababu, 2013) and, on some

accounts, a variety of additional conditions (Bratman, 1984; Yaffe, 2011). This multi-component structure of intentionality may contribute to the contrast sensitivity of "intentionally" because the components may fit together in different ways or may themselves have multiple different meanings. For example, the philosophical literature describes two different forms of desire (Davis, 1986; Tooming, 2019), either of which may contribute to intention.

Second, laypeople's concept of acting intentionally also appears to have an evaluative component, which may contribute to its contrast-sensitivity. One of the ways this evaluative component manifests is that people more readily say that an agent brought about a harmful, versus helpful, side effect "intentionally" (Adams & Steadman, 2007; Guglielmo & Malle, 2010; Knobe, 2003, 2004, 2010; Lindauer & Southwood, 2021; Nichols & Ulatowski, 2007). The same is true for more versus less harmful side effects (Kneer & Bourgeois-Gironde, 2017). Cova and colleagues (2012) explain these effects by arguing that evaluative expectations play a key role in making different meanings of "intentionally" salient, which is why the word is contrast-sensitive.[2]

Finally, "intentionally" has long had a wide variety of different meanings. In common law legal systems, the term has meant acting with any culpable *mens rea* (Kadish et al., 2012; *Regina v. Cunningham*, 1957), acting with foreknowledge (Moore, 1997; *United States v. United States Gypsum Co.*, 1978), or acting with a specific or direct intent to bring about the criminalized harm (Hart, 1967; *People v. Hood*, 1969). The contrast-sensitivity of "intentionally"

[2] Laurent and colleagues (2019, 2021) have argued that the effect of morality on intentionality attributions does not arise from participants understanding "intentionally" to have different meanings in different contexts, but from contextual effects on what participants understand "intentionally" to modify in the questions used in the studies showing these effects. On that account, "intentionally" would not be context-sensitive, since the word itself would not have multiple potential meanings. The evidence this paper provides for the claim that "knowingly" and "recklessly" are context-sensitive are not subject to the same objections, however.

may arise from the long historical development of the term, during which time it may have accumulated many different meanings.

These features, which appear plausibly related to the contrast-sensitivity of "intentionally," are not distinctive to "intentionally"; they are also true of "knowingly."[3] First, knowledge, like intention, has multiple components, some of which are themselves complex. On classic philosophical accounts, for example, knowledge is something like justified true belief, perhaps with additional conditions to rule out cases in which the justification is coincidental (Ichikawa & Steup, 2018). Empirical evidence suggests that laypeople think about knowledge along the same lines (Machery et al., 2017; Starmans & Friedman, 2012, 2020). Whether a belief counts as knowledge may also depend on additional components, such as stakes (DeRose, 2009; Francis & Beaman, 2023; Porter et al., 2024) and whether the agent's evidence is sufficient for action (Turri et al., 2016). Several of components of knowledge may themselves take different forms depending on context, such as belief (e.g., "thick" and "thin" belief; Buckwalter et al., 2015).

Second, a large body of research has shown that knowledge attributions also have an evaluative component, such that people are more likely to say that an agent knew that a bad consequence would occur than a good consequence, on equivalent evidence (Beebe &

[3] Note that Schaffer and Knobe (2012) argue that knowledge is *contrastive*, which is different than the contrast-sensitivity for which we argue here. They propose that people evaluate knowledge statements by contrasting the propositional content of the knowledge (i.e., the fact potentially known) with a relevant counterfactual – i.e., people evaluate whether the agent knew proposition A as compared to proposition B (Gerken, 2017). They do not propose that people evaluate statements about whether an agent caused an outcome "knowingly" by comparing acting "knowingly" to acting with some other possible mental state – i.e., by evaluating whether the agent knowingly, as opposed to some other mental state, caused outcome C. Indeed, Knobe has agreed that the contrast sensitivity of "knowingly," which this paper addresses, is distinct from the contrastivity of knowledge, which Knobe and Schaffer investigated (personal communication, 2024).

Buckwalter, 2010; Beebe & Jensen, 2012; Maćkiewicz, 2022; Turri, 2014). Although much of this research has addressed knowledge attributions (i.e., judgments about whether an agent "knew" that a consequence would arise), a similar effect appears for judgments about whether someone caused an outcome "knowingly" (Nadelhoffer, 2006). Plausibly, evaluative expectations that vary with context give rise to these effects, in the same way they may give rise to the parallel effects for "intentionally." If so, then "knowingly" is likely contrast-sensitive.

Third, "knowingly" also has multiple different legal definitions (Macleod, 2016). For example, in the criminal context, "knowingly" can include forms of constructive knowledge, like willful ignorance (ALI, 1962, § 2.02(7); *Global-Tech Appliances, Inc. v. SEB SA*, 2011). In addition, federal courts sometimes conflate acting "knowingly" with the common law *mens rea* category of general intent (*Carter v. United States*, 2000; Johnson, 2016).

In addition, "to know" has multiple possible non-technical meanings in common usage, which contrasts can trigger. As a result, researchers studying knowledge attributions often ask participants whether a particular character "knows" some proposition or (by contrast) "only thinks [he/she] knows" (Buckwalter, 2014; Rose et al., 2017), "feels like she knows" (Nagel et al., 2013), "only believes" (Machery et al., 2017; Starmans & Friedman, 2012; Weinberg et al., 2001; for discussion see Cullen, 2010), or "only thinks" (Starmans & Friedman, 2012, 2020). This question wording is necessary because laypeople may otherwise use "know" more loosely or engage in protagonist projection – i.e., assessing knowledge claim by taking the perspective of the agent (Buckwalter, 2014; Nagel et al., 2013; Rose et al., 2017; Starmans & Friedman, 2020) – which can produce different responses than assessing knowledge from the participant's own perspective (Rose et al., 2017; for a discussion see Beebe & Shea, 2013). That is, it is the meaning of "to know" that is triggered by contrast that has central theoretical importance.

Finally, “knowingly” also has multiple possible antonyms. Consider the antonym formed by adding “un-” to the beginning of the word to obtain “unknowingly.” The meaning of “unknowingly” is, roughly, “unwittingly” or “ignorantly” (Oxford University Press, n.d.). On the Model Penal Code definition, by contrast, a person knowingly brings about a harm by acting with awareness that there is a “high probability” or that it is “practically certain” that the harm will result (American Law Institute (ALI), 1962, § 2.02; Macleod, 2016). A person acts non-knowingly on this definition by acting without certainty or with awareness of a low probability, not by acting ignorantly.

Based on these arguments, we hypothesize that “knowingly” is also contrast sensitive, which raises a further question: Which other concept(s) can serve as a contrast? Contrast sensitivity does not imply that every other mental state term will work. Again, the term “intentionally” provides an illustration. Although Cova and colleagues (2012) propose that “intentionally” is contrast sensitive, they do not suggest that “knowingly” is a potential contrast concept. Contrary to Hart’s (1967) argument above, the empirical evidence suggests that people view acting knowingly as a way of acting intentionally, rather than a contrast to it (Adams & Steadman, 2007). For example, laypeople judge that agents *best* described as bringing about a side effect “knowingly” (Gugliemo & Malle, 2010) can still be *acceptably* described as doing so “intentionally” (Adams & Steadman, 2007; Knobe, 2003).

For the purposes of the studies reported here, we use “recklessly” as the contrast with “knowingly.” We do so for two reasons. First, that is the contrast built into the Model Penal Code’s mental state categories, where “recklessly” is the mental state term immediately below “knowingly” in culpability. Second, there is some evidence that “knowingly” can be used with a broad scope to mean something like “with knowledge of a probability” (Macleod, 2016),

suggesting that the contrast with recklessness would exclude an available meaning of "knowingly."

## Is "Recklessly" Contrast Sensitive?

Next, we consider whether "recklessly" is contrast sensitive. There is less research on the term (Flick et al., 2024), so there is less to say on either side.

Favoring the contrast sensitivity of "recklessly," the term has multiple legal meanings. The Model Penal Code defines it narrowly as bringing about a harmful result while "consciously disregard[ing] a substantial and unjustifiable risk" that the harm would occur (ALI, 1962, § 2.02). By contrast, Massachusetts criminal law – which still follows the common law in several respects – defines a similar term, "wanton and reckless," more broadly to include conduct that caused a harm where "an ordinary normal [person] under the same circumstances would have realized the gravity of the danger" the conduct created (*Commonwealth v. Pugh*, p. 496). In other words, the Model Penal Code requires actual awareness of a risk that one's conduct will cause harm, while the Massachusetts definition only requires that an ordinary person would recognize such a risk. The Massachusetts definition is closer to what the Model Penal Code calls negligence (ALI, 1962, § 2.02), as well as what laypeople sometimes think of as negligence – i.e., a failure of thought (Sarin & Cushman, 2024). The persistence of these two definitions of recklessness in the criminal law suggests that people may be able to understand the term in different ways. The overlap of one of those meanings with negligence also suggests that "recklessly" may be sensitive to the contrast with "negligently" in particular.

## Existing Research on Contrastive Concepts

In this paper, we investigate the contrast sensitivity of two specific concepts that play an important role in moral and legal judgments: "knowingly" and "recklessly." We do not propose

that contrast sensitivity is a general feature of all concepts. Indeed, prior work on related phenomena has produced mixed results.

Existing research has shown that some sets of concepts exhibit *contrast category effects*, which may be related to the contrast sensitivity we investigate here. Contrast category effects appear when the target concept and the contrast concept "are mutually exclusive terms organized under an inclusive covering term" (Verheyen, De Deyne, et al., 2011, p. 1519). In these cases, the more an item has features of one of the concepts (the contrast concept), the less likely participants are to categorize it as the other concept (the target concept), regardless of whether the contrast concept is explicitly made salient. For example, participants asked to make categorization judgments about whether various items were *fruits* were less likely to endorse membership for fruits that shared more features with the category *vegetable* (Verheyen, De Deyne, et al., 2011), even though the experimenters did not draw any explicit contrast with vegetables in their questioning.

This phenomenon does not extend uniformly to concepts with different structures or that pick out more abstract categories. For example, participants who made categorization judgments about whether items belonged to the categories *art forms*, *crimes*, *diseases*, *emotions*, *media*, *sciences*, and *virtues* showed an influence of contrast category features only for two of the categories, *diseases* and *virtues* (Verheyen, Stukken, et al., 2011). Based on this research, it is unclear what exactly gives rise to contrast category effects in abstract concepts.

This existing research therefore does not make a clear prediction about whether "knowingly" and "recklessly" would show contrast category effects. These concepts are unlike the *fruits-vegetables* pair in that they are not "mutually exclusive terms organized under an inclusive covering term" (Verheyen, De Deyne, et al., 2011, p. 1519). Instead, the law makes

"knowingly" a subset of "recklessly" (ALI, 1962, § 2.02). Existing research does not indicate whether laypeople think of these concepts the same way, though the studies reported in this paper shed light on the question. These concepts are also unlike *diseases* and *virtues*, which both have a single, obvious contrast concept (*health* and *vices*, respectively). As discussed above, "knowingly" and "recklessly" both have multiple possible antonyms.

Thus, the theoretical framework described in this existing literature suggests that "recklessly" and "knowingly" may not demonstrate contrast category effects. However, they may still exhibit contrast sensitivity – that is, a change in meaning depending on whether a specific contrast becomes salient. Indeed, one might posit that concepts exhibiting contrast category effects do so because the structure of the concept – i.e., the fact that the concept is one of a set of "mutually exclusive terms organized under an inclusive covering term" (Verheyen, De Deyne, et al., 2011, p. 1519) – makes contrast concept salient automatically. Thus, contrast sensitivity may only arise for concepts that *do not* exhibit contrast category effects.

## Relevance to Law

Understanding whether "knowingly" and "recklessly" are contrast sensitive also has practical significance. In United States criminal law, the distinction between acting intentionally and acting knowingly is "inconsequential for most purposes of liability" (ALI, 1985, pp. 269-274). As legal philosophers have observed, intentionally and knowingly causing harms generally gives rise to equal culpability (Hart, 1967; Moore, 1997). By contrast, the distinction between knowingly and recklessly causing a harm can be the difference between murder and manslaughter (Vilares et al., 2017), and the difference between recklessly and negligently causing a harm can be the difference between some criminal liability and none at all (Kadish et al., 2012; Moore, 1997). How lay jurors answer yes-or-no questions about whether a defendant

caused a harm "knowingly" or "recklessly" can therefore dictate whether or not that defendant spends decades in prison.

For that reason, understanding the factors that shape lay jurors' understanding and application of terms like "knowingly" and "recklessly" could make a difference to legal practice (Sommers, 2021). If these terms are contrast sensitive, then telling jurors that "knowingly" contrasts with "recklessly" rather than "unknowingly" (a potential default contrast) and that "recklessly" contrasts with "negligently" rather than "accidentally" (another potential default contrast) could result in jurors using meaningfully narrower versions of these concepts. Such a finding would be particularly significant because interventions to change juror concepts by providing them with more precise definitions of mental state terms tend to have small or null effects (Devine, 2012; Ginther et al., 2014). Moreover, legal practitioners could make use of such a finding without awaiting a change in the law: In real criminal cases, judges decide on jury instructions based on the proposals and arguments of both parties. Parties generally have latitude to propose, and judges to accept, instructions that help clarify the relevant legal concepts (e.g., *McDowell v. Calderon*, 1997).

There is also reason to think that the narrower concepts jurors might use if informed about the relevant contrast are the concepts used by other legal actors. A jury trial is, fundamentally, an opportunity for the state to ask six to twelve members of the community to answer a question. In a second-degree homicide case, for example, that question is: "Did the defendant cause the death of the victim knowingly?" To answer this question, jurors need to understand the key concepts appearing in that question the same way the state does. The state – whether conceptualized as the legislators who wrote the laws, the prosecutors who decided what charges to bring, or the judges running the criminal trial – is aware of the full range of criminal

mental states, including the relevant contrasts. As one would predict if mental state terms are contrast sensitive, the evidence suggests that people with legal education apply narrower mental state concepts than those without (Tobia, 2024), perhaps because they are aware of the contrasts that the law makes salient.

## The Current Studies

The present studies provide evidence that both "knowingly" and "recklessly" are contrast sensitive. Studies 1a and 1b demonstrated a difference in attributions of "knowingly" between a yes-or-no question and a multiple-choice question (Study 1a) or a combination of a yes-or-no knowledge attribution and a yes-or-no recklessness attribution (Study 1b). Study 2 replicated Study 1b and extended its finding to two conditions in which the participants did not receive definitions of "knowingly" and "recklessly," implying that laypeople's ordinary understandings of these terms are contrast sensitive. Study 3 showed a difference in yes-or-no attributions of "knowingly" between a condition where participants only heard about knowledge and a condition where they also heard about recklessness but could not attribute it, suggesting that motivational and pragmatic factors do not explain the effect observed in Studies 1-2. Finally, Study 4 demonstrated that the effect observed in Study 3 extends to attributions of "recklessly," when contrasted with "negligently." Together, these studies provide evidence that both "knowingly" and "recklessly" are contrast sensitive.

## Study 1

This first study investigated whether laypeople give different judgments about whether an agent caused a harm "knowingly" depending on the phrasing of the knowledge question, even when the facts of the scenario do not change. In Study 1a, we compared responses to a yes-or-no question with responses to a multiple-choice question. Prior research on "intentionally" has

shown large differences in mental state attributions between similar conditions (Guglielmo & Malle, 2010; Mueller et al., 2012). In Study 1b, we compared responses to a yes-or-no question about whether the agent acted "knowingly" with responses to two separate yes-or-no questions, one about "knowingly" and one about "recklessly." These response options do not force a contrast between "knowingly" and "recklessly" – unlike in Study 1a, where the choices are exclusive – and therefore help shed light on whether the concept itself is contrast-sensitive.

To elicit these judgments, we asked participants to imagine themselves as jurors in a criminal case, described the admitted facts of the case, and told them that the only outstanding issue was the mental state of the defendant. In both conditions, participants learned that conviction for the charged crime required that the defendant have acted knowingly. This stipulation controlled the expected effect of the attribution across conditions.

The facts of the vignette participants read were a simplified version of the facts of a real Massachusetts case, *Commonwealth v. Subenko* (2001), which upheld a conviction for reckless homicide.[4] The facts did not include any explicit characterization of the defendant's mental states as to the death he caused. Instead, they included only neutral descriptions of the defendant's observable behavior that caused the death. Participants therefore had to draw mental state inferences entirely from those neutral descriptions. The absence of strong, direct mental state evidence – such as the agent's statements about their own mental states in the side-effect effect paradigm (Knobe, 2003) – both made the vignette more realistic, as a majority of criminal defendants do not testify (Eisenberg & Hans, 2008), and made it more likely that participants

[4] As discussed above, Massachusetts does not follow the Model Penal Code and the recklessness standard used in the state, called "wanton and reckless," differs from the Model Penal Code standard. Nevertheless, the facts of this case would have been legally sufficient sustain a conviction for reckless homicide even under the Model Penal Code definition.

saw the vignette as a case that they could accurately describe as either knowing or reckless. We expected the contrast sensitivity of "knowingly" to make the biggest difference in such cases.

## Methods

### *Participants*

**Study 1a.** Prior research has detected a large difference between two similar conditions (involving "intentionally") in a within-subjects sample of 101 participants (Guglielmo & Malle, 2010). For this between-subjects study, we therefore recruited 213 adult participants on Amazon's Mechanical Turk.[5] Eleven participants completed the entire survey in less than one minute, as measured using the Duration variable recorded by Qualtrics, and were excluded or did not complete the survey, leaving an analysis set of 202 participants[6] ($M_{age}$ = 35 years; $SD_{age}$ = 10.8 years; 47% female, 52% male, 1% chose an option not otherwise listed). The sample's highest level of education was distributed as follows: 11% high school or GED; 33% some college or university; 42% college or university degree; 12% master's or professional degree; and 2% PhD. [7] Although we did not conduct an a priori power analysis when we performed this

[5] The following applies to all studies reported in this paper and the supplement: Informed consent was obtained from all participants at the beginning of the online surveys. These studies were conducted under an exemption approved by the Yale University Institutional Review Board. Samples recruited on Amazon's Mechanical Turk (mTurk) are generally more diverse than in-person samples recruited at universities and, historically, of similar quality (e.g., Buhrmester et al., 2011). Since summer 2018, however, researchers have detected a decline in quality of mTurk data, either due to an increase in bots or a decrease in the sincerity of responses (Ahler et al., 2020; Chmielewski & Kucker, 2020), though it remains possible to draw meaningful conclusions from these responses despite this increase in noise (Johnson & Ryan, 2020). All the studies reported in this paper were conducted in Fall 2017, before this decline in quality.

[6] Here and in all subsequent studies, except where noted, the patterns of results reported in the main text also emerged when analyzing data from all respondents.

[7] For all studies reported in this paper, we also collected demographics on whether participants had previous experience with the legal system, including whether they had served on a jury of any kind, served on a jury in a criminal case, participated in a criminal case in some other capacity (e.g., criminal defendant, clerk, attorney), and had obtained or were in the process of

study in Fall 2017, a sensitivity analysis conducted in G*Power 3.1, using an alpha level of 0.05, a power of 0.80, and features of the sample described below (i.e., a probability of attributing knowledge in the target-only condition if the null hypothesis is true equal to the probability of attributing knowledge in the contrast condition), shows that this study could reliability detect effects larger than odds ratio 2.54 (for the binary outcome) and effect size $d = 0.40$ (for the Likert scale outcome).

**Study 1b.** We recruited 216 adult participants on Amazon's Mechanical Turk. Eighteen participants were excluded because they completed the entire survey in less than one minute, as measured using the Duration variable recorded by Qualtrics, or did not complete the survey, leaving an analysis set of 198 participants ($M_{age} = 35$ years; $SD_{age} = 11$ years; 51% female, 48% male, 1% did not answer or chose an answer not otherwise listed). The sample's highest level of education was distributed as follows: 12% high school or GED; 32% some college or university; 43% college or university degree; 11% master's or professional degree; and 2% PhD. A sensitivity analysis conducted in G*Power 3.1, using an alpha level of 0.05, a power of 0.80, and features of the sample described below, shows that this study could reliability detect effects larger than odds ratio 2.26 (for the binary outcome) and effect size $d = 0.40$ (for the Likert scale outcome).

---

obtaining a law degree. However, including variables reflecting the number of juries on which participants had served, the amount of legal education they had received, and whether they had been participants in criminal cases as covariates in the models run to analyze the data in this study and all other studies made no meaningful difference to the effect of condition in any study. As a result, we do not report these demographics for the studies, though the variables remain available in the public data (Mott, 2023).

***Procedure***

Participants read one of two versions of the following vignette (bracketed and underlined sections appeared in the contrast condition only):

> Imagine that you are a juror in a criminal case. The defendant, Joseph, has been charged with second-degree murder. You have heard evidence that establishes the following:
>
> One day last year, Joseph was in a public parking lot in a medium-sized city stealing a car. He had already opened the door and was working on starting the ignition, when a police officer saw him, realized what he was doing, and ordered him to exit the vehicle. Instead, Joseph started the car and drove out of the parking lot onto the street. The police officer got into his own vehicle and pursued Joseph. Joseph increased his speed, and the police officer did as well. Soon, multiple police cars were chasing Joseph down a highway, and Joseph was traveling around ninety miles per hour. It was the middle of the day, and there were few cars on the road. At a break in the highway divider, Joseph crossed on to the other side of the highway, so that he was driving the wrong way down the road. Coming around a curve in the road, he collided head-on with a vehicle driven by a man named Chris. Chris died in the accident; Joseph survived.
>
> The judge instructs you that in your jurisdiction, a defendant is guilty of second-degree murder if he knowingly caused the death of another human being. Joseph has conceded that his actions caused Chris's death. The only question before the jury, therefore, is whether Joseph knowingly caused Chris's death.
>
> The judge instructs you that a person knowingly causes a result if the person is aware that it is practically certain that his conduct will cause such a result.
>
> [The judge instructs you that a person recklessly causes a result if he consciously disregards a substantial and unjustifiable risk that his conduct will cause such a result.]
>
> Finally, the judge instructs you that if you find that Joseph knowingly caused Chris's death, then Joseph will be guilty of second-degree murder. If you do not find that Joseph knowingly caused Chris's death, then[, even if Joseph recklessly caused Chris's death,] Joseph will not be guilty of second-degree murder.

Participants then received the mental state attribution questions. In Study 1a, participants in the target-only condition selected one answer to the following question:

> In what way did Joseph cause Chris's death?
> - Knowingly

- Not knowingly

Participants in the contrast condition received the same question but with different response options, of which they could only select one:

In what way did Joseph cause Chris's death?
- Knowingly
- Recklessly
- Neither knowingly nor recklessly

In Study 1b, participants in the target-only condition selected one answer to the following question:

Did Joseph knowingly cause Chris's death?
- Yes
- No

Participants in the contrast condition received both questions and response options listed below. The instructions did not tell participants how they should or could answer these questions, and they were not prevented from answering "yes" or "no" to both.

Did Joseph knowingly cause Chris's death?
- Yes
- No

Did Joseph recklessly cause Chris's death?
- Yes
- No

In both studies, all participants then advanced to a new page and rated on a scale from 1 ("completely disagree") to 7 ("completely agree") the extent to which they agreed with the statement that Joseph caused Chris's death knowingly. Participants in the contrast condition also rated on a 7-point scale the extent to which they agreed with the statement that Joseph caused Chris's death recklessly.

### *Transparency and Openness*

The following applies to all studies described in this paper and the Supplemental Materials: We report all data exclusions, all manipulations, and all measures in the study. All data and analysis code are available at https://osf.io/mn32g/?view_only=399413ebc7284ab88cea1ddd3a70a9cd (Mott, 2023). The text of the study materials and measures are set out in the methods section. Data were analyzed using R, version 4.2.2 (R Core Team, 2022) and visualized using *ggplot2* (Wickham, 2016).

The studies' design and analysis were not pre-registered, as we completed this work before pre-registration was common in psychology. However, the simple design of the studies and the small number of dependent variables collected, all of which are reported, make the planned analyses clear. In addition, we did not perform any analyses until all data for a given study were collected.

## Results

### *Study 1a*

Responses to the categorical outcome variable showed a significant difference between conditions. In the target-only condition, 73% of participants said that Joseph caused the death knowingly and 27% said he did not. In the contrast condition, 17% of participants said that Joseph caused the death knowingly, 80% said he caused the death recklessly, and 3% said neither. A binary logistic regression of condition (target only = 1; contrast = 0) on the knowledge attribution question (attributing knowledge = 1; any other response = 0) showed a significant effect of condition on knowledge attributions, $z(200) = 7.39$, $p < 0.001$. The coefficient on the condition variable was 2.56 (95% CI: 1.90 to 3.26), yielding an odds ratio of 12.92 (95% CI: 6.69 to 26.15). This effect size exceeded the output of the sensitivity analyses.

Responses on the 7-point Likert-type scale to the question about whether Joseph caused the death knowingly also showed a significant difference. An independent samples t-test comparing mean responses in the target-only condition ($m$ = 5.18, $sd$ = 1.74, 95% CI: 4.84 to 5.52) to mean responses in the contrast condition ($m$ = 3.73, $sd$ = 1.82, 95% CI: 3.36 to 4.09) showed a significant effect of condition, $t(200) = 5.82$; $p < 0.001$, $d = 0.82$ (95% CI: 0.52 to 1.11). This effect size exceeded the output of the sensitivity analyses.

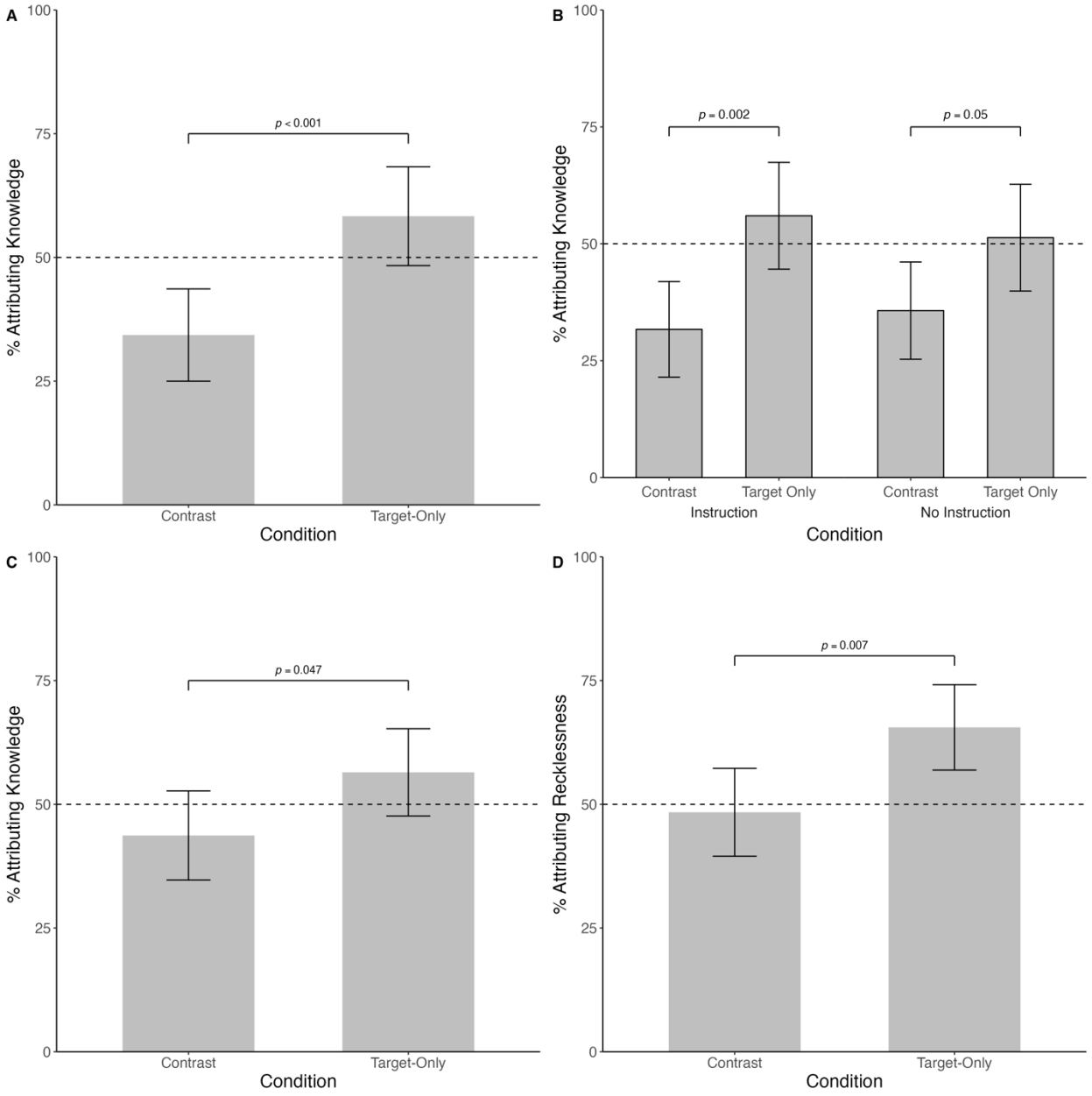
***Study 1b***

Responses to the categorical outcome variable showed a significant difference between conditions. In the target-only condition, 58% of participants (95% CI: 48% to 68%) said that Joseph caused the death knowingly. In the contrast condition, 34% of participants (95% CI: 25% to 44%) attributed knowledge, while 94% of participants (95% CI: 89% to 99%) attributed recklessness. In addition, in the contrast condition, 86% (95% CI: 74% to 98%) of participants who attributed knowledge also attributed recklessness. The percentage of participants attributing knowledge in each condition appear in Figure 1A. A binary logistic regression of condition (target only = 1; contrast = 0) on the knowledge attribution question (yes = 1; no = 0) showed a significant effect of condition on knowledge attributions, $z(196) = 3.36$, $p < 0.001$. The coefficient on the condition variable was 0.99 (95% CI: 0.42 to 1.57), giving an odds ratio of 2.68, and the coefficient on the intercept was -0.65 (95% CI: -1.07 to -0.25). Based on these estimates, moving a participant from the target-only condition to the contrast condition would decrease their probability of attributing knowledge by 24 percentage points, from 58% to 34%.[8] The size of this effect exceeded the output of the relevant sensitivity analysis.

[8] To put it another way, assuming independent probabilities (i.e., no effect of deliberation), there is a 4% chance that in a randomly chosen jury containing six people, all six would attribute

**Figure 1**

*Percentage of participants attributing knowledge in Studies 1b-4*




knowledge in the target-only condition and a 0.1% chance that all six would attribute knowledge in the contrast condition.

*Note*. Percentage of participants attributing knowledge in (A) Study 1b, (B) Study 2, and (C) Study 3. (D) Percentage of participants attributing recklessness in Study 4. Error bars show 95% confidence intervals.

Responses on the 7-point Likert-type scale to the question about whether Joseph caused the death knowingly also showed a significant difference. An independent-samples t-test comparing mean responses in the target-only condition ($m$ = 4.42, $sd$ = 2.21, 95% CI: 3.97 to 4.86) to mean responses in the contrast condition ($m$ = 3.62, $sd$ = 2.06, 95% CI: 3.21 to 4.02) showed a significant effect of condition, $t(196) = 2.63$; $p = 0.009$, $d = 0.37$ (95% CI: 0.09 to 0.65). The size of this effect was similar in magnitude to the output of the relevant sensitivity analysis. The average responses to the knowledge question on the Likert-style 1-to-7 scale appear in Figure 2A.

**Discussion**

Studies 1a and 1b provided initial evidence that "knowingly" is contrast sensitive. In Study 1a, a substantial majority (73%) of participants in the target-only condition thought that "knowingly" was an acceptable way to describe Joseph's action, but an even larger majority (80%) of participants in the contrast condition preferred "recklessly." This scenario therefore shows the same pattern for "knowingly" that was previously shown for "intentionally" (Guglielmo & Malle, 2010; Mueller et al., 2012). This finding is consistent with contrast-sensitivity.

However, there are alternative explanations for the results of Study 1a. One explanation is that the difference in participants' responses reflects a difference between the calculations used to answer yes-or-no questions and multiple-choice questions, rather than a difference in the

concept to which "knowingly" refers in the two conditions. Computational models of categorization judgments posit that answering a yes-or-no question about whether an item falls into a category involves determining whether the item's similarity to the category exceeds some threshold, while answering a multiple choice question about which category is the best fit for an item involves determining which category out of the available options is most similar to the item (Hampton, 2007; Minda & Smith, 2011; Minda et al., 2024; Murphy, 2002; Nosofsky, 2011; Verheyen et al., 2010; Verheyen et al., 2011).

Study 1b provided evidence against this alternative explanation. Participants still responded that the agent acted "knowingly" in the target-only condition significantly more than in the contrast condition, even when they answered yes-or-no questions in both conditions, which should rest on the same calculations. Moreover, the data from Study 1b suggest that participants understood that the "knowingly" and "recklessly" questions were not mutually exclusive: Of the 35 participants who attributed knowledge in the contrast condition, 30 (86%) also attributed recklessness. That is, participants in the contrast condition demonstrated awareness that they could attribute both mental states and were not supposed to choose a single best option.

There is, however, another alternative explanation for the findings of Studies 1a and 1b: The difference between conditions could reflect something particular to the legal definitions of "knowingly" and "recklessly," rather than participants' ordinary understandings of those concepts. This explanation has a facial plausibility: The legal definitions of "knowingly" and "recklessly" are similar to one another, which could cause participants to represent these concepts more similarly than they would by default. Moreover, existing research shows that laypeople assign different meanings to "knows" and "is certain" (Turri et al., 2017), which may

mean that the legal definition of "knowingly" – defined as "practically certain" – causes participants to use a different mental state concept than they otherwise would. In that case, these results might say something about how jurors reason after receiving legal definitions but nothing about whether "knowingly" is contrast-sensitive in general or about how jurors reason in the circuits that provide no instruction on the meaning of "knowingly" (Macleod, 2016). The next study investigated whether the presence of the legal definitions of these mental state concepts was necessary for the effect to emerge.

## Study 2

This study replicated and extended Study 1b by varying the stimuli along another dimension: Whether or not participants received legal definitions of "knowingly" and "recklessly." The study therefore investigated whether the effect observed in Study 1b was specific to the mental state concepts' legal definitions or a more general fact about the way people understand the term "knowingly" and its contrast, "recklessly."

### Methods

#### *Participants*

We recruited 338 adult participants on Amazon's Mechanical Turk. Twenty-one participants were excluded because they completed the entire survey in less than one minute, as measured using the Duration variable recorded by Qualtrics, or failed to complete the study, leaving an analysis set of 317 participants ($M_{age}$ = 34 years; $SD_{age}$ = 9.3 years; 43% female, 57% male). The highest level of education for the sample was distributed as follows: 9% high school or GED; 27% some college or university; 49% college or university degree; 13% master's or professional degree; and 1% PhD. A sensitivity analysis conducted in G*Power 3.1, using an alpha level of 0.05, a power of 0.80, and features of the sample described below, shows that this

study could reliability detect effects larger than odds ratio 1.91 (for the binary outcome) and effect size partial $\eta^2 = 0.024$ (for the Likert scale outcome).

***Procedure***

Participants were randomly assigned to one of four groups, using a 2 (mental states: target-only vs. contrast) x 2 (instructions: instructions vs. no instructions) between-subjects design. All four groups received the preamble and vignette used in Studies 1a and 1b. The two groups in the instructions condition saw jury instructions identical to the target-only and contrast conditions in Studies 1a and 1b. The other two groups saw the same preamble and fact summary, followed by one version of the following paragraphs (with bracketed and underlined sections appearing only in the contrast-and-no-definition condition), which differ from the materials for Studies 1a and 1b only by omitting the legal definitions of the mental states:

> The judge instructs you that in your jurisdiction, a defendant is guilty of second-degree murder if he knowingly caused the death of another human being. Joseph has conceded that his actions caused Chris's death. The only question before the jury, therefore, is whether Joseph knowingly caused Chris's death.
>
> Finally, the judge instructs you that if you find that Joseph knowingly caused Chris's death, then Joseph will be guilty of second-degree murder. If you do not find that Joseph knowingly caused Chris's death, then[, even if Joseph recklessly caused Chris's death,] Joseph will not be guilty of second-degree murder.

The two target-only groups then responded to the following question:

> Did Joseph knowingly cause Chris's death?
> - Yes
> - No

The two contrast groups instead responded to the following questions:

> Did Joseph knowingly cause Chris's death?
> - Yes
> - No
>
> Did Joseph recklessly cause Chris's death?
> - Yes

- No

After answering these questions, all groups then advanced to a new page where they indicated their level of agreement on a 7-point scale with the statement that Joseph knowingly caused Chris's death. Participants in the knowledge and recklessness condition also rated on a 7-point scale the extent to which they agreed with the statement that Joseph recklessly caused Chris's death.

**Results**

Responses to the categorical outcome variable showed a significant difference between conditions. In the target-only condition, 54% of participants (95% CI: 46% to 62%) said that Joseph caused the death knowingly; in the contrast condition, only 34% of participants (95% CI: 27% to 41%) attributed knowledge. These percentages did not differ substantially based on instructions condition. The percentages of participants attributing knowledge by condition appear in Figure 1B. A binary logistic regression of mental states condition (target only = 0.5; contrast = -0.5), instruction condition (instruction provided = 0.5; no instruction = -0.5), and their interaction on the knowledge attribution question (yes = 1; no = 0) showed a significant effect of condition on knowledge attributions, $z(313) = 3.56$, $p < 0.001$, but did not show a significant effect of instruction condition, $z(313) = 0.02$, $p = 0.984$, or a significant interaction, $z(313) = 0.79$, $p = 0.427$. The coefficient on the mental state condition variable was 0.82 (95% CI: 0.37 to 1.28), giving an odds ratio of 2.28, and the coefficient on the intercept was -0.27 (95% CI: -0.49 to -0.04), giving an odds ratio of 1.00. Based on these estimates, moving a group of participants, evenly split between the instruction and no-instruction conditions, from the target-only condition to the contrast condition would decrease their probability of attributing knowledge by 20

percentage points, from 54% to 34%.[9] The size of the effect of the knowledge condition exceeded the output of the relevant sensitivity analysis.

A significant effect of knowledge condition also appeared within each instruction condition. A logistic regression fitted on just the participants who received the legal definitions of the mental state terms showed a significant effect of mental state condition, $z(155) = 3.04$, $p = 0.002$, as did a model fitted just on the participants who did not receive the definitions, $z(158) = 1.98$, $p = 0.048$.

To confirm that the effect of instruction condition was too small to be practically important, we also ran an equivalence test on a logistic regression model containing both condition variables as predictors and no interaction. This model, like the first model reported above, showed a significant effect of mental state condition on knowledge attributions, $z(314) = 3.55$, $p < 0.001$, and did not show a significant effect of instruction condition, $z(314) = 0.02$, $p = 0.981$. The coefficient on the instruction condition was less than 0.01 and the intercept was -0.26. That is, given a group of participants, half of whom were in the target-only condition and half of whom were in the contrast condition (i.e., a group corresponding on average to a 0 value on a contrast-coded binary predictor for mental state condition), the model predicts that 43.5% would attribute knowledge if they received the instructions and 43.4% would do so if they did not. A change in the coefficient of plus-or-minus 0.4 corresponds to a plus-or-minus 5% chance of attributing knowledge for participants in both conditions. If the coefficient were 0.41, the model would predict that 48.5% of participants in the instruction condition and 38.5% of participants in

[9] To put it another way, assuming independent probabilities (i.e., no effect of deliberation), there is a 2% chance that in a randomly chosen jury containing six people, all six would attribute knowledge in the target-only condition and a 0.1% chance that all six would attribute knowledge in the contrast condition.

the no-instruction condition would attribute knowledge. If the coefficient were -0.39, the model would predict that 38.7% of participants in the instruction condition and 48.3% of participants in the no-instruction condition would attribute knowledge. We treated plus-or-minus 5% as the smallest practically significant effect, since it corresponds to a less than 1% difference in the probability that six independent jurors would all attribute knowledge. Based on an equivalence test conducted using the *marginaleffects* package (version 0.25.0; Arel-Bundock, et al., 2025), the coefficient on instruction condition in this model is greater than -0.39, $p = .040$, lower than 0.41, $p = .044$, and equivalent to 0 within these bounds, $p = .044$.

As in Study 1b, participants appeared to understand that acting "knowingly" and acting "recklessly" were not mutually exclusive. Out of all the participants in the contrast condition, including those who attributed knowledge, 96% (95% CI: 93% to 99%) attributed recklessness. Out of the 34% of participants in the contrast condition who attributed knowledge, 88% (95% CI: 79% to 96%) also attributed recklessness.

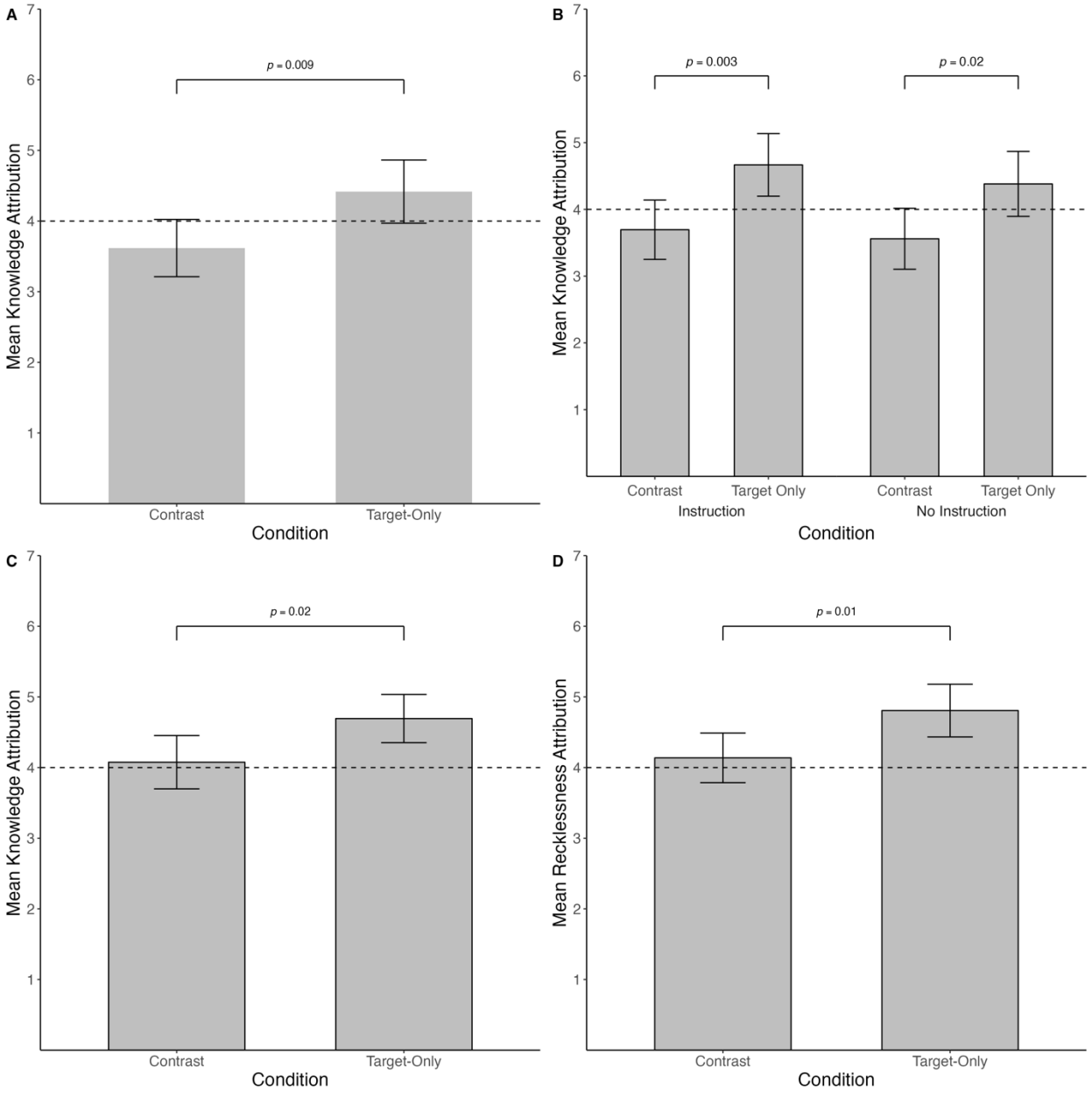
The responses to the 7-point Likert-type scale also showed an effect of the mental states condition but not the instruction condition. A 2 (Mental state condition: target-only vs. contrast) x 2 (Instruction condition: instruction vs. no instruction) between-participants ANOVA showed a significant main effect of mental state condition, $F(1, 313) = 14.80$, $p < 0.001$, partial $\eta^2 = 0.05$ (95% CI: 0.01 to 0.10), but did not show a significant main effect of instruction condition, $F(1, 313) = 0.81$, $p = 0.368$, partial $\eta^2 < 0.01$ (95% CI: 0 to 0.03), or a significant interaction, $F(1, 313) = 0.10$, $p = 0.749$, partial $\eta^2 < 0.01$ (95% CI: 0 to 0.02).[10] The size of the effect of the

[10] An ordinal linear regression, conducted using the *ordinal* package (version 2022.11-16; Christensen, 2015), also showed a significant effect of the mental state condition, $z(313) = 3.69$, $p < 0.001$, $\beta = 0.74$ (95% CI = 0.35 to 1.14), no effect of instruction condition, $z(313) = 0.98$, $p = 0.325$, $\beta = 0.19$ (95% CI = -0.19 to 0.58), and no significant effect of the interaction, $z(313) = 0.12$, $p = 0.901$, $\beta = 0.05$ (95% CI = -0.73 to 0.82).

knowledge condition exceeded the output of the relevant sensitivity analysis. The average responses to the Likert-style knowledge question appear in Figure 2B.

**Figure 2**

*Mean responses to Likert-style knowledge questions in Studies 1b-4*




*Note*. Mean knowledge attributions on ordinal scale in (A) Study 1b, (B) Study 2, (C) Study 3, and (D) Study 4. Error bars show 95% confidence intervals.

**Discussion**

Study 1's results were consistent with "knowingly" being contrast sensitive but left open the possibility that this effect arose from the legal definition of "knowingly" in particular. Study 2 provided strong evidence against that alternative explanation, as a difference between conditions in attributions of "knowingly" appeared whether or not participants received legal definitions of "knowingly" and "recklessly."

This study therefore provided evidence that laypeople's default understanding of "knowingly," without using the legal definition, is contrast sensitive. One explanation for this finding is that the legal definition matches the way laypeople ordinarily understand "knowingly." However, the evidence that laypeople distinguish "to know" and "to be certain" makes this possibility less likely (Turri et al., 2017). A more likely explanation is that laypeople have difficulty understanding jury instructions, including mental state jury instructions (Devine, 2012; Ginther et al., 2014; Shen et al., 2011), and so participants reverted to their default definitions, even when they received the dictionary definitions.

Although Studies 1b and 2 provided evidence against two alternative explanations for the observed differences between the target-only and contrast conditions, another alternative explanation remains possible: Participants may have been motivated to say that Joseph acted "knowingly" in the target-only condition, even though they did not really believe "knowingly" was an accurate description. This motivational explanation takes two primary forms.

First, participants may have been motivated to punish Joseph for causing harm. The design of the studies reported in this paper makes this form of the motivational explanation implausible. In both the target-only and the contrast conditions, all the vignettes informed participants that *only attributing knowledge* would lead to conviction. Thus, the motivation to

punish the agent legally would have caused participants to attribute knowledge in both conditions and could not have given rise to the difference we observe between the target-only and contrast conditions.

Second, participants may have been motivated to communicate Joseph's culpability by describing him in a way that sounded culpable. On this explanation, participants did not really think Joseph acted knowingly in either condition. But in the target-only condition, when asked whether Joseph caused the death "knowingly," they did not want to answer "no," since that answer failed to describe Joseph in a way that reflected his culpability and may have implied that he was not blameworthy. However, in the contrast condition, participants could describe Joseph as having acted "recklessly," which they felt communicated his culpability to a sufficient degree. The next study tests whether this alternative explanation is correct.

## Study 3

The alternative explanation described at the end of the last section predicts that participants in the contrast condition agreed that Joseph acted "knowingly" less often only because they could apply the label "recklessly." This explanation therefore does not predict a difference between conditions if the contrast condition merely makes "recklessly" salient as an alternative to "knowingly" without allowing participants to attribute it. The contrast-sensitivity hypothesis, however, does predict that participants will reduce their attributions of "knowingly" whether or not they can attribute "recklessly" in the contrast condition, since merely making "recklessly" salient makes them interpret "knowingly" to refer to a narrower version of the concept. This study tests these competing predictions.

## Methods

### *Participants*

We recruited 254 adult participants on Amazon's Mechanical Turk. Eleven participants were excluded because they completed the entire survey in less than one minute, as measured using the Duration variable recorded by Qualtrics, or did not complete the survey, leaving an analysis set of 243 participants ($M_{age}$ = 34.7 years; $SD_{age}$ = 11.0 years; 45% female, 55% male). The sample's highest level of education was distributed as follows: 15% high school or GED; 23% some college or university; 49% college or university degree; 12% master's or professional degree; and 1% PhD. A sensitivity analysis conducted in G*Power 3.1, using an alpha level of 0.05, a power of 0.80, and features of the sample described below, shows that this study could reliability detect effects larger than odds ratio 2.07 (for the binary outcome) and effect size $d$ = 0.36 (for the Likert scale outcome).

### *Procedures*

Participants were randomly assigned to one of two conditions, which were identical to the two conditions in Studies 1a and 1b, except that participants in both the target-only and contrast conditions only answered a single question:

> Did Joseph knowingly cause Chris's death?
> - Yes
> - No

Participants in both conditions then advanced to a new page and used a 7-point scale to indicate how much they agreed with the statement that Joseph knowingly caused Chris's death.

## Results

Responses to the categorical outcome variable showed a significant difference between conditions. In the target-only condition, 57% of participants (95% CI: 48% to 65%) said that

Joseph caused the death knowingly. In the contrast condition, only 44% of participants (95% CI: 35% to 53%) attributed knowledge. The percentages of participants giving each response by condition appear in Figure 1C. A binary logistic regression of condition (target only = 1; contrast = 0) on the knowledge attribution question (yes = 1; no = 0) showed a significant effect of condition on knowledge attributions, $z(241) = 1.98$, $p = 0.047$.[11] The coefficient on the condition variable was 0.51 (95% CI: 0.01 to 1.02), giving an odds ratio of 1.67, and the coefficient on the intercept was -0.25 (95% CI: -0.62 to 0.11). Based on these estimates, moving a participant from the target-only condition to the contrast condition would decrease their probability of attributing knowledge by 12 percentage points, from 56% to 44%.[12]

The size of this effect was smaller than the output of the sensitivity analysis. Therefore, to confirm the effect was reliable, we combined the data from Study 3 and the Supplement to Study 3. These studies were identical through the binary knowledge attribution question and were conducted within two days of one another. The combined sample contained responses from 536 participants. A sensitivity analysis conducted in G*Power 3.1, using an alpha level of 0.05, a power of 0.80, and features of the sample described below, shows that this study could reliability detect effects larger than odds ratio 1.63 (for the binary outcome). In this sample, 59% of participants (95% CI: 53% to 65%) attributed knowledge in the target-only condition and 46%

[11] This is the only reported p-value that rises above the significance threshold, to $p = 0.060$, when participants who completed the study in less than 1 minute are included. However, we also collected data on the yes-or-no outcome part of this study in a larger sample; see the Supplement to Study 3 in the Supplemental Materials. That study also found a significant effect of condition on the percentage of participants attributing knowledge, $z(291) = 2.18$, $p = 0.025$. That finding remains significant when all participants are included. The study in the Supplement also collected blame ratings after collecting the knowledge attribution. We discuss these below.

[12] To put it another way, assuming independent probabilities (i.e., no effect of deliberation), there is a 3% chance that in a randomly chosen jury containing six people, all six would attribute knowledge in the target-only condition and a 1% chance that all six would attribute knowledge in the contrast condition.

(95% CI: 40% to 52%) did so in the contrast condition. A logistic regression showed a significant effect of condition on knowledge attributions, $z(534) = 2.99$, $p = 0.003$. The coefficient on the condition variable was 0.52 (95% CI: 0.18 to 0.87), giving an odds ratio of 1.69. This odds ratio exceeded the output of the sensitivity analysis for the combined sample.

Responses on the 7-point Likert-type scale to the question about whether Joseph caused the death knowingly also showed a significant difference. An independent-samples t-test comparing mean responses in the target-only condition ($m = 4.69$, $sd = 1.92$, 95% CI: 4.35 to 5.03) to mean responses in the contrast condition ($m = 4.08$, $sd = 2.08$, 95% CI: 3.70 to 4.45) showed a significant effect of condition, $t(241) = 2.41$; $p = 0.017$, $d = 0.31$ (95% CI: 0.06 to 0.56). The size of this effect was similar in magnitude to the output of the relevant sensitivity analysis. The average responses to the knowledge question on the Likert-style 7-point scale appear in Figure 2C.

**Discussion**

Participants were significantly more likely to describe the agent as acting knowingly when "knowingly" was the only salient mental state term than when a similar contrast term, "recklessly," was also salient. This difference appeared when participants could attribute recklessness in the contrast condition, as shown in Studies 1-2, but also when they merely learned about recklessness in the contrast condition, as shown in Study 3. Although the findings of Studies 1-2 were explicable based on a motivational explanation – i.e., people may have wanted to describe someone who caused harm as culpable, so they over-attributed knowledge when that was the only culpable mental state available but attributed it less when recklessness was also available – this study's findings were not. In both conditions, participants could only express the agent's culpability by saying he acted knowingly, but a difference in knowledge

attributions between conditions still appeared. These findings were, however, compatible with the hypothesis that "knowingly" is contrast sensitive.

In addition, we note that the size of the effect observed in this study is similar to the effect sizes observed in Studies 1 and 2. In this study, the odds ratio for the condition variable in the model of the binary outcome was 1.67 (95%CI: 1.01 to 2.78) and the effect size of that variable in the model for the ordinal outcome was $d = 0.31$ (95% CI: 0.06 to 0.56). In Study 1a, those numbers were 12.92 (95% CI: 6.69 to 26.15) and $d = 0.82$ (95% CI: 0.52 to 1.11). In Study 1b, they were 2.68 (95% CI: 1.52 to 4.80) and $d = 0.37$ (95% CI: 0.09 to 0.65). In Study 2, predicting outcomes using only the contrast factor, they were 2.28 (95% CI: 1.45 to 3.61) and $d = 0.43$ (95% CI: 0.21 to 0.66). The confidence intervals of the condition variable in both models overlap for Studies 1b, 2, and 3, implying that they do not differ significantly between studies. The inability to statistically distinguish the effect sizes of these studies suggests that any marginal effect of the pragmatic factors that could have been at work in Studies 1b and 2 was small in size, and that the contrast sensitivity of "knowingly" primarily drove these effects.

Finally, the Supplement to Study 3 reports an additional study that uses an identical design to Study 3 with a single exception: A Likert-scale question about how much blame the agent deserves replaces the Likert-scale knowledge question. As noted in footnote 11, the study in the Supplement replicates the finding of Study 3 with respect to the yes-or-no question about whether agent acted knowingly.

In addition, responses to the blame question can help shed light on how participants understood the study. One potential objection to the vignette used in Studies 1-3 is that creative participants without much understanding of the legal system may have made an incorrect assumption about the law that affected their responses. For example, they may have assumed that

defendants who cause a death recklessly are guilty of a more serious crime than second-degree murder and that some other proceeding would convict the defendant of that crime if they answered "no" to the knowledge question. As reported in the Supplement, however, attributing knowledge predicts significantly greater blame than not attributing knowledge, controlling for condition, which is inconsistent with this objection. Indeed, 93.2% (95% CI: 89.2% to 97.1%) of participants who attributed knowledge gave one of the two highest blame ratings (6 or 7), while only 81.1% (95% CI: 74.3% to 87.8%) of participants who did not attribute knowledge did so.

Study 3 and the Supplement to Study 3 also help address a different alternative explanation for the results Studies 1-2. On this explanation, participants in the contrast condition made an assumption beyond the text of the vignette: that Joseph could still be guilty of some other, lesser crime if they said he acted "recklessly." Participants in the contrast condition who wanted to ensure Joseph received some punishment therefore had a means to do so without saying he acted "knowingly," but participants in the target-only condition did not. This series of assumptions could lead to a difference between conditions, even if participants in both conditions agreed about whether Joseph really acted knowingly.

This alternative explanation is implausible for two reasons. First, it is unlikely that participants made this series of assumptions with no prompting from the vignette, which stated only that Joseph had been charged with second-degree murder and did not mention any other crime. This alternative explanation would have required participants to assume that recklessly causing the death constituted another crime, that the defendant could be convicted of that crime even if not convicted of second-degree murder (e.g., that there was a possibility of convicting on a lesser included offense not mentioned in the vignette or that double jeopardy somehow did not apply), and that conviction would lead to what they viewed as an appropriate level of

punishment. Second, it is particularly unlikely that participants in Study 3 assumed Joseph could be guilty of a lesser offense if he acted recklessly, since they never provided their judgments about whether Joseph acted recklessly.

In addition, this explanation assumes participants on average wanted Joseph to receive less punishment than they thought he would receive if they said he acted "knowingly." In the Supplement to Study 3, however, mean blame judgments were close to ceiling (7) in both the target-only condition ($m$ = 6.51, $sd$ = 0.92, 95% CI: 6.37 to 6.66) and the contrast condition ($m$ = 6.56, $sd$ = 0.80, 95% CI: 6.43 to 6.69), and an equivalence test shows that they do not differ by a meaningful amount. Although blame and punishment judgments are not identical, they are correlated, and blame judgments are generally more sensitive to inferences about an agent's mental state (Kneer & Skoczeń, 2023). The high and similar blame judgments observed across conditions therefore make it less plausible that participants wanted to punish Joseph for some lesser crime and assumed that could happen in the contrast condition but not the target-only condition.

## Study 4

Prior research suggested that "intentionally" had multiple semantic meanings associated with different conceptual contrasts (Cova et al., 2012). Studies 1-3 provide evidence that the same is true of "knowingly." These two examples suggest that contrast sensitivity may be a general feature of mental state terms – i.e., their meanings shift depending on what one understands the relevant contrast mental state to be.

A contrary view, however, is that acting intentionally and acting knowingly are special in some way. Both, for example, have received substantial philosophical interest over a long period of time (e.g., Austin, 1966; Hart, 1967; Mele & Moser, 1994; Moore, 1997). Perhaps the

conceptual complexity that makes these concepts philosophically interesting also gives rise to the multiple semantic meanings. Both concepts may also have unique cognitive roles. Perceiving intentionality in oneself and others is a basic building block of social and cultural cognition (Tomasello et al., 2005). Perhaps a more complex concept of intentionality exists to support these important social functions. Knowledge is also cognitively special: Knowledge appears to be a basic way of representing others' mental states, to an even greater degree than belief (Phillips et al., 2021). Perhaps "knowingly" has multiple meanings because of this property, rather than because of a more general feature of all mental state concepts.

To test whether contrast sensitivity appears for another mental state concept, Study 4 tested recklessness as the target mental state and the next-most-serious Model Penal Code mental state, negligence, as the contrast. This study used a new vignette in which "recklessly" and "negligently" were both plausible ways to describe the agent's action. This vignette drew from another Massachusetts criminal case (*Commonwealth v. Sostilio*, 1949), which involved reckless driving that created a lower level of risk than in *Subenko* but nevertheless caused a death.

## Methods

### *Participants*

We recruited 260 adult participants on Amazon's Mechanical Turk. Seventeen participants were excluded because they completed the entire survey in less than one minute, as measured using the Duration variable recorded by Qualtrics, or did not complete the survey, leaving an analysis set of 243 participants ($M_{age}$ = 37 years; $SD_{age}$ = 11.5 years; 47% female, 51% male, 2% did not answer or chose an answer not otherwise listed). The sample's highest level of education was distributed as follows: 12% high school or GED; 28% some college or university; 48% college or university degree; 11% master's or professional degree; and 1% PhD. A

sensitivity analysis conducted in G*Power 3.1, using an alpha level of 0.05, a power of 0.80, and features of the sample described below, shows that this study could reliability detect effects larger than odds ratio 2.09 (for the binary outcome) and effect size $d = 0.36$ (for the Likert scale outcome).

***Procedure***

Participants were randomly assigned to one of two conditions. Participants received one of two versions of the following vignette (bracketed and underlined passages appeared only in the Contrast condition):

> Imagine that you are a juror in a criminal case. The defendant, Joseph, has been charged with voluntary manslaughter. You have heard evidence that establishes the following:
>
> Joseph regularly participated in amateur car races at a local stadium. The race course was elliptical in shape, with long straight sections joined by curved end sections. Last year, Joseph participated in a race with five other drivers. At the start of the race, the cars were lined up in three rows of two. As one of the most experienced drivers in the race, Joseph started the race in the last row. When the cars had travelled somewhat more than half a lap, Joseph found himself directly behind the car driven by Chris. The left wheels of Chris's car were only two feet from the inside of the track. Joseph tried to drive to the left of Chris's car, although there was not sufficient space for him to pass on the left. Joseph was driving upwards of a hundred miles an hour. There was evidence that he struck Chris's car, and that it skidded off the track and turned over three times. Chris died in the crash.
>
> The judge instructs you that in your jurisdiction, a defendant is guilty of voluntary manslaughter if he recklessly caused the death of another human being. Joseph has conceded that his actions caused Chris's death. The only question before the jury, therefore, is whether Joseph recklessly caused Chris's death.
>
> The judge instructs you that a person recklessly causes a result if he consciously disregards a substantial and unjustifiable risk that his conduct will cause such a result.
>
> [The judge instructs you that recklessly causing a result differs from negligently causing a result. A person negligently causes a result if he should be aware of a substantial and unjustifiable risk that his conduct will cause such a result.]

> Finally, the judge instructs you that if you find that Joseph recklessly caused Chris's death, then Joseph will be guilty of voluntary manslaughter. If you do not find that Joseph recklessly caused Chris's death, then[, even if Joseph negligently caused Chris's death,] Joseph will not be guilty of voluntary manslaughter.

Participants in both conditions answered the following question:

> Did Joseph recklessly cause Chris's death?
> - Yes
> - No

On the following page, all participants then rated on a 7-point scale their agreement with the statement that Joseph recklessly caused Chris's death,[13] following by a demographic questionnaire.

**Results**

Responses to the categorical outcome variable showed a significant difference between conditions. In the target-only condition, 66% of participants (95% CI: 57% to 74%) said that Joseph caused the death recklessly; in the contrast condition, only 48% of participants (95% CI: 40% to 57%) attributed recklessness. The percentages of participants attributing recklessness by condition appear in Figure 1D. A binary logistic regression of condition (target only = 1; contrast = 0) on the recklessness attribution question (yes = 1; no = 0) showed a significant effect, $z(241) = 2.68$, $p = 0.007$. The coefficient on the condition variable was 0.71 (95% CI: 0.19 to 1.23), giving an odds ratio of 2.03, and the coefficient on the intercept was -0.06 (95% CI: -0.42 to 0.29). Based on these estimates, moving a participant from the knowledge-only condition to the contrast condition would decrease their probability of attributing knowledge by 18 percentage

[13] In this study, the page with the Likert-type scale questions also asked participants in the contrast condition to rate whether Joseph negligently caused Chris's death. Thus, it is possible that pragmatic factors could have affected responses to the ordinal, but not the categorical, dependent variable.

points, from 66% to 48%.[14] The size of this effect was similar in magnitude to the output of the relevant sensitivity analysis.

Responses on the 7-point Likert-type scale to the question about whether Joseph caused the death recklessly also showed a significant difference. An independent-samples t-test comparing mean responses in the target-only condition ($m$ = 4.81, $sd$ = 2.06, 95% CI: 4.43 to 5.18) to mean responses in the contrast condition ($m$ = 4.14, $sd$ = 1.98, 95% CI: 3.79 to 4.49) showed a significant effect of condition, $t(241) = 2.59$; $p = 0.010$, $d = 0.33$ (95% CI: 0.08 to 0.59). The size of this effect was similar in magnitude to the output of the relevant sensitivity analysis.

**Discussion**

This study found that recklessness attributions dropped when negligence was made salient, just as knowledge attributions dropped when recklessness was made salient in Study 3. In both cases, learning about a contrast mental state concept affected attributions of the target mental state. Importantly, this result shows that the effect observed in Studies 1-3 generalizes beyond the specific mental state used in those studies.

## General Discussion

Mental state attributions shape social interactions by contributing to blame and punishment judgments (Cushman, 2008; Cushman et al., 2013; Flick et al., 2024; Young & Saxe, 2008, 2011; for reviews, see Cushman, 2015; Heiphetz & Young, 2014), inferences about obligations (Turri, 2017; Turri et al., 2017), and judgments about desires, dispositions, and future

[14] To put it another way, assuming independent probabilities (i.e., no effect of deliberation), there is an 8% chance that in a randomly chosen jury containing six people, all six would attribute recklessness in the target-only condition and a 1% chance that all six would attribute recklessness in the contrast condition.

behavior (Jara-Ettinger et al., 2016, 2020; Ouellette & Wood, 1998). The legal system formalizes this role, such that the label jurors use to describe a harm – "purposely," "knowingly," "recklessly," "negligently" – plays a substantial role in determining whether the act that caused the harm is a crime and, if so, how serious (ALI, 1962, § 2.02; Ginther et al., 2014; Shen et al., 2011; Vilares et al., 2017). Thus, understanding how laypeople define and use these terms has important practical implications for legal decisions.

Defining mental state terms also has theoretical significance. A large body of research on intentionality attributions suggests that the term "intentionally" may have multiple meanings – that is, it can pick out different concepts in different contexts. Many of these types of intentionality correspond to different contrasts with other mental state concepts – e.g., "intentionally" contrasted with "carelessly," "unwillingly," or "accidentally" (Cova et al., 2012, p. 388). The present studies investigated whether "knowingly" and "recklessly" also have multiple meanings. Four studies showed that people understand and apply "knowingly" differently when "recklessly" is salient than when it is not. A fifth study showed the same is true of "recklessly" contrasted with "negligently." These findings suggest that the meaning of "knowingly" and "recklessly," like the meaning of "intentionally," is not fixed and depends on what alternative mental state the context makes salient.

These studies followed a similar experimental logic to a series of studies on intentionality (Adams & Steadman, 2007; Guglielmo & Malle, 2010b; Knobe, 2003; Lindauer & Southwood, 2021; Nichols & Ulatowski, 2007). To minimize the possibility that any difference in judgments was due to some change in the facts of the scenario that caused the mental state concept to apply in one condition but not the other, we did not vary those facts of the vignettes between conditions in any of the studies. Study 1a provided evidence that motivational or pragmatic factors,

differences in calculations, or differences in meaning led to a difference in knowledge attributions between a yes-or-no question and a multiple-choice question that included "recklessly," a plausible contrast concept, as an option. Study 1b provided evidence that a difference in calculations could not fully explain the effect. Study 2 extended Study 1b's findings to a context where the legal definition was not provided, suggesting that laypeople's ordinary understanding of "knowingly" is contrast sensitive. Study 3 provided evidence that motivational or pragmatic factors could not explain this effect. Study 4 provided evidence that "recklessly" is also contrast sensitive.

Even in light of this evidence, one may posit an additional, alternative explanation for the findings, which is a variant of the motivational explanation based on communicative intentions considered above. The communicative motivational explanation suggests that, in Studies 1b and 2, participants in the target-only condition did not really think Joseph acted "knowingly" but said he did because they wanted to communicate his culpability, while in the contrast condition they could do so by saying he acted "recklessly." Study 3 provided evidence against that explanation because it found a difference between conditions even when participants could not attribute recklessness in the contrast condition.

In response, one might propose that participants in Study 3 were motivated not to *communicate* Joseph's culpability but merely to have some way to label him as culpable *in their own minds*. On this explanation, participants in the target-only condition did not really think Joseph acted knowingly but labeled him that way in their own minds so they could think of him as culpable and then communicated that label in their response to the attribution question. Participants in the contrast condition, however, saw the word "recklessness" in the vignette,

which gave them a different way to label Joseph as culpable in their own minds and then answered the attribution question accordingly.

We have three responses to this argument. First, we note that the practical implications of all the motivational explanations point in the same direction as the contrast-sensitivity hypothesis – if anything, they provide an even stronger argument for instructing juries using contrasts. After all, the motivational explanations all posit that laypeople in the target-only condition are misapplying their mental state concepts. Second, this new motivational explanation rests on two assumptions that seem implausible. One assumption is that participants in the target-only condition who were motivated to label Joseph as culpable in their own minds could not come up with any other label to use for that internal purpose. Another assumption is that participants motivated to misuse a label internally to satisfy their own desire to think of an agent as culpable would use that label to respond to a question about the agent but were not also motivated to communicate the agent's culpability. (After all, if they were motivated to communicate that culpability, then Study 3 would not have found a difference between conditions.) That combination of motivations seems implausible.

The findings of the present studies may have broader theoretical implications. The existence of multiple meanings for mental state terms, which depend on what conceptual contrast the context makes salient, may be an instance of a more general phenomenon. When children learn certain groups of closely related words, they use "lexical contrast" – that is, the assumption that even similar words have slight differences in meaning – to gradually acquire adult understandings of the words (Tomasello, 2005, p. 72). In this way, groups of similar words come to inter-define one another, such that people understand them in part by how they differ from one another (Saji et al., 2020; Wagner et al., 2018). Contrast sensitivity may arise for groups of

concepts learned in this way, which will often have fuzzy boundaries and an overlapping structure.[15] By contrast, mutually exclusive terms that completely partition a covering term may lead to a different phenomenon – contrast category effects (Verheyen, De Deyne, et al., 2011).

The present findings also have practical implications for the law. Unlike the side-effect effect (Knobe, 2003) or the closely related severity effect (Kneer & Bourgeois-Gironde, 2017), both of which involve differences in intentionality attributions between different factual scenarios, the effects described in this paper depend on the phrasing of the mental state attribution question, even holding the facts fixed. Thus, these findings suggest it may be possible to change jurors' mental state attributions in an individual case by making a particular contrast concept salient or not, though, for real juries, deliberation may impact the effects described in this paper. If these effects continue to appear after deliberation, they present a potential way to adjust jury instructions to pick out more precisely the mental state concepts jurors should use – a type of conceptual engineering (Lewis, 2025). This possibility is noteworthy because jury instructions often have no or little effect on laypeople's judgments (Devine, 2012; Diamond & Rose, 2005; Ginther et al., 2014).

[15] Although we explain these results in terms of semantics – i.e., the mental state terms have multiple concepts as possible referents – it may also be possible to explain them in terms of lexical pragmatics – i.e., mental state terms generally refer to *ad hoc* concepts constructed on the fly to be relevant in context (Padilla Cruz, 2023; Sperber & Wilson, 2012). Working out such an explanation in full is beyond the scope of this article. However, we do not think the possibility of such an explanation undercuts the importance of the present results, either theoretically or practically. On the theoretical side, it is not the case that context can give words *any* meaning at all. The lexically mapped concept places constraints on the scope of possible pragmatic enrichment (Sperber & Wilson, 2012). Thus, these results reveal something about available meanings of "knowingly" and "recklessly," either semantic or pragmatic. On the practical side, there is no obvious reason to think one of our two conditions involves the form of the concept mapped to a lexical item – which one might regard as the true meaning of the word – and the other one involves lexical narrowing or broadening.

The present studies are also relevant to the jury context because they involved a task similar to a jury's task. Jurors making mental state attributions generally must answer yes-or-no questions, rather than multiple-choice questions, about whether a mental state applies (Beatty & Fondacaro, 2018). Even when a party requests and the judge provides a lesser included offense instruction – e.g., an instruction that the jury may convict a defendant of manslaughter instead of second-degree murder – the judge defines each crime separately, such that the jury's task more closely resembles answering two yes-or-no questions, as in Study 1b, than answering a multiple-choice question, as in Study 1a. The similar size of the effects observed in Study 1b and Study 3 suggests that being able to answer a separate question about the contrast mental state has a minimal effect on yes-or-no answers about the target mental state.

However, there is one aspect of the jury's task in some cases that may more closely resemble the multiple-choice question used in Study 1a and some prior studies (Guglielmo & Malle, 2010; Mueller et al., 2012). In cases with lesser included offense instructions, the ultimate verdict form can look like a multiple-choice question with three possible answers: (1) the greater offense; (2) the lesser offense; and (3) acquittal. In these cases, as described above, jurors are instructed to think about each crime separately and to answer the yes-or-no mental state attribution questions separately. But, in the end, they must choose between three case outcomes. The effect of contrast on mental state attributions may operate differently in these types of cases, in which both the target concept and the contrast concept are associated with a crime. Indeed, existing research suggests the effect may be larger in such cases, since the contrast-sensitivity of the mental state terms would push jurors in the same direction as the compromise effect – i.e., the bias towards the middle options in ordered choice sets (Kelman et al., 1996).

Finally, the present studies cannot resolve which type of instruction – none, the legal definition, or an instruction containing information about contrasts – would lead to more accurate judgments by jurors. The vignettes were designed so that the (more culpable) target mental state was a less plausible description of the action than the (less culpable) contrast mental state, but either one was arguably acceptable. However, as we argued in the introduction, providing jurors with information about contrast concepts may better serve the purpose of a jury trial by causing lay jurors to use narrower mental state concepts, more similar to those used by legal actors (Tobia, 2024). The present studies demonstrate that informing jurors about contrast concepts does indeed cause their mental state concepts to narrow.

**Limitations and Future Directions**

As noted above, the current work makes several theoretical and translational contributions. Like all work, it also raises questions for future research.

First, the present studies involved only participants from the United States, a culture that is particularly punitive (Alexander, 2012; Barkow, 2019; Forman, 2017) and where mental states play a crucial role in attributions of both moral and legal responsibility (Heiphetz & Young, 2014; Shen et al., 2011). This single-culture approach allowed us to focus on understanding how people in the United States reason about others' minds in a criminal context, which is important for evaluating and reforming legal institutions that affect the lives of many residents (Carson, 2021). However, there is likely substantial cultural variation in this type of cognition. For instance, mental states play a larger role in moral reasoning in Western industrialized societies than they do in some agrarian, small-scale, and non-Western societies (Barrett, et al., 2016; McNamara, et al., 2019). People from these cultures may not represent the different mental state concepts used in these studies as alternatives to each other to the same extent as people from

cultures that use them to ascribe different levels of legal responsibility. Additional research will be necessary to provide a definitive answer to questions about cross-cultural generalization.

Second, the present research focused on participants' mental state attributions rather than the downstream consequences of those attributions (e.g., punishment). We took this approach because, in the United States criminal legal system, juries generally decide on guilt but not on punishment, so deciding whether to attribute a mental state without being able to control the punitive consequences of any resulting conviction resembles the task real jurors face. Indeed, one of the arguments for bifurcating the guilt and punishment decisions between the jury and the judge is that jurors might otherwise make results-oriented decisions about guilt (Moss, 2021; *Shannon v. United States*, 1994). As a result, understanding the factors that affect mental state attributions in isolation is also practically important, even if laypeople would ultimately evaluate or punish similarly across the two conditions. However, the experiment reported in the Supplement does provide some suggestive evidence about whether the contrast-sensitivity of mental state terms may make a difference to blame judgments – which are related to punishment judgments, though not identical (Cushman, 2008; Kneer & Skoczeń, 2023). The marginally significant mediation reported in the Supplement suggests that the difference in mental state attribution does lead to a difference in blame judgments, but that effect is cancelled out by another process, perhaps the recognition that the judgments involved in the target-only condition are imprecise. Further research would be required to probe the interaction of contrast-sensitivity and downstream evaluations and responses.

Finally, these studies used mock jurors, rather than real jurors, and abstracted from some of the complexities involved in real cases, including deliberation. However, future research may be able to explore the effect of some forms of contrast sensitivity on real juries using records

from federal criminal cases. For federal crimes, many United States Courts of Appeals have adopted model jury instructions, which parties can use as the bases for their own proposals (Macleod, 2016). Some – but not all – Courts of Appeals have adopted definitions of "knowingly" in their model jury instructions that contrast acting "knowingly" with acting by "mistake" or "accident." For example, the Third Circuit Court of Appeals (2024, p. 6) defines "knowingly" as follows:

> A person acts "knowingly" if that person acts voluntarily and intentionally and not because of mistake or accident or other innocent reason. This means that the government must prove beyond a reasonable doubt that *(name)* was conscious and aware of the nature of *(his) (her)* actions and of the surrounding facts and circumstances, as specified in the definition of the offense*(s)* charged.

The contrast in this instruction may or may not make a difference to laypeople's understanding of "knowingly" – that depends on whether "mistake" and "accident" can act as contrasts for that term. This circuit-to-circuit variation in federal model jury instructions provides a possible way to investigate this question.

## Conclusion

Mental state terms like "knowingly" and "recklessly" play a crucial role in moral and legal judgments about the culpability of a harmful action. In the legal context, the difference between knowledge and recklessness can make a substantial difference in the seriousness of the crime a person has committed and the length of the sentence they receive (Vilares et al., 2017). Given the practical effects of these mental state attributions, understanding the factors that drive them has both moral and legal importance. The present studies provided evidence that these terms are contrast sensitive – that is, their meanings shift depending on the other concepts that are salient as contrasts. In particular, people understand these mental state terms to have a narrower meaning when contrasted with terms for mental states that United States criminal law

treats as less serious. On the theoretical side, these results suggest that people do not understand legally relevant mental state terms in isolation but as inter-defined regions of a single conceptual space. On the practical side, these results suggest that making clear the relevant contrast could cause jurors to use narrower mental state concepts, which the legal actors are also likely using based on their legal education (Tobia, 2024).

## References


Adams, F. & Steadman, A. (2007). Folk concepts, surveys and intentional action. In C. Lumer & S. Nannini (Eds.), *Intentionality, deliberation and autonomy: The action-theoretic basis of practical philosophy* (pp. 17-33). Ashgate.

Ahler, D. J., Roush, C. E., & Sood, G. (2019). The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. *Political Science Research and Methods*, 1-20, https://doi.org/10.1017/psrm.2021.57

Alexander, M. (2012). *The New Jim Crow* (Rev. Ed.). The New Press.

American Law Institute (ALI). (1962). *Model penal code*.

American Law Institute (ALI). (1985). *Model penal code and commentaries*.

Arel-Bundock, V., Greifer, N., Bacher, E., McDermott, G., & Heiss, A. (2025). marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests. R package version 0.25.0. https://cran.r-project.org/package=marginaleffects.

Austin, J.L. (1962). *Sense and sensibilia* (2nd ed.). Oxford.

Austin, J.L. (1966). Three ways of spilling ink. *The Philosophical Review, 75*(4), 427-440. https://doi.org/10.2307/2183222

Beattey, R.A. & Fondacaro, M.R. (2018). The misjudgment of criminal responsibility. *Behavioral Sciences & the Law, 36*, 457- 469. https://doi.org/10.1002/bsl.2354

Barkow, R.E. (2019). *Prisoners of politics: Breaking the cycle of mass incarceration*. Harvard University Press.

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., Henrick, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B. A., Stich, S., von Rueden, C., Zhao, W., & Laurence, S. (2016). Small-scale societies exhibit fundamental variation

in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences, 113*, 4688-4693, https://doi.org/10.1073/pnas.1522070113

Beebe, J. R., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language, 25*, 474-498, https://doi.org/10.1111/j.1468-0017.2010.01398.x

Beebe, J. R., & Jensen, M. (2012). Surprising connections between knowledge and action: The robustness of the epistemic side-effect effect. *Philosophical Psychology, 25*, 689-715, https://doi.org/10.1080/09515089.2011.622439

Beebe, J. R., & Shea, J. (2013). Gettierized Knobe effects. *Episteme, 10*, 219-240, https://doi.org/10.1017/epi.2013.23

Bratman, M. (1984). Two faces of intention. *The Philosophical Review, 93*(3), 375-405. https://doi.org/10.2307/2184542

Buckwalter, W. (2014). Factive verbs and protagonist projection. *Episteme, 11*, 391-409, https://doi.org/10.1017/epi.2014.22

Buckwalter, W., Rose, D., & Turri, J. (2015). Belief through thick and thin. *Noûs, 49*(4), 748-775. https://doi.org/10.1111/nous.12048

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?. *Perspectives on Psychological Science, 6*, 3-5, https://psycnet.apa.org/doi/10.1037/14805-009

Carson, E.A. (2021). Prisoners in 2020 – Statistical tables, *Bureau of Justice Statistics: Statistical Tables*, https://bjs.ojp.gov/content/pub/pdf/p20st.pdf

*Carter v. United States*, 530 U.S. 255 (2000).

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11*, 464-473, https://doi.org/10.1177/1948550619875149

Christensen, R. H. B. (2015). ordinal: Regression Models for Ordinal Data. R package version 2022.11-16. https://cran.r-project.org/package=ordinal.

*Commonwealth v. Pugh*, 969 N.E.2d 672 (Mass. 2012)

*Commonwealth v. Sostilio*, 89 N.E.2d 510 (Mass., 1949)

*Commonwealth v. Subenko*, 1999-P-1404, 2001 Mass. App. LEXIS 1076 (Mass. App. Ct. Nov. 20, 2001)

Cova, F. (2016). The folk concept of intentional action: Empirical approaches. In J. Sytsma & W. Buckwalter (Ed.s) *A companion to experimental philosophy* (pp. 117-141). Wiley Blackwell.

Cova, F., Dupoux, E., & Jacob, P. (2012). On doing things intentionally. *Mind & Language, 27*(4), 378-409. https://doi.org/10.1111/j.1468-0017.2012.01449.x

Cullen, S. (2010). Survey-driven romanticism. *Review of Philosophy and Psychology, 1*, 275-296, https://doi.org/10.1007/s13164-009-0016-1

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127*(1), 6–21. https://doi.org/10.1016/j.cognition.2012.11.008

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology, 6*, 97–103. https://doi.org/10.1016/j.copsyc.2015.06.003

Davis, W. A. (1984). A causal theory of intending. *American Philosophical Quarterly, 21*(1), 43-54.

Davis, W. A. (1986). The two senses of desire. In J. Marks (Ed.), *The ways of desire: New essays in philosophical psychology on the concept of wanting* (pp. 63–82). Precedent Publishing.

DeRose, K. (2009). *The case for contextualism: Knowledge, skepticism, and context* (Vol. 1). Oxford University Press.

Devine, D.J. (2012). *Jury decision making*. New York University Press.

Diamond, S. & Rose, M. (2005) Real juries. *Annual Review of Law & Social Science, 1*, 255-284. https://doi.org/10.1146/annurev.lawsocsci.1.041604.120002

Eisenberg, T., & Hans, V. P. (2008). Taking a stand on taking the stand: The effect of a prior criminal record on the decision to testify and on trial outcomes. *Cornell L. Rev., 94*, 1353.

*Elonis v. United States*, 135 S. Ct. 2001 (2015)

Flick, C., Nuñez, N., & Laurent, S. M. (2024). If negligence is intentionality's cousin, recklessness is it's sibling: Differentiating negligence and recklessness from accidents and intentional harm. *Journal of Experimental Social Psychology, 110*, 104529. https://doi.org/10.1016/j.jesp.2023.104529

Forman, J. Jr. (2017). *Locking up our own: Crime and punishment in Black America*. Farrar, Straus and Giroux.

Gardner, J. (2012). How law claims, what law claims. In M. Klatt (ed.), *Institutionalized reason: The jurisprudence of Robert Alexy* (pp. 29-44). Oxford.

Gerken, M. (2017). *On folk epistemology: How we think and talk about knowledge*. Oxford University Press.

Ginther, M.R., Shen, F.X., Bonnie, R.J., Hoffman, M.B., Jones, O.D., Marois, R., & Simons, K.W. (2014). The language of *mens rea*. *Vanderbilt Law Review, 67*, 1327-1372.

*Global-Tech Appliances, Inc. v. SEB SA*, 563 U.S. 754 (2011)

Guglielmo, S. & Malle, B. (2010). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin, 26*, 1635-1647, https://doi.org/10.1177/0146167210386733

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31*, 355-384, https://doi.org/10.1080/15326900701326402

Hart, H.L.A. (1967). Intention and punishment. In H.L.A. Hart, *Punishment and responsibility: Essays in the philosophy of law*. Clarendon Press.

Heiphetz, L., & Young, L. L. (2014). A social cognitive developmental perspective on moral judgment, *Behaviour, 151*, 315-335, https://doi.org/10.1163/1568539X-00003131

Ichikawa, J. J. & Steup, M. (2018). The analysis of knowledge. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/aechives/sum2018/entries/knowledge-analysis

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences, 20*(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The Naïve Utility Calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology, 123*, 101334. https://doi.org/10.1016/j.cogpsych.2020.101334

Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences, 29*, 105–110. https://doi.org/10.1016/j.cobeha.2019.04.010

Johnson, E.A. (2016). Understanding general and specific intent: Eight things I know for sure. *Ohio State Journal of Criminal Law, 13*(2), 521-541.

Johnson, D., & Ryan, J. B. (2020). Amazon mechanical turk workers can provide consistent and economically meaningful data. *Southern Economic Journal, 87*, 369-385, https://doi.org/10.1002/soej.12451

Kadish, S. H., Schulhofer, S. J., Steiker, C.S., Barkow, R.E. (2012). *Criminal law and its processes* (9th ed.). Wolters Kluwer.

Kelman, M., Rottenstreich, Y., & Tversky, A. (1996). Context-dependence in legal decision making. *The Journal of Legal Studies, 25*(2), 287-318. https://doi.org/10.1086/467979

Kneer, M. & Bourgeois-Gironde, S. (2017). Mens rea ascription, expertise and outcome effects: Professional judges surveyed. *Cognition, 169*, 139-146. https://doi.org/10.1016/j.cognition.2017.08.008

Kneer, M., & Skoczeń, I. (2023). Outcome effects, moral luck and the hindsight bias. *Cognition, 232*, 105258. https://doi.org/10.1016/j.cognition.2022.105258

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*, 190-194. https://doi.org/10.1093/analys/63.3.190

Knobe, J. (2004). Intention, intentional action and moral considerations. *Analysis, 64*, 181-187, https://doi.org/10.1111/j.1467-8284.2004.00481.x

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*, 315-329. https://doi.org/10.1017/S0140525X10000907

Laurent, S. M., Reich, B. J., & Skorinko, J. L. M. (2019). Reconstructing the side-effect effect: A new way of understanding how moral considerations drive intentionality asymmetries. *Journal of Experimental Psychology: General, 148*(10), 1747–1766. https://doi.org/10.1037/xge0000554

Laurent, S. M., Reich, B. J., & Skorinko, J. L. M. (2021). Understanding Side-Effect Intentionality Asymmetries: Meaning, Morality, or Attitudes and Defaults? *Personality and Social Psychology Bulletin, 47*(3), 410–425. https://doi.org/10.1177/0146167220928237

Lewis, J. (2025). Competing conceptual inferences and the limits of experimental jurisprudence. In: K. P. Tobia (ed.) *The Cambridge Handbook of Experimental Jurisprudence*. Cambridge University Press.

Lindauer, M., & Southwood, N. (2021). How to cancel the Knobe effect: the role of sufficiently strong moral censure. *American Philosophical Quarterly, 58*(2), 181-186. https://doi.org/10.2307/48614004

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., & Hashimoto, T. (2017). Gettier across cultures. *Noûs, 51*(3), 645–664. https://doi.org/10.1111/nous.12110

Maćkiewicz, B., Kuś, K., Paprzycka-Hausman, K., & Zaręba, M. (2022). Epistemic side-effect effect: A meta-analysis. *Episteme*, 1-35. https://doi.org/10.1017/epi.2022.21

Macleod, J. (2016). Belief states in criminal law. *Oklahoma Law Review, 68*, 497-554.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*(2), 101-121. https://doi.org/10.1006/jesp.1996.1314

*McDowell v. Calderon*, 130 F. 3d 833 (9th Cir. 1997)

McNamara, R. A., Willard, A. K., Norenzayan, A., & Henrich, J. (2019). Weighing outcome vs. intent across societies: How cultural models of mind shape moral reasoning. *Cognition, 182*, 95-108, https://doi.org/10.1016/j.cognition.2018.09.008

Mele, A.R. & Moser, P.K. (1994). Intentional action. *Nous, 28*(1), 39-68. https://doi.org/10.2307/2215919

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 40-64). Cambridge University Press.

Minda, J. P., Roark, C. L., Kalra, P., & Cruz, A. (2024). Single and multiple systems in categorization and category learning. *Nature Reviews Psychology, 3*, 536–551. https://doi.org/10.1038/s44159-024-00336-7

Moore, M. (1997). *Placing blame*. Oxford University Press.

Moss, S. (2021). Pragmatic encroachment and legal proof. *Philosophical Issues, 31*(1), 258–279. https://doi.org/10.1111/phis.12207

Mott, C. (2023, July 28). Mental State Concepts. Retrieved from osf.io/mn32g

Mueller, P. A., Solan, L. M., & Darley, J. M. (2012). When does knowledge become intent? Perceiving the minds of wrongdoers. *Journal of Empirical Legal Studies, 9*(4), 859-892. https://doi.org/10.1111/j.1740-1461.2012.01269.x

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations, 9*, 203-219, https://doi.org/10.1080/13869790600641905

Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay denial of knowledge for justified true beliefs. *Cognition, 129*(3), 652-661. https://doi.org/10.1016/j.cognition.2013.02.008

Nichols, S., & Ulatowski, J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language, 22*, 346-365, https://doi.org/10.1111/j.1468-0017.2007.00312.x

Nobes, G., & Martin, J. W. (2022). They should have known better: The roles of negligence and outcome in moral judgements of accidental actions. *British Journal of Psychology, 113*(2), 370-395. https://doi.org/10.1111/bjop.12536

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 18-39). Cambridge University Press.

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin, 124*(1), 54–74. https://doi.org/10.1037/0033-2909.124.1.54

Oxford University Press. (n.d.). Unknowingly, adv., 2. In *Oxford English dictionary*. Retrieved July 26, 2025, from https://doi.org/10.1093/OED/2766806266

Padilla Cruz, M. (2023). Paralanguage and ad hoc concepts. *Pragmatics, 33*(3), 343–367. https://doi.org/10.1075/prag.21059.pad

*People v. Hood*, 462 P. 2d 370 (Cal., 1969)

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L. & Knobe, J. (2021). Knowledge before belief. *Behavioral and Brain Sciences, 44*, https://doi.org/10.1017/S0140525X20000618

Porter, B., Barr, K., Bencherifa, A., Buckwalter, W., Deguchi, Y., Fabiano, E., Hashimoto, T., Halamova, J., Homan, J., Karasawa, K., Kanovsky, M., Kim, H., Kiper, J., Lee, M., Liu, X., Mitova, V., Bhaya, R., Pantovic, L., Quintanilla, P., . . . Machery, E. (2024). A puzzle about knowledge ascriptions. *Noûs*. https://doi.org/10.1111/nous.12515

Quillien, T. & German, T. C. (2021). A simple definition of 'intentionally'. *Cognition, 214*, 104806. https://doi.org/10.1016/j.cognition.2021.104806

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

*Regina v. Cunningham*, 2 Q.B. 396 (1957).

Robinson, P. H., & Dubber, M. D. (2007). The American model penal code: A brief overview. *New Criminal Law Review, 10*, 319-341. https://doi.org/10.1525/nclr.2007.10.3.319

Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., Buchtel, E. E., Chatterjee, A., Cheon, H., Cho, I., Cohnitz, D., Cova, F., Dranseika, V., Lagos, A. E., Ghadakpour, L., Grinberg, M., Hannikainen, I., Hashimoto, T., Horowitz, A.,... & Zhu, J. (2017). Nothing at stake in knowledge. *Noûs, 53*, 224-247, https://doi.org/10.1111/nous.12211

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*, 1-36. doi:10.18637/jss.v048.i02.

Saji, N., Imai, M., & Asano, M. (2020). Acquisition of the Meaning of the Word Orange Requires Understanding of the Meanings of Red, Pink, and Purple: Constructing a Lexicon as a Connected System. *Cognitive Science, 44*(1), e12813. https://doi.org/10.1111/cogs.12813

Sarin, A., & Cushman, F. (2024). One thought too few: An adaptive rationale for punishing negligence. *Psychological Review, 131*(3), 812–824. https://doi.org/10.1037/rev0000476

Schaffer, J. & Knobe, J. (2012). Contrastive knowledge surveyed. *Nous, 46*, 675-708. https://doi.org/10.1111/j.1468-0068.2010.00795.x

*Shannon v. United States*, 512 US 573 (1994).

Shen, F. X., Hoffman, M. B., Jones, O. D., Greene, J. D., Marois, R. (2011). Sorting guilty minds. *New York University Law Review, 86*, 1306-1360.

Sinhababu, N. (2013). The desire-belief account of intention explains everything. *Noûs, 47*(4), 680-696. https://doi.org/10.1111/j.1468-0068.2012.00864.x

Sommers, R. (2021). Experimental jurisprudence. *Science, 373*(6553), 394-395. https://doi.org/10.1126/science.abf0711

Sperber, D. & Wilson, D. (2012). The mapping between the mental and the public lexicon. In D. Wilson & D. Sperber (Eds.), *Meaning and relevance* (pp. 31–46). Cambridge University Press.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition, 124*, 272-283, https://doi.org/10.1016/j.cognition.2012.05.017

Starmans, C., & Friedman, O. (2020). Expert or esoteric? Philosophers attribute knowledge differently than all other academics. *Cognitive Science, 44*, e12850, https://doi.org/10.1111/cogs.12850

Tobia, K. (2024). Legal concepts and legal expertise. *Synthese, 203*, 107. https://doi.org/10.1007/s11229-024-04512-3

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675-691, https://doi.org/10.1017/S0140525X05000129

Tooming, U. (2019). Active desire. *Philosophical Psychology, 32*(6), 945-968. https://doi.org/10.1080/09515089.2019.1629406

Turri, J. (2014). Problem of ESEE Knowledge. *Ergo, 1*(4), 101-127. https://doi.org/10.3998/ergo.12405314.0001.004

Turri, J., Buckwalter, W., & Rose, D. (2016). Actionability judgments cause knowledge judgments. *Thought: A Journal of Philosophy, 5*(3), 212-222. https://doi.org/10.1002/tht3.213

Turri, J. (2017). Knowledge attributions and behavioral predictions. *Cognitive Science, 41*(8), 2253–2261. https://doi.org/10.1111/cogs.12469

Turri, J., Friedman, O., & Keefner, A. (2017). Knowledge central: A central role for knowledge attributions in social evaluations. *Quarterly Journal of Experimental Psychology, 70*(3), 504–515. https://doi.org/10.1080/17470218.2015.1136339

*United States v. United States Gypsum Co.*, 438 U.S. 422 (1978).

United States Attorney's Office for the Central District of California (2014). Two Men Who Helped Set Illegal Campfire In Angeles National Forest That Became Colby Fire Sentenced To Federal Prison. *Department of Justice*, https://www.justice.gov/usao-cdca/pr/two-men-who-helped-set-illegal-campfire-angeles-national-forest-became-colby-fire

United States Court of Appeals for the Third Circuit (Jan. 2024). Final instructions: Consideration of mental state. *Model Criminal Jury Instructions*. https://www.ca3.uscourts.gov/model-criminal-jury-table-contents-and-instructions

Verheyen, S., Hampton, J. A., & Storms, G. (2010). A probabilistic threshold model: Analyzing semantic categorization data with the Rasch model. *Acta Psychologica, 135*, 216-225, https://doi.org/10.1016/j.actpsy.2010.07.002

Verheyen, S., De Deyne, S., Dry, M. J., & Storms, G. (2011). Uncovering contrast categories in categorization with a probabilistic threshold model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1515,-1531 https://psycnet.apa.org/doi/10.1037/a0024431

Vilares, I., Wesley, M. J., Ahn, W. Y., Bonnie, R. J., Hoffman, M., Jones, O. D., Morse, S.J., Yaffe, G., Lohrenz, T., & Montague, P. R. (2017). Predicting the knowledge–recklessness distinction in the human brain. *Proceedings of the National Academy of Sciences, 114*, 3222-3227, https://doi.org/10.1073/pnas.1619385114

Wagner, K., Jergens, J., & Barner, D. (2018). Partial Color Word Comprehension Precedes Production. *Language Learning and Development, 14*(4), 241–261. https://doi.org/10.1080/15475441.2018.1445531

Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics, 29*, 429-460.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Yaffe, G. (2011). *Attempts*. Oxford University Press.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage, 40*(4), 1912–1920. https://doi.org/10.1016/j.neuroimage.2008.01.057

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120*(2), 202–214. https://doi.org/10.1016/j.cognition.2011.04.005