

# MORAL EVALUATIONS OF POLICE AND CIVILIAN USE OF FORCE

Bradley D. Celestin

Submitted to the faculty of the University Graduate School in partial fulfillment of the requirements for the degree Doctor of Philosophy in the Department of Psychological & Brain Sciences, Indiana University  
May 2019

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

---

John K. Kruschke, Ph. D

---

Edward R. Hirt, Ph. D

---

Eliot R. Smith, Ph. D

---

Robert J. Rydell, Ph. D

April 17th, 2019

Copyright © 2019

Bradley D Celestin

## Acknowledgments

This research, and my entire journey toward the degree in which it culminates, would not have been possible without all of the incredibly supportive people in my life. Like most protracted challenges, success in a Ph.D. program is largely the result of being surrounded by a thoughtful and generous group of mentors, colleagues, friends, and family members. These people have propped me up, improved my ideas and graciously given away their own, and enabled me to persist when things have been really hard and I would otherwise have failed. Whatever degree of success or impact ultimately results from this work is attributable to their amazing efforts on my behalf and their presence throughout this intellectual pilgrimage. I am a better thinker and scientist, and a better person, because of their influence.

I particularly want to thank the members of my dissertation committee for their persistent guidance and advice, as well as their graciousness and feedback throughout the entirety of graduate school and the dissertation process. My lab colleagues, Seung Joo Yang, Sam Regas, and Binyan Li have endured countless hours of discussion about these topics and provided many thoughtful insights. I am also especially grateful to my former lab mate and partner in crime for several years, Dr. Torrin Liddell, for his enduring friendship and his willingness to have enthusiastic conversations about morally-charged and often offensive topics that would have been off-putting to anyone else. And then, there is my advisor, Dr. John Kruschke, for whom I find it difficult to generate words that adequately express my gratitude. I have learned more from John than I have from anyone. He is the sort of mentor that is exemplified in books and movies about prodigious scientists and teachers, the type of person who exudes an intelligence and curiosity that infects everyone around him. As much as I admire John's vast intellectual contributions and capabilities and appreciate his capacity to push me to expand the bounds of my own knowledge and abilities, the quality that I most aspire to emulate is his genuine kindness and deeply-driven care for his students. I have never doubted that John is in my corner, and that has made all the difference.

Finally, none of this would have been possible without my amazing family. For my entire life, my Mom and Dad (Marsha and Frank) have always encouraged me to pursue the things I am passionate about, even when my goals were lofty (aspiring to become a famous musician) or weird (starting a business performing "professional" magic shows in 5th grade). They, and my other family members, have relinquished many things to facilitate this pursuit, most notably time with me, their daughter-in-law, and their grandson. And that leaves my innermost circle for the closing and most important acknowledgment. The people who make my life meaningful moment to moment are my wife, Amber, and my son, Huxley. Words cannot sufficiently express how much Amber and Hux have sacrificed to push me forward on this path or how fiercely I love them. Without their constant support and enduring love, I would never achieve success in my goals, academic or otherwise. I am exceedingly glad that we three have gone on this adventure together, and I am eternally grateful for their constant presence and unending support; they perpetually inspire me to squeeze the happiness out of every moment in life.

Bradley D Celestin

Moral evaluations of police and civilian use of force

At least 80% of law enforcement agencies in the United States rely on some form of a force “continuum”, which is a scale that categorizes forceful actions systematically according to their severity, and enables officers to make expeditious choices about appropriate responsive force in real time. It is unclear, however, how much civilian evaluations align with or diverge from such a severity scale. The studies in this dissertation derive precise severity estimates for a representative set of forceful actions that span the entire range of possibilities. These scale values reveal that lay perceptions of police and civilian force substantively differ from formal models; non-lethal actions cluster together homogeneously, while actions at the extremes are relatively spread out. Moreover, actions that are not role-normative for police are perceived as especially morally severe, while more normative actions are less morally severe. Participant ratings of morality and physical magnitude are also moderated by participant beliefs about police legitimacy and other individual differences (e.g., race, gender, political affiliation, etc.). Additional studies demonstrate that lay evaluators generally expect police to use less force than that with which they are confronted, even when they are confronted with lethal force. This expectation is in direct conflict with the “one level above” heuristic rule commonly used by law enforcement, and also with the supreme court guidelines for evaluating when lethal police force is justified. Taken together, these results imply that the groundwork for disagreement about the legitimacy of police and civilian actions may be partially rooted in the differential way that action severity is perceived by law enforcement relative to civilian observers. They also offer insight into specific ways that educational outreach and policy revision efforts can seek to better align lay and professional moral evaluations force.

---

John K. Kruschke, Ph. D

---

Edward R. Hirt, Ph. D

---

Eliot R. Smith, Ph. D

---

Robert J. Rydell, Ph. D

## Contents

<b>Chapter 1: Introduction</b>	<b>1</b>
<b>Chapter 2: Lay evaluations of police and civilian use of force: Action severity estimates</b>	<b>4</b>
Ordinal Levels of Force and Heuristics For Decision-Making . . . . .	5
Recent Criticism and Disparate Evaluations of Force . . . . .	8
Previous Research on Moral Evaluations of Use Of Force . . . . .	10
The Need For Scale Values For Representative Forceful Actions . . . . .	12
<b>Method</b>	<b>13</b>
Participants . . . . .	14
Stimuli . . . . .	15
Set of actions . . . . .	15
Vignette . . . . .	19
Measures . . . . .	20
Procedure and Design . . . . .	21
Software . . . . .	23
<b>Results</b>	<b>24</b>
Data Analysis . . . . .	24
Bayesian method . . . . .	26
Comparing Latent Scale Estimates and SOLOF Categories . . . . .	27
Comparing Physical and Moral Scale Estimates . . . . .	31
<b>Discussion</b>	<b>34</b>
Implications . . . . .	35

Normativity in Lethal Force . . . . .	36
Dealing with the Disconnect . . . . .	38
Limitations and Future Directions . . . . .	39
Conclusion . . . . .	40
<b>Chapter 3: Lay evaluations of police and civilian use of force: Moral and physical magnitude ratings of officer and civilian actions</b>	<b>42</b>
<b>Police and civilian perspectives on use of force</b>	<b>43</b>
Six ordinal levels of force (SOLOF) . . . . .	45
Individual differences: legitimacy beliefs . . . . .	46
Assaying lay perceptions of specific forceful actions . . . . .	47
<b>Method</b>	<b>47</b>
Participants . . . . .	47
Design . . . . .	48
Procedure . . . . .	50
<b>Results</b>	<b>55</b>
Data transformation and regression model . . . . .	55
Bayesian Methods . . . . .	58
Estimates of regression coefficients . . . . .	59
Officer punishment ratings . . . . .	61
Officer acceptability ratings . . . . .	63
Officer appropriateness ratings . . . . .	65
Officer force ratings . . . . .	69
Summary of officer action ratings . . . . .	71
Civilian punishment ratings . . . . .	72

Civilian acceptability ratings . . . . .	74
Civilian appropriateness ratings . . . . .	77
Civilian force ratings . . . . .	79
Summary of civilian action ratings . . . . .	80
<b>Discussion and Conclusions</b>	<b>81</b>
<b>Chapter 4: Lay evaluators expect police to use less force than civilians and more normative force</b>	<b>85</b>
<b>Force Continua and Heuristic Rules</b>	<b>85</b>
<b>Action Normativity and Moral Severity</b>	<b>86</b>
<b>Method: Studies 1a, 1b, and 1c</b>	<b>88</b>
Participants . . . . .	88
Design . . . . .	89
Procedure . . . . .	90
<b>Results</b>	<b>93</b>
Study 1a . . . . .	93
Study 1b . . . . .	96
Study 1c . . . . .	98
Extracting action normativities from inferential models . . . . .	100
<b>Method: Study 2</b>	<b>100</b>
Participants . . . . .	100
Design . . . . .	101
Procedure . . . . .	103

<b>Results</b>	<b>106</b>
Explicit Task . . . . .	106
Implicit Task . . . . .	109
<b>Discussion</b>	<b>113</b>
<b>Chapter 5: Conclusions</b>	<b>117</b>
<b>References</b>	<b>121</b>
<b>Curriculum Vitae</b>	

## **Chapter 1: Introduction**

Policing in the United States has changed dramatically since the first publicly-funded police agency was formed in Boston in 1838. Following President Hoover's Wickersham Commission in 1929, police agencies across the U.S. were driven (albeit slowly) toward professionalization, and the goals of acting and being perceived as politically neutral organizations who imposed the rule of law fairly became increasingly prominent (Waxman, 2017). Alongside this general movement, law enforcement use of force procedures and public expectations have evolved over the past century, and they continue to evolve into the future as new technologies for applying force (e.g., tazers) and observing force (e.g., body worn cameras) develop and become sufficiently cost effective to be broadly implemented. As technology and culture expand and refine the abilities of police to more precisely apply force in tandem with the abilities of the public to observe instances of applied force in high fidelity, the importance of ensuring that the perspectives of police and the public regarding what constitutes justifiable force relative to excessive force is increasingly acute (Boivin, Gendron, Faubert, & Poulin, 2017; Culhane, Boman, & Schweitzer, 2016).

Despite this growing need to ensure that police and citizen evaluations of use of force maximally converge, very little empirical research has investigated how lay individuals evaluate police and civilian use of force in direct comparison to the procedures and rules commonly used by law enforcement to categorize and select appropriate forceful actions on the ground (Gerber & Jackson, 2017). Since the 1980s (Stetser, 2001, pp. 36-37), police in the U.S. and Canada (and many other western countries) have increasingly relied on binning forceful actions into levels according to their severity. For example, low-level actions such as the mere presence of police or polite verbal interactions are generally categorized at the lowest levels of severity, while high-level actions such as using a handgun or rifle are classified at the highest severity levels. These classification structures are fre-

quently referred to as a "force continuum" in law enforcement communities and policies. Although there is considerable variety across agencies regarding the precise details of these force continua, the basic structure remains similar and is used by the vast majority of law enforcement agencies in the U.S. (Terrill, Paoline, III, & Ingram, 2011). Accompanying these rank-ordered action bins, law enforcement agencies commonly use a heuristic rule such as "officers can employ one level of force above civilian resistance" in order to facilitate expeditious decision-making and quick resolution of violent encounters. Despite the ubiquity of this continuum schema among law enforcement, it is unclear whether lay citizens use a similar structure to categorize the severity of actions, or whether their expectations tend to agree with the "one above" heuristic or favor some other trend (e.g., using equivalent or less force).

Thus, the primary goal of the research presented in this dissertation is to measure lay perceptions of police and civilian use of force using representative actions that span the entire range of force severity, and to assess the extent to which those lay perceptions are similar to and/or different from the framework of formal force continua used by law enforcement. The bulk of this dissertation is comprised of three distinct but conceptually unified manuscripts that have been written in collaboration with my graduate advisor, Dr. John K. Kruschke. The first and second manuscripts are presented in Chapters 2 and 3, respectively, and they describe unique analyses of a novel experimental paradigm that combines separate undergraduate and Amazon Mechanical Turk (MTurk) samples. In Chapter 2 (i.e., manuscript 1), the analyses focus on precisely estimating the perceived latent moral and physical severities of a representative set of officer and civilian actions that span the entire range of continuum severity. This analysis yields intriguing comparisons with the continuum framework and a novel psychometric scale. Chapter 3 (i.e., manuscript 2) focuses on the ratings of morality and physical magnitude, and their interaction with participant beliefs regarding police legitimacy, as well as with one another. These findings shed light on

how moral and physical magnitude judgments change according to different levels of officer and civilian action severity and different levels of legitimacy beliefs. Finally, Chapter 4 (i.e., manuscript 3) describes the results from several studies designed to measure which actions lay evaluators select as the most appropriate officer responses to varied levels of civilian resistance, as well as the extent to which the role-normativity of specific actions corresponds to differential perceptions of moral relative and physical severity. These results provide insight into whether the perceptions of lay evaluations align or diverge from the "one above" heuristic rule, as well as whether action normativity can partially explain discrepancies in moral severity judgments.

## **Chapter 2: Lay evaluations of police and civilian use of force: Action severity estimates**

Criminal violence is an unfortunate but perennial aspect of human behavior within societies. And until the root causes of criminal violence are fully understood and resolved, control and mitigation of criminal violence require some form of socially-sanctioned counter-violence. But average citizens lack the ability and equipment to apply violence in a measured and precise manner, and even if citizens possessed the necessary skills and tools, doing so requires that they expose themselves to elevated risk and eliminates the neutrality of third-party intervention. Consequently, the bargain that citizens of modern democratic societies make is to willingly cede the authority to use physical force to a ruling government that exercises a moral monopoly on the use of violence (with limited exceptions for immediate defense of self or others). In exchange, citizens gain the ability to call on specialized government agents to use force on their behalf for two primary purposes: (1) to maintain law and social order (e.g., to effect arrests of resisting criminal suspects); and (2) to provide skilled protection from other violent threats that exist in society (e.g., active shooters, hostage takers, terrorists, etc.). Ultimately, the result of this social agreement is a more peaceful and just society than one controlled by cyclical revenge and domination by the largest, strongest, and most brutal (Pinker, 2011). Nevertheless, the tragic nature of any use of violence demands that those who employ it professionally do so with maximum solemnity, care, and public accountability. The legitimate use of violence is one of the most awesome powers democratic citizens give to their governments (Friedman, 2017), and in modern societies, this responsibility is exclusively entrusted to commissioned law enforcement officers by municipalities, states, and federal governments, at the behest of the citizens they represent.

## **Ordinal Levels of Force and Heuristics For Decision-Making**

Across the United States (as of 2008), there are approximately 18,000 state and municipal police agencies, employing in excess of 750,000 full-time law enforcement officers (Reaves, 2011). In addition, there are at least 24 federal agencies that employ roughly 120,000 full-time law enforcement officers (Reaves, 2012). At least 80% of those agencies utilize a detailed list of rank-ordered force options (commonly referred to as a "force continuum") as part of their official use-of-force policy (Terrill et al., 2011). These formal classifications provide policy-level guidance for how the officers employed at a given government agency are allowed to use force (Terrill et al., 2011).

Rank-ordered classifications of force are also utilized within the social scientific literatures, where police and civilian actions are sometimes separated into distinct lists of role-specific categories. Researchers have used these lists to calculate the proportionality of police use of force relative to civilian resistance by quantifying a mathematical difference score, dubbed a "force factor" (Terrill, Alpert, Dunham, & Smith, 2003; Alpert & Dunham, 2004). An advantage of this method is that it explicitly quantifies the moral evaluation of police and civilian violence (moral in the sense that it indicates whether force is insufficient, appropriate, or excessive), and there is evidence that such an algorithmic approach can yield reasonable inter-rater reliability between two groups of evaluators (Hickman, Atherley, Lowery, & Alpert, 2015). However, the core of the force factor approach requires a clear and uncontested assignment of actions to categorical severity levels, which is a condition that does not reflect the variety of actual law enforcement classification policies. Moreover, the courts have explicitly ruled that evaluations of specific instances of police use of force are so idiosyncratic that they must be evaluated individually with all relevant facts and circumstances taken into account (*Bell v. Wolfish*, 441 U.S. 520, 1979). Thus, while strictly algorithmic approaches to the moral evaluation of force may be useful for

researchers, they are insufficient tools for retrospective courtroom decision making. Moreover, the legal precedent the courts have set requiring legitimate police use of force to be "objectively reasonable" (*Graham v. Connor*, 490 U.S. 3, 1989) is too ambiguous to be useful for officers who must make prospective decisions in real time under time pressure and threat.

Acknowledging that law enforcement agencies and researchers have approached force classification and moral evaluation similarly due to the nature of use-of-force decision-making, it is, nevertheless, unclear whether lay people making evaluations of force are framing force in terms of a sequence of categories (like police) or doing anything resembling a force-factor calculation (like researchers). Even among professionals, there is variation in the precise structure of use-of-force policies. There is no absolute consensus between law enforcement agencies regarding a uniformly accepted list of rank-ordered force options, and policies can differ in terms of the exact placement of some actions (Terrill et al., 2011). For example, conducted energy devices (e.g., tazers) are placed below hard empty hand techniques (e.g., punches or kicks) by some agencies, but the order is reversed by other agencies. Additionally, the number of rank-ordered categories contained within a given use-of-force policy varies between 3 and 9 levels, depending on the law enforcement agency (Terrill et al., 2011). Still, nearly three quarters of all law enforcement agencies use a 5 or 6 level classification (Terrill et al., 2011), and many researchers (Terrill et al., 2003) recommend a six-category list for evaluation of police use of force and civilian resistance. Thus, despite the existence of minor variations, most use-of-force policies exhibit a very similar structure and categorization scheme that is mirrored in research paradigms. Therefore, the current study utilized a six-category list which is depicted in Table 1 and described more completely in the method section. As noted above, it is common parlance among law enforcement professionals and in the criminal justice literature to refer to these force option lists as a "force continuum," but in actuality they are ordered categories that

Table 1

*Six ordinal levels of force (SOLOF) scale.*

<b>Level</b>	<b>Description</b>
1	Cooperation, physical proximity, polite dialogue.
2	Strong verbal interaction.
3	Open hand control, defensive resistance, non-deadly weapon use with minimal injury potential.
4	Closed hand techniques.
5	Intermediate or aggravated force, non-deadly weapon used with moderate injury potential.
6	Deadly force.

do not contain information about the distance between levels of force nor about differences within categories. For the purposes of the current research, we refer to this list as the *six ordinal levels of force (SOLOF) scale*.

Law enforcement agencies routinely use a SOLOF-style scale in combination with a heuristic rule to train their officers to quickly identify appropriate levels of responsive force (Garner, Schade, Hepburn, & Buchanan, 1995; Terrill & Paoline, 2013), though it is worth noting that there is considerably less inter-agency consensus about these rules than about the SOLOF-scale itself (Terrill et al., 2011). For example, a policy might dictate that officers are allowed to be one level of force above a resisting civilian in order to maintain control and to minimize the duration of the encounter, as well as to reduce the potential for officer and civilian injury that could result from an extended interaction (e.g., if a civilian punches an officer, an agency might authorize the officer to respond by striking the civilian with a collapsible baton rather than engaging in a protracted punch-for-punch exchange). By memorizing their agency’s SOLOF scale and heuristic rule, officers can presumably make consistent choices about the most appropriate level of force to use in real-time encounters with resisting civilians. This strategy is especially useful for professional law enforcement officers who can converse with colleagues about hypothetical situations (a.k.a.

"if-then dialogues") and thereby enhance intra-agency consistency.

One problem with the formal SOLOF measure endorsed by Terrill et al. (2003), however, is that it contains ambiguous categories. For example, it is unclear what distinguishes "active resistance" from "aggravated active resistance." Consequently, there is room for disagreement regarding which SOLOF category best accommodates a specific forceful action. It is possible that the widespread conflict surrounding police use of force that is observable in the current national discourse is rooted in this sort of disagreement about action binning, and it is also possible that utilizing a SOLOF list and heuristic rule are largely unrepresentative of what lay people are actually doing when they make moral judgments about use of force. Because there is so little systematic research investigating lay evaluations of specific forceful actions, the extent to which civilians utilize something like a SOLOF-style list and heuristic rule is currently unknown.

### **Recent Criticism and Disparate Evaluations of Force**

Because of the delegation of legitimate use of violence to law enforcement agencies, police use of force in the United States is perennially under scrutiny. In recent years, police use of force has been heavily criticized (Friedersdorf, 2015, 2017b). Videos of deadly force involving police and minorities have been prominent in the national news (Handelman, 2014), and harsh criticism of police has consequently proliferated (Coates, 2015). Some of this criticism is clearly warranted and is valuable as a catalyst for reforming unjust or insufficiently skilled practices; law enforcement agents are rightly held to a high standard of excellence, especially when employing violence (Selby, 2016). However, some argue that a disproportionate amount of police violence in the media has resulted in public perceptions that the use of force by police is a common occurrence (Sullivan, 2017) when it is actually quite rare relative to the total number of police-citizen encounters (Hyland, Langton, & Davis, 2015; Lemoine, 2017). Moreover, a careful examination of the rhetoric

(both anti-law enforcement and pro-law enforcement) in dialogues about police use of force reveals a regular disconnection from the facts of particular cases, and an unwillingness to wade into the full nuance of the details (Loury, 2015; McWhorter, 2016). Because police use of force is such a highly moralized topic, it is especially vulnerable to ideological distortions, extreme levels of conviction, and confirmation biases that influence the way it is evaluated. In general, moralized attitudes have unique characteristics that include increased intolerance of others with dissimilar views, reduced cooperation, and an inhibited ability to resolve disagreements (Skitka, Bauman, & Sargis, 2005). So, to the extent that moral polarization surrounding police and civilian violence is increasing, third-party observers on opposite sides are prone to oversimplify in favor of their prior moral mandates despite the complexity and nuanced details that are frequently present at the level of individual use-of-force cases.

There are many ways to parse instances of police use of force. For example, observational and experimental researchers have routinely used dichotomous (e.g., "reasonable" vs. "excessive") or trichotomous (e.g., "no force," "necessary force," "improper force") categories as measurement tools (Worden, 2015; Terrill & Mastrofski, 2002). In other domains, it has been argued that police use of force can be rationally divided into categories that include (1) excessive: cases of egregious, unlawful abuse of authority, (2) negligent: cases of unintentional/non-malicious officer overreaction, often coupled with a lack of sufficient training, and (3) necessary: cases in which law enforcement actions yield tragic outcomes but are, nevertheless, appropriate (Harris & Loury, 2016). Notice that each of these various categorization strategies decomposes use of force in moral terms (i.e., in terms of perceived right and wrong) rather than in more descriptive/objective terms like raw physical severity, potential to cause injury, etc. In other words, although at face value it seems like questions such as "how many times did police in the U.S. use excessive force in the past year?" are fact-based questions demanding a precise numeric answer, in reality they are value-

driven questions that implicitly require binning complex sets of facts into sharply divided and mutually exclusive moral categories (Balko, 2016). This sort of categorization process presupposes a particular set of shared values and experiences (Graham, Haidt, & Nosek, 2009; Haidt, 2012; Friesdorf, Conway, & Gawronski, 2015). In ambiguous cases, whether or not an individual judges an officer to have acted, for example, "too quickly" or "at the appropriate moment," likely depends, in part, on many individual differences including the evaluator's political leaning, gender, personal and vicarious experiences with officers, race, ethnicity, etc. As a result, for many of the cases of police violence recently highlighted by the media, there is widespread social disagreement about whether the details of that particular event belong in the "necessary" or "excessive" bin; people watching the same video or reading the same set of facts seem to be coming to starkly different moral conclusions (Kahan, Hoffman, & Braman, 2009; Pew, 2014).

Parsing the extent to which these disparate evaluations are driven by individual differences or by the actions themselves is a challenging but important goal. To the extent that a society values improving trust between citizens and police, and increasing police efficacy and perceived legitimacy (goals that are worth pursuing, Tyler, 2004; Tyler, Goff, & MacCoun, 2015), it is crucial to develop a clear and detailed empirical understanding of how third-party observers make moral evaluations of use of force, and the ways in which their perceptions align or differ from one another. Despite this pressing need, experimental research into third-party evaluations of use of force is limited in both quantity and scope.

### **Previous Research on Moral Evaluations of Use Of Force**

A variety of previous research has investigated different aspects of police use of force. For example, one body of work has focused on describing the actual rates and correlates of use of force that exist in U.S. policing (Klinger, 1995; Fryer Jr., 2016). These studies analyze real-world data sets (i.e., reports of actual use-of-force incidents) that are usually

sampled from one or a small number of police agencies, and they often model relationships between a selection of independent variables (e.g., policing style, agency culture, officer/suspect race, etc.) and rates of "excessive" force. Other research has investigated general public attitudes toward police use of force using a variety of large-scale methods. Some studies make inquiries about first-person experiences (e.g., the Police-Public Contact Survey, Eith & Durose, 2011), while others rely on a small subset of questions extracted from routinely collected, non-specific datasets such as the general social survey (D. Johnson & Kuhns, 2009; Silver & Pickett, 2015). While all of these studies are scientifically valuable, they do not speak directly to how observers make moral evaluations of use of force, or how observers perceive the severity of specific forceful actions.

In another corpus of research using experimental paradigms, participants have been presented with forceful actions that are divided, a priori, into justified and unjustified categories (Bradford, Milani, & Jackson, 2016; Gerber & Jackson, 2017). These studies tend to model associations between ratings of acceptability and individual-difference measures. One variant of these studies restricts the focus to evaluations of deadly force exclusively (Fridell & Binder, 1992; Skinner & Haas, 2016). Such an approach is useful for understanding deadly force but is unlikely to generalize to all levels of force, in part because deadly force is extreme and rare (as rare as a person being struck by lightning, Eith & Durose, 2011; Lemoine, 2017)). Other experimental research in psychology has used time-restricted decision-making tasks (e.g., the "shooter task"), but these studies have almost exclusively focused on biases by police and civilians in the application of force (Correll et al., 2007; Hall, Hall, Perry, & Hall, 2016) rather than on moral evaluations of force from the perspective of lay observers.

All of these previous approaches were driven by top-down strategies that assumed stimuli reflect a moral consensus about the reasonableness or excessiveness of a particular forceful action, for example, that if a civilian is armed then lethal force is acceptable but if the

civilian is unarmed then lethal force is excessive (Nix, Campbell, Byers, & Alpert, 2017). However, this method of dividing up force evaluations is likely to be misaligned with the nuance and complexities of real assessment (Klinger & Slocum, 2017). As far as we are aware, none of the studies in the relevant literature has systematically measured perceptions of representative actions across the entire spectrum of possibilities. Because official classifications of forceful actions have a rank-ordered structure, and heuristic rules place legitimate actions relative to other actions within that structure, it is important to assess how lay perceptions of specific actions across the entire range of force severity do or do not align with law enforcement classifications.

Thus, our primary focus for the current research was estimating the underlying moral and physical severity of a wide variety of representative forceful actions in order to generate an emblematic set of psychometric scale values. These scales enable subsequent comparisons between lay evaluations and the structure of SOLOF models. They also provide a rigorous foundation that makes it possible for future research to ask questions about the general level of severity lay evaluators expect police to use, and, thus, they generate a framework for evaluating common heuristic rules used by police and comparing them to lay expectations.

### **The Need For Scale Values For Representative Forceful Actions**

To measure whether lay force evaluations resemble a SOLOF-style scale, and whether lay evaluations agree with a heuristic such as "officers can use force that is one level above civilian resistance," it is important to know the perceived severity of a wide range of representative forceful actions. Such baseline knowledge is required because these heuristics are, by nature, *relative* references to other possible actions. To know that people generally expect officer force to be above civilian force (or below or equivalent to), one must know where a wide range of actions are located in terms of their perceived severity. It

cannot merely be presumed that actions officially categorized at a given SOLOF level are, in fact, perceived by lay evaluators the same way. Thus, if we want to know whether the underlying general tendency of lay evaluators judging a specific action sequence matches or diverges from a particular law enforcement heuristic, we must first have detailed knowledge about the precise perceived severity of the actions in question and a sufficient number of surrounding actions with which to make relative comparisons.

The underlying severity inherent in any given action has at least two components: a raw physical component that reflects descriptive severity, and a moral component that is overtly prescriptive. These components roughly correspond to David Hume's famous is-ought distinction (Hume, 1738/2003) in that the descriptive component reflects how much force there is, while the prescriptive component reflects how much force there ought to be. For this reason, in the current study, we estimated psychometric scale values for a wide range of forceful actions using both descriptive and prescriptive dimensions.

As explained in detail below, we presented participants with minimal vignettes describing an officer and civilian encounter that involved one officer action and one civilian action. The actions in these paired vignettes spanned the complete range of force options. Participants rated the officer and civilian actions on moral dimensions (i.e., acceptability, appropriateness, and punishability of the action) and a physical dimension (i.e., how much force). These ratings were used to derive underlying moral and physical scale values for each action.

## **Method**

All of the experiments presented in this article were approved by the Indiana University Institutional Review Board in Bloomington, Indiana.

## Participants

To obtain a diverse sample, we collected data from both the Indiana University (IU) undergraduate subject pool and Amazon Mechanical Turk (MTurk). To verify that the two different subject pools did not differ in their ratings, we gathered reasonably large samples from IU and MTurk and conducted the analysis described below separately for each sample. Results showed virtually identical model descriptions, and therefore we combined the samples together in order to increase the precision of the estimates. Separate analyses for IU and MTurk subject pools are available in the supplementary material (<https://osf.io/h6jg3>).

A total of 411 participants were recruited through MTurk and 395 participants were recruited through the IU undergraduate subject pool. Each participant was presented with a consent form explaining the procedure for the study and the participation requirements, which included being a current resident of the United States, at least 18 years old, and a fluent speaker of the English language. Thirty eight participants were excluded from the MTurk sample for failing to participate from within the US (as verified by IP address geolocation), and 23 were additionally excluded because they failed more than one of the random manipulation checks or because the mean absolute deviation (MAD) of their ratings for the entire experiment exceeded 3 standard deviations below the MAD for all MTurk participants (i.e., their responses hardly varied across questions or items, suggesting they may have left the response slider at its default level for most trials). Participation within the US was guaranteed with the IU sample, but 78 participants were excluded for failure of more than one manipulation check or if the MAD exceeded 3 standard deviations below the MAD for all IU participants. This yielded a total of 350 MTurk participants and 317 IU participants, for a combined total of 667 participants.

Of the included three hundred and fifty MTurk participants, 196 (56%) identified as female, 152 (43.4%) identified as male, and 2 (0.6%) identified as other. 2 (0.6%) identified

as American Indian/Alaska Native, 23 (6.6%) identified as Asian, 27 (7.7%) identified as Black or African American, 20 (5.7%) identified as Hispanic or Latino, 269 (76.9%) identified as White, 8 (2.3%) identified as More than one, and 1 (0.3%) identified as Unknown.

Of the included three hundred and seventeen IU participants, 208 (65.6%) identified as female, 109 (34.4%) identified as male, and 0 (0%) identified as other. 1 (0.3%) identified as American Indian/Alaska Native, 48 (15.1%) identified as Asian, 27 (8.5%) identified as Black or African American, 12 (3.8%) identified as Hispanic or Latino, 217 (68.5%) identified as White, 11 (3.5%) identified as More than one, and 1 (0.3%) identified as Unknown.

## **Stimuli**

**Set of actions.** We began by crafting a role-neutral SOLOF scale that combined Terrill et al. (2003)'s *suspect resistance levels* and *officer force levels* (see Figure 1). Using this unified SOLOF scale, we generated a large list of actions (approximately 100) that could plausibly, or at least possibly, be committed by both an officer and a civilian, being sure to include multiple actions that rationally fit into each of the six SOLOF levels. It is worth noting that this initial list was generated loosely in the sense that it functioned as an initial brainstorm of options and as a mechanism to automatically exclude any obviously implausible actions. Importantly, the initial list did not attempt to generate a set of actions that are equally plausible for both an officer and a civilian because generating such a list is likely to be impossible; in the context of a police-civilian interaction, some actions are intrinsically more associated with police (e.g., handcuffing), and the same is true for other actions that are probably more associated with civilians (e.g., calling the other person a "pig" or pulling away).

After composing this larger list, we set out to narrow it down to a more manageable

<b>SOLOF Level</b>	<b>Action Statement</b>	<b>Action Name</b>
1	the [officer/civilian] stood near the [civilian/officer]	stoodNear
	the [officer/civilian] asked the [civilian/officer] how he was doing	asked
	the [officer/civilian] waved at the [civilian/officer] in greeting	waved
2	the [officer/civilian] yelled “get back” at the [civilian/officer]	getBack
	the [officer/civilian] shouted “get your fucking hands up!” / “I don't have to fucking listen to you!” at the [civilian/officer]	handsUp/ dontListen
	the [officer/civilian] said “if you touch me, you're gonna get hurt” to the [civilian/officer]	hurtThreat
	the [officer/civilian] called the [civilian/officer] a “thug” / “pig”	nameCall
3	the [officer/civilian] put handcuffs on/pulled away from the [civilian/officer]	handcuff/ pullAway
	the [officer/civilian] sprayed the [civilian/officer] with pepper spray	pepperSpray
	the [officer/civilian] shocked the [civilian/officer] with a tazer	tazer
	the [officer/civilian] twisted the arm of the [civilian/officer], forcing him to the ground	armbar
4	the [officer/civilian] punched the [civilian/officer] in the face	punchFace
	the [officer/civilian] kicked the [civilian/officer] in the stomach	kickStom
	the [officer/civilian] applied a chokehold to the [civilian/officer]	choke
	the [officer/civilian] headbutted the [civilian/officer] in the face	headbutt
5	the [officer/civilian] struck the [civilian/officer] on the leg with a collapsible baton/metal pipe	batonLeg/ metalPipe
	the [officer/civilian] shot/struck the [civilian/officer] in the chest with a beanbag shotgun round/baseball bat	beanbag/ ballBat
	the [officer/civilian] struck the [civilian/officer] in the stomach with a metal flashlight	flashlight
6	the [officer/civilian] shot the [civilian/officer] in the chest with a handgun	handgunChest
	the [officer/civilian] shot the [civilian/officer] in the head with a handgun	handgunHead
	the [officer/civilian] struck the [civilian/officer] with a motor vehicle	vehicle
	the [officer/civilian] slammed the [civilian/officer]'s head into the concrete curb	headSlam

*Figure 1.* All 22 actions are listed in the middle column, with their SOLOF level in the left column and abbreviated name in the right column. Colors indicate the SOLOF level, and will be used in subsequent graphs.

set of 22 actions per role while retaining at least three actions in each of the six SOLOF categories. The first thing we observed was that, although we were able to generate a large number of actions, there seemed to be a reasonably small core set of violent actions that can plausibly occur within a dyadic pair. Beyond the core set, additional actions are most readily generated by specifying severity modifiers such as the target of the action or the number of times an action occurred. For example, an officer could kick a civilian in the face, stomach, leg, etc., and could do so once or more than once. We immediately eliminated actions that occurred more than once in order to avoid the confound of comparing multiple actions (e.g., punching repeatedly) with one action. To deal with action target, we initially tried to specify actions in their most generic form (e.g., *the civilian punched the officer*), but realized that target clearly matters for judgments of action severity (e.g., kicking someone in the head versus kicking them in the hip), and that leaving the target ambiguous would likely increase participant confusion and decrease measurement precision (e.g., some participants might infer a default target if left unspecified). For that reason, we chose to specify action target (e.g., *the civilian punched the officer in the face*) in all cases except those in which the target was obvious (e.g., handcuffing) or seemed largely irrelevant. This target specification is congruent with law enforcement training procedures that incorporate target into judgments of force severity for intermediate severity levels (Jenkinson, Neeson, & Bleetman, 2006).

To further hone the set of experimental actions, we considered selecting officer actions according to their actual frequency of use, but there are several problems with that approach. First, police in the United States only use force in about 1-2% of all police-citizen encounters, so it is a relatively rare occurrence overall. Secondly, the distribution of force severity among those 1-2% of encounters is heavily skewed because most police use of force is comprised of low severity actions (Hyland et al., 2015; Davis, Whyde, & Langton, 2018). As a result, selecting actions according to how commonly they are used by police

in the field does not yield a set of actions that include representative options across the full spectrum of severity; instead, such an approach would produce many actions at low levels of severity and few or no actions at high levels of severity. Because we are attempting to map actions across the entire range of severity, selecting actions based on their raw or proportional frequency of use was not appropriate.

With all of these considerations in mind, we ultimately selected actions that were broadly distributed across all severity levels, that we believed were widely representative of real-life force options, and that also reflected within-category variety and maximal role-plausibility. Specifically, we first chose officer actions that are in common use among law enforcement agencies and frequently appear on the majority of use-of-force policies and in police training materials in the U.S. (Terrill et al., 2011). For example, it is standard practice at most police agencies to issue collapsible metal batons to officers for use as impact weapons (although some agencies distribute other impact weapons such as nunchucks or PR-24 rigid batons, collapsible batons are issued by the majority of modern law enforcement agencies because of their compact size that facilitates ease of carry without significantly reducing efficacy). It is worth noting that, prior to conducting this research, the first author (BDC) spent more than a decade as a commissioned law enforcement officer at municipal and federal levels. During a majority of that time, he functioned as a use-of-force instructor for a joint set of law enforcement agencies, as a subject matter expert on police use of force, and as a member of internal review boards evaluating the legality of specific instances of force application in the field. Thus, our action selection process was also informed by his prior experience and expertise.

Seventeen of the officer actions we chose could also be plausibly performed by a civilian. However, five of the 22 selected actions differed between the officer and the civilian. For example, one of the officer actions at level 5 of the SOLOF indicates that "the officer shot the civilian in the chest with a beanbag shotgun round." Because it is implausible

that any civilian would have access to such a specialized weapon, we matched this officer action with "the civilian hit the officer in the chest with a baseball bat," which makes the action more plausible for a civilian but is reasonably balanced with the general level of force caused by a beanbag shotgun round. We retained these five actions because we wanted to maximize diversity of action within each SOLOF severity level. For a full list of all actions, including the five actions that differ between officer and civilian, please see Figure 7.

Because there were five unique civilian actions matched with five other officer actions, the total number of *unique* actions was 27. Nevertheless, to avoid confusion and to reinforce that these five actions were matched across roles, we refer to the full set of "22 actions" throughout this article. Importantly, although we tried to match the five differing actions a priori in terms of severity, it is not crucial to the experimental design, statistical analyses, or results that these actions ultimately turn out to be well-matched between officer and civilian because each of the 27 actions has its own moral and physical severity estimates that are relative to all of the other actions. As we explain later, results show that the estimates for these five differing actions are remarkably similar.

**Vignette.** A minimal vignette was constructed to state each action pair without introducing superfluous contextual details:

While working a patrol shift, a police officer was dispatched to a reported crime, and in the course of the investigation, he came into contact with a civilian. During their subsequent interaction, the following actions took place \*in the order listed below\* (other actions may have occurred before, between, or after):

Two specific actions were then listed, in bold font. There was one action by the civilian and one action by the officer. Figure 8 shows an example of the display seen by participants.

**Measures.** For each vignette with its two actions, participants were asked to make four ratings of the second action, one rating per screen. Each rating was made using a slider with anchors marked as indicated in parentheses:

- Please rate how much **force** is involved in the officer's/civilian's action: (0 = No Force, 100 = Maximum Force)
- Please rate how **morally acceptable** the officer's/civilian's action is: (0 = Completely Unacceptable, 100 = Completely Acceptable)
- Please rate the extent to which the officer/civilian used an **appropriate** level of force (0 = Insufficient, 50 = Appropriate, 100 = Excessive)
- Please rate the extent to which the officer/civilian should be **punished** for his action: (0 = Not at all, 100 = Severely)

The rating of force was designed to assess evaluations of the severity of the raw physical force of the actions. The other three rating questions were intended to assess moral evaluations. Thus, these dependent measures allowed for a comparison between the physical

While working a patrol shift, a police officer was dispatched to a reported crime, and in the course of the investigation, he came into contact with a civilian. During their subsequent interaction, the following actions took place \*in the order listed below\* (other actions may have occurred before, between, or after).

1. **the civilian headbutted the officer in the face**
2. **the officer put handcuffs on the civilian**

Please rate how **morally acceptable** the **officer's** action is:

0 10 20 30 40 50 60 70 80 90 100

Completely Unacceptable Completely Acceptable

Submit Answer

Figure 2. Example screenshot from the experiment.

and moral dimensions of force severity. Only the appropriateness slider had its center point labeled, because we wanted to provide participants with an opportunity to rate officer and civilian actions as insufficient as well as excessive. It is worth noting that we consciously opted to probe physical severity with only one question in order to keep the total number of questions brief enough that the experiment duration remained tolerable for participants. Intuitively, physical severity is a less complex concept than moral severity, and we collected enough data to make precise estimates of scale values. Figure 8 shows an example of the display with the response slider. Although the structure of SOLOF-style policies is generally ordinal, we purposely used a continuous measurement scale in order to maximize precision and to allow participants to respond in an unconstrained way. Importantly, our continuous response scale does allow for actions to cluster into categories (e.g., six categories) if they are perceived that way, but it is also flexible if the underlying perceived severity does not map well into discrete categories.

### **Procedure and Design**

After informed consent, participants read a set of instructions explaining that they would be presented with a scenario that was representative of a real life interaction involving a police officer and a civilian, accompanied by a pair of actions that occurred during the encounter: one action by the police officer, and one by the civilian. Participants were also told that the order in which any given pair of actions had occurred was the order in which they were presented, and that other actions may have occurred before, between, or after the two listed actions. This latter information was included in order to allow for some ambiguity in the scenario as well as to increase the coherence of actions pairs that might otherwise be perceived as nonsensical. For example, if the civilian strikes the officer with a motor vehicle and subsequently the officer places handcuffs on the civilian, there must have been other things that occurred between the actions. Participants were informed that

they would be asked to make ratings of force, as well as appropriateness, moral acceptability, and punishment, for the second action that occurred during the encounter (i.e., always the action that occurred subsequent to the first action). Participants were also advised that there were no right or wrong answers and that we only wanted their intuitive, thoughtful responses. Following the instructions, participants provided ratings of general attitudes toward police (J. A. Johnson, Hogan, & Zonderman, 1981). Complete details are provided at <https://osf.io/h6jg3>.

Participants then viewed a series of 22 vignettes, each with a pair of actions. Officer-civilian action pairings were manipulated within subjects. Because the 22 actions could be performed by both the officer and the civilian in either order, the total number of possible action pairs was 968. Rather than present participants with all 968 action pairs, which would have caused fatigue and inattentiveness, we instead permuted the officer and civilian actions separately and then paired them together to form uniquely randomized action pairs for each participant. By using this method, we were able to present participants with only 22 action pairs while still ensuring that every participant saw the civilian and officer performing all 22 actions without any action repetition within a given role. Once the 22 action pairs had been selected, 11 random pairs were assigned to the civilian evaluation condition (i.e., the officer action was listed first followed by the civilian action, and participants were asked to rate the civilian action), and the remaining 11 pairs were assigned to the officer evaluation condition (i.e., the civilian action was listed first followed by the officer action, and participants were asked to rate the officer action). Finally, all 22 action pairs were randomly ordered so that ratings of civilian and officer actions would be interspersed as the experiment progressed. Participants were not informed about the set of possible actions or the constraints on their pairings. For each vignette, the four measures (force, moral acceptability, appropriateness, punishment) were presented in a random order, one at a time.

Eleven attentiveness checks were also inserted randomly for some of the action pairs.

A random number of these catch questions instructed participants to move the slider all the way to the left, and the remainder all the way to the right. These catch trials were crucial for identifying participants who were not providing thoughtful answers or not following instructions, a concern that is especially important when using crowd-sourced participant pools (Paolacci & Chandler, 2014).

Following the scenario ratings, participants answered a series of questions designed to assess their beliefs about police legitimacy (Gerber & Jackson, 2017), as well as questions measuring the number of personal and vicarious negative experiences with police (Weitzer, 2005). The focus of the current analysis is on developing precise psychometric estimates of the actions, so these individual-difference measures will not be discussed at length here (but see the Appendix for a detailed list of these measures). Further discussion of the individual difference measures can be found in Celestin and Kruschke (2019b). The experiment concluded with a series of demographic questions, including a single item measure of political affiliation (Iyer, Koleva, Graham, Ditto, & Haidt, 2012), and a debriefing. Complete details are available in the Appendix and the supplementary material (<https://osf.io/h6jg3>).

## **Software**

The experimental sequence was programmed using the JavaScript, html, and CSS programming languages, with the open-source libraries available in the jsPsych package (de Leeuw, 2015). The experiment was made available to participants online using an open-source psiTurk (Gureckis et al., 2015) or Apache server running on a desktop computer in combination with a SQL database or server-side PHP scripts. In addition, data cleaning, pre-processing, and analysis were done using the R programming language, RStudio (Team, 2016), JAGS (Plummer, 2003), runjags (Denwood, 2016), and Stan (Stan Development Team, 2016) software, as well as variants of the Bayesian analysis programs provided by Kruschke (2015).

## Results

### Data Analysis

In our analyses, the key dependent variables are the moral ratings of acceptability, punishment, excessiveness, and the rating of physical force. Upon visual inspection of the raw data, we observed that the distributions of responses on the 1-100 sliding scale often piled at the extreme high or low end of the scale. This is an indication that the scale itself was limiting participant responses, and that extreme responses would often exceed the scale limits if measured in an unconstrained manner. That is, piling of responses at the extremes indicates that participants would have sometimes responded even greater than 100 or less than 1 if they could have done so. Therefore, we treated responses less than 6 or greater than 94 as censored, meaning that they indicated some uncertain value at least that extreme. To make the distributions of uncensored values nearly normal, we applied a logit transformation to all responses (after dividing by 100). Complete details are in the R code available at (<https://osf.io/h6jg3>).

We modeled the (censored and logit transformed) ratings using linear regression in which each trial's rating was described as an additive influence of the latent scale values of civilian and officer actions. In other words, the ratings of the actions are a weighted combination of the underlying scale values of the actions. As an intuitive example, consider rating the acceptability of an officer action. As the severity of the officer action increases, the action's acceptability should (intuitively) decrease, implying that the regression coefficient on officer-action severity would have negative sign. But as the preceding civilian action increases in severity, the acceptability of the officer action should increase, implying that the regression coefficient on civilian-action severity would have positive sign. The analysis estimates the regression coefficients and simultaneously estimates scale values for the actions.

Formally, the regression model states that the predicted rating,  $\hat{y}$ , is a weighted combination of the civilian action severity,  $\gamma$ , and the officer action severity,  $\chi$ , as follows:

$$\hat{y}_{q,c,f} = \underbrace{\beta_{0,q}}_{\text{baseline}} + \underbrace{\beta_{1,q}}_{\text{influence of civilian action}} \cdot \underbrace{\gamma_c}_{\text{civilian action severity}} + \underbrace{\beta_{2,q}}_{\text{influence of officer action}} \cdot \underbrace{\chi_f}_{\text{officer action severity}} \quad (1)$$

where  $\hat{y}$  represents the predicted value of a rating (logit transformed),  $q$  indexes the rating question (1–8, i.e., the four types of ratings of both officer and civilian actions),  $c$  indexes the civilian action (1–22),  $f$  indexes the officer action (1–22),  $\gamma$  denotes the estimated civilian-action scale value, and  $\chi$  denotes the estimated officer-action scale value. Each of the  $\beta$  variables is an estimated regression coefficient. The regression coefficients and action scale values were estimated simultaneously. We analyzed the three social/moral questions (i.e., moral acceptability, appropriateness, and punishment) separately from the force question, and derived separate latent scale values for perceived moral and physical severities of the actions. Neither the moral nor the force questions were aggregated; rather, they were included in their respective analyses simultaneously. For example, the moral severity scale estimates are informed by all six social/moral questions (three officer, three civilian).

The estimation of scale values is analogous to structural equation modeling in which observed ratings covary with an underlying latent factor. Because of algebraic indeterminacies inherent in linear regression, the latent scale must be pinned down at two arbitrary points. We chose to fix the "waved" and "handgunHead" actions (see Figure 7 for full action descriptions) at 1.0 and 6.0, respectively, because they were likely to be the most extreme. The model did not constrain the other action estimates which were allowed to exceed the bounds of the two pinned action values if warranted.

We also analyzed a variety of other regression models that involved individual-difference

variables (e.g., beliefs about police legitimacy, negative police experience, political affiliation, participant race, participant gender, etc.) and various interactions of predictors. These alternative analyses can be found in supplementary material (<https://osf.io/h6jg3>). The resulting estimated scale values of actions were remarkably consistent across all of the various models. Because the focus of this article is the scale values of the actions, we report results from the simplest model that derives robustly stable scale estimates in Equation 1. A more elaborate model is reported in a companion manuscript (Celestin & Kruschke, 2019b) along with extensive discussion of the regression coefficients and trends in the ratings. Notably, although this companion manuscript and the results we report here are based on the same data, the analyses are mathematically different, focus on unique aspects of the models, and have substantively distinct implications. The current results focus on the latent severity scales implied by the data and how those scales relate to the SOLOF categories used by law enforcement. Whereas, in Celestin and Kruschke (2019b), the focus is on the ratings of appropriateness, punishability, etc., and the relation of those ratings to the predictors. For our purposes here, it is of primary importance to note that, despite the use of more elaborate models, the latent severity scales remain essentially the same, and so the results we report currently are robust to the inclusion of additional relevant predictors.

**Bayesian method.** We used Bayesian estimation of the parameters (Kruschke, 2015; Kruschke & Liddell, 2018a, 2018b). Bayesian methods are particularly useful in this application because they yield a complete posterior probability distribution over all of the parameter values in our models, including the regression coefficients and the scale values of the actions. We summarize the posterior distribution on a parameter by its mode (i.e., most probable value) and its *95% highest density interval* (HDI), which spans the 95% most probable parameter values. A narrower HDI indicates a more precisely estimated value.

Bayesian analysis requires specification of a prior probability distribution on the pa-

rameters. We used a very broad, vague prior relative to the scale of the data. Consequently, the prior distribution had negligible influence on the results.

Posterior distributions were computed using Markov chain Monte Carlo (MCMC) methods, in the R programming language, along with JAGS (Plummer, 2003) and runjags (Denwood, 2016) software, using variants of the Bayesian analysis programs provided by Kruschke (2015). All MCMC chains converged well, and the effective sample size for all parameter chains approaches 10,000.

## Comparing Latent Scale Estimates and SOLOF Categories

Table 2  
*Estimated scale values for actions*

	Action (officer/civilian)	Officer Force	Officer Moral	Civilian Force	Civilian Moral
1	asked	0.98 (0.80, 1.16)	0.99 (0.79, 1.17)	1.11 (0.91, 1.28)	1.01 (0.88, 1.16)
2	waved	1 (1, 1)	1 (1, 1)	1 (1, 1)	1 (1, 1)
3	stoodNear	1.06 (0.89, 1.24)	1.19 (1.00, 1.38)	1.24 (1.07, 1.41)	1.52 (1.40, 1.65)
4	nameCall	1.77 (1.62, 1.92)	2.56 (2.38, 2.70)	1.97 (1.82, 2.12)	2.60 (2.48, 2.70)
5	getBack	1.85 (1.70, 2.00)	1.51 (1.34, 1.70)	2.14 (2.00, 2.29)	2.47 (2.36, 2.57)
6	hurtThreat	2.18 (2.02, 2.30)	2.25 (2.07, 2.39)	2.40 (2.25, 2.53)	3.08 (2.97, 3.18)
7	handsUp/dontListen	2.58 (2.45, 2.72)	2.38 (2.19, 2.52)	2.45 (2.32, 2.59)	2.99 (2.88, 3.08)
8	handcuff/pulledAway	2.81 (2.69, 2.95)	2.05 (1.88, 2.22)	2.36 (2.22, 2.50)	2.44 (2.34, 2.55)
9	pepperSpray	3.52 (3.39, 3.64)	3.19 (3.03, 3.33)	3.82 (3.70, 3.96)	4.19 (4.09, 4.29)
10	armbar	3.72 (3.58, 3.84)	2.96 (2.80, 3.11)	4.14 (4.02, 4.28)	4.28 (4.18, 4.38)
11	flashlight	3.79 (3.67, 3.92)	3.84 (3.70, 4.00)	4.12 (4.00, 4.26)	4.24 (4.14, 4.34)
12	batonLeg/pipeLeg	3.82 (3.69, 3.96)	3.52 (3.36, 3.66)	4.36 (4.22, 4.48)	4.49 (4.38, 4.58)
13	punchFace	3.91 (3.77, 4.03)	4.02 (3.88, 4.17)	4.14 (4.01, 4.26)	4.27 (4.17, 4.37)
14	tazer	3.94 (3.82, 4.07)	3.27 (3.11, 3.41)	4.29 (4.16, 4.43)	4.53 (4.42, 4.62)
15	headbutt	3.98 (3.85, 4.10)	4.00 (3.86, 4.16)	4.19 (4.06, 4.31)	4.28 (4.18, 4.38)
16	kickStom	3.99 (3.86, 4.12)	3.98 (3.84, 4.13)	4.05 (3.93, 4.18)	4.25 (4.13, 4.34)
17	choke	4.07 (3.94, 4.20)	3.80 (3.64, 3.94)	4.30 (4.18, 4.45)	4.46 (4.36, 4.56)
18	beanbag/ballBat	4.20 (4.08, 4.35)	4.01 (3.86, 4.16)	4.44 (4.31, 4.58)	4.63 (4.53, 4.74)
19	headSlam	4.52 (4.39, 4.66)	4.97 (4.81, 5.12)	4.80 (4.66, 4.94)	4.93 (4.82, 5.03)
20	vehicle	5.07 (4.93, 5.22)	5.83 (5.69, 6.03)	5.25 (5.11, 5.39)	5.28 (5.17, 5.39)
21	handgunChest	5.43 (5.27, 5.58)	5.50 (5.35, 5.66)	5.70 (5.55, 5.87)	5.60 (5.49, 5.73)
22	handgunHead	6 (6, 6)	6 (6, 6)	6 (6, 6)	6 (6, 6)

*Numerical entries indicate the modal estimate and 95% HDI. Rows are ordered by ascending estimates of the "Officer Force" column. Please see Figure 7 for expanded descriptions of the action names.*

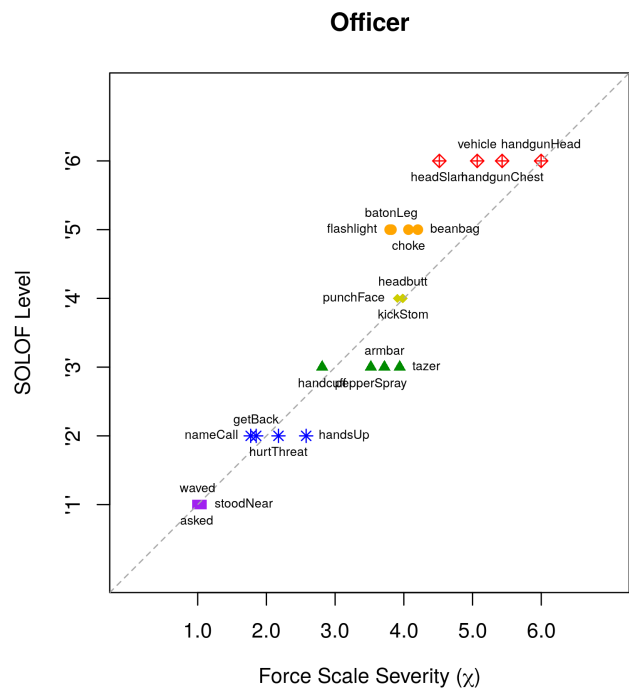
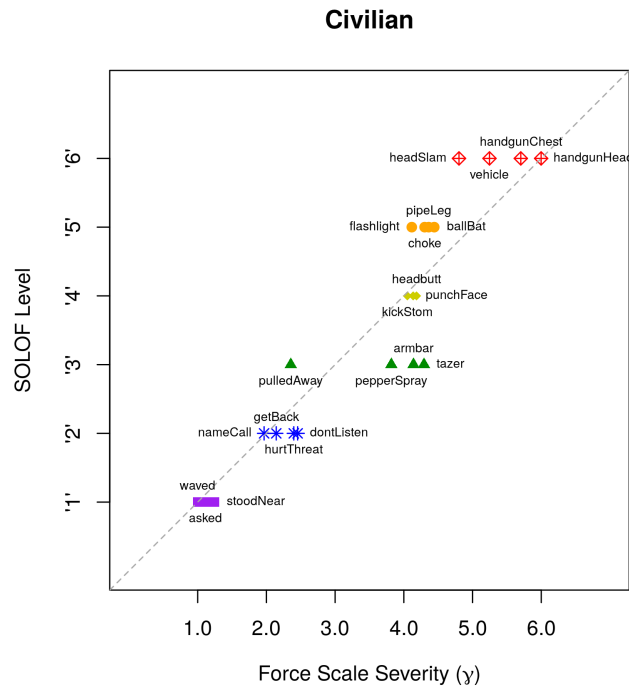


Figure 3. Latent scale values from ratings of force. The upper panel plots the latent action estimates for the civilian, and the horizontal axis denotes  $\gamma$  in Equation 1. The lower panel plots the latent action estimates for the officer, and the horizontal axis denotes  $\chi$  in Equation 1.

The latent scale values derived from the ratings of *physical force* are plotted in Figure 3 along the horizontal axis, while the vertical axis represents the SOLOF levels. Detailed numerical values of the latent scale values are shown in Table 2. From Figure 3 it can be seen that the latent severity of physical force tends to increase with the SOLOF levels, but there are also clear discrepancies between the latent severity of physical force and the SOLOF levels. Importantly, actions in SOLOF levels '3', '4', and '5' have latent force severities that overlap substantially, indicating that participants are not strongly distinguishing these actions from one another in terms of physical force. This lack of differentiation among the SOLOF levels '3', '4', and '5' is evident for both civilian and officer actions.

The latent scale values derived from the ratings of *moral* attributes are plotted in Figure 4. Detailed numerical values of the latent scale values are shown in Table 2. From Figure 4 it can be seen that the latent moral severity tends to increase with the SOLOF levels, but there are also clear discrepancies between the latent moral severity and the SOLOF levels. Importantly, for civilians, actions in SOLOF levels '3', '4', and '5' have latent moral severities that overlap substantially, indicating that participants are not distinguishing these actions from one another in terms of moral severity. For officers, SOLOF levels '4' and '5' overlap, with most actions of level '5' actually falling below the moral severity of actions in level '4'. It is notable that the perceived severities of SOLOF level '4' and '5' actions completely overlap. This suggests that once non-lethal actions reach a certain severity threshold, they are not morally very distinguishable from one another for officers or civilians.

Another prominent feature of the scale values is that one action that is in SOLOF level '3' has scale values among the SOLOF level '2' actions. The action in question is handcuffing for officers and its matched action, pulling away, for civilians (see Figure 7 for full action descriptions). The fact that handcuffing by a police officer and pulling away by a civilian are scaled similarly provides some confirmation that those actions were indeed

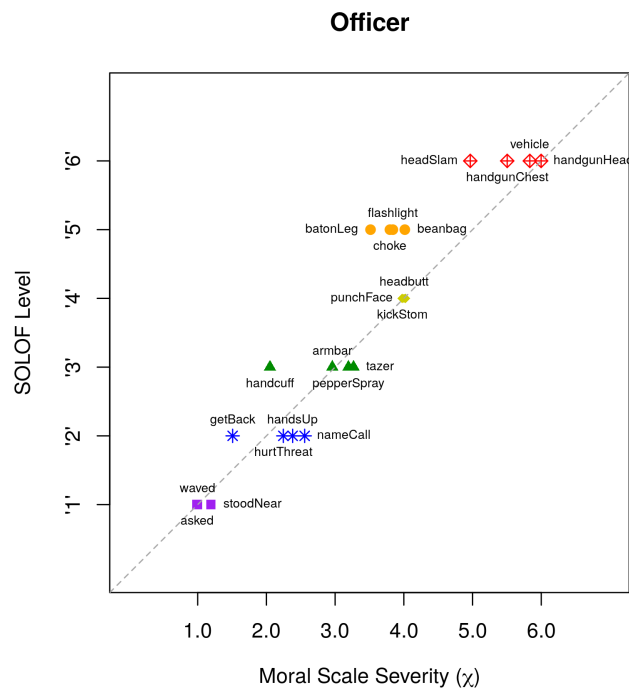
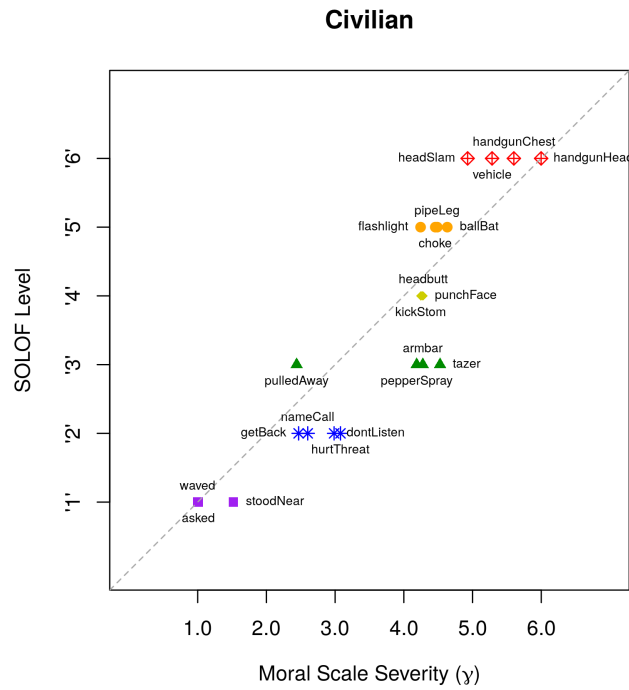


Figure 4. Latent scales for moral ratings (i.e., appropriateness, acceptability, and punishment). The upper panel plots the latent action estimates for the civilian, and the horizontal axis denotes  $\gamma$  in Equation 1. The lower panel plots the latent action estimates for the officer, and the horizontal axis denotes  $\chi$  in Equation 1.

well-matched as comparably severe actions. Handcuffing and pulling away are physical actions, not verbal actions, and therefore would be situated in SOLOF level '3' (see Table 2). Nevertheless, ratings of physical forcefulness and moral appropriateness place those actions less severely, in the range of the verbal actions of SOLOF level '2'.

### **Comparing Physical and Moral Scale Estimates**

The moral scale values are plotted against the force scale values in Figures 5 and 6. Figure 5 shows the estimates for the officer's actions, and Figure 6 for the civilian's actions. Recall that the least severe action was arbitrarily set at a scale value of 1.0 and the most severe action was arbitrarily set at a scale value of 6.0, and that all other actions were estimated relative to these two fixed points, so it is the relative not the absolute relationships that are informative in these figures. In particular, the diagonal line is displayed merely as a visual guide to proportionality and ordinality. The diagonal line should not be treated as a line of equality across scales because the scale values could be shifted or stretched on either axis. The segments intersecting the plotted points in Figure 5 and Figure 6 represent 95% HDIs. As indicated by the relatively short length of the HDI segments, all of the scale values were estimated with considerable precision on both force and moral dimensions.

One notable pattern in Figures 5 and 6 is non-monotonicities in moral severity as force severity increases. If moral severity were proportional to force severity, then an increase in force severity would always produce an increase in moral severity. But this does not always occur. For example, in Figure 5 the moral severity of handCuff is *less* than the moral severity of nameCall, even though the force severity of handCuff is *greater* than the force severity of nameCall (see Figure 7 for full action descriptions). These non-monotonicities reveal actions that are either perceived as more morally than physically severe or vice-versa.

The lack of differentiation for non-lethal categories observed in Figures 3 and 4 is also prominent in the relatively tight clustering of all three physical but non-lethal actions

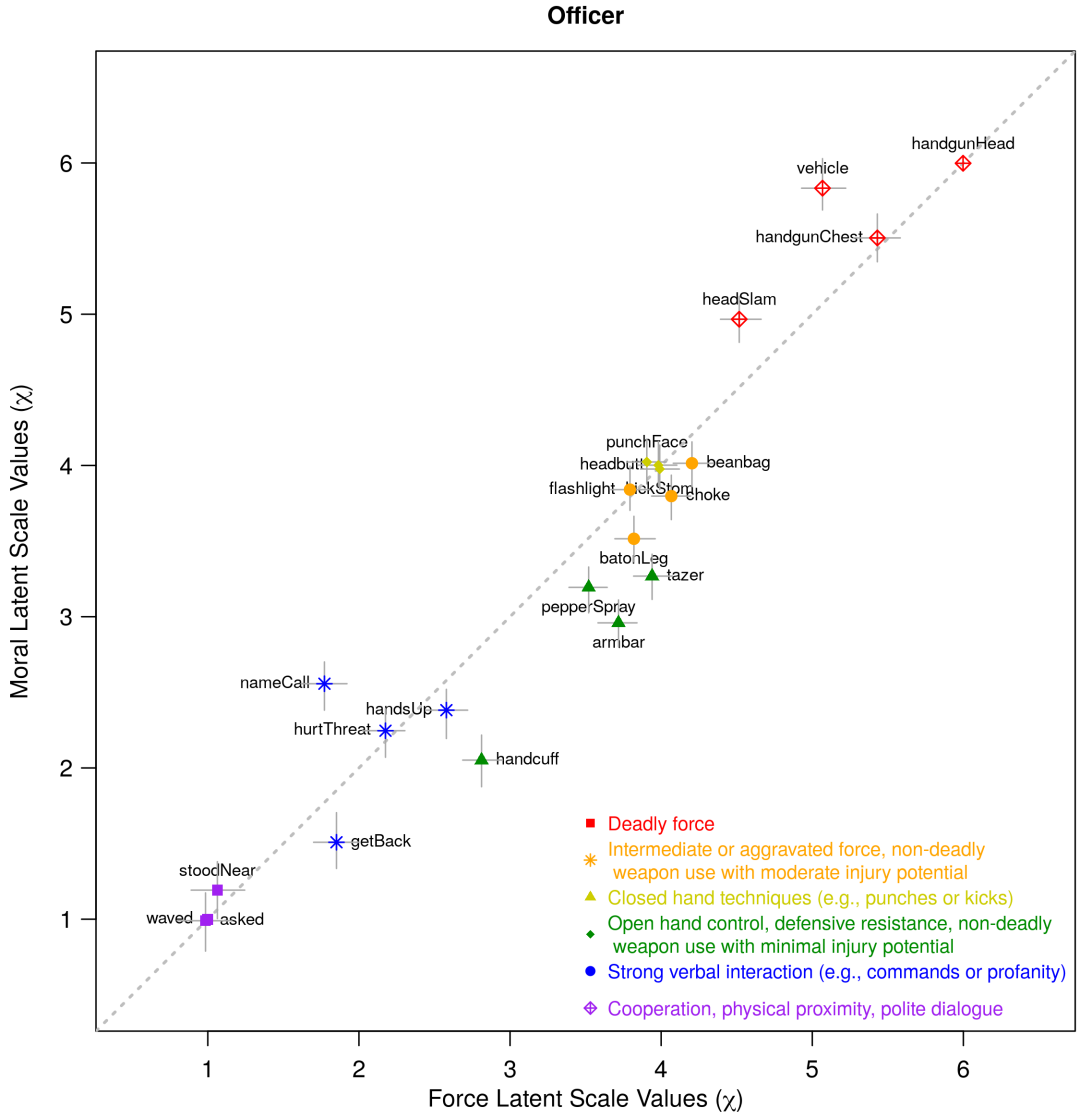


Figure 5. Posterior estimates of latent action scale values for the *officer*. The scale values for the moral questions are plotted on the vertical axis as a function of the estimated scale values for the force questions on the horizontal axis. Segments through each point represent the 95% HDI's. Each point is labeled by the abbreviated action name in Table 7.

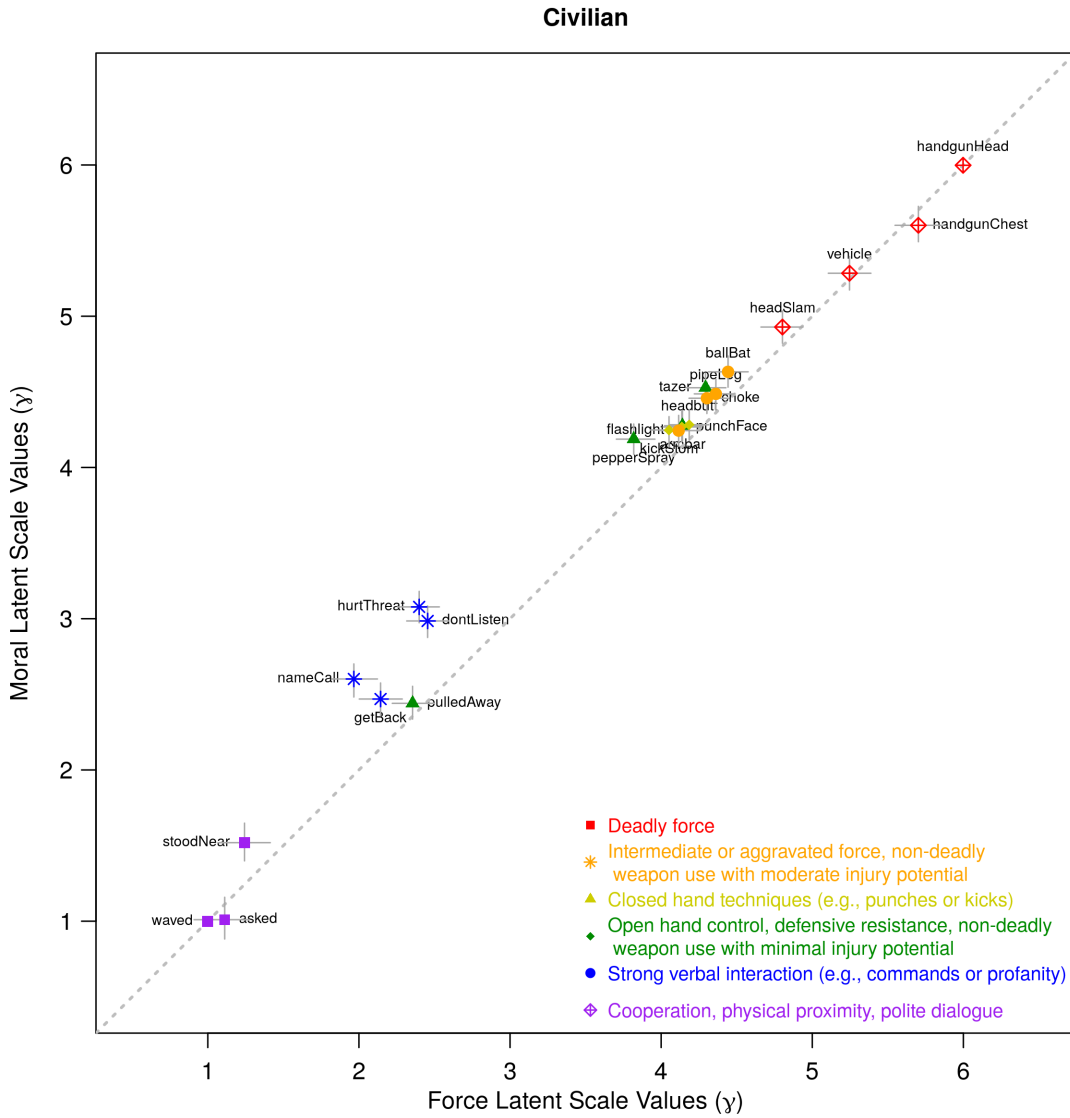


Figure 6. Posterior estimates of latent action scale values for the *civilian*. The scale values for the moral questions are plotted on the vertical axis as a function of the estimated scale values for the force questions on the horizontal axis. Segments through each point represent the 95% HDI's. Each point is labeled by the abbreviated action name in Figure 7.

in Figures 5 and 6. In contrast, the lethal and verbal actions (SOLOF levels 6 and 2, respectively) are relatively spread out – especially the lethal actions. That is, the severity of verbal and lethal actions exhibits considerable variability relative to non-lethal actions.

Because of the potential concern associated with implausible disparate action pairs to skew the results, we applied the model to a subset of data that excluded any pairs of SOLOF level 1 with SOLOF level 6, which comprised about 4.9% of all trials. We observed that responses to these extreme pairings were consistent and reasonable. Additionally, we re-ran our analyses on a subset of data that excluded these extreme pairs and observed that the latent severity estimates were virtually unchanged. As mentioned in the Method, we also ran a variety of analyses including all of the individual difference measures we collected, and the latent severity estimates are virtually identical to those reported currently. That is, the latent severity estimates we report are robust, and the patterns we describe in this article do not change when extreme action pairings are excluded or when additional relevant individual difference predictors are included in our models.

## **Discussion**

Our results provide a detailed initial map of the latent moral and physical severity of a set of realistic actions that represent the complete range of force options. Because these actions are broadly representative and include many of the most common forceful actions actually employed by and against police, they yield a comprehensive view of action severity perceptions, and offer insight into discrepancies between lay and formal evaluations of force. Moreover, 17 of our 22 actions were identical for the officer and civilian, and the remaining 5 were matched a priori based on our intuitions which were confirmed empirically by the remarkably similar severity estimates when comparing the matched actions in Figures 3 and 4 and Figures 5 and 6. Finally, the combination of a large number of participants and Bayesian methods provided precisely estimated scale values.

This new set of scale values reveals that lay perceptions of police and civilian use of force are substantially disconnected from the way force is traditionally categorized in most law enforcement agencies (i.e., using a SOLOF-style classification). Psychologically, non-lethal physical (beyond verbal) actions cluster together and effectively merge half of the canonical SOLOF categories (i.e., all three levels of non-lethal physical actions). But within the categories of verbal and lethal force, lay participants distinguish much more variable action severities.

## **Implications**

Of the discrepancies revealed in Figures 3 to 6, handcuffing stands out as a unique example. While it is clearly on the physical side of the verbal-physical boundary, it is also a routine police procedure in nearly all arrests, regardless of whether a civilian resists, and therefore might be perceived as especially normal and acceptable relative to other physical actions. Likewise, pulling away (the matched civilian action for handcuffing) may be viewed as maintaining personal autonomy and as more defensive than offensive relative to other physical actions. Moreover, while some handcuffing and pulling away may be highly forceful, it is possible to handcuff or pull away in a relatively gentle manner, which is something that does not seem to be true for any of the other physical actions.

The general pattern of non-monotonicities can largely be explained, we believe, in terms of the intuitive normativities of the actions for the roles. For the role of police officer, intuitively normative actions involve official tools such as handcuffs, pepper spray, and tazers, while intuitively *non*-normative actions (i.e., actions that are not primarily or uniquely associated with police officers) include name calling and hitting someone with a motor vehicle. In Figure 5 it can be seen that the moral severity of normative actions tends to fall well below the diagonal reference line, while the moral severity of non-normative actions tends to fall above the diagonal reference line.

Thus, for police officers, actions that are not normative (e.g., striking with a motor vehicle) may be evaluated as less acceptable than other actions that are more strongly associated with police (e.g., handgun shot to chest), even if both of those actions are categorized as lethal force according to the SOLOF system. This distinction by lay people is particularly notable because it diverges from the way that courts evaluate actions and the way that police officers train to identify when lethal force is appropriate under realistic time and resource constraints.

There are relatively few non-monotonicities for the civilian actions in Figure 6, but the few that occur may also be explained by normativity. In some sense, any action by a civilian more severe than cooperation is not prescriptively normative, but we are concerned here with moral severity (of civilian actions) that is not proportional to physical severity (of those civilian actions) in response to forceful actions by police officers. Specifically, for civilians it is not normative to call a police officer names or to threaten a police officer, and these actions have moral severities that fall well above the diagonal reference line.

### **Normativity in Lethal Force**

In general, police rely on legal precedent that treats police use of force as a seizure under the 4th amendment, and only affirms lethal force as lawful if a "citizen poses a threat of serious physical harm, either to the officer or others" (*Tennessee v. Garner*, 471 U.S. 1, 1985); thus, lawful lethal force must be "objectively reasonable" (*Graham v. Connor*, 490 U.S. 3, 1989). Many law enforcement agencies use this criterion to establish a general threshold, and when that threshold is exceeded, any form of lethal force is permitted in order to accomplish the goal of stopping the threat, regardless of how role-normative it is. Thus, lethal force actions are formally evaluated in a binary manner as either justified or not. In contrast, the psychological separation of different lethal force actions observed in Figure 5 suggests that lay perceptions of moral severity are not categorical, and that lay

evaluators have a more nuanced view of lethal actions than formal evaluative criteria.

This disconnect between professional and lay evaluations of non-normative lethal force application was especially evident in the case of a sniper attack in Dallas, Texas on July 7, 2016. During this attack, Micah Xavier Johnson killed five police officers, injured nine other officers and two civilians, and then barricaded himself inside of an El Centro College building where he engaged surrounding police with gunfire (Wikipedia Contributors, 2017). After negotiations failed, Dallas police maneuvered a robot carrying an explosive device near Johnson and killed him. This event was the first time a law enforcement agency in the U.S. had used a robot to deploy lethal force, making it highly non-normative. Consistent with the pattern observed in our results, considerable moral outrage was subsequently voiced by the public, not only about the use of lethal force but about the specific and unusual way it was employed (Fountain & Schmidt, 2016; King, 2016).

Our theory, that normative actions are perceived as less morally severe than their physical severity would imply, is post hoc and we will investigate it further in the future. Meanwhile, we note that it is consistent with other recent research that suggests people think of frequently observed actions as morally acceptable actions. Lindström, Jangard, Selbing, and Olsson (2018) showed that people more strongly endorse free riding on a public good when free riding is common. Those authors suggested that people use a general judgment heuristic: *common is moral*. In work using multiple cognitive tasks, Eriksson, Strimling, and Coultas (2014, p. 60) showed that people have strong associations between what is common and what is moral, and cited earlier work showing that “moral judgments of socially undesirable behavior tend to be less harsh when the behavior is perceived to be common (McGraw, 1985; Trafimow, Reeder, & Blising, 2001; Welch et al., 2005)”. Thus, we believe it is very plausible that actions perceived to be normative and common for police officers should also be judged to be less morally severe than those actions’ physical force would suggest. Importantly, our conceptualization of normativity only includes actions that

are common and also not routinely punished as we discuss below in the limitations.

### **Dealing with the Disconnect**

The patterns in our data point to questions that may promote understanding among all parties invested in police-civilian interactions. For example, the difference in moral severity between an officer yelling "get back" at the civilian and pepper spraying the civilian is quite large according to lay evaluation, but the difference is only one level of force in SOLOF-style scales. Does this imply that official policies should make finer distinctions at the boundary between verbal and physical actions? As another example, it appears that the perceived normativity of a police action influences the way that lay people judge the action's moral severity. Should this be treated as a cognitive bias in lay evaluation that should be corrected because what counts as normative may be an accident of history? Or should this influence of normativity be incorporated into the official classification of actions to emphasize normative actions and de-emphasize nonnormative actions? Finally, it is clear that lay people do not strongly distinguish among SOLOF levels '3,' '4,' and '5' (i.e., the non-lethal beyond-verbal actions). Does this failure to distinguish levels reflect a lack of expertise, implying that law enforcement agencies need to educate the public about important gradations in physical force? Or does the lay perception imply that official classifications should collapse or reorganize some levels?

Importantly, the disconnect does not necessarily imply that SOLOF-style scales should be abandoned. Many activists who are currently calling for reformative changes to how the police use force are explicitly recommending the ubiquitous adoption of SOLOF-style rules by all law enforcement agencies in the US (Campaign Zero, 2018). The widespread adoption of SOLOF-style categorizations presumably means they are useful, and careful consideration of the challenges surrounding use-of-force decision-making reveals the need for some sort of specific structure to guide officers in the field. But our results may suggest

ways to bring law-enforcement and lay perceptions into closer alignment. For example, educational outreach programs might help lay people understand how police and the legal system classify use of force, especially targeting the discrepancies from lay intuition revealed by the present research. As another example, knowing that more role-normative actions are viewed by the public as less morally severe than non-normative actions may guide officers when choosing a force option within a particular SOLOF level. Additionally, the public might be educated to be wary of becoming morally de-sensitized (or hyper-sensitized) to the forcefulness of actions merely because of their normalization (or lack thereof).

### **Limitations and Future Directions**

As with any empirical research, there are limitations that apply to our findings. One is that, although we selected our 22 actions in order to maximize representativeness and intra-SOLOF-category diversity across the whole spectrum of police and civilian force options, those we selected are merely a sample of all possible actions. Nevertheless, because of our selection procedure, we fully expect that alternative sets of actions will replicate the general patterns we observe even as there are likely to be idiosyncratic differences for some actions.

Our vignettes only used two actions, but real-life conflict between police and civilians is comprised of a set of many interdependent actions that occur over the time-course of an encounter. Future research could explore how moral evaluations are affected by a sequence of interdependent actions between officers and civilians, perhaps paying special attention to how increasing or decreasing the severity of force over time alters perceptions about officer or civilian intentions to escalate or de-escalate conflict. In addition, our use of vignettes allows for enhanced control over stimuli, but may be more limited than video in terms of realism.

In this article we have focused on the scale estimates of actions without describing the pattern of moral and physical ratings for specific pairs of actions. In particular, the

scale values alone do not reveal whether lay expectations generally align with common law enforcement heuristics such as the "one level above" heuristic. These ratings are described in a companion paper by Celestin and Kruschke (2019b). Moreover, future studies can use the scale estimates we provide to precisely compare the severity of actions chosen by lay evaluators as appropriate police responses to specific civilian violence.

Currently, it is unclear to what extent the disconnect we observe between formal use-of-force policies and lay intuitions emerges as a result of differential expertise rather than moral disagreement. To investigate this issue, future research should attempt to disentangle the influence of descriptive and prescriptive effects on lay evaluations of force. One approach would be to teach lay evaluators about SOLOF-style force categorization and common law enforcement heuristic rules and then observe whether and how that affects subsequent judgments. Care would need to be taken to avoid experimenter demand effects, but such an approach could provide useful information for law enforcement agencies regarding how to allocate resources to target educational public outreach efforts or to hone their use-of-force policies to maximize alignment with public morality.

Finally, future research should clarify what conditions are needed for common (i.e., normative) actions to be perceived as moral (or less immoral) actions. We suspect that actions become perceived as less immoral especially when those actions are not routinely punished or disparaged. Actions that are perceived to be frequent but are also routinely punished may still be perceived to be immoral.

## **Conclusion**

Because police use of force is a highly moralized topic, evaluating specific instances of applied force is likely to always involve some contention between police and civilian observers. In fact, reasonable people — both lay citizens and officers — can and do disagree about whether specific instances of force are necessary or excessive, and even whether in-

stances that are identified as excessive are isolated incidents or whether they imply a broad problem and the need for systemic change (Morin, Parker, Stepler, & Mercer, 2017). Regardless of how one makes such inferences, it has become clear that police actions are promptly evaluated in the court of public opinion, and it matters little to lay evaluators whether a particular set of actions falls within the law or within agency protocol if the intuitive public reaction to viewing the episode evokes sufficient moral outrage. It is our hope that the results presented here will provide an empirically-grounded starting point for helping to bring together lay and professional moral evaluations of use of force.

### **Chapter 3: Lay evaluations of police and civilian use of force: Moral and physical magnitude ratings of officer and civilian actions**

In the United States, police use of force is constrained by the fourth amendment of the Constitution. The U.S. Supreme Court declared that judging whether or not police use of force is reasonable is so idiosyncratic that it "is not capable of precise definition or mechanical application" (*Bell v. Wolfish*, 441 U.S. 520, 1979). Nevertheless, law enforcement agencies must provide guidance to officers in the field, and during the past century agencies have developed more detailed use-of-force policies for that purpose. Most policies include categorical levels of force, commonly referred to as force "continua" even though they are not actually continuous scales, that bin specific actions into categories according to their severity. The levels of the categories are generally described without specific situational details; for example a level might be labeled as *deadly force* and refer to any action in nearly any situation that may result in great bodily harm or death. Binning actions into these categorical levels of force is intended to provide a framework for decision-making, with which officers can act at a level of force that is "objectively reasonable" (*Tennessee v. Garner*, 471 U.S. 1, 1985).

If lay evaluations of force are not well aligned with law-enforcement policies, there may be serious disputes between public and official judgments of appropriateness. But research focused on how lay observers evaluate instances of police violence is scant (Jefferis, Butcher, & Hanley, 2011; Gerber & Jackson, 2017). In particular, we are not aware of any previous research that has investigated the detailed correspondence of lay evaluations of force with the categorical structure used by law enforcement across the entire severity range of force options. Moreover, very little research has experimentally studied the effects of beliefs about police legitimacy on evaluations of civilian resistance toward police; instead, the focus has tended to be on police actions or on motivating civilian cooperation (Thompson & Daniel Lee, 2004; Sunshine & Tyler, 2003; Gerber & Jackson, 2017).

A general relationship between high-legitimacy and reduced support for vigilante violence has been demonstrated (Jackson, Huq, Bradford, & Tyler, 2013), but we are unaware of any research on legitimacy modulating evaluations of a wide array of specific civilian reactions to police — especially not in a manner that corresponds to the way actions are described in police training and policy. The present research provides a map of lay evaluations of actions that comprehensively span the force categories, and offers insight into how legitimacy affects judgments of both police and civilian violence.

### **Police and civilian perspectives on use of force**

There are raging debates about whether force used by police is morally appropriate. Take, for example, the case of Terence Crutcher who was shot by Tulsa, Oklahoma police officer Betty Jo Shelby who believed Crutcher was reaching for a weapon when, in fact, he was unarmed (Salinger, 2016). Or the case of Daniel Shaver — an intoxicated but unarmed man who was shot by Mesa, Arizona police officer Phillip Brailsford while crying and begging officers "please do not shoot" (Friedersdorf, 2017a, 2017c). In both of these cases the officers (i.e., Shelby and Brailsford) were criminally prosecuted, but juries ultimately acquitted them of all charges. Nevertheless, these cases (and others like them) are extraordinarily tragic, and seem to many people — including the families of Crutcher and Shaver and protesters who rallied around them — to represent unnecessary and immoral levels of violence inflicted by police (Hogan, 2017; Friedersdorf, 2017a). On the other hand, consider the case of a female Chicago police officer whose face was repeatedly slammed into the pavement by a male suspect until she lost consciousness and "thought she was gonna die," but even in that moment resolved not to use her firearm "because she didn't want her family or department to go through the scrutiny the next day on the national news" (Gorner & Dardick, 2016). Or the recent mass shooting at Marjory Stoneman Douglass high school, during which multiple Broward County sheriff's deputies — includ-

ing school resource deputy Scot Peterson — remained stationary outside the building for four to six minutes while Nikolas Cruz was actively murdering the students and teachers inside (Stewart, 2018). It is unknowable how many of the 17 people killed (or 17 others wounded) might have been saved from harm if the first responders on the scene had engaged Cruz immediately, but given that the entire shooting only lasted about six minutes (Blinder & Mazzei, 2018; Wikipedia Contributors, 2018), the number of victims would almost certainly have been dramatically reduced.

In all of the aforementioned cases, the facts lend themselves to a relatively extreme and singular moral interpretation (i.e., either excessive or insufficient/negligent levels of force used by police). But, notably, even in these cases, there is considerable disagreement about whether the police acted appropriately; different members of the public (and even different law enforcement officers and agencies) have very different attitudes about what constitutes appropriate police use of force. In more ambiguous cases, where some of the facts are more reasonable and uncertainty is amplified, this problem is exacerbated.

We point out these controversies regarding police use of force to emphasize a key premise: Public perceptions of police use of force often seem to be misaligned with law-enforcement implementations (and, ultimately, judicial/legal outcomes) of use of force. (If they did match, there wouldn't be such vociferous disputes.) On the one hand, law enforcement policies usually categorize forceful actions into discrete levels and agencies provide heuristics regarding which level of force is appropriate relative to the level of civilian resistance confronting officers. On the other hand, lay perceptions of the appropriateness of actions might be quite different than the organization and heuristics of official policies. Moreover, lay perceptions are likely to be modulated by relevant individual differences. Given the apparent conflict between public outrage and legal assessment police use of force, research investigating how they align and diverge is crucial for maximizing police efficacy and public trust. In the next two sections, we describe law enforcement policies and some

<b>Level</b>	<b>Description</b>
1	Cooperation, physical proximity, polite dialogue.
2	Strong verbal interaction.
3	Open hand control, defensive resistance, non-deadly weapon use with minimal injury potential.
4	Closed hand techniques.
5	Intermediate or aggravated force, non-deadly weapon used with moderate injury potential.
6	Deadly force.

Table 3  
*Six ordinal levels of force (SOLOF) scale.*

important individual differences among people.

### **Six ordinal levels of force (SOLOF)**

At least 80% of law enforcement agencies use a list of rank-ordered force options as part of their official use-of-force policy (Terrill et al., 2011). These formal classifications provide guidance for how the officers employed at a given government agency are allowed to use force. Rank-ordered classifications of force are also prescribed within the social scientific literatures, where police and civilian actions are separated into distinct lists of role-specific categories (Terrill et al., 2003; Alpert & Dunham, 2004). These lists are used to calculate the proportionality of police use of force relative to civilian resistance by quantifying a difference score, dubbed a "force factor". There is some evidence that this method for evaluating the reasonableness or excessiveness of force is reliable (Hickman et al., 2015) and useful for professionals.

Nearly three quarters of all law enforcement agencies use a 5 or 6 level classification (Terrill et al., 2011), and Terrill et al. (2003) recommend a six-category list for evaluation of police use of force and civilian resistance. Therefore, the current study used a six-category list shown in Table 3 and described more completely in the Method section. Because this

type of force categorization is not actually continuous, we refer to this list as the *six ordinal levels of force* (SOLOF) scale (rather than using the common "force continuum" language). Notably, the descriptions of the levels in Table 3 do not specify contextual details such as location, time of day, or state of mind (e.g., intoxication). Instead, the levels are decontextualized because specifying all possible relevant contextual variables would be unwieldy and intractable, and also maximize correspondence with law enforcement policies.

### **Individual differences: legitimacy beliefs**

A considerable amount of research has investigated the factors that influence public evaluations of police. Although salient demographic factors such as race (D. Johnson & Kuhns, 2009) and gender (Chapman, 2012) do affect attitudes about police, the strongest and most salient individual difference that predicts differential evaluations of police conduct are people's existing beliefs about police legitimacy (Tyler, 2006). That is, legitimacy beliefs are the key individual difference to observe when measuring attitudes about the justness of police actions, including the use of force.

Beliefs about police legitimacy are driven by current and prior perceptions of police acting in a procedurally just manner (e.g., behaving respectfully toward citizens and allowing them to voice their perspective) (Mazerolle, Antrobus, Bennett, & Tyler, 2013), and these perceptions are grounded in the social context to which evaluators are or have been exposed (Braga, Winship, Tyler, Fagan, & Meares, 2014). Legitimacy beliefs include at least two subcomponents (Gerber & Jackson, 2017): (1) felt obligation, which is the degree to which a person feels obligated to obey the police; and (2) norm alignment, which is the extent to which a person believes that their own values align with the values of the police. These two dimensions are drawn from the theoretical conceptualization of the legitimacy of authority rooted in a process-based model of procedural justice (Tyler, 2006; Hamm, Trinkner, & Carr, 2017). This model prioritizes perceptions of fairness over

outcome-driven evaluations as the primary predictor of legitimacy beliefs, and posits that the extent to which citizens' values align with those in authority (i.e., norm alignment) and the extent to which they grant authority figures the ability to dictate acceptable behavior (i.e., felt obligation) together comprise individual perceptions of fairness and, thus, of the legitimacy or illegitimacy of authority figures, including police.

### **Assaying lay perceptions of specific forceful actions**

In the present research, we generated a list of specific actions for each of the SOLOF levels. These actions could be performed by either a police officer or a civilian with whom the officer was interacting. Lay participants rated the morality and physical severity of police or civilian actions when performed in response to an action by the other party. Trends in ratings were captured by Bayesian regression models that predicted ratings as a function of the actions and the rater's legitimacy beliefs, gender, race, and political affiliation.

## **Method**

### **Participants**

To maximize the diversity and generalizability of our sample, we collected data from the Indiana University (IU) undergraduate subject pool in the Department of Psychological and Brain Sciences and from Amazon Mechanical Turk (MTurk). Analyses were initially run separately for the IU and MTurk samples, and then combined after we observed that the model descriptions for each sample were virtually identical. Supplementary material presents the separate analyses (<https://osf.io/h6jg3>).

We recruited 411 MTurk participants and 395 IU subject pool participants, all of whom completed the study through a web browser. Participation requirements constrained participation to individuals residing in the United States who were 18 or older at the time of

participation and who were fluent in English. Although we set up MTurk requirements to disallow non-US participation, we also recorded geolocation information associated with the Internet Protocol addresses used by participants, and subsequently excluded 38 MTurk participants for participating from outside the US. We excluded participants who missed more than one of the 11 attention checks. We also excluded participants who had a mean absolute deviation (MAD) in their ratings that was more than three standard deviations below the mean MAD of all participants because very small MAD probably indicates leaving the response slider at its default level for most trials. This data cleaning procedure resulted in a final N of 350 MTurk participants and 317 IU subject pool participants. The combined total N was 667.

A demographic breakdown of our 667 participants, according to the categories they selected on the questionnaire at the conclusion of the experiment, included 404 females, 261 males, and 2 participants who selected other. There were 3 American Indian/Alaska Native participants, 71 Asian participants, 54 Black or African American participants, 32 Hispanic or Latino participants, 486 White participants, 19 participants who selected More than one, and 2 participants who selected Unknown.

## **Design**

Using a combined version of Terrill et al. (2003)'s separated civilian resistance and officer force levels, we constructed a list of forceful actions that could reasonably be performed by either an officer or a civilian. We then selected 22 representative actions from this list, ensuring that at least three actions were located, a priori, in each of the six SOLOF categories. In addition to this primary goal of choosing broadly representative actions, we also tried to choose actions that were diverse and realistic. These actions were presented de-contextualized (i) to match the de-contextualized nature of the SOLOF levels and (ii) because specifying contextual details would introduce myriad additional factors. Seventeen

<b>SOLOF Level</b>	<b>Action Statement</b>
1	the [officer/civilian] stood near the [civilian/officer]
	the [officer/civilian] asked the [civilian/officer] how he was doing
	the [officer/civilian] waved at the [civilian/officer] in greeting
2	the [officer/civilian] yelled “get back” at the [civilian/officer]
	the [officer/civilian] shouted “get your fucking hands up!” / “I don't have to fucking listen to you!” at the [civilian/officer]
	the [officer/civilian] said “if you touch me, you're gonna get hurt” to the [civilian/officer]
	the [officer/civilian] called the [civilian/officer] a “thug” / “pig”
3	the [officer/civilian] put handcuffs on/pulled away from the [civilian/officer]
	the [officer/civilian] sprayed the [civilian/officer] with pepper spray
	the [officer/civilian] shocked the [civilian/officer] with a tazer
	the [officer/civilian] twisted the arm of the [civilian/officer], forcing him to the ground
4	the [officer/civilian] punched the [civilian/officer] in the face
	the [officer/civilian] kicked the [civilian/officer] in the stomach
	the [officer/civilian] applied a chokehold to the [civilian/officer]
	the [officer/civilian] headbutted the [civilian/officer] in the face
5	the [officer/civilian] struck the [civilian/officer] on the leg with a collapsible baton/metal pipe
	the [officer/civilian] shot/struck the [civilian/officer] in the chest with a beanbag shotgun round/baseball bat
	the [officer/civilian] struck the [civilian/officer] in the stomach with a metal flashlight
6	the [officer/civilian] shot the [civilian/officer] in the chest with a handgun
	the [officer/civilian] shot the [civilian/officer] in the head with a handgun
	the [officer/civilian] struck the [civilian/officer] with a motor vehicle
	the [officer/civilian] slammed the [civilian/officer]'s head into the concrete curb

Figure 7. Actions with their *a priori* SOLOF severity level from Table 3. Colors in SOLOF column correspond to colors used in subsequent plots.

of the actions were exactly the same for the officer and civilian, while 5 of the actions were unique for those roles to ensure that they seemed plausible for a civilian or officer but were rationally matched in terms of severity. For example, the officer action of handcuffing was not performed by the civilian because it seems implausible that a civilian would handcuff an officer. Instead, the action of pulling away was matched with handcuffing and assigned exclusively to the civilian. Figure 7 displays all of the actions and the SOLOF category in which we located them.

The current study employed a partially-crossed within-subjects design that randomly permuted the separate lists of officer and civilian actions and then paired them together. Eleven pairs were assigned to the "civilian" condition, in which the civilian action occurred second and was the focus of the ratings, while the remaining 11 actions were assigned to the "officer" condition in which the officer action occurred second and was the focus of the ratings. The officer and civilian trials were randomly interspersed. This method assured that in only 22 pairs of actions every participant saw all of the officer and civilian actions (providing ratings of 11 civilian actions and 11 officer actions).

## **Procedure**

Participants gave informed consent to take part in the study and were shown the following instructions:

"On the following screens, you will be presented with scenarios that represent real life interactions involving police officers and civilians. Officer and civilian names have been purposefully excluded. For each scenario, you will be presented with a pair of actions that occurred during the encounter. Each action pair will include one action by the officer and one action by the civilian, though other actions may have occurred before, between, or after the two ac-

tions that are listed. The order in which the presented actions occurred during the encounter is the order in which they are listed. It is very important that you read each action description carefully, because the actions will change for each scenario."

"You will be asked to make several ratings of some of the actions. You will always be asked to rate the second action listed, but the specific actions and the order of the questions will change between scenarios, so it is very important that you read each question carefully prior to responding."

Participants were also instructed about the four rating questions (see details below) and pairs of officer and civilian actions. We emphasized that the order in which the actions occurred was the order in which they were presented to mitigate a potential confound of action order that might arise if action order was ambiguous. Order alone clearly matters for moral inference (Schwitzgebel & Cushman, 2012). And it seems reasonable for participants to infer that whoever acts first in our scenario is initiating aggression and likely more blameworthy as a moral agent, while whoever acts subsequently is responding as a moral patient (Schein & Gray, 2017). The nature of the officer role potentially complicates this dynamic, because a participant (especially one with high levels of trust in police) might assume that police officers are always (or nearly always) responding to some prior violent act or threat, even if it isn't explicitly mentioned. Citizens, on the other hand, do not have any such intrinsic benefit of the doubt, because the category of "citizen" is very general and does not describe a set of professional responsibilities that include the legitimate use of violence. For this reason, we always inquired about the second action which was explicitly preceded by an action from the other role. Thus, ratings were always of "reactions" to the other party, to be sure there was no confounding of actor role (civilian vs officer) with action order (action vs reaction).

While working a patrol shift, a police officer was dispatched to a reported crime, and in the course of the investigation, he came into contact with a civilian. During their subsequent interaction, the following actions took place \*in the order listed below\* (other actions may have occurred before, between, or after).

1. **the civilian headbutted the officer in the face**
2. **the officer put handcuffs on the civilian**

Please rate how **morally acceptable** the **officer's** action is:

0 10 20 30 40 50 60 70 80 90 100

Completely Unacceptable Completely Acceptable

Submit Answer

Figure 8. Example screenshot from the experiment.

Within each action rating trial, participants read a short vignette as shown in the screenshot of Figure 8. As can be seen in both, actions were displayed as bulleted points which were emphasized using bold font. The key word in every rating question was also displayed in bold font. The rating questions were presented one at a time in random order, and participants made each rating by adjusting the location of a slider on a bar spanning the range of 0 to 100. The rating questions and their semantic anchors are listed below:

- Please rate how much **force** is involved in the officer's/civilian's action: (0 = No Force, 100 = Maximum Force)
- Please rate how **morally acceptable** the officer's/civilian's action is: (0 = Completely Unacceptable, 100 = Completely Acceptable)
- Please rate the extent to which the officer/civilian used an **appropriate** level of force (0 = Insufficient, 50 = Appropriate, 100 = Excessive)
- Please rate the extent to which the officer/civilian should be **punished** for his action: (0 = Not at all, 100 = Severely)

In order to ensure thoughtful responding and to provide a mechanism by which to exclude participants who were not engaged in the task, 11 attention checks were inserted in random places in the sequence for some action pairs. Each attentiveness question explicitly advised participants to move the slider to the leftmost or rightmost position on the scale.

Following the action rating trials, participants answered six, randomly ordered questions measuring the two subcomponents of police legitimacy beliefs (Gerber & Jackson, 2017) which are (1) felt obligation: the extent to which someone believes they should obey police, and (2) norm alignment: the extent to which someone believes that their own values overlap those of the police. The specific questions capturing these two dimensions were as follows:

- Felt Obligation

- You should support the decisions made by police officers even when you disagree with them.
- You should do what the police tell you even if you do not understand or agree with the reasons.
- You should do what the police tell you to do even if you do not like how they treat you.

- Norm Alignment

- The police generally have the same sense of right and wrong that you do.
- The police stand up for values that are important to you.
- You and the police want the same things for your community.

Each of these legitimacy questions was answered by adjusting a slider ranging from 0 ("completely disagree") to 100 (completely agree).

Participants also answered yes/no questions measuring personal or vicarious negative experiences with police (Weitzer, 2005):

- Personal
  - Have you ever been stopped by police on the street without good reason?
  - Have the police ever used insulting language toward you?
  - Have you ever seen a police officer engage in any corrupt activities (e.g., taking bribes or involvement in drug trade)?
  - Have police ever used excessive force against you?
- Vicarious
  - Have the police ever used insulting language toward anyone else in your household?
  - Have police ever used excessive force against anyone else in your household?
  - Has anyone else in your household ever been stopped on the street by police without good reason?

In subsequent data analyses we considered regression models that used the negative-experience answers as covariates, but ultimately found that they were not particularly useful. See discussion of the models below for further information and see also the supplementary materials (<https://osf.io/h6jg3>) for specific results that include these variables.

The experiment concluded with a series of demographic questions including race and gender, as well as a single item measure of political affiliation (Iyer et al., 2012), and a debriefing.

The experiment interface was programmed using JavaScript, html, and CSS, as well as the libraries available in the jsPsych package (de Leeuw, 2015). We employed either a

psiTurk (Gureckis et al., 2015) or apache server alongside a SQL database and PHP scripts to run the experiment and save the data. Additionally, all data cleaning, pre-processing, and analyses were conducted via the R programming language, RStudio (Team, 2016), JAGS (Plummer, 2003), and Stan (Stan Development Team, 2016), and the programs available in Kruschke (2015).

## Results

### Data transformation and regression model

The dependent variables are ratings of acceptability, excessiveness, punishment, and force. Recall that these ratings were all made using a slider that ranged from 0 to 100. Examining the raw distributions of each of these ratings revealed that a notable proportion of the ratings fell at both ends of the scale, and the other ratings tended to form a unimodal distribution that was skewed toward the middle. Therefore, we treated any response more extreme than 5 or 95 as censored, meaning its value is not modeled as the specific slider value but is instead modeled as an unknown value at least as extreme as 5 or 95, respectively. To approximately normalize the uncensored data, we logit transformed all of the ratings (after dividing by 100).

A key independent variable was participants' beliefs in police legitimacy as measured by their ratings of felt obligation (FO) and norm alignment (NA). FO and NA ratings were highly reliable and strongly correlated. Specifically, the three items for FO had a Cronbach's  $\alpha = 0.82$ , the three items comprising NA had  $\alpha = 0.87$ , and all six items together had  $\alpha = 0.87$ . The correlation between the combined FO and combined NA items was  $\rho = 0.61$ . Because all of the items were so strongly related, the three questions measuring felt obligation were averaged together, as were the three questions measuring norm alignment. These averaged ratings were then standardized. Finally, the two standardized ratings

were averaged together to form one composite individual difference measure of police legitimacy beliefs, denoted  $\ell$ . Notice that this legitimacy score is inherently standardized, which implies that the mean  $\ell$  is zero and that most values of  $\ell$  fall between  $-2.0$  and  $+2.0$ .

We constructed Bayesian linear regression models for each of the rating questions. Ratings were described as a weighted combination of civilian action severity, officer action severity, and legitimacy beliefs (i.e.,  $\ell$ ), along with selected interactions. The severity of the officer and civilian actions were simultaneously estimated, yielding latent moral and physical severity scales for the entire set of actions. These scale estimates are described in detail in a separate report (Celestin & Kruschke, 2019a), and the focus of this report is the regression coefficients and corresponding trends in the ratings.

The regression model is expressed formally as

$$\hat{y}_{q,c,f} = \underbrace{(\beta_{0,q} + \beta_{3,q} \cdot \ell)}_{\text{baseline}} + \underbrace{(\beta_{1,q} + \beta_{4,q} \cdot \ell)}_{\text{influence of civilian action}} \cdot \underbrace{\widehat{\gamma}_c}_{\text{civilian action severity}} + \underbrace{(\beta_{2,q} + \beta_{5,q} \cdot \ell)}_{\text{influence of officer action}} \cdot \underbrace{\widehat{\chi}_f}_{\text{officer action severity}} + \underbrace{\beta_{6,q} \cdot \gamma_c \cdot \chi_f}_{\text{civilian/officer severity interaction}} \quad (2)$$

$$= \beta_{0,q} + \beta_{1,q} \cdot \gamma_c + \beta_{2,q} \cdot \chi_f + \beta_{3,q} \cdot \ell + \beta_{4,q} \cdot \ell \cdot \gamma_c + \beta_{5,q} \cdot \ell \cdot \chi_f + \beta_{6,q} \cdot \gamma_c \cdot \chi_f \quad (3)$$

where  $\hat{y}$  represents the predicted value of a rating (logit transformed),  $q$  indexes the rating question (1–8, i.e., the four types of ratings of both officer and civilian actions),  $c$  indexes the civilian action (1–22),  $f$  indexes the officer action (1–22),  $\ell$  denotes the participant's legitimacy beliefs,  $\gamma$  denotes the estimated civilian-action scale value, and  $\chi$  denotes the estimated officer-action scale value. Each of the  $\beta$  variables is an estimated regression coefficient. Equation (3) arranges the terms in a traditional form that accentuates multiplicative interaction terms, while Equation (2) arranges the terms to emphasize the conceptual framework of the model: A rating is generated from a baseline plus a weighted influence of the civilian action plus a weighted influence of the officer action, with the weightings and baseline modulated by the legitimacy beliefs ( $\ell$ ) of the participant, and by the interaction

of civilian and officer action severities.

To aid interpretation of Equation (2), consider rating the *acceptability* of officer actions, so  $\hat{y}$  is the predicted acceptability rating. Intuitively, as officer actions  $\chi_f$  become more severe, their acceptability should decline, and so the coefficient  $\beta_2$  should be negative. The magnitude of this decline might be modulated by the respondent's legitimacy beliefs  $\ell$ : it might be the case that respondents with high  $\ell$  would be more accepting of severe officer actions, meaning that coefficient  $\beta_5$  might be positive. Analogously, as *civilian* actions  $\gamma_c$  become more severe, the acceptability of severe officer actions should increase, meaning that coefficient  $\beta_1$  should be positive. It might be the case that respondents with high legitimacy beliefs  $\ell$  should be more sensitive to civilian action severity, meaning that coefficient  $\beta_4$  might be positive. Additionally, the rate at which officer action acceptability declines as officer action severity increases should be greater for low severity civilian actions than it is for high severity civilian actions. That is, intuitively, officer acceptability should decrease less rapidly when the civilian action is very severe compared to when it is not severe, meaning that coefficient  $\beta_6$  should be positive. Finally, the baseline acceptability is established by the intercept,  $\beta_0$ , which might be higher for respondents with higher legitimacy beliefs  $\ell$ , that is, coefficient  $\beta_3$  might be positive.

The ratings for moral evaluations (i.e., acceptability, appropriateness, and punishability) were modeled together using the same estimated latent scale values for the actions, but separately from the ratings for physical evaluations (i.e., force) which had its own estimated latent scale values for the actions. This separation was made *a priori* because we assumed that the scale value for the physical force of an action might be quite different than the scale value for the moral severity of an action. Estimated scale values are discussed at length in Celestin and Kruschke (2019a).

We made the conventional assumption that ratings were normally distributed around the predicted value in Equation 2, with distinct standard deviation  $\sigma_q$  for each type of question.

The data were reasonably normal because of our logit transformation and censoring.

Several variations of the model were also considered. In one variation, we added negative experience with police as a predictor, including an interaction term for both officer and civilian actions. In another variation, we removed legitimacy beliefs  $\ell$  as a predictor and replaced it with negative experience with police, again including interaction terms. In yet another variation, we added participant sex, race, and political affiliation, as well as negative experiences with police as predictors. All of these models yielded estimates that reflected weak or non-existent effects and showed increased uncertainty in the parameter estimates as a result of including additional predictors. Moreover, any effects that were present were unreliable across dependent measures. For these reasons, we are not confident that any effects observed for these additional individual difference measures represent meaningful effects in reality. Importantly, the inclusion of these additional predictors did not substantively alter the conclusions we present here. Thus, the analyses we present reflect the model of Equation 2. However, see the supplementary materials (<https://osf.io/h6jg3>) for additional results from the model variations.

**Bayesian Methods.** We used Bayesian methods to estimate the parameters Kruschke (2015); Kruschke and Liddell (2018a, 2018b). Bayesian analysis produces a joint probability distribution across all the model parameters, including all regression coefficients and latent scale values. This probability distribution is called a “posterior” distribution because it represents the credibilities of parameter values after taking the data into account. The posterior distribution on a parameter is summarized by its most probable value (i.e., its mode) and the interval that spans the 95% most probable values, referred to as the 95% highest density intervals (HDI). The width of an HDI represents the uncertainty of the parameter’s estimate, so smaller HDIs reflect greater precision.

A unique element of Bayesian analysis is that it requires establishing prior probability

distributions for every estimated parameter. We chose priors which were diffuse in relation to the scale of the data, ensuring that our prior choices had very little influence on the posterior estimates.

We coded and calculated all of our models using the programming language R and JAGS software (Plummer, 2003) via the `runjags` package (Denwood, 2016). All MCMC chains exhibited convergence, and the effective sample size for our parameter chains approaches 10,000.

### **Estimates of regression coefficients**

The modal posterior estimates for all of the regression coefficients, and their 95% HDIs, are presented in Table 4. As an example, recall from our previous discussion after introducing Equation 2 that intuition suggests the coefficient  $\beta_2$  for acceptability of officer action should be negative because the acceptability of actions should decline as their severity goes up. Table 4 shows that the estimated value of  $\beta_2$  for acceptability of officer actions is in fact strongly negative. The value of the coefficient is how much the predicted rating changes (on the logit transformed scale) when the action severity changes by one unit on its scale anchored at 1.0 for the least severe action and 6.0 for the most severe action. Graphical representations of the coefficients are presented in the following sections.

[Blank space intentionally inserted for readability of subsequent sections.]

Parameter	Question	Posterior Mode (95% HDI)
$\beta_0$	punish, officer action	-1.57(-1.74, -1.41)
$\beta_1$	punish, officer action	-1.13(-1.19, -1.07)
$\beta_2$	punish, officer action	1.59(1.52, 1.67)
$\beta_3$	punish, officer action	-0.51(-0.58, -0.44)
$\beta_4$	punish, officer action	0.01(-0.04, 0.06)
$\beta_5$	punish, officer action	-0.03(-0.08, 0.03)
$\beta_6$	punish, officer action	-0.28(-0.32, -0.25)
$\sigma$	punish, officer action	2.31(2.25, 2.37)
$\beta_0$	acceptable, officer action	0.45(0.30, 0.61)
$\beta_1$	acceptable, officer action	1.21(1.14, 1.27)
$\beta_2$	acceptable, officer action	-1.38(-1.45, -1.30)
$\beta_3$	acceptable, officer action	0.50(0.42, 0.57)
$\beta_4$	acceptable, officer action	0.04(-0.02, 0.09)
$\beta_5$	acceptable, officer action	0.05(0.00, 0.10)
$\beta_6$	acceptable, officer action	0.41(0.36, 0.45)
$\sigma$	acceptable, officer action	2.54(2.48, 2.60)
$\beta_0$	appropriate, officer action	0.42(0.34, 0.49)
$\beta_1$	appropriate, officer action	-0.53(-0.56, -0.5)
$\beta_2$	appropriate, officer action	0.69(0.66, 0.73)
$\beta_3$	appropriate, officer action	-0.07(-0.11, -0.03)
$\beta_4$	appropriate, officer action	-0.03(-0.05, 0.01)
$\beta_5$	appropriate, officer action	0.06(0.03, 0.09)
$\beta_6$	appropriate, officer action	0.02(0, 0.04)
$\sigma$	appropriate, officer action	1.58(1.55, 1.62)
$\beta_0$	force, officer action	0.06(-0.09, 0.21)
$\beta_1$	force, officer action	-0.22(-0.25, -0.2)
$\beta_2$	force, officer action	1.6(1.54, 1.67)
$\beta_3$	force, officer action	-0.09(-0.13, -0.05)
$\beta_4$	force, officer action	0.04(0.01, 0.06)
$\beta_5$	force, officer action	0.08(0.05, 0.11)
$\beta_6$	force, officer action	-0.02(-0.04, 0)
$\sigma$	force, officer action	1.35(1.32, 1.38)
$\beta_0$	punish, civilian action	-0.86(-0.98, -0.73)
$\beta_1$	punish, civilian action	1.72(1.67, 1.78)
$\beta_2$	punish, civilian action	-0.43(-0.46, -0.39)
$\beta_3$	punish, civilian action	0.38(0.33, 0.44)
$\beta_4$	punish, civilian action	0.11(0.07, 0.16)
$\beta_5$	punish, civilian action	0.08(0.04, 0.11)
$\beta_6$	punish, civilian action	-0.12(-0.15, -0.09)
$\sigma$	punish, civilian action	1.7(1.66, 1.74)
$\beta_0$	acceptable, civilian action	-0.29(-0.42, -0.17)
$\beta_1$	acceptable, civilian action	-1.56(-1.63, -1.51)
$\beta_2$	acceptable, civilian action	0.62(0.57, 0.66)
$\beta_3$	acceptable, civilian action	-0.51(-0.58, -0.45)
$\beta_4$	acceptable, civilian action	-0.15(-0.2, -0.11)
$\beta_5$	acceptable, civilian action	-0.03(-0.07, 0.01)
$\beta_6$	acceptable, civilian action	0.21(0.18, 0.25)
$\sigma$	acceptable, civilian action	2.13(2.08, 2.18)
$\beta_0$	appropriate, civilian action	0.49(0.42, 0.56)
$\beta_1$	appropriate, civilian action	0.69(0.66, 0.73)
$\beta_2$	appropriate, civilian action	-0.33(-0.36, -0.3)
$\beta_3$	appropriate, civilian action	0.22(0.17, 0.27)
$\beta_4$	appropriate, civilian action	0.08(0.05, 0.12)
$\beta_5$	appropriate, civilian action	0.03(0, 0.07)
$\beta_6$	appropriate, civilian action	-0.02(-0.04, 0.01)
$\sigma$	appropriate, civilian action	1.8(1.77, 1.84)
$\beta_0$	force, civilian action	-0.19(-0.33, -0.04)
$\beta_1$	force, civilian action	1.58(1.52, 1.64)
$\beta_2$	force, civilian action	-0.15(-0.18, -0.13)
$\beta_3$	force, civilian action	0.05(0.01, 0.09)
$\beta_4$	force, civilian action	0.12(0.09, 0.15)
$\beta_5$	force, civilian action	0.04(0.01, 0.07)
$\beta_6$	force, civilian action	-0.02(-0.04, 0)
$\sigma$	force, civilian action	1.36(1.33, 1.39)

Table 4

Parameter estimates for the regression model of Equation 2. HDI = highest density interval.

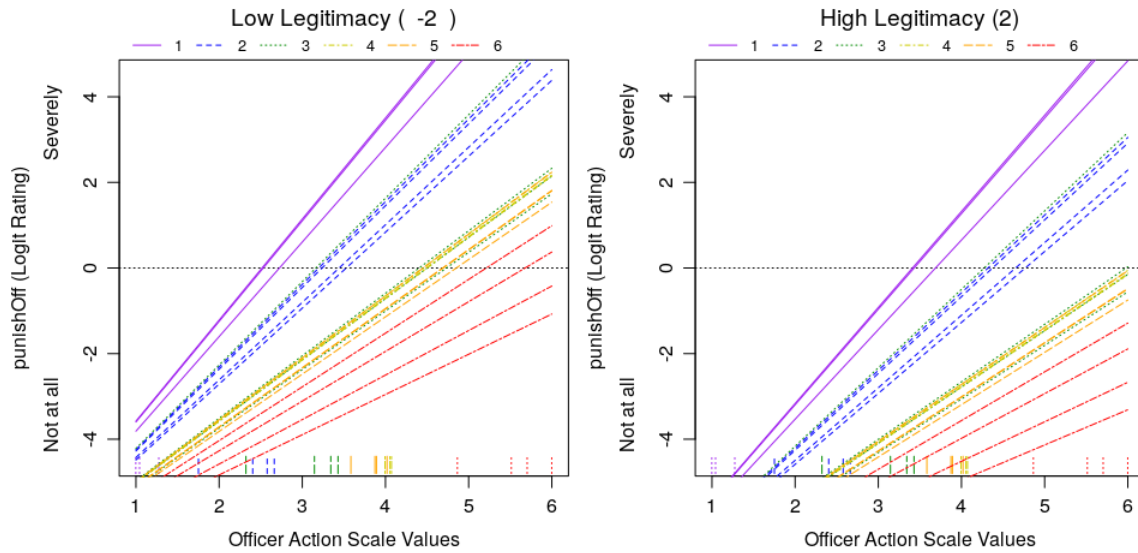


Figure 9. Regression model trend lines for officer punishment ratings.  $\text{punishOff}$  = punishability rating of officer action. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

**Officer punishment ratings.** The model’s trend lines for officer punishment are plotted in Figure 9. There are two panels in Figure 9, with the left panel showing model trend for participants with very low legitimacy belief ( $\ell = -2$ ) and the right panel showing model trend for participants with very high legitimacy belief ( $\ell = +2$ ). In each panel, the vertical axis is the logit transformed rating, where 0 on the logit transformed scale corresponds to 50 on the original rating scale, and  $-4$  and  $+4$  on the transformed scale correspond to approximately 2 and 98, respectively, on the original rating scale. The horizontal axis is the latent action severity, where 1.0 is anchored at the least severe action and 6.0 is anchored at the most severe action. The “rug plot” along the horizontal axis (i.e., the small vertical lines) indicates the estimated scale values of the 22 officer actions (i.e., modal posterior  $\chi$  from Eq. 2, reported in Celestin & Kruschke, 2019a), with their SOLOF categories coded by color line type. The main body of the panels plot a separate line for each civilian action.

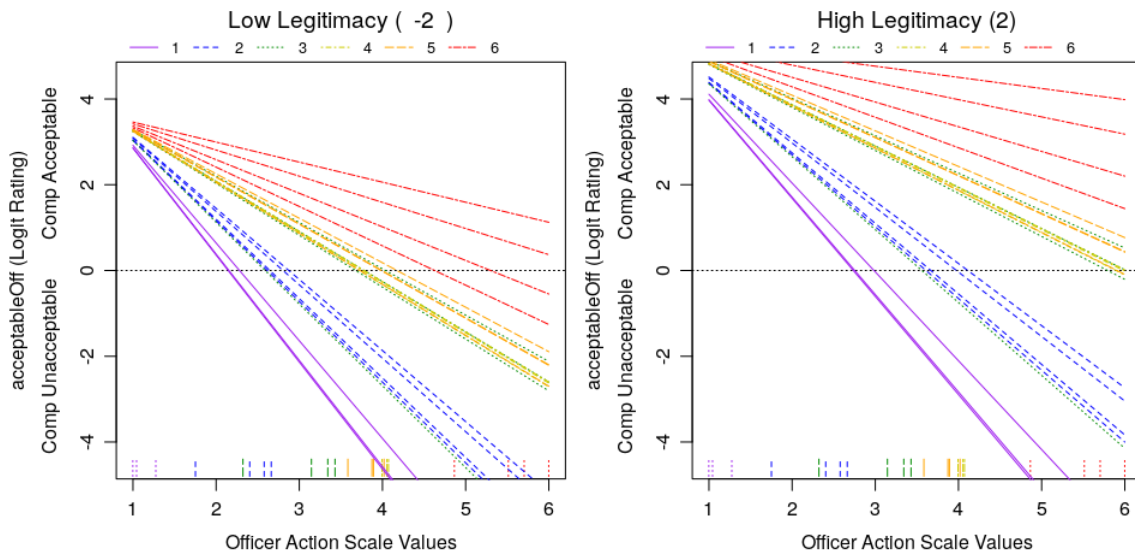
The trend lines in Figure 9 show the punishability of officer actions. Notice that the

lines go up, indicating that punishability increases as officer action severity increases. The slope of the lines is the slope coefficient on officer-action severity from Equation 2, namely  $\beta_2 + \beta_5 \cdot \ell + \beta_6 \cdot \gamma$ , using  $\ell = -2$  in the left panel and  $\ell = +2$  in the right panel. Table 4 shows that  $\beta_2$  for punishability of officer action is indeed positive, with a posterior mode of +1.59 (and 95% HDI from 1.52 to 1.67). Table 4 shows that  $\beta_5$  for punishability of officer action had a posterior mode of nearly zero (with an HDI including zero), which means that the slope for punishability is essentially unaffected by a participant's legitimacy belief, and therefore the slopes of the lines in the two panels of Figure 9 are visually very similar. However, the interaction between civilian and officer action severity controlled by regression coefficient  $\beta_6$  had a negative posterior mode of  $-0.28$  (and 95% HDI from  $-0.32$  to  $-0.25$ ). This indicates that the rate at which officer punishment increases as officer action severity increases, changes (i.e., decreases) as civilian action severity increases. This interaction is visible in both panels of Figure 9 by observing the fan of non-parallel lines, that is, that the slope of low-severity civilian actions is much steeper than the slope of high-severity civilian actions.

In Figure 9 there is a separate line for each civilian action, and the intercept of each line is given by Equation 2 as  $(\beta_{0,q} + \beta_3 \cdot \ell) + (\beta_1 + \beta_4 \cdot \ell) \cdot \gamma$  where  $\gamma$  is the latent severity of the civilian action (for discussion of the values of  $\gamma$  see Celestin & Kruschke, 2019a). Intuitively, as a civilian action becomes more severe, any specific officer reaction should be less punishable, suggesting that  $\beta_1$  should be negative. Indeed, Table 4 shows that  $\beta_1$  had a posterior mode of  $-1.13$  (with 95% HDI from  $-1.19$  to  $-1.07$ ). The large magnitude of  $\beta_1$  is represented visually in Figure 9 by the large spread between the lines. Note that the highest lines are the least severe civilian actions (hence an officer re-action is more punishable) and the lowest lines are the most severe civilian actions (hence an officer re-action is less punishable).

Visually comparing the two panels of Figure 9 suggests that the main influence of le-

gitimacy beliefs is the overall rating of punishability; that is, the lines in the  $\ell = -2$  panel are higher than the lines in the  $\ell = +2$  panel. In other words, participants with very low legitimacy beliefs tended to rate officer actions as much more punishable than participants with very high legitimacy beliefs. The regression coefficient that controls the overall level as a function of  $\ell$  is  $\beta_3$ , which had a posterior mode of  $-0.51$  (with 95% HDI from  $-0.58$  to  $-0.44$ ). In particular, notice that when officer actions are most severe, the lines in the  $\ell = -2$  panel representing all of the physical but non-lethal actions, as well as two of the lethal action lines, fall above the 0 midpoint. In contrast, all of the physical but non-lethal and the lethal action lines fall below the midpoint in in the  $\ell = +2$  panel.



*Figure 10.* Regression model trend lines for officer acceptability ratings. `acceptableOff` = acceptability rating of officer action. `Comp` = completely. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

**Officer acceptability ratings.** Acceptability ratings for officer actions showed an inverse pattern relative to punishment. This makes intuitive sense, as more punishable actions should be less acceptable. The trend lines for acceptability of officer action are plotted in Figure 10. Acceptability declines as action severity increases; the coefficient  $\beta_2$  had a

posterior mode of  $-1.20$  (with 95% HDI from  $-1.27$  to  $-1.14$ ). The decline is virtually unaffected by legitimacy beliefs as coefficient  $\beta_5$  is very near zero (with an HDI including zero), meaning that the slopes of the lines are nearly the same in the two panels of Figure 10. However, the interaction between civilian and officer action severity controlled by regression coefficient  $\beta_6$  had a positive posterior mode of  $0.41$  (and 95% HDI from  $0.36$  to  $0.45$ ). This indicates that as civilian action severity increases, acceptability judgments of officer actions decrease less steeply when officer actions become more severe. This interaction is visible in both panels of Figure 9 by observing that the slope of low-severity civilian actions (i.e., the purple solid lines) is much steeper than the slope of high-severity civilian actions (i.e., the red dot-dash lines).

Acceptability of officer reactions increased with the severity of the preceding civilian action. This can be seen in Figure 10 as the spread between the lines representing different civilian actions, with the solid (purple) lines for mild actions well below the dash-dot (red) lines of lethal actions. The regression coefficient  $\beta_1$  had a posterior mode of  $1.21$  (with 95% HDI from  $1.14$  to  $1.27$ ). The regression coefficient  $\beta_4$  was nearly zero, meaning that the effect of civilian action was similar across the range of legitimacy beliefs.

The primary influence of legitimacy can be seen in the overall height of the lines across panels: acceptability of officer action tends to be higher for high legitimacy than for low legitimacy. The regression coefficient  $\beta_3$  had a posterior mode of  $0.50$  (with 95% HDI from  $0.42$  to  $0.57$ ) indicating that rated acceptability of officers increased as legitimacy  $\ell$  increased.

Notice in the left panel of Figure 10, for participants with very low legitimacy beliefs, officer actions at SOLOF level 6 (above 5.0 on the latent severity scale) are rated as somewhat *unacceptable* (i.e., two are below the 0 midline, and two just above it) even when the civilian actions are also lethal. By contrast, in the right panel, for participants with very high legitimacy beliefs, lethal officer actions are acceptable, well above the midline, when

civilian actions are also lethal. This suggests that there may be a willingness among people with high legitimacy beliefs to accept lethal officer force as a viable response to lethal civilian force, but people with low legitimacy beliefs may rarely view lethal force as acceptable, even when they are confronting very severe civilian actions.

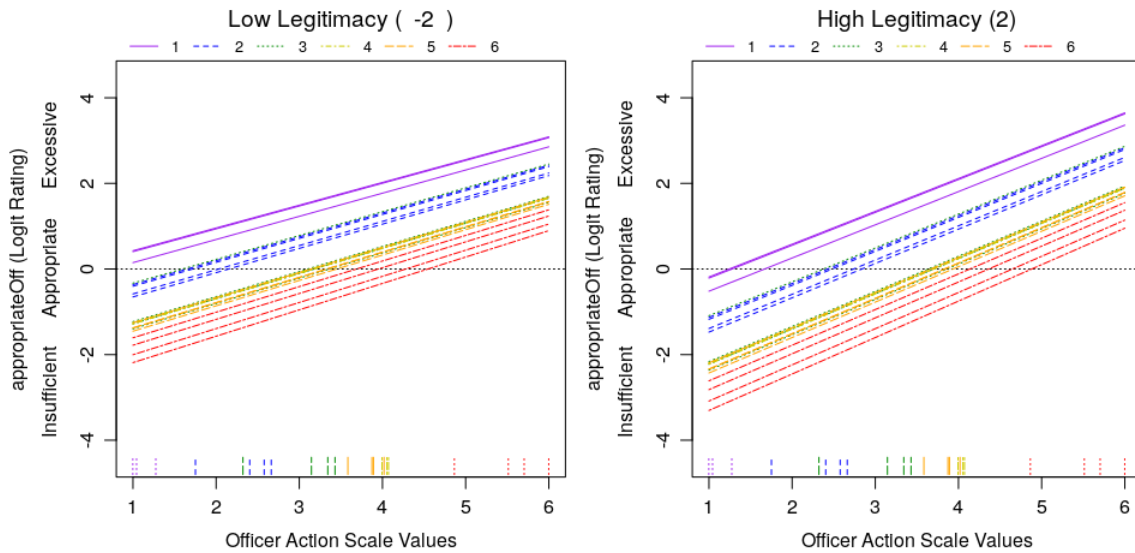


Figure 11. Regression model trend lines for officer appropriateness ratings. appropriateOff = appropriateness rating of officer action. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

**Officer appropriateness ratings.** Recall that for appropriateness of an action, the response slider was labeled “appropriate” at its midpoint, “insufficient” at its low end, and “excessive” at its high end. Therefore an increase on this scale is not always an increase in appropriateness, because an increase on the high end of the scale is actually away from appropriate and toward excessive. Figure 11 has a horizontal dotted line marking the midpoint of the scale, that is, ratings at the point labeled “appropriate.” Trend lines for appropriateness of officer action are plotted in Figure 11. The lines increase as officer action severity increases, meaning that ratings transition from insufficient through appropriate to excessive as officer actions become more severe. The regression coefficient  $\beta_2$  had a posterior mode

of 0.69 (with 95% HDI from 0.66 to 0.73).

The regression coefficient on civilian-action severity,  $\beta_1$ , had a posterior mode of  $-0.53$  (with 95% HDI from  $-0.56$  to  $-0.50$ ), indicating that ratings of officer appropriateness transitioned from excessive to appropriate to insufficient as civilian-action severity increased. The negative regression coefficient is reflected in Figure 11 by the decreasing intercepts of the the civilian-action lines as they become more severe.

Intriguingly, visual inspection of Figure 11 indicates that lethal officer actions tended to be rated on the excessive side of “appropriate.” That is, in both panels of Figure 11 the trend lines for the most severe officer actions (scale values of 5.0 to 6.0 on the horizontal axis) fall above the 0 midline, even when the civilian actions are also lethal (i.e., red dot-dash lines). Thus, even when the civilian used lethal force, a lethal response by the officer is rated toward excessive rather than squarely appropriate.

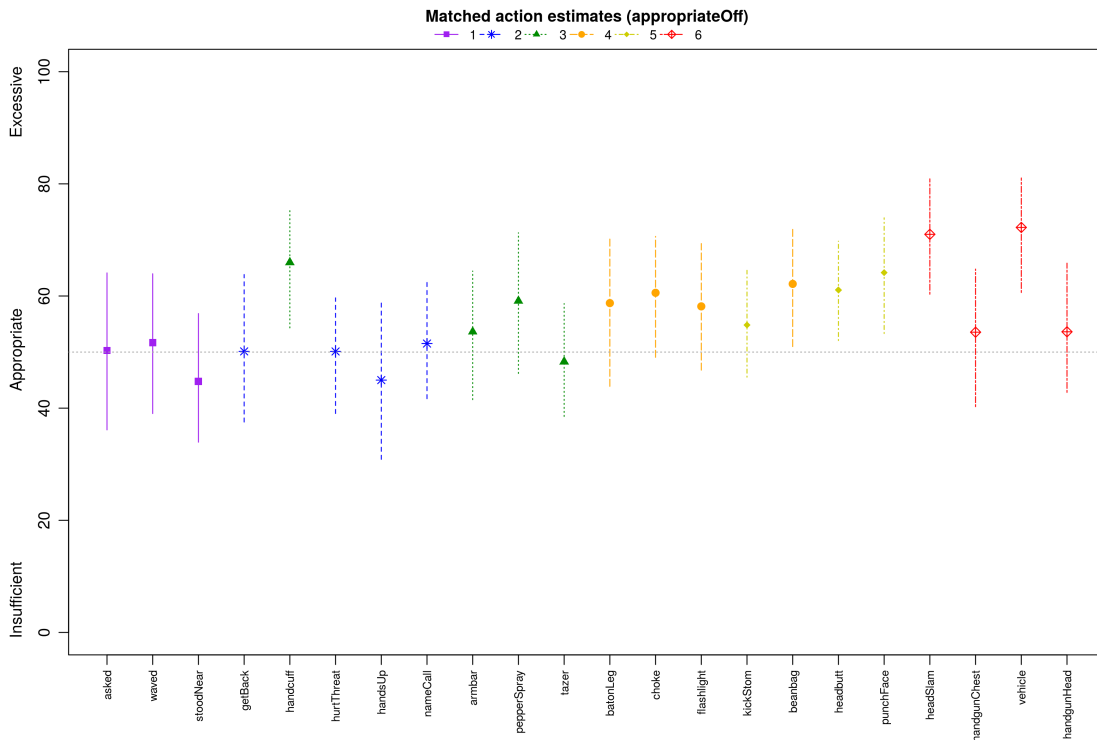
To further investigate this trend, we isolated the subset of trials in which the officer and civilian actions exactly matched (or corresponded in the case of differing officer/civilian actions). For example, matched trials for "handgunChest" would include only trials in which the civilian and the officer both shot each other in the chest. We then modeled ratings of officer appropriateness for each action while including legitimacy beliefs as a covariate:

$$\hat{y}_a = \alpha_{0,a} + \alpha_1 \cdot \ell \quad (4)$$

where  $\hat{y}_a$  is the predicted rating (logit transformed and censored) and where  $a$  (1–22) indexes the action. This analysis yielded intercept coefficients ( $\alpha_{0,a}$  in Eqn. 4) that reflect the estimated appropriateness rating for each action when legitimacy beliefs are at their mean (recall that legitimacy beliefs are standardized, so 0 is the mean). Notice that this analysis does not involve estimating latent action severities. See supplementary material for detailed

analysis code and further details (<https://osf.io/h6jg3>).

Importantly, the appropriateness question was unique among our dependent measures in that it provided participants with a middle anchor labeled “appropriate” in addition to the two end point anchors of “insufficient” and “excessive.” As a result, analyzing departures from the midpoint is meaningful for this question, and matched actions would be expected to fall exactly at the mid-point if participants general expectations are that officers should be meeting civilian resistance with the same level of force.



*Figure 12.* Mean ratings of officer appropriateness using only matched civilian/officer actions ( $\alpha_{0,a}$  in Eqn. 4). Vertical segments show 95% HDI’s. Vertical axis shows ratings on original scale, whereas analysis was performed on logit-transformed (and censored) ratings. Actions on the horizontal axis are ordered according to their latent severity (Celestin & Kruschke, 2019a). Symbols and colors indicate SOLOF level (as in Figure 7).

As seen in Figure 12, appropriateness ratings for low-force actions fall very close to the midpoint of the scale. However, as actions become more severe (toward the right side of

Figure 12), the predicted rating falls toward the excessive side of the scale. That is, even when the officer responds with the same action as the civilian, physically forceful officer actions still tend to be rated as somewhat excessive. In particular, the officer reactions rated as especially excessive include headButt, punchFace, headSlam, and vehicle, even in response to the identical action by the civilian. These actions are not intuitively normative for police officers, and therefore may be judged to be particularly excessive. On the other hand, it is normative for police officers to use their officially-issued hand gun, which may explain why handgunChest and handgunHead have mean ratings near the “appropriate” midline. An unexpected result in Figure 12 is that handcuff is rated as excessive, despite it being a normative police action. We believe this result is idiosyncratic for the particular matched civilian action, pulling away. Whereas pulling away has physical force comparable to handcuffing, pulling away by itself may seem not to merit handcuffing. Celestin and Kruschke (2019a) discuss the normativity of actions in terms of the estimated scale values of the actions.

Overall, this matched-action analysis is largely congruent with the primary analysis and suggests that lay evaluators may judge an officer reaction that matches a civilian action to be somewhat excessive. However, there are some inconsistencies between analyses. In Figure 11, all four lethal-action lines seem to fall above the midpoint at their matched action rug-plot points, and to roughly the same extent. This contrasts with the result in Figure 12 where the two gun-related actions fall closer to the midpoint. A second discrepancy is visible in the regression lines of Figure 11 which suggest that the ratings for handcuffing (in response to the matched civilian action "pullingAway") should be roughly at the midpoint or slightly below (depending on  $\ell$ ); this stands in contrast to Figure 12 which indicates that handcuffing falls above the midline. These discrepancies result from the flexibility of Equation 4 relative to the more restrictive linearity of Equation 3. On the other hand, the analysis of matched actions involved only a small subset of data, while the full regression

analysis involved all the data. The general trend of physically forceful officer actions being rated on the excessive side of appropriate remains supported by both analyses.

Returning now to the primary analysis in Figure 11, the effect of legitimacy beliefs can be discerned from visual comparison of the two panels in Figure 11. At the right edges of the two panels, the trend lines rise to approximately the same heights of excessiveness. But at the left edges of the two panels, the trend lines are lower, toward “insufficient,” for the high-legitimacy panel. That is, participants with very high legitimacy beliefs tended to rate mild officer actions as more insufficient than participants with very low legitimacy beliefs. This difference between panels is expressed mathematically by the combined effects of the regression coefficients involving legitimacy. The coefficient  $\beta_3$  had a posterior mode of  $-0.07$  (with 95% HDI from  $-0.11$  to  $-0.03$ ) indicating that higher legitimacy participants tended to have lower ratings overall (i.e., more insufficient, less excessive), but the coefficient  $\beta_5$  had a posterior mode of  $0.06$  (with 95% HDI from  $0.03$  to  $0.09$ ) indicating that higher legitimacy participants tended to have steeper slopes on officer severity. The combined effect of the coefficients is that high-legitimacy participants tend to rate mild officer actions as more insufficient than low-legitimacy participants (left edge of the panels), but high-legitimacy participants tend to rate severe officer actions about the same as low-legitimacy participants (right edge of the panels). The interaction of civilian and officer action severity represented by the  $\beta_6$  coefficient was close to zero with an HDI that included zero, indicating that appropriateness ratings changed at a consistent rate across actions.

**Officer force ratings.** The trend lines for ratings of the forcefulness of an officer reaction are plotted in Figure 13. Perhaps the most prominent feature of the graphs, in contrast to previous graphs, is that the lines for different civilian actions are close together, especially for participants with high legitimacy (right panel). This dense packing of the

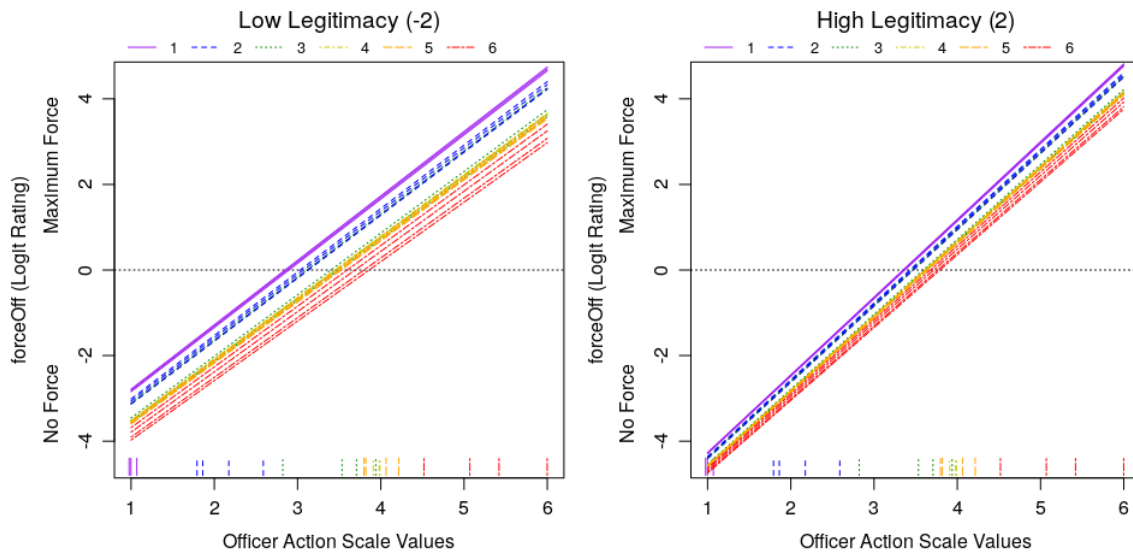


Figure 13. Regression model trend lines for officer force ratings. forceOff = force rating of officer action. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

lines makes intuitive sense: The physical forcefulness of an action is inherent in the action itself, regardless of what may have happened before the action. The small separation of the lines is determined mathematically by the value of regression coefficient  $\beta_1$ , which had a posterior mode of only  $-0.22$  (95% HDI from  $-0.25$  to  $-0.20$ ). The negative sign indicates that people rated the physical forcefulness of an officer reaction less severely as the civilian-action severity increased. But the rate of decline was small compared to moral ratings (cf. for punishment  $\beta_1 = -1.13$ ). The slope on civilian-action severity was modulated slightly by legitimacy beliefs, as can be seen in the tighter packing of the lines for high legitimacy (right panel) than for low legitimacy (left panel). This interaction with legitimacy is governed mathematically by regression coefficient  $\beta_4$ , which had a posterior mode of  $0.04$  (and 95% HDI from  $0.01$  to  $0.06$ ).

Legitimacy also had small modulating effects on the baseline and slope with respect to officer-action severity. The baseline decreased slightly as legitimacy beliefs increased, with

$\beta_3$  having a posterior mode of  $-0.09$  (and 95% HDI from  $-0.13$  to  $-0.05$ ). Simultaneously, the slope on officer-action severity increased slightly as legitimacy belief increased, with  $\beta_5$  having a posterior mode of  $0.08$  (and 95% HDI from  $0.05$  to  $0.11$ ). Together, these influences of legitimacy are most noticeable at the left edges of the panels in Figure 13: participants with low legitimacy beliefs tended to rate mild officer actions as having a bit more forcefulness than the ratings from participants with high legitimacy beliefs.

**Summary of officer action ratings.** Punishment ratings increased as officer-action severity increased, and punishment ratings decreased as civilian-action severity increased. Consistent with those trends, acceptability ratings lessened as officer actions became more severe, while acceptability ratings escalated as civilian actions became more severe. Punishment and acceptability ratings were modulated by the interaction of civilian and officer actions such that they increased/decreased at greater rates when civilian actions were less severe. Therefore, participants were judging the moral aspects of officer actions in the context of the preceding civilian action.

As participant legitimacy beliefs increased, overall punishment ratings decreased and acceptability ratings increased. Notably, participants with very low legitimacy beliefs viewed lethal officer actions as *unacceptable* even when the civilian actions were also lethal. This finding suggests it is possible that lethal force is not viewed as a viable police response for people with very low legitimacy beliefs about police.

With respect to ratings of appropriateness, lethal force by an officer was viewed as slightly excessive even after lethal civilian force and regardless of participants' legitimacy beliefs. On the other hand, participants with high legitimacy beliefs tended to rate mild officer reactions as insufficient more than participants with low legitimacy beliefs.

Finally, perceived physical force of officer actions increased with action severity, but decreased slightly as civilian actions became more severe. In principle, a judgment of an

action’s physical forcefulness should depend only on the action itself and not on any preceding action. The participants in these studies showed an influence of the preceding action, with low-legitimacy participants showing a slightly larger influence, but the influence was much smaller than in the ratings of moral consequences.

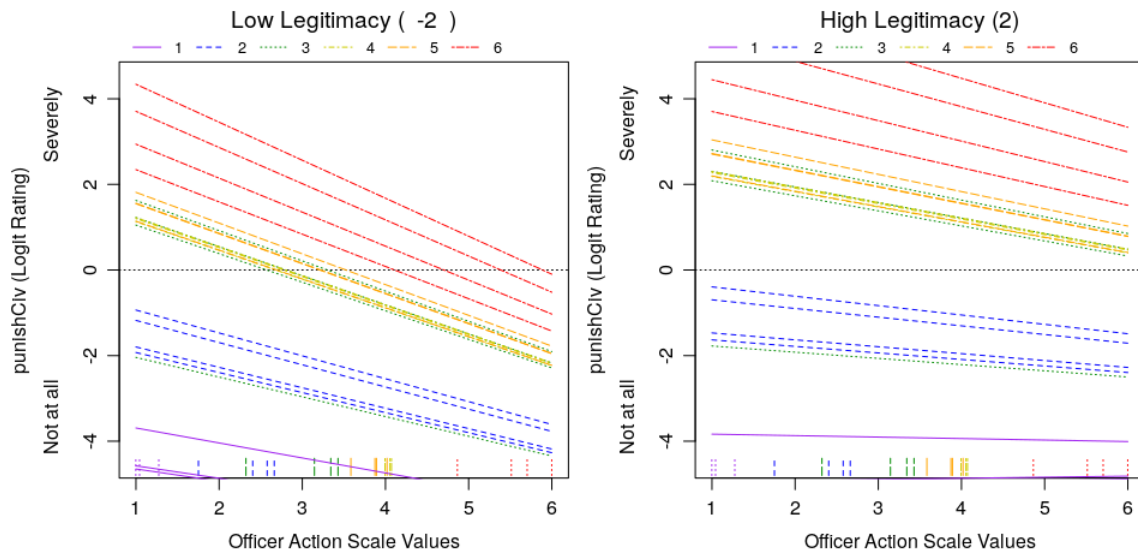


Figure 14. Regression model trend lines for civilian punishment ratings.  $\text{punishCiv}$  = punishability rating of civilian action. Notice the horizontal axis is *officer* action severity. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

**Civilian punishment ratings.** Trend lines for punishment ratings of civilian actions (after a preceding officer action) are plotted in Figure 14. Notice that the horizontal axis continues to represent *officer* action severity, as in previous graphs. And, as before, the effect of *civilian* action severity is represented by the spread between the lines for different civilian actions. In can be seen in both panels of Figure 14 that the lines are quite spread out, indicating a large influence of civilian-action severity on the punishability of the action. Regression coefficient  $\beta_1$  had a posterior mode of 1.72 (95% HDI from 1.67 to 1.78). Moreover, the spread of the lines is larger as legitimacy beliefs increase; coefficient  $\beta_4$  had

a posterior mode of 0.11 (95% HDI from 0.07 to 0.16).

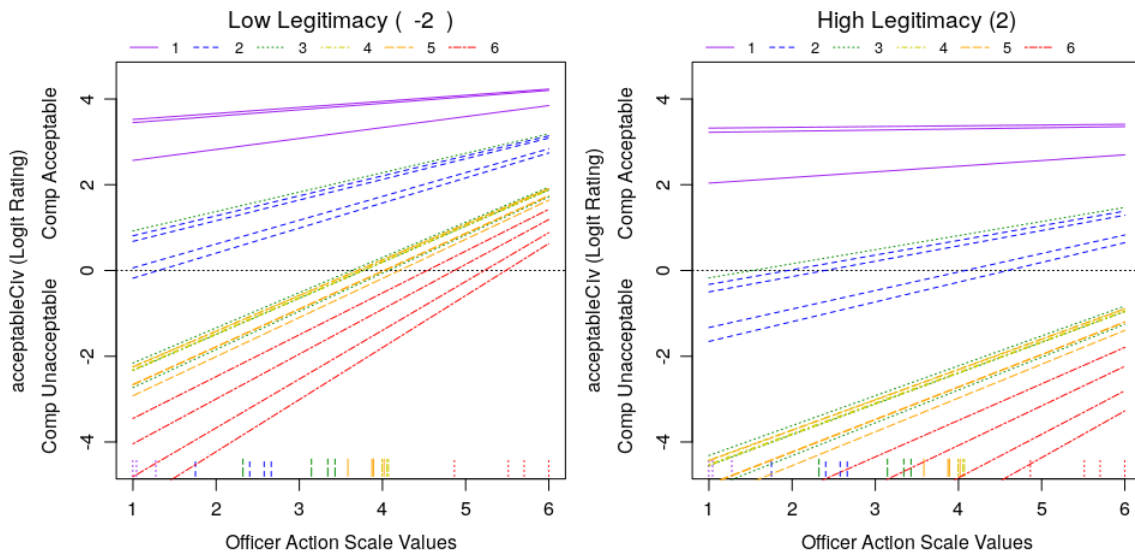
In both panels of Figure 14 the lines descend, indicating that the punishability of a civilian action declines as the severity of the preceding officer action increases. This decline is captured by coefficient  $\beta_2$  which had a posterior mode of  $-0.43$  (and 95% HDI from  $-0.46$  to  $-0.39$ ). Interestingly, the magnitude of the decline depended on the legitimacy belief of the participant, with low-legitimacy beliefs (left panel of Figure 14) endorsing a steeper decline in punishability than high-legitimacy beliefs (right panel of Figure 14). This change in slopes is captured by coefficient  $\beta_5$  which had a posterior mode of 0.08 (95% HDI from 0.04 to 0.11).

The interaction between civilian and officer action severity controlled by regression coefficient  $\beta_6$  had a negative posterior mode of  $-0.12$  (and 95% HDI from  $-0.15$  to  $-0.09$ ). This indicates that as officer action severity increases, punishments judgments of civilian actions increase less steeply when officer actions are more severe. This interaction is visible in both panels of Figure 14 by noticing that the slope of low-severity civilian actions (i.e., the purple solid lines) is much less steep than the slope of high-severity civilian actions (i.e., the red dot-dash lines).

Visual inspection of the two panels of Figure 14 also suggests that punishment ratings tended to be higher overall for participants with high legitimacy beliefs. This is robustly verified by the regression coefficient  $\beta_3$  which had a posterior mode of 0.38 (and 95% HDI from 0.33 to 0.44).

It is interesting that the influence of legitimacy on baseline punishability is almost as strong (in magnitude) for ratings of civilian actions as for officer actions (i.e.,  $\beta_3 = 0.38$  for civilian actions and  $\beta_3 = -0.51$  for officer actions). Moreover, the influence of legitimacy on the effect of civilian action severity is stronger for ratings of civilian punishability than ratings of officer punishability (i.e.,  $\beta_4 = 0.11$  for civilian actions and  $\beta_4 = -0.01$  for officer actions), and the influence of legitimacy on the effect of officer action severity

is stronger for ratings of civilian punishability than ratings of officer punishability (i.e.,  $\beta_5 = 0.08$  for civilian actions and  $\beta_5 = -0.03$  for officer actions). The differences in magnitudes of the  $\beta_4$  and  $\beta_5$  coefficients across civilian and officer ratings are marginal, but similar trends appear in judgments of acceptability and appropriateness. In other words, the influence of legitimacy is at least as strong for judgments of civilian reactions to police as for judgments of police reactions to civilians.



*Figure 15.* Regression model trend lines for civilian acceptability ratings. `acceptableCiv` = acceptability rating of civilian action. Notice the horizontal axis is *officer* action severity. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

**Civilian acceptability ratings.** Acceptability ratings of civilian actions were consistent with punishment ratings: the patterns were inverses of each other. Trend lines for acceptability ratings of civilian actions (i.e., reactions to officer actions) are shown in Figure 15. Again the lines are quite spread out, indicating a large influence of civilian-action severity on the acceptability of the action. Regression coefficient  $\beta_1$  had a posterior mode of  $-1.56$  (95% HDI from  $-1.63$  to  $-1.51$ ). Moreover, the spread of the lines is larger (more

negative) as legitimacy beliefs increase; coefficient  $\beta_4$  had a posterior mode of  $-0.15$  (95% HDI from  $-0.20$  to  $-0.11$ ).

In both panels of Figure 15 the lines ascend, indicating that the acceptability of a civilian action increases as the severity of the preceding officer action increases. This increase is captured by coefficient  $\beta_2$  which had a posterior mode of  $0.62$  (95% HDI from  $0.57$  to  $0.66$ ). There is marginal evidence that the magnitude of the increase depended on the legitimacy belief of the participant, with low-legitimacy beliefs (left panel of Figure 15) endorsing a steeper increase in acceptability than high-legitimacy beliefs (right panel of Figure 15). This change in slope is captured by coefficient  $\beta_5$  which had a posterior mode of  $-0.03$  (95% HDI from  $-0.07$  to  $0.01$ ).

The interaction between civilian and officer action severity controlled by regression coefficient  $\beta_6$  had a positive posterior mode of  $0.21$  (and 95% HDI from  $0.18$  to  $0.25$ ). This indicates that as officer action severity increases, acceptability judgments of low severity civilian actions increases only slightly (if at all), while acceptability of more severe civilian actions increases much more. This interaction is visible in both panels of Figure 14 by noticing that the slope of low-severity civilian actions (i.e., the purple solid lines) is much less steep than the slope of high-severity civilian actions (i.e., the red dot-dash lines).

Visual inspection of the two panels of Figure 15 also suggests that ratings of acceptability of civilian actions tended to be lower overall for participants with high police-legitimacy beliefs. This is robustly verified by the regression coefficient  $\beta_3$  which had a posterior mode of  $-0.53$  (95% HDI from  $-0.59$  to  $-0.46$ ).

The influence of legitimacy on baseline acceptability is as strong (in magnitude) for civilian actions as for officer actions (i.e.,  $\beta_3 = -0.50$  for civilian actions and  $\beta_3 = 0.51$  for officer actions). Moreover, the influence of legitimacy on the effect of civilian action severity is stronger for ratings of civilian acceptability than ratings of officer acceptability (i.e.,  $\beta_4 = -0.15$  for civilian actions and  $\beta_4 = 0.04$  for officer actions), and the influence

of legitimacy on the effect of officer action severity is no weaker for ratings of civilian acceptability than ratings of officer acceptability (i.e.,  $\beta_5 = -0.03$  for civilian actions and  $\beta_5 = 0.05$  for officer actions, HDIs both include 0). In other words, the influence of legitimacy is at least as strong for judgments of civilian reactions to police as for judgments of police reactions to civilians.

Interestingly, the severe civilian action lines in the low legitimacy panel ( $\ell - 2$ ) appear well above the midpoint of acceptability when officer action severity is high, while they remain well below the midpoint even at extreme officer severity levels in the high legitimacy panel ( $\ell + 2$ ). That is, the right panel of Figure 15 shows that for participants with the highest legitimacy beliefs, violent civilian actions are not acceptable even when in response to lethal officer actions. In contrast, participants with the lowest legitimacy beliefs rate lethal civilian actions well above the acceptability midpoint when officer actions are lethal.

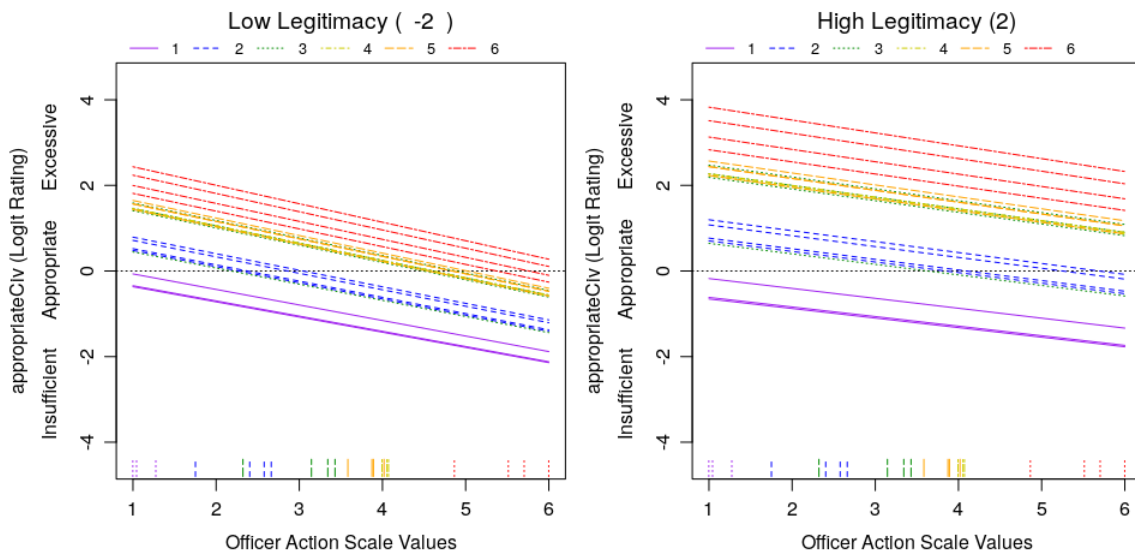


Figure 16. Regression model trend lines for civilian appropriateness ratings. appropriateCiv = appropriateness rating of civilian action. Notice the horizontal axis is officer action severity. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

**Civilian appropriateness ratings.** The trend lines for appropriateness of civilian action are plotted in Figure 16. Recall that the anchors for the response scale went from “insufficient” through “appropriate” to “excessive.” The regression coefficient on civilian action,  $\beta_1$ , had a posterior mode of 0.69 (with 95% HDI from 0.66 to 0.73), indicating that judgments of civilian action transitioned from insufficient to appropriate to excessive as their severity increased. This slope on civilian action depended on legitimacy belief, as coefficient  $\beta_4$  had a posterior mode of 0.08 (with 95% HDI from 0.05 to 0.12). This effect can be seen by comparing the two panels of Figure 16, in which the lines are more spread out for  $\ell = +2$  than for  $\ell = -2$ .

Ratings declined with the severity of the preceding officer action, as shown by the descending lines in Figure 16. Coefficient  $\beta_2$  had a posterior mode of  $-0.33$  (with 95% HDI from  $-0.36$  to  $-0.30$ ). The slope of the lines did not depend much on legitimacy ( $\beta_5$  had a posterior mode of 0.03 with 95% HDI from 0.00 to 0.07), as can be seen by the only subtle change in slope across panels.

Perhaps the most prominent feature of Figure 16 is that severe civilian actions (dash-dot red lines) tend to be rated toward excessive (above the midline) for very-high legitimacy respondents, even when the preceding officer action is also severe. But severe civilian actions (dash-dot red lines) tend to be rated as appropriate (near the midline) for very-low legitimacy respondents when the preceding officer action is also severe. This difference is captured in part by the overall influence of legitimacy on the intercept, given by coefficient  $\beta_3$ , which had a posterior mode of 0.22 (with 95% HDI from 0.17 to 0.27). The effect of legitimacy on baseline appropriateness is stronger for judgments of civilian actions than for judgments of officer actions ( $\beta_3 = 0.22$  for civilian actions and  $\beta_3 = -0.07$  for officer actions).

We also analyzed responses to matched civilian reactions, just as we did for officer reactions in Equation 4. Figure 17 plots the intercept estimates of the model, which indicate

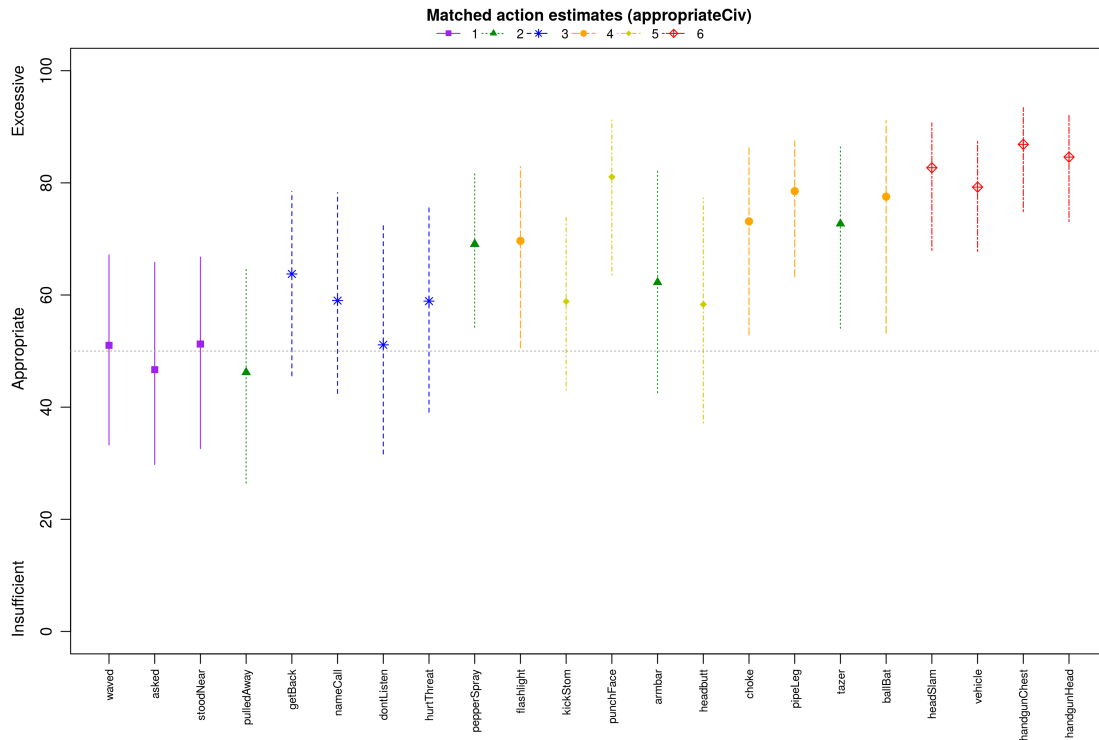


Figure 17. Mean ratings of civilian appropriateness using only matched civilian/officer actions. Vertical segments show 95% HDI's. Vertical axis shows ratings on original scale, whereas analysis was performed on logit-transformed (and censored) ratings. Actions on the horizontal axis are ordered according to their latent severity (Celestin & Kruschke, 2019a). Symbols and colors indicate SOLOF level (as in Figure 7).

the predicted appropriateness rating at mean legitimacy ( $\ell = 0$ ) for each civilian action. It is clear in Figure 17 that the majority of civilian actions fall above the appropriateness midpoint toward the excessive end of the scale, with the more severe civilian actions being closer to the excessive end of the scale.

A visual comparison of Figures 12 and 17 shows that the rated excessiveness of officer matched reactions tends not to be as extreme as civilian matched reactions. We suspect this difference arises from different expectations of police and civilian behavior. Police are legally authorized to use force but civilians are rarely authorized to use force against police. That is, any violent civilian action is excessive by default unless accompanied by

some additional justification.

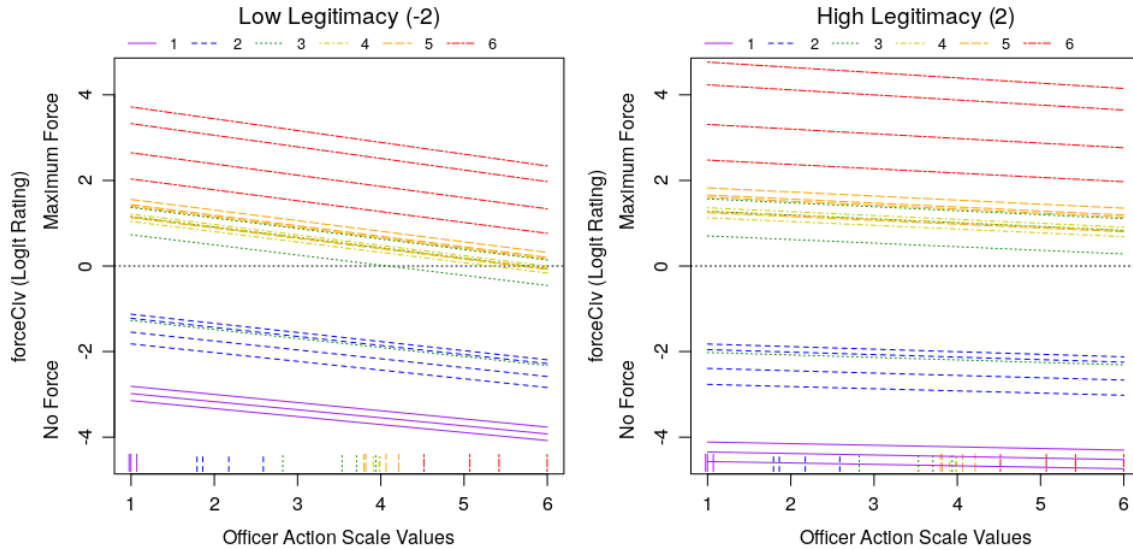


Figure 18. Regression model trend lines for civilian force ratings.  $\text{forceCiv}$  = force rating of civilian action. Notice the horizontal axis is *officer* action severity. Solid purple lines denote the least severe civilian actions, and dot-dash red lines denote the most severe civilian actions.

**Civilian force ratings.** Trend lines for ratings of forcefulness of civilian action are shown in Figure 18. As indicated by large spread between lines, there was a strong influence of civilian action severity, with  $\beta_1$  having a posterior mode of 1.58 (and 95% HDI from 1.52 to 1.64). The spread grew somewhat with legitimacy, as  $\beta_4$  had a posterior mode of 0.12 (with 95% HDI from 0.09 to 0.15).

The overall ratings of civilian force increased slightly with legitimacy, as  $\beta_3$  had a posterior mode of 0.05 (with 95% HDI from 0.01 to 0.09). This can be discerned visually in Figure 18 by the subtle overall increase in the height of lines in the right panel relative to the left panel.

Perhaps the most interesting aspect of the civilian force ratings is that they depend (to a small degree) on the preceding officer action. The trend lines in Figure 18 descend; coeffi-

cient  $\beta_2$  had a posterior mode of  $-0.15$  (with 95% HDI from  $-0.18$  to  $-0.13$ ). Notice that the magnitude of the influence of preceding action on this physical attribute is far smaller than for moral attributes. Interestingly, the influence of the preceding officer action is less for high-legitimacy participants than for low-legitimacy participants ( $\beta_5$  had a posterior mode of  $0.04$  with 95% HDI from  $0.01$  to  $0.07$ ), as can be seen in Figure 18 by the less steep lines in the right panel.

**Summary of civilian action ratings.** Ratings of civilian punishment decreased as the severity of preceding officer action increased, but this decrease was stronger for participants with low legitimacy beliefs.

Acceptability of civilian action increased as the severity of preceding officer action increased, and this increase was stronger for participants with low legitimacy beliefs. Participants with very high legitimacy beliefs tended to rate severe civilian actions as unacceptable even when preceded by a severe officer action, unlike participants with low legitimacy beliefs. Civilian punishment and acceptability ratings were modulated by the interaction of civilian and officer actions such that they decreased and increased, respectively, at greater rates when civilian actions were more severe. That is, low-severity civilian actions are always highly acceptable and not deserving of punishment, but high-severity civilian actions become much more acceptable and much less deserving of punishment as officer actions become severe—especially for participants with low legitimacy beliefs.

Appropriateness of civilian action went from insufficient to excessive as action severity increased, but had the reverse trend as officer action severity increased. Participants with very high legitimacy beliefs tended to rate severe civilian actions as excessive even when preceded by a severe officer action, unlike participants with low legitimacy beliefs.

In general, the influence of legitimacy beliefs was just as strong on ratings of civilian reactions to police as on ratings of officer reactions to civilians. This is interesting because

the questions for assaying legitimacy are explicitly focused on attitudes about police, not about civilians.

Finally, perceived physical force of civilian actions depended to a small degree on the severity of the preceding officer action, and this influence of officer action was stronger for low-legitimacy participants.

### **Discussion and Conclusions**

We measured lay evaluations of police and civilian use of force using a unique paradigm incorporating actions from the complete range of law enforcement force options. The results provide a detailed view of the way lay observers make moral and physical evaluations of representative police and civilian actions, and the effects of legitimacy beliefs on those evaluations.

Results showed that as officer actions become more severe, they are viewed as more physically forceful, and officer punishment and excessiveness ratings increase while acceptability ratings decrease. As civilian actions become more severe, they are also viewed as more physically forceful, and civilian punishment and excessiveness ratings increase while civilian acceptability ratings decrease. Moreover, both evaluations depend on the severity of the preceding action, such that the severity of the preceding action increases acceptability and decreases punishability of subsequent severe actions. Evaluations of physical force also depend, to a lesser magnitude, on the severity of the preceding action.

Beyond these primary effects, individual differences in legitimacy beliefs altered both moral and physical evaluations of officer and civilian violence. As legitimacy beliefs decreased, officer punishment ratings were higher and acceptability ratings were lower. Thus, as one might expect, low legitimacy individuals are generally skeptical of the moral legitimacy of police use of force across the entire spectrum of actions (i.e., not only for extreme police actions). Intriguingly, low-legitimacy individuals also exhibit increased per-

ceptions of moral legitimacy for civilian violence, even for lethal actions, although the rate of increase is attenuated relative to low-legitimacy individuals. Legitimacy beliefs also affected judgments of physical force (though not as much as for moral judgments), with low-legitimacy participants showing a stronger influence of preceding civilian action on judgments of officer force, and a stronger influence of preceding officer action on judgments of civilian force. The influence of legitimacy was just as strong for judgments of civilian reactions to police as for judgments of police reactions to civilians.

One possible explanation for the tendency of legitimacy to increase the morality of either police or civilian violence is that legitimacy shifts the extent to which each participant in a police-civilian encounter is perceived as a moral agent deserving of judgment, or a moral patient deserving of sympathy. Gray and Wegner (2009) proposed that moral judgments possess a fundamental structure in which parties involved are pigeon holed into either a perpetrator (agent) or a victim (patient), a process known as "moral typecasting" (see also Gray, Schein, and Ward (2014) and Schein and Gray (2017)). Viewed in this light, perhaps the inverse effects of legitimacy on police and civilian evaluations of violence reflect the tendency of low-legitimacy individuals to ascribe patiency primarily to citizens, while high-legitimacy individuals ascribe patiency more to police. This proposed role of moral typecasting is speculative and requires additional research.

From this rich array of results, perhaps the most surprising and important finding is that ratings of lethal officer force are consistently on the "excessive" side of appropriate, even when the preceding civilian action is also lethal, and even for high-legitimacy participants. This result was demonstrated by the analysis involving simultaneous estimation of latent scale values (see Figure 11) and by the analysis involving ratings of matched-severity actions (see Figure 12). The trend lines of Figure 11 suggest that when a preceding civilian action is lethal (in SOLOF category 6), an officer response that is squarely "appropriate" should be less severe, as low as SOLOF category 5 for low-legitimacy participants. Qual-

ifying this general observation that lethal officer force always tends to be perceived as excessive are the observed differences in acceptability and punishment of officer and civilian actions across different levels of participant legitimacy beliefs. Specifically, low-legitimacy participants view civilian violence as acceptable and much less worthy of punishment when officer actions become sufficiently severe, while high-legitimacy participants seem much less willing to legitimate civilian violence. Importantly, these results demonstrate the need to take both civilian and officer actions jointly into account when formulating models of lay evaluations of police use of force, because the actions of both have meaningful effects on judgments of punishment, acceptability, appropriateness, and even physical magnitude.

One reasonable conclusion from these results is that law-enforcement policies that allow officers to use force that matches or exceeds the severity of civilian resistance may be perceived as excessive, even by lay observers with strong beliefs in the legitimacy of police. It seems that lay observers expect police to react to civilian force with somewhat less severity than the civilian force. This finding directly conflicts with typical policies that license police officers to use somewhat greater force in order to quickly control the situation, mitigate collateral harm, and achieve compliance. Attaining resolution between police and civilian perspectives may require some combination of education and tuning of use-of-force policy. Specifically, educating the public about SOLOF structure and the need for "one-level-up" police heuristics may increase the extent that they are sufficiently knowledgeable about the details of police policy and also convinced of the moral justifiability of the current use-of-force models. Simultaneously, re-evaluation of SOLOF-style categorization and accompanying heuristics by police policy-makers in order to emphasize police reactions that maximize alignment with public expectations according to the data we provide may prove useful. In the limit, attenuating the fundamental conflict we observe between lay perceptions and police policy is likely to be a complicated pursuit. Nevertheless, democratic principles demand that police decisions about appropriate levels

of force have broad public support, and merging these two strategies (i.e., educating the public and re-evaluation of current use-of-force policy) with the current results in mind may be a good empirically-informed starting point to narrow the gap between police and civilian perceptions of what constitutes legitimate police violence.

## **Chapter 4: Lay evaluators expect police to use less force than civilians and more normative force**

Police use of force in the United States has been increasingly scrutinized in recent years. Most notably, cases of lethal police-civilian encounters have been thrust into the national spotlight (McWhorter, 2016), but public discourse regarding all levels of applied force has become prominent, and a wide range of citizen objections have emerged as video recording technology such as cell phone (civilian) and body-worn (police) cameras are now commonplace in western society (Culhane et al., 2016). Incidents of police-civilian conflict are now immediately and routinely broadcast across social media, and so the importance of developing a scientific understanding of how lay observers evaluate forceful police actions in relation to the way police and the criminal justice system approach force selection and evaluation is acute. Nevertheless, very little empirical research has compared lay evaluations with the methods police use to make choices about appropriate forceful actions.

### **Force Continua and Heuristic Rules**

Over the past several decades, law enforcement agencies in the United States have developed a general protocol for categorizing forceful actions according to their assessed severity. Specifically, actions are commonly sorted into rank-ordered bins and are situated in relation to similar levels of civilian resistance (Terrill et al., 2003; Jefferis et al., 2011; Terrill et al., 2011). These lists of actions sorted according to severity are commonly called a “force continuum,” although we have previously referred to them as the six ordinal levels of force (SOLOF) scale in order to highlight their ordinal nature. There is considerable variety among different police agencies regarding many aspects of these scales, but more than 80% of U.S. agencies employ them in some form, and the majority of U.S. agencies use a six-level version (Terrill et al., 2011). Accompanying a SOLOF scale, police agencies and educators frequently employ a heuristic rule dictating that officers may use one level of

force above the corresponding level of civilian resistance in order to minimize the length of conflict and risks to all involved in volatile situations, as well as to maximize the likelihood of inducing compliance. All over the country, individual police officers are constantly using these tools (i.e., a SOLOF scale and the "one-above" rule) to make choices about which forceful action to employ on the street. And yet, we are unaware of any systematic empirical research that has investigated whether the moral intuitions of lay members of society exhibit consensus with or divergence from these decision-making tools and practices. That is, it is unknown what level of force severity citizens generally expect police to use in response to subject resistance.

### **Action Normativity and Moral Severity**

In prior work (Celestin & Kruschke, 2019a, 2019b), we have developed precise scales for the moral and physical magnitude of a representative set of forceful actions. These scales reveal that some actions are perceived to be much more morally than forcefully severe, while for other actions the inverted pattern is true. Intuitively, the actions that are disproportionately morally severe are those that seem non-normative for police (e.g., slamming someone's head into a curb), while actions that exhibit the inverted pattern seem to be especially normative for police (e.g., handcuffing). Based on these findings, we have hypothesized that role-normativity (i.e., the strength of unique association between a given action and police) substantively drives moral evaluations of police use of force at the level of decontextualized actions.

In the social psychology literature, the terms "norm" and "normativity" are frequently used to describe social norms, but it has been rightfully pointed out that these terms are somewhat ambiguous and can have multiple meanings (Shaffer, 1983). In particular, the concept of normativity has been decomposed into two primary types of norms: injunctive norms are those that reflect moral judgment and describe what people ought to do (i.e.,

what is morally appropriate), while descriptive norms are those that reflect behavior that is common or typical for others (i.e., what people actually do) (Cialdini, Reno, & Kallgren, 1990). Both injunctive and descriptive norms occur in the context of specific groups or settings. For example, it can be normative for grocery shoppers to return their carts to a collection location when they are done unloading in the parking lot—both because it is what they should do (i.e., injunctively normative) and/or because it is what most other people do (i.e., descriptively normative). Violating social norms is often perceived by observers as a moral violation, in part because norm violations are more informative regarding the mental states and intentions of the violator than normative actions which are perceived as default behaviors (Uttich & Lombrozo, 2010).

Congruent with our normativity hypothesis regarding police actions, previous research has demonstrated the existence of a “common is moral” heuristic (Eriksson et al., 2014; Lindström et al., 2018). Specifically, this research has found that people tend to infer the morality of a social behavior according to its relative frequency, and that actions are judged as more moral and merit less punishment when they are common than when they are rare. Moreover, the findings of Eriksson et al. (2014) suggest that injunctive and descriptive norms are not well distinguished psychologically, and that people frequently infer normative judgments from perceived descriptive norms and vice versa. For this reason, we use the term “normativity” in a more general way that includes both injunctive and descriptive meanings because the current research does not have a mechanism to distinguish descriptive from injunctive normativity, and also because both types of normativity frequently co-occur (Eriksson et al., 2014). Moreover, we believe it is likely that both are contributing to the moral/physical severity discrepancies we observe for different police actions.

To our knowledge, no studies have extended the “common is moral” heuristic into the domain of criminal justice. In particular, no prior research has directly assessed the extent to which forceful actions are differentially associated with police or whether any observed

differences are directly associated with differences in moral judgment. Thus, the current studies are broadly designed to assess the level of force lay evaluators judge as appropriate for police when responding to civilian resistance, and also to measure the differential strength of association of different forceful actions with police and the relationship of those normativities with previously observed discrepancies in moral severity judgments.

### **Method: Studies 1a, 1b, and 1c**

All of the experiments presented here were approved by the Indiana University Institutional Review Board in Bloomington, Indiana.

#### **Participants**

In Study 1a, we recruited 478 participants from the Indiana undergraduate subject pool. 377 of these participants passed experimental manipulation checks and were included in all subsequent analyses. According to participant demographic responses, 268 were female, 108 male, and 1 participant selected other. There was 1 American Indian/Alaska Native participant, 46 Asian participants, 22 Black or African American participants, 13 Hispanic or Latino participants, 277 White participants, 17 participants who selected More than one, and 1 participants who selected Unknown. Twenty eight participants identified as very liberal, 93 as liberal, 48 as slightly liberal, 70 as moderate/middle of the road, 42 as slightly conservative, 32 as conservative, 12 as very conservative, 37 as don't know/not political, 11 as libertarian, and 4 as other.

In Study 1b, we recruited 553 participants from the Indiana undergraduate subject pool. 418 of these participants passed experimental manipulation checks and were included in all subsequent analyses. According to participant demographic responses, 213 were female, 203 male, and 2 participant selected other. There were 47 Asian participants, 29 Black or African American participants, 24 Hispanic or Latino participants, 292 White participants,

Twenty four participants who selected More than one, 1 Native Hawaiian or Other Pacific Islander, and 1 participants who selected Unknown. Thirty one participants identified as very liberal, 101 as liberal, 51 as slightly liberal, 69 as moderate/middle of the road, 39 as slightly conservative, 62 as conservative, 8 as very conservative, 41 as don't know/not political, 7 as libertarian, and 9 as other.

In Study 1c, we recruited 378 participants from the Indiana undergraduate subject pool. 317 of these participants passed experimental manipulation checks and were included in all subsequent analyses. According to participant demographic responses, 188 were female, 125 male, and 4 participant selected other. There were 39 Asian participants, 25 Black or African American participants, 21 Hispanic or Latino participants, 213 White participants, 15 participants who selected More than one, and 4 participants who selected Unknown. Twenty participants identified as very liberal, 81 as liberal, 61 as slightly liberal, 61 as moderate/middle of the road, 33 as slightly conservative, 21 as conservative, 2 as very conservative, 30 as don't know/not political, 5 as libertarian, and 3 as other.

## **Design**

The studies in this manuscript utilize a set of actions initially reported in Celestin and Kruschke (2019a, 2019b). These actions are representative of the full range of plausible actions in the context of police-civilian encounters. The research of Celestin and Kruschke (2019a, 2019b) derived precise scale estimates for the moral and physical severities of the actions.

Studies 1a, 1b, and 1c were all constructed using the same set of actions used by Celestin and Kruschke (2019a, 2019b), but each study asked participants to respond to a different prompt. In the design, participants were presented with a decontextualized vignette:

While working a patrol shift, a police officer was dispatched to a reported crime, and in the course of the investigation, he came into contact with a civilian. During their subsequent interaction...

During the first block of all three studies, participants were presented with each civilian action at the end of the vignette and the full set of 22 officer actions. Participants were asked to choose an officer action according to the prompt. Civilian actions were highlighted in bold font to maximize ease of comprehension, and prompts were italicized. See Figure 19 for an example. Each civilian action was presented in a randomized order. Thus, from a participant's perspective, every screen displayed a complete vignette that included one civilian action and required the selection of one officer response. Four manipulation check questions were randomly inserted in order to test whether participants were responding thoughtfully. These check questions asked participants to select either the very first or very last action in order to calibrate the survey.

In studies 1a, and 1b, a subsequent block presented the same vignette but included an officer rather than a civilian action appended to each vignette. This subsequent block instructed participants to choose a civilian action that "makes the officer action an appropriate response." Although originally intended as an inverse framing of the first experimental block, highly variable responses indicated that this second block was challenging for participants to interpret. For this reason, the second block was removed from Study 1c and is not discussed further here (but see the supplementary material for analyses of the second block data).

## **Procedure**

Following informed consent and initial instructions, participants were presented with a screen formatted similarly to the experiment accompanied by instructions for resizing

While working a patrol shift, a police officer was dispatched to a reported crime, and in the course of the investigation, he came into contact with a civilian. During their subsequent interaction, **the civilian punched the officer in the face**. Please choose the *most appropriate officer response* (assuming he is able):

- stand near the civilian
- ask the civilian how he is doing
- wave at the civilian in greeting
- yell "get back!" at the civilian
- shout "get your fucking hands up!" at the civilian
- say "if you touch me, you're gonna get hurt" to the civilian
- call the civilian a "thug"
- put handcuffs on the civilian
- spray the civilian with pepper spray
- shock the civilian with a tazer
- twist the arm of the civilian, forcing him to the ground
- punch the civilian in the face
- kick the civilian in the stomach
- apply a chokehold to the civilian
- headbutt the civilian in the face
- strike the civilian on the leg with a collapsible baton
- shoot the civilian in the chest with a beanbag shotgun round
- strike the civilian in the stomach with a metal flashlight
- shoot the civilian in the chest with a handgun
- shoot the civilian in the head with a handgun
- slam the civilian's head into the concrete curb
- purposefully strike the civilian with a motor vehicle

Figure 19. Screenshot from Study 1a.

the screen. Participants were asked to affirm complete visibility of the content in order to ensure that the vignette and all of the actions were simultaneously visible. Participants were then provided with the following instructions:

*On the following screens, you will be presented with scenarios that represent real life interactions involving police officers and civilians. Officer and civilian names have been purposefully excluded. For each scenario, you will be presented with a civilian action and asked to select an officer action. Please be sure to read everything carefully as you go because the civilian actions will change each time. There are no right or wrong answers; we would like your*

*intuitive, thoughtful responses.*

For Study 1a, participants were then advised "For this set of scenarios, please choose the most appropriate officer response (assuming he is able)". For Study 1b, they were advised "Please choose the most severe but still appropriate and not excessive officer action (assuming he is able)". For Study 1c, they were advised "Please choose what you think an officer would do (assuming he is able, and regardless of what you think is the most appropriate officer response)". Thus, Study 1a was intended to measure participant perceptions of the most appropriate officer response, Study 1b the most extreme but still appropriate officer response, and Study 1c the action that participants believe an officer would actually choose.

Participants were then presented with each vignette and asked to respond to the associated prompt by choosing an action. If the action that a participant selected had a latent severity difference greater than 1 (i.e., 20% of the range of the severity scale) (Celestin & Kruschke, 2019a), a warning message was presented in red advising the following:

*Are you certain this is the action you want to choose? Please double check, and when you're sure, click "Submit Answer" again to confirm your choice.*

This warning message was only displayed one time, and participants were allowed to proceed without changing their selection if they chose to do so. The purpose of this warning flag were to encourage thoughtful responding and to discourage selecting actions randomly as a method to complete the experiment faster.

Following these experimental blocks, participants answered six questions assessing their felt obligation to obey police and their personal alignment with police values using a slider. They also answered binary questions about personal and vicarious negative experiences. Finally, participants completed a brief demographic questionnaire and were debriefed.

<b>SOLOF Level</b>	<b>Civilian Action (presented)</b>	<b>Officer Action (selected)</b>
1	stood near the officer	stand near the civilian
1	asked the officer how he was doing	ask the civilian how he is doing
1	waved at the officer in greeting	wave at the civilian in greeting
2	yelled "get back!" at the officer	yell "get back!" at the civilian
2	shouted "I don't have to fucking listen to you!" at the officer	shout "get your fucking hands up!" at the civilian
2	said "if you touch me, you're gonna get hurt" to the officer	say "if you touch me, you're gonna get hurt" to the civilian
2	called the officer a pig	call the civilian a thug
3	pulled away from the officer	put handcuffs on the civilian
3	sprayed the officer with pepper spray	spray the civilian with pepper spray
3	shocked the officer with a tazer	shock the civilian with a tazer
3	twisted the arm of the officer, forcing him to the ground	twist the arm of the civilian, forcing him to the ground
4	punched the officer in the face	punch the civilian in the face
4	kicked the officer in the stomach	kick the civilian in the stomach
4	applied a chokehold to the officer	apply a chokehold to the civilian
5	headbutted the officer in the face	headbutt the civilian in the face
5	struck the officer on the leg with a metal pipe	strike the civilian on the leg with a collapsible baton
5	hit the officer in the chest with a baseball bat	shoot the civilian in the chest with a beanbag shotgun round
5	struck the officer in the stomach with a metal flashlight	strike the civilian in the stomach with a metal flashlight
6	shot the officer in the chest with a handgun	shoot the civilian in the chest with a handgun
6	shot the officer in the head with a handgun	shoot the civilian in the head with a handgun
6	slammed the officer's head into the concrete curb	slam the civilian's head into the concrete curb
6	purposefully struck the officer with a motor vehicle	purposefully strike the civilian with a motor vehicle

Figure 20. Civilian and officer actions used in studies 1a, 1b, and 1c.

## Results

### Study 1a

Figure 21 plots the frequency with which participants selected each officer action (plotted on the vertical axis) relative to each civilian action that was presented (plotted on the horizontal axis). The total frequency with which each officer action was chosen across all civilian actions is plotted using squares in the right margin. In particular, the larger the

Please choose the most appropriate officer response (assuming he is able):

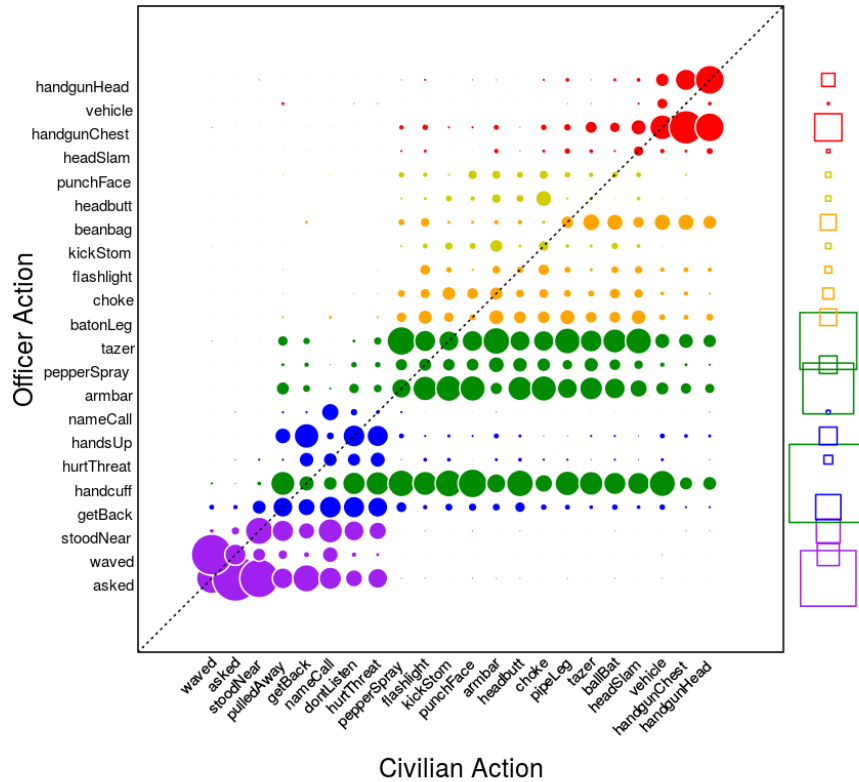


Figure 21. Frequency counts of participant action selection of the “most appropriate officer response” (Study 1a). Actions are color coded according to officer SOLOF level (see Figure 20). Marginal squares represent the total frequency with which each officer action was selected and are sized accordingly. All actions on both vertical and horizontal axes are spaced at equal intervals.

circle or square relative to the others, the more frequently participants selected that action. Both officer and civilian actions are plotted at equal intervals from one another, and they are plotted in ascending order of their latent severity estimates (Celestin & Kruschke, 2019a). A diagonal dashed line is also visible as a reference for proportionality between officer and civilian action severity.

One pattern that is visible in the marginal squares is that participants chose lower severity actions with much greater frequency than higher severity actions. Moreover, several of-

Officer actions were commonly chosen across a wide range of civilian actions; these include handcuffing, armbar, and tazer (see Figure 20 for full action descriptions). These actions correspond to the most normative actions measured in Study 2 (see Figures 27 and 28). Conversely, some actions such as nameCall, headSlam, and punchFace were selected very infrequently. These actions correspond to non-normative actions measured in Study 2.

Overall, the pattern of responses indicates a large portion of the actions selected fall below the line of proportionality. This suggests that participants are most frequently selecting officer actions that are less severe than civilian actions as the "most appropriate" response.

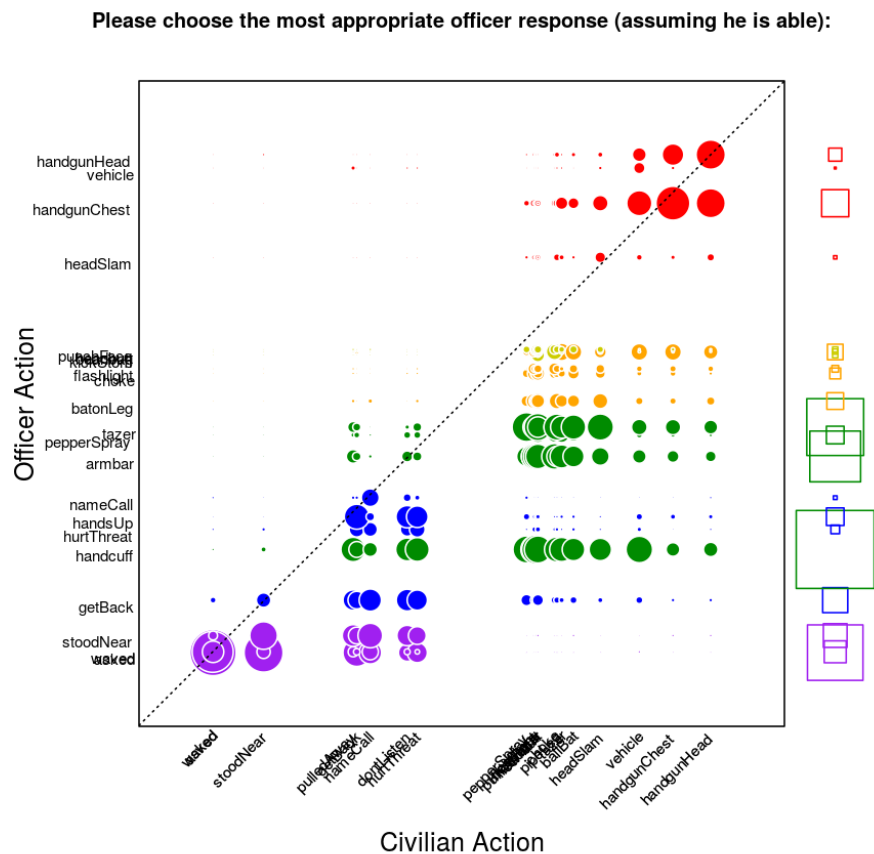


Figure 22. The information presented here is the same as in Figure 21 (Study 1a). However, the actions on the vertical and horizontal axes are plotted at their latent moral severities for officer and civilian, respectively (Celestin & Kruschke, 2019a).

Figure 21 is useful for displaying the data for all possible action pairs while minimizing visible overlap, but it is somewhat misleading due to the equal spacing of the actions on the axes. Placing actions at equal intervals is problematic because our prior research demonstrates that the underlying moral and physical severities of this set of actions is quite different from a standard force "continuum" that might organize them in that manner (i.e., equidistantly). Thus, Figure 22 plots the same information as Figure 21 but with the officer and civilian actions shifted to their respective latent moral severity estimates on the vertical and horizontal axes (Celestin & Kruschke, 2019a).

In Figure 22, it can be seen that officer actions at extreme levels (low and high) are selected in proportion to civilian severity to some degree; but for all of the verbal and non-lethal actions, participants are very clearly selecting officer actions that are less severe than the civilian actions they are confronting.

### **Study 1b**

The prompt for Study 1a solicits general judgments of the most appropriate action. To the extent that action appropriateness is normally distributed with excessive actions in the high tail and insufficient actions in the low tail, the "most appropriate" prompt is likely extracting action choices from the peak of the distribution. However, it is possible that some actions on the excessive side of the distribution would still be evaluated as appropriate as a holistic lay judgment, even if they are not the "most appropriate" choice. That is, it could be the case that more extreme officer actions that are largely proportional to civilian severity might still be deemed appropriate. For this reason, the prompt for Study 1b asked participants to select the "most severe but still appropriate and not excessive" officer action. This prompt is more specific than the prompt in Study 1a and was intended to measure the most extreme officer action that was still morally justified relative to each civilian action in the minds of participants.

Please choose the most severe but still appropriate and not excessive officer action (assuming he is able):

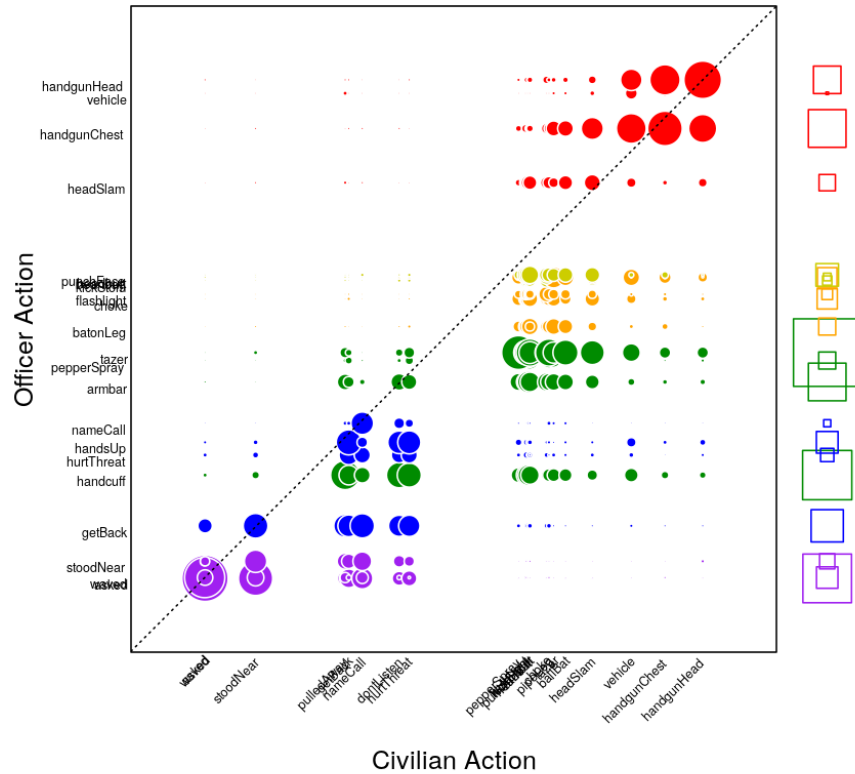


Figure 23. Frequency counts of participant action selection of the “most severe but still appropriate and not excessive” officer action (Study 1b). Actions on the vertical and horizontal axes are plotted at their latent moral severities for officer and civilian, respectively (Celestin & Kruschke, 2019a).

As in Figure 22, Figure 23 plots the frequency of each officer action selected by participants at their relative latent action severities on the vertical and horizontal axes. Visual inspection reveals that the individual and marginal frequencies have shifted up vertically, indicating that participants did tend to choose more severe actions than in Study 1a. Lethal actions were chosen more frequently and are more proportionately distributed around the diagonal line. However, despite the tendency to choose more severe officer actions, the bulk of non-lethal and verbal actions still fall below the diagonal line. This suggests that even the most extreme officer actions that are still perceived as morally acceptable are less

severe than the civilian force confronting the officer.

### **Study 1c**

Studies 1a and 1b assessed participant choices for what they believed an officer should do. But these prescriptive prompts raised the question of whether what participants thought an officer should do differs from what they believe an officer would actually do when confronted with a specific civilian action. To assess what participants thought an officer would actually do, Study 1c prompted participants to "choose what you think an officer would do (assuming he is able, and regardless of what you think is the most appropriate officer response)".

As before, Figure 24 plots the frequency counts of the officer actions participants selected, and the vertical and horizontal axes are plotted at their respective latent severity estimates (Celestin & Kruschke, 2019a). Visual inspection of Figure 24 indicates that it looks much more like Figure 23 than Figure 22. That is, lethal officer actions are relatively proportionate, and the bulk of the actions are shifted up slightly indicating participants are choosing somewhat more severe actions. However, as in both Figure 22 and Figure 23, the bulk of non-lethal and verbal actions fall below the diagonal line. One possible explanation is that participants are generally expecting officers to use the most severe level of force that is legally justified (as they interpreted Study 1b's prompt). Another possibility is that the "would do" question is weird or otherwise hard for participants to think about, but this explanation does not seem likely as the prompt is rationally straightforward and participant responses were consistent and reasonable (i.e., did not exhibit signs of random responding).

Thus, we interpret these results as indicating that participant's prescriptive judgments about what police should do when confronted with civilian resistance are quite similar to their descriptive judgments of what they believe officers would actually do. That is, it seems that participants generally believe that police officers will choose forceful actions that are

Please choose what you think an officer would do  
(assuming he is able, and regardless of what you think is  
the most appropriate officer response):

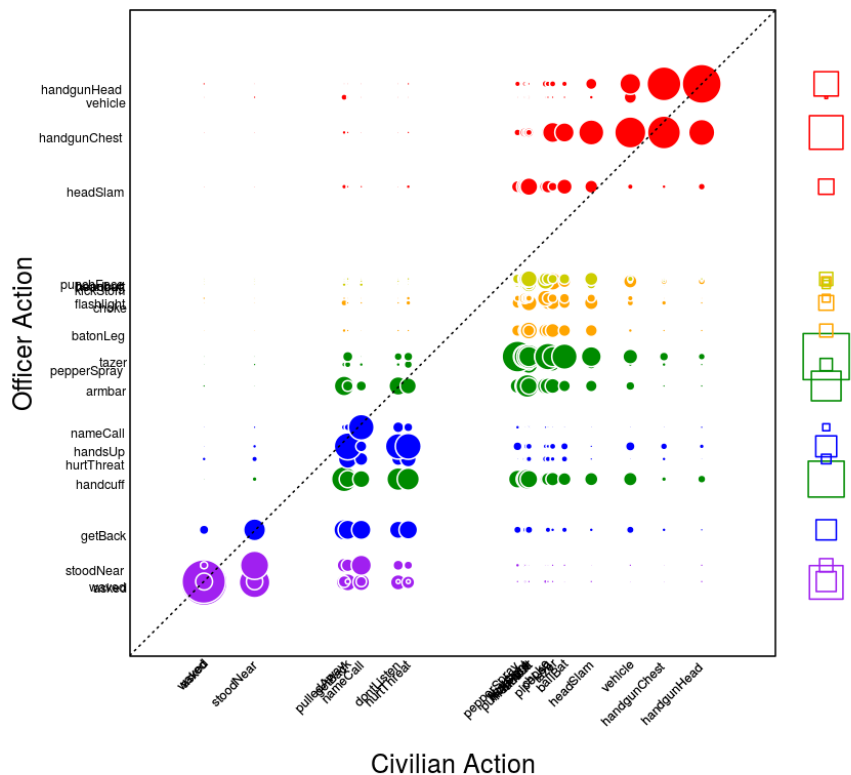


Figure 24. Frequency counts of participant action selection for what an officer “would do (assuming he is able, and regardless of what you think is the most appropriate officer response)”. Actions on the vertical and horizontal axes are plotted at their latent moral severities for officer and civilian, respectively (Celestin & Kruschke, 2019a).

slightly more severe than those they view as most appropriate, but still within participant’s range of acceptable actions which tend to be less severe than civilian resistance.

## **Extracting action normativities from inferential models**

Further analyses of the data from Studies 1a, 1b, and 1c are planned. In particular, an inferential model using a skew-normal noise distribution offers the potential to extract the role-normativity of the actions for police. The model, in progress, describes the chosen officer action as a linear or quadratic trend of the civilian action severity, weighted by the normativity of the officer action. These normativities are estimated in the model and should correlate with the normativities estimated in (Celestin & Kruschke, 2019a).

### **Method: Study 2**

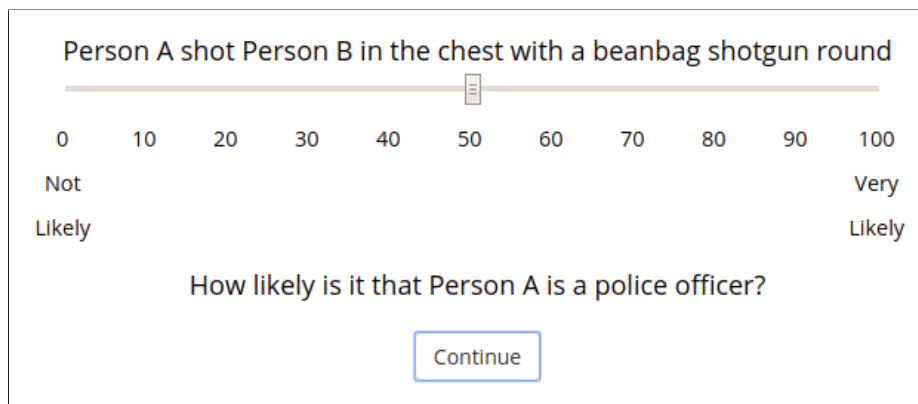
Study 2 was designed to directly assay the normative strength of association between all 27 actions initially reported in Celestin and Kruschke (2019a). It was comprised of two components that were both administered to every subject, resulting in a fully crossed within-subjects design.

### **Participants**

For Study 2, we recruited 426 participants from the Indiana undergraduate subject pool. 399 of these participants passed experimental manipulation checks and were included in all subsequent analyses. According to participant demographic responses, 267 were female, 130 male, and 2 participant selected other. There were 3 American Indian/Alaska Native participant, 55 Asian participants, 26 Black or African American participants, 18 Hispanic or Latino participants, 280 White participants, and 17 participants who selected More than one. Forty participants identified as very liberal, 90 as liberal, 51 as slightly liberal, 69 as moderate/middle of the road, 38 as slightly conservative, 32 as conservative, 9 as very conservative, 63 as don't know/not political, 2 as libertarian, and 5 as other.

## Design

The experimental paradigm consisted of two blocks: An explicit ratings block and an implicit reaction time block. In the first block (the explicit task), participants were presented with each of the 27 actions described in Figure 20 (i.e., 17 actions that are exactly the same for the officer and civilian, plus 5 unique officer actions and 5 unique civilian actions of matching severity) in a random order. Rather than explicitly describing who was performing the action as we have in prior paradigms, we used generic roles of "Person A" and "Person B" to describe the actors. Actions were presented one at a time (i.e., one per screen) and participants were asked to respond to the prompt "How likely is it that Person A is a police officer?". Person A always represented the actor, while Person B represented the person being acted upon. Participants responded using a sliding scale ranging from 0 to 100 and labeled in increments of 10, with the 0 endpoint also anchored as "Not Likely" and the 100 endpoint also anchored as "Very Likely". See Figure 25 for a screenshot of the first experimental block. Additionally, three manipulation check questions were randomly



The screenshot shows a rating scale interface. At the top, the text reads "Person A shot Person B in the chest with a beanbag shotgun round". Below this is a horizontal slider with a vertical bar in the center, positioned at the 50 mark. The scale is labeled with numbers from 0 to 100 in increments of 10. Underneath the numbers, "Not Likely" is written under 0 and "Very Likely" is written under 100. Below the scale, the question "How likely is it that Person A is a police officer?" is displayed. At the bottom center, there is a blue-bordered button labeled "Continue".

Figure 25. Screenshot from the first block of Study 2.

inserted to allow us to gauge thoughtless responding post-hoc. These check questions advised participants to set the slider to one extreme or the other, with the specific requested endpoint randomly selected across trials.

In the second block (the implicit task), participants were asked to familiarize themselves with a list of 12 possible and 12 impossible actions, and then rapidly categorize them as possible or impossible for either a police officer or criminal. The 12 possible actions were culled from the larger list in Figure 20. Specifically, we chose actions that were all physical and also exhibited highly disproportionate or proportionate severity (Celestin & Kruschke, 2019a). Selecting a subset of actions was necessary in order to keep the total number of trials manageable, and our intention was to test the most proportionate relative to the most disproportionate actions in our original set. Whether participants first categorized actions for an officer or criminal was randomly assigned across participants, but all actions were categorized for one role at a time. The same 12 possible actions were presented to every participant. However, we generated a larger list of 36 impossible actions and randomly selected 12 of them to match our 12 possible actions for each participant. The primary reason for sampling from this larger set of impossible actions was to ensure that any idiosyncratic impossible action would not unduly influence our results. Knowing that our primary dependent measure for this implicit task was reaction time, we took great care to ensure that the character length and number of syllables across our lists of possible and impossible actions were matched as closely as was feasible. Table 5 presents all of the possible and impossible actions.

During the task itself, participants were presented with the prompt "Is this something that is possible for [a criminal | a police officer] to do? (Please press "P" or "I")", accompanied by the action in bold font. For each trial, the action was preceded by a 400 millisecond prime that displayed the role (i.e., either **police** or **criminal**) which was replaced with the action when the prime duration expired. A 500 millisecond inter-trial interval was also used between each trial. Instructions at the bottom of the screen advised that a red X would appear if the wrong response was selected and that the correct key must be pressed to proceed. A screenshot of the experiment is presented in Figure 26.

Table 5

*Table of possible and impossible actions.*

Possible Actions	Impossible Actions
yell "get back!"	rip arms off
call thug	grow claws
apply handcuffs	become vampire
use a tazer	cast spell
twist someone's arm	use "the force"
choke hold	fly away
headbutt	walk on water
baton strike leg	summon zombie
hit with bat	control weather
shoot in chest	become invisible
slam head on curb	use telepathy
hit with car	become giant
–	channel lightning
–	become immortal
–	regrow a limb
–	walk through wall
–	breathe fire
–	teleport
–	travel back in time
–	become liquid
–	dimension travel
–	become dragon
–	lightsaber
–	catch bullet
–	punch in half
–	raise dead
–	frozen touch
–	destroy soul
–	read minds
–	become younger
–	become stone
–	x-ray vision
–	summon demon
–	bulletproof skin
–	laser eye rays
–	grow wings

All 12 possible actions were presented to every participant, but 12 impossible actions were randomly selected from the full list of 36 for each participant.

## Procedure

Following informed consent, participants were provided with the following instructions:

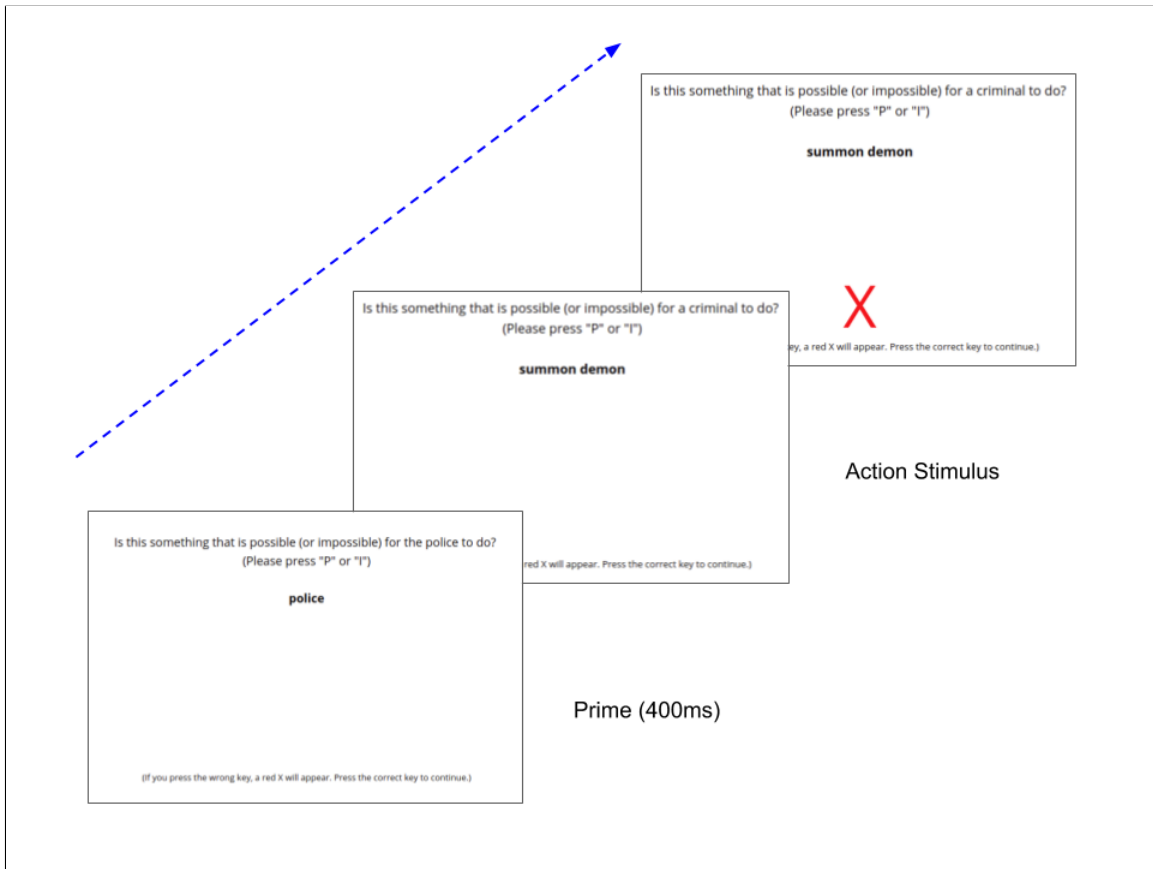


Figure 26. Screenshot of the sequence of an impossible action trial for the implicit task of Study 2. The final panel in the sequence represents the screen presented to participants if they select the incorrect response.

*On the following screens, you will be presented with different actions that are described as being performed by Person A on Person B. For each action, you will be asked to rate how likely it is that Person A is a police officer from 0 = 'Not Likely' to 100 = 'Very Likely'. There are no right or wrong answers, but please respond thoughtfully.*

Participants were then presented with the following instructions:

*In a moment, you will be asked to rapidly categorize actions according to whether or not it is possible for someone in a given role to perform them.*

*Some of the actions are possible (e.g., headbutting someone), while others are impossible (e.g., breathing fire)*

They were then presented with a list of possible and impossible actions and asked to familiarize themselves with the actions. These instructions were intended to mitigate participant confusion and ensure clarity regarding which actions are possible and which are impossible (though we chose the actions to maximize their apparent impossibility). Participants were specifically instructed, "For each action, you will be asked whether it is something that [a criminal police officer] could do (i.e., whether it is possible or impossible)." They were instructed to press the "P" key with their left index finger for possible actions and the "I" key with their right index finger for impossible actions. Finally, we reinforced the importance of responding thoughtfully, but also as quickly and accurately as possible. Participants were then presented with the block of trials for the officer and criminal in a random order. Each possible and impossible action was presented to participants four times. Thus, the total number of action trials for police primes/prompts was 96, and for criminal primes/prompts was also 96, for a total of 192 trials. In order to minimize participant fatigue, breaks were inserted every 48 trials. During these breaks, participants were instructed to take a moment to rest and, when ready, place their index fingers on the "P" and "I" keys and press the spacebar to continue. Responding as quickly as possible was also re-emphasized at the conclusion of the breaks. At the conclusion of the experimental blocks, participants responded to the same measures of legitimacy beliefs and demographics as in Studies 1a, 1b, and 1c, and debriefed.

## Results

### Explicit Task

We modeled the likelihood ratings of Person A being a police officer by estimating a mean and standard deviation of rating for each action using Bayesian estimation. Ratings were described as a function of each action, allowing every action to have its own variability parameter. Because responses for some of the actions piled at the extremes of the scale, indicating that participants would have responded beyond the scale bounds if allowed, we treated extreme values greater than 5 and 95 as censored data. Formally, the model states

$$y \sim dnorm(\mu_a, \sigma_a) \quad (5)$$

where  $y$  represents a rating, and  $a$  indexes each action. The  $\mu$  parameter is an estimated coefficient for the corresponding action, and the  $\sigma$  parameter an estimated coefficient for the noise distribution around each  $\beta$ . There was no hierarchical structure imposed on the  $\mu_a$  or  $\sigma_a$  parameters. Priors for all estimated parameters were diffuse relative to the scale of the data, so they had negligible influence on the resulting posterior estimates.

Figure 27 displays the modal estimates and highest density intervals (HDIs) for each action. Actions are ordered on the horizontal axis from least to greatest, and the vertical axis represents the likelihood that Person A is a police officer. Actions are also color-coded according to their categorization a priori into a six-level force continuum (Celestin & Kruschke, 2019a).

Visual inspection of Figure 27 reveals that actions across all continuum categories are distributed over the likelihood range. That is, it is not the case that more severe actions are perceived systematically as more likely to be performed by a police officer, or vice versa for less severe actions. Additionally, there is wide variability in the degree to which

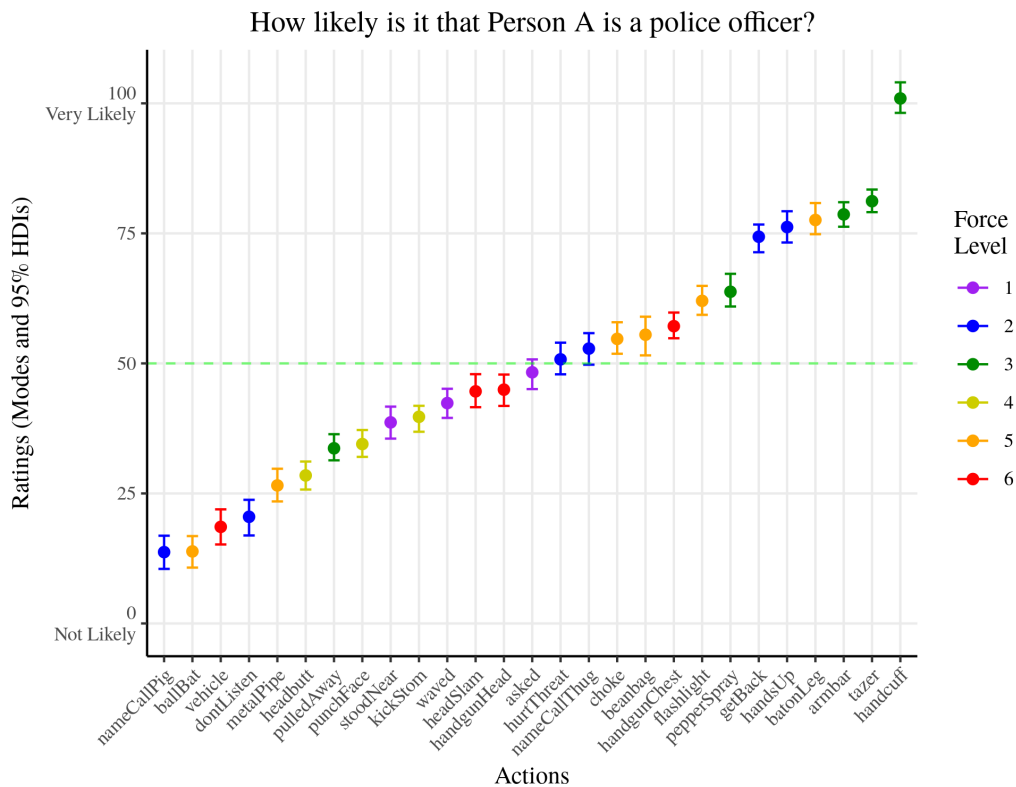


Figure 27. Mode and HDI estimates for each action in the explicit task. Actions are ordered by mode, and color-coded according to their a priori force continuum category (Celestin & Kruschke, 2019a).

a given action is likely to be performed by a police officer. For some actions, such as calling someone a “pig”, hitting someone with a baseball bat, or striking someone with a vehicle, participants rated the likelihood that the actor is a police officer as very low. For other actions, such as twisting someone’s arm, tazing someone, or handcuffing someone, participants rated the likelihood that the actor is a police officer as very high. Actions of middling likelihood reflect those that are perceived to be roughly equally likely to involve a police actor or an actor who is not a police officer. As we anticipated, the actions at the extreme levels of likelihood correspond to those that exhibit non-monotonicities of moral relative to physical severity in our prior research (Celestin & Kruschke, 2019a).

In order to directly compare the ratings of likelihood that Person A is a police officer

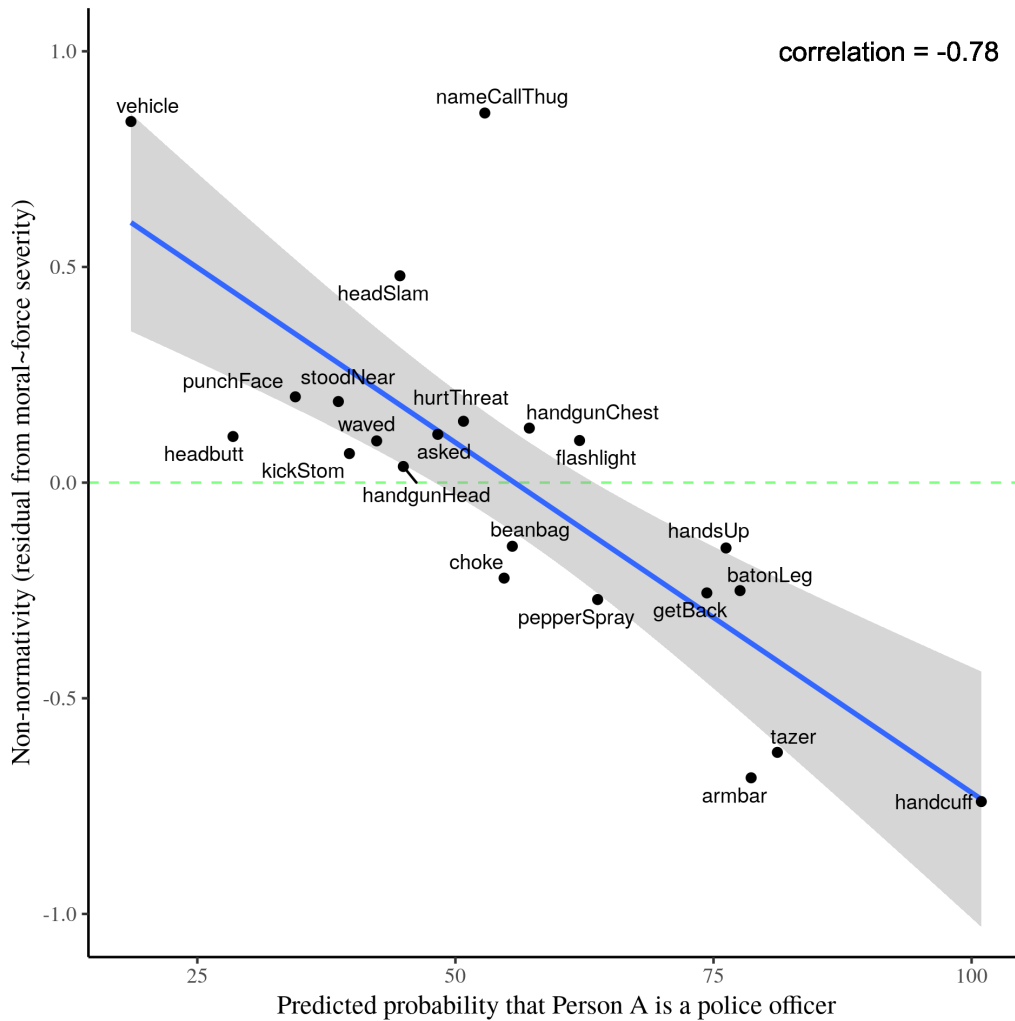


Figure 28

with the non-monotonicities observed in the moral and physical severity scales we derived previously, we first fit a least-squares linear regression model predicting the moral severity estimates as a function of the force severity estimates. This approach allowed us to quantify the non-monotonicities using a linear model that best fit all of the data points (i.e., all of the action scale estimates) together. From this model, we extracted the residuals which represent the non-normativity of each action. That is, positive residual values indicate actions with scale values that are more morally than physically severe, while negative residual values reflect actions with scale values that are more physically than morally severe. Residual

values that are close to 0 indicate actions whose moral and physical severity scale values are approximately proportionate.

Figure 28 plots these residual values on the vertical axis, and the predicted probability that Person A is a police officer on the horizontal axis. The correlation between residuals and predicted probability is  $-0.78$ , indicating that as actions become more likely to have been performed by a police officer, the residuals move from positive to negative. In other words, actions that are highly non-normative (i.e., have relatively large positive residuals) are also rated as very unlikely to have been performed by a police officer, while actions that are highly normative (i.e., have relatively large negative residuals) are rated as quite likely to have been performed by a police officer.

### **Implicit Task**

To analyze the reaction time (i.e., implicit task) data, we first extracted the response trials, excluding the primes and inter-trial intervals. Subsequently, we removed all trials that had reaction times greater than 2500 milliseconds or less than 200 milliseconds, as these trials were very uncommon and indicative of erroneous responses. Commonly, reaction time data exhibit skewed distributions, and our data were no exception. For this reason, we log transformed all response times jointly which yielded a distribution that was approximately normal. The primary actions of interest for the purposes of the current analysis were the possible actions because we were interested in evaluating reaction time differences between those actions for the criminal relative to the police prime. For this reason, we excluded the impossible actions and analyzed the possible actions in isolation.

We modeled these cleaned and transformed data using Bayesian regression. Formally, the model was specified as

$$\hat{y} = \beta_0 + \beta_a + \beta_r + \beta_{a \times r} + \sigma_a \quad (6)$$

where  $\hat{y}$  represents the predicted value of a response time. The  $\beta_0$  regression coefficient represents the baseline response time, the  $\beta_a$  coefficients represent the response time deflections for each action from baseline across roles, the  $\beta_r$  coefficient represents the response time deflection from one role to the other across actions, and the  $\beta_{a \times r}$  coefficients represent the response time deflections for one role and each action relative to the corresponding action for the other role (i.e., the interaction of action and role). As indicated by the subscript, the model also allowed a separate  $\sigma_a$  for each action. In essence, this model estimates a mean for each action  $\times$  role combination.

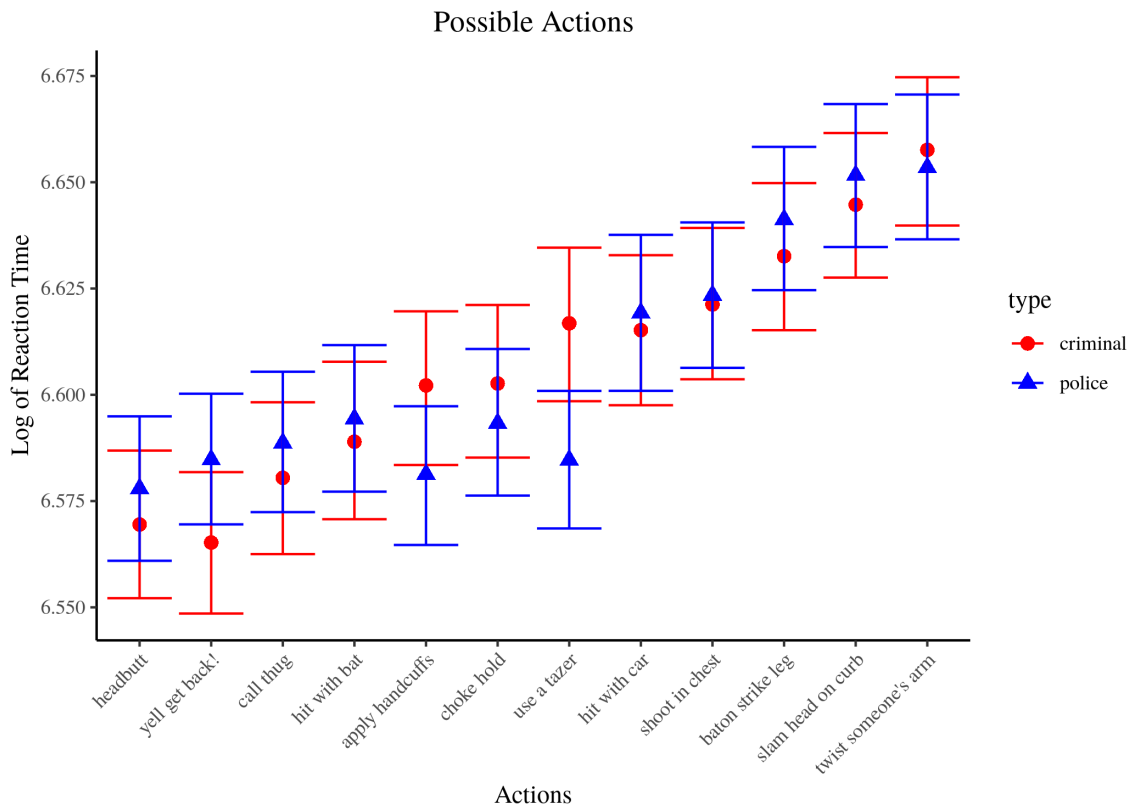


Figure 29. Log reaction times for the 12 possible actions.

Figure 29 plots the reaction time model predictions and 95% highest density intervals (HDIs) from equation 6. Actions are plotted on the horizontal axis, and log reaction time is plotted on the vertical axis. Criminal and police roles are differentiated by point (circle for

criminal, triangle for police) and color (red for criminal, blue for police). Examining the pattern displayed reveals that some actions (e.g., twist someone’s arm) did exhibit slower reaction times than others (e.g., headbutt) collapsing across police and criminal prime conditions. Moreover, a few of the actions also exhibit differences in the predicted direction for criminals relative to police. In particular, “apply handcuffs” and “use a tazer” demonstrate faster reaction times for police relative to criminals indicating that they are more strongly associated with police. Overall, however, the observed differences between police and criminal across all actions are negligible which is reflected in very small or zero estimates for the  $\beta_a$ ,  $\beta_r$ , and  $\beta_{a \times r}$  coefficients in the model. To follow up on these results, we ran frequentist models that correspond to equation 6 using the raw response times as the predicted values and observed that the average difference in response time between different role/action combinations was approximately 26 milliseconds, while the average standard deviation across those combinations was approximately 300 milliseconds. Thus, the reaction time data we have appear to be much too noisy to adequately measure the small effects between role/action pairs that we attempted at the level of individual actions.

However, in order to assess the relationship of the overall pattern of the possible reaction time differences with the moral and physical severity discrepancies observed in Celestin and Kruschke (2019a), we conducted an additional analysis that is depicted in Figure 30. Specifically, we calculated the estimated reaction time difference for criminal relative to police prime for each possible action and then calculated the correlation between those difference values and the moral/physical residual discrepancies. As shown in Figure 30, there was a negative correlation of  $-0.56$  indicating that actions exhibiting slower reaction times for criminal relative to police primes (i.e., those that demonstrate stronger association with police) tended to have greater latent physical than moral severity. Conversely, actions that had slower reaction times for police relative to criminal primes tended to have greater latent moral relative to physical severity. Visual inspection of Fig-

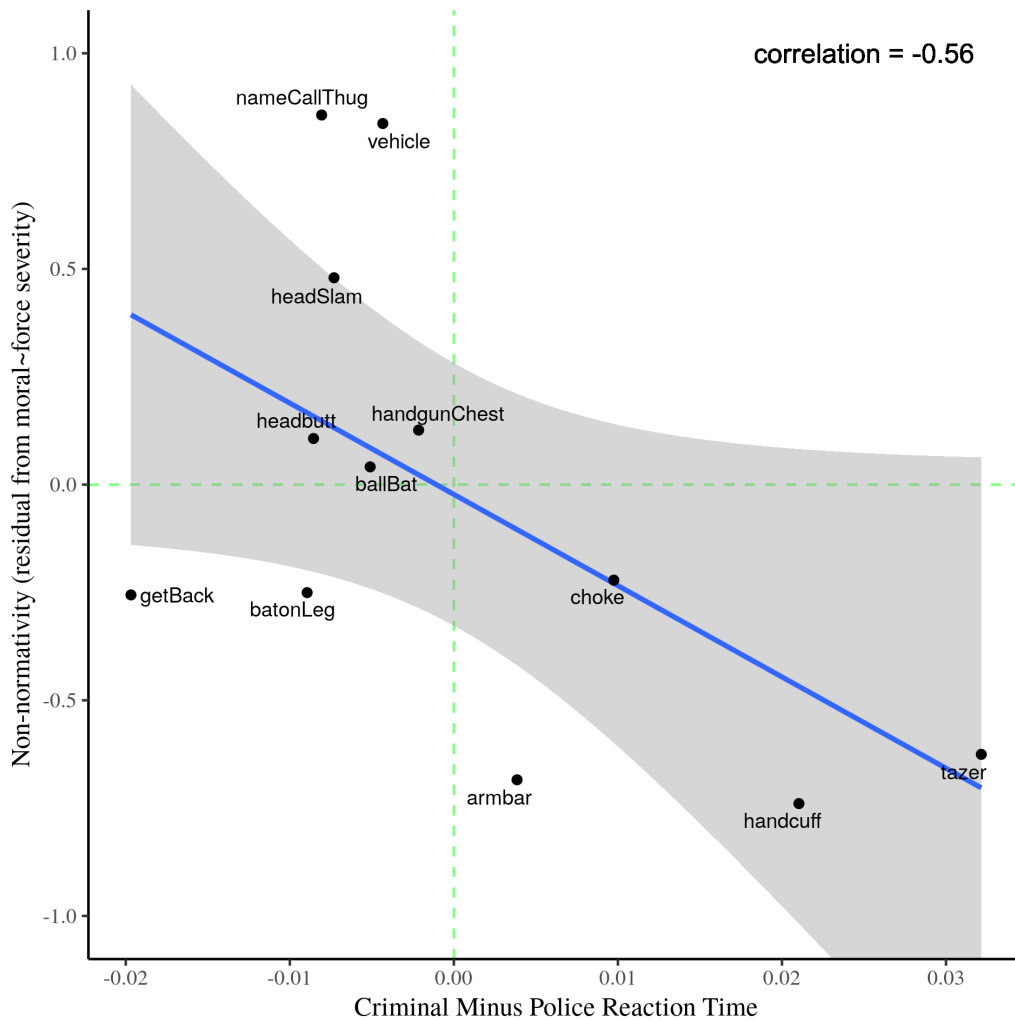


Figure 30. The reaction time difference for criminal prime minus police prime is plotted on the horizontal axis, while the residuals from a linear model of moral severity as a function of physical severity (Celestin & Kruschke, 2019a) are plotted on the vertical axis.

Figure 30 indicates that, although this correlation is of reasonable magnitude, it is driven by only a few data points (e.g., tazer). Moreover, some of the actions that were anticipated to show a stronger police association have slightly slower reaction times for police than criminal primes (e.g., handgunChest and batonLeg). Taking these observations into account means that, although this correlation does offer support for the overall relationship between normative association and moral severity perceptions, these reaction time data are very noisy, and so conclusions must be tentative. Thus, the current reaction time data can

only contribute modest additional information about the strength of association between our possible actions and police relative to criminals. Future iterations of this task should keep the limitations of our reaction time data in mind when modifying our task or designing a new task to measure these associations.

## **Discussion**

The current studies directly assessed lay perceptions of appropriate police responses to civilian resistance in the context of previously derived severity scale estimates. They also measured the extent to which specific actions are more or less uniquely associated with police to explain discrepancies in perceived moral relative to physical severity.

Study 1a demonstrated that lay evaluators generally tend to select officer actions that are less severe than the civilian action with which they are confronted. Study 1b focused participant responses on the most extreme but still acceptable officer responses, and confirmed that even the most extreme allowable actions still tend to be less severe than civilian resistance. These results stand in stark contrast to the general heuristic rule that police agencies use allowing them to use force that is greater than (i.e., one level above) civilian action severity (Garner et al., 1995; Terrill & Paoline, 2013). Like most aspects of use-of-force policy, the heuristic rules that law enforcement use were not chosen frivolously, and there are good reasons to be hesitant to alter them to match citizen expectations. The supreme court in United States treat the legal evaluation of use of force as something that must be done by someone with expertise (i.e., by an officer with similar training and experience, without the benefit of hindsight (*Tennessee v. Garner*, 471 U.S. 1, 1985; *Graham v. Connor*, 490 U.S. 3, 1989)), and it is evident that police require sufficient discretion to make choices about use of force that maximize personal and bystander safety and also the probability of subject compliance. Nevertheless, it is clear from our data that police and civilians have quite different expectations and are, in a sense, speaking different languages

about what constitutes appropriate use of force.

Rather than assessing lay evaluations of what officers should do, Study 1c measured what participants thought officers would actually do when confronted with different levels of civilian force. These results appear very similar to the most extreme actions that are still morally acceptable measured in Study 1b, but they still indicate participants are expecting that officers would actually select less severe actions than civilians. This finding may reflect the extent to which civilian and police expectations are descriptively as well as prescriptively disconnected. That is, it does not seem to be the case the civilian evaluators understand that police are trained to use more force than they are confronting but disagree with that training (i.e., a largely moral disagreement). Rather, lay evaluators believe that the most appropriate police responses to civilian resistance are less severe than said resistance, and that police would actually select less severe actions in real life, which is a more fundamentally knowledge-based misunderstanding. This is a profound discrepancy, but it also suggests a hopeful counterpoint: if the public is largely ill informed about police training and the reasons underlying the structure of force continua and accompanying heuristic rules, it seems likely that educational outreach efforts aimed at broadening citizen knowledge about these issues and the rationale underlying law enforcement policies holds promise for uniting police and civilian perspectives. Moral disagreements can be particularly intractable, so it could be especially challenging (though not impossible) to facilitate alignment of police and citizen evaluations of use of force if citizens believe police should act one way but would actually act in another. Given that our participants underestimate the severity of actual police responses relative to policy recommendations, it seems possible that mere exposure to those policy recommendations could ameliorate some of the conflict in the current national discourse about police use of force.

Study 2 assessed the normativity of actions in relation to the strength of their association with police. In particular, the explicit task presented participants with each action

and inquired about the probability that the actor was a police officer, while the subsequent implicit task assessed reaction time differences for categorizing actions following primes of police or criminal groups. The explicit task revealed clear differences between the extent to which particular actions are associated with police. Some actions were very unlikely to indicate that the actor was a police officer, while others were highly likely to mean that the actor was a police officer. Moreover, the correlation between the likelihood that an action was a strong signal for police and the discrepancy between moral and physical action severities previously measured (Celestin & Kruschke, 2019a) was strongly negative. Additionally, although the reaction time measure of implicit associations was too noisy to yield useful data at the level of individual actions, the overall correlation of reaction time differences and moral/physical action severities supported the general pattern observed in the explicit ratings. Taken together, these results provide direct evidence that some actions are strongly and, perhaps, uniquely associated with police, and that those actions alter moral judgments of severity relative to the perceived physical magnitude of force. Conversely, the results also indicate that actions that are not associated with police are more likely to yield moral judgments that are especially severe. Because our paradigms are decontextualized to mimic the structure of official policies, these effects are occurring at the level of the actions themselves, independently of other details that might systematically alter judgments. Thus, these normativity effects are an important additional variable to consider by stakeholders working at all levels of criminal justice and criminal justice reform.

Because police can only function effectively when they are broadly supported by the public (Friedman, 2017), these findings suggest an important and potentially overlooked consideration that agencies and academics could focus on to develop educational materials and methods and policy recommendations. Namely, it is worth paying attention to the perceived normativity of actions, and even choosing maximally normative force options if all other outcomes are likely to be equal.

In sum, the studies we report indicate that public evaluations of police use of force are disconnected from the decision-making tools and heuristics that law enforcement in the United States routinely use to make decisions on the streets. Citizens are expecting police to act in ways that highly role-normative using actions that are generally less severe than those they are confronting. Meanwhile, it is common knowledge among police that appropriate, legally justified police actions can and often should be of higher severity than civilian resistance in order to maximize safety and quickly achieve compliance. Police are also instructed (at least indirectly) that action legitimacy does not depend on normativity—especially in the case of lethal force(*Tennessee v. Garner*, 471 U.S. 1, 1985; *Graham v. Connor*, 490 U.S. 3, 1989). The first step toward ameliorating police-civilian conflict regarding the morality of specific instances of use of force is precisely identifying where conflict exists, and so we are hopeful that the findings we present here can initiate useful convergence of police and civilian perspectives that will mutually benefit both groups.

## Chapter 5: Conclusions

The studies presented in this dissertation measured lay perceptions of police and civilian use of force using novel paradigms that allow direct comparison to the force continuum structure and "one above" heuristic rule that are commonly used by law enforcement to make decisions about appropriate force. Globally, these results provide evidence that citizen perceptions are substantively different from law enforcement evaluation strategies and tools. Lay evaluators do not seem to meaningfully distinguish the severity of actions that are non-lethal but physically forceful, unlike the nuanced structure and fine gradations that police impose on these actions. Conversely, lay evaluators seem to view verbal and lethal actions with considerably more nuance than police, who bundle these actions into one respective force category (i.e., one for verbal and one for lethal), and who evaluate the justifiability of all lethal actions in a binary (i.e., justified or not justified) manner.

Lay evaluators also tend to rate police and civilian actions of increasing severity as more forceful and excessive, as well as less acceptable and more deserving of punishment, but there are marked differences in the judgments of individuals who view the police as highly legitimate authorities compared to those who do not. High-legitimacy individuals view more severe police actions as somewhat justified in response to sufficiently severe civilian actions, but low-legitimacy individuals appear reluctant to judge severe police actions as justified, even when civilian actions are also severe. Moreover, low-legitimacy individuals tend to perceive lethal civilian actions as somewhat acceptable and much less deserving of punishment when police actions are also severe, while high-legitimacy individuals never endorse severe civilian actions.

When asked to select appropriate police responses to civilian violence, lay evaluators overwhelmingly choose actions that are less severe than those confronting officers, even when they are asked to choose the most extreme action that would still be justified. Further-

more, when selecting actions that they believe police would actually choose (irrespective of what seems appropriate), lay evaluators still select actions that are less severe than civilian resistance. In ambiguous scenarios, the strength of association between specific actions and police varies considerably, demonstrating that some actions are highly role-normative for police, while others are highly non-normative. The non-normative actions are strongly correlated with increased moral severity perceptions, while normative actions are associated with decreased moral severity.

Taken together these results demonstrate that lay observers and police officers are routinely speaking different languages when they evaluate instances of police and civilian use of force. It is true that, in some ways, these results demonstrate alignment between lay intuitions and police policy; like force continua, lay moral judgments of police actions depend on the severity of civilian actions, and vice versa. And so, at a very coarse level, there is overlap between lay and professional assessments that bolsters justification for the broad organization of actions along some continuum of severity. However, police commonly parcel non-lethal force into three increasingly severe categories that are practically invisible to a lay audience, and the expectations of lay people for police to use less force than civilian resistance stand in stark contrast to the "one-level-above" heuristic that guides police choices. These differences help to explain why people reading the same news stories or watching the same videos of violent interactions between police and civilians often come to starkly different conclusions than police or even the decisions that are ultimately handed down by the courts.

In the limit, moral evaluations of police and civilian violence are just a special case of moral evaluations generally. That is, the same psychological processes are operating when human beings evaluate forceful actions in the context of a police-civilian encounter that are operating when they evaluate any other morally charged situation—especially those in which there are potentially aversive tradeoffs that must be resolved in the moment but

are psychologically problematic to think about (Tetlock, Kristel, Elson, Green, & Lerner, 2000; McGraw & Tetlock, 2005). In some ways, police use of physical force is analogous to surgery; both require incurring some level of localized physical damage but with the end goal of mitigating greater global harm, and both must balance high-stakes risks relative to rewards under considerable time and resource constraints. In addition, it is unclear whether lay individuals observing two different surgeries—one in which surgeons used current best-practices and one in which they were negligent and potentially guilty of malpractice—could reliably distinguish between them, perhaps, in part, because the visceral experience of observing any surgery tends to evoke unpleasant and unsettling feelings for many people. Likewise, the application of justifiable versus criminal violence by police may be challenging for lay observers to disintegrate for similar reasons, and the selection and training processes that separate police and surgeons from lay audiences may represent a kind of fundamental difference in constitution and/or expertise that cannot be easily overcome. On the other hand, policing in the U.S. has a tragic history of enforcing the will of the majority against the minority in ways that are now rightly judged as unfair and cruel (Leovy, 2015), and police agencies must grapple with this history and with the constant need to continue to progress toward the goal of impartiality in the application of criminal justice. In-group dynamics and self-interest affect police officers and agencies just as much as any other group, and police (as well as surgeons) are rightly held to a very high standard given the awesome powers that we entrust to them. For these reasons, law enforcement officers and agencies must be especially vigilant about using violence only when and to the degree that it is necessary, and they must do so in a way that reduces the influences of bias as much as possible and takes lay perceptions into adequate account. All of this is to say that morally evaluating police and civilian use of force is a highly complex endeavor that is certain to be riddled with high-stakes tradeoffs and intrinsic psychological biases, and it will likely remain so for the foreseeable future. Lay and professional evaluations

of force may never fully converge, but the results presented in this dissertation shed new light on the specific ways that lay perceptions diverge from official models of force, and they offer insight into mechanisms that might be leveraged to facilitate better alignment of police and civilian perspectives about what constitutes legitimate use of force. Ultimately, police can only operate effectively when they are underwritten by broad societal support. We are hopeful and cautiously confident that the current findings can help unify divergent perspectives and contribute to more strongly and widely supported policing, and thus to a more peaceful and just society.

## References

- Alpert, G. P., & Dunham, R. G. (2004). *Understanding Police Use of Force: Officers, Suspects, and Reciprocity*. Cambridge University Press. doi: 10.1017/CBO9780511499449
- Balko, R. (2016, jul). *Why it's impossible to calculate the percentage of police shootings that are legitimate* - *The Washington Post*. Retrieved from [https://www.washingtonpost.com/news/the-watch/wp/2016/07/14/why-its-impossible-to-calculate-the-percentage-of-police-shootings-that-are-legitimate/?utm{\\\_}term=.2116c367c05f](https://www.washingtonpost.com/news/the-watch/wp/2016/07/14/why-its-impossible-to-calculate-the-percentage-of-police-shootings-that-are-legitimate/?utm{\_}term=.2116c367c05f)
- Bell v. Wolfish*, 441 U.S. 520. (1979).
- Blinder, A., & Mazzei, P. (2018, feb). *As Gunman Rampaged Through Florida School, Armed Deputy 'Never Went In'*. New York. Retrieved from <https://www.nytimes.com/2018/02/22/us/nikolas-cruz-florida-shooting.html>
- Boivin, R., Gendron, A., Faubert, C., & Poulin, B. (2017). The malleability of attitudes toward the police: immediate effects of the viewing of police use of force videos. *Police Practice and Research*, 18(4), 366–375. Retrieved from <http://dx.doi.org/10.1080/15614263.2016.1230063> doi: 10.1080/15614263.2016.1230063
- Bradford, B., Milani, J., & Jackson, J. (2016). Identity, Legitimacy and 'Making Sense' of Police Violence. *Social Science Research Network*, 1–14.
- Braga, A. A., Winship, C., Tyler, T. R., Fagan, J., & Meares, T. L. (2014). The Salience of Social Contextual Factors in Appraisals of Police Interactions with Citizens: A Randomized Factorial Experiment. *Journal of Quantitative Criminology*, 30(4), 599–627. doi: 10.1007/s10940-014-9216-7
- Campaign Zero. (2018). *Campaign Zero*. Retrieved 2018-02-03, from <https://www.joincampaignzero.org/solutions/>

- Celestin, B. D., & Kruschke, J. K. (2019a). *Lay evaluations of police and civilian use of force: Action severity scales*. Retrieved from <https://osf.io/h6jg3/>
- Celestin, B. D., & Kruschke, J. K. (2019b). *Lay evaluations of police and civilian use of force: Moral and physical magnitude ratings of officer and civilian actions*. Retrieved from <https://osf.io/aqfeh/>
- Chapman, C. (2012). Use of force in minority communities is related to police education, age, experience, and ethnicity. *Police Practice and Research, 13*(5), 421–436. doi: 10.1080/15614263.2011.596711
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A Focus Theory of Normative Conduct. *Journal of Personality and Social Psychology, 58*(6), 1015–1026. doi: 10.1037/0022-3514.58.6.1015
- Coates, T.-N. (2015, apr). The myth of police reform. *The Atlantic*. Retrieved from <http://www.theatlantic.com/politics/archive/2015/04/the-myth-of-police-reform/390057/>
- Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: police officers and racial bias in the decision to shoot. *Journal of personality and social psychology, 92*(6), 1006–1023.
- Culhane, S. E., Boman, J. H., & Schweitzer, K. (2016). Public Perceptions of the Justifiability of Police Shootings: The Role of Body Cameras in a Pre- and Post-Ferguson Experiment. *Police Quarterly, 19*(3), 251–274. doi: 10.1177/1098611116651403
- Davis, E., Whyde, A., & Langton, L. (2018). *Contacts Between Police and The Public, 2015* (Tech. Rep.). Washington, DC: U.S. Department of Justice Bureau of Justict Statistics.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*, 1–12. doi: 10.3758/s13428-014-0458-y
- Denwood, M. J. (2016). runjags: An R package providing interface utilities, model tem-

- plates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, 71(9), 1–25. doi: 10.18637/jss.v071.i09
- Eith, C., & Durose, M. R. (2011). *Contacts Between Police and The Public, 2008* (Vol. BJS 2008; Tech. Rep. No. 234599).
- Eriksson, K., Strimling, P., & Coultas, J. C. (2014). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes*, 129, 59–69. doi: 10.1016/j.obhdp.2014.09.011
- Fountain, H., & Schmidt, M. S. (2016, jul). *'Bomb Robot' Takes Down Dallas Gunman, but Raises Enforcement Questions*. New York. Retrieved from [https://www.nytimes.com/2016/07/09/science/dallas-bomb-robot.html?smid=tw-nytimes&smtyp=cur&{\\\_}r=0](https://www.nytimes.com/2016/07/09/science/dallas-bomb-robot.html?smid=tw-nytimes&smtyp=cur&{\_}r=0)
- Fridell, L. a., & Binder, A. (1992). Police officer decisionmaking in potentially violent confrontations. *Journal of Criminal Justice*, 20(5), 385–399.
- Friedersdorf, C. (2015, jun). It's Time for Good Cops to Do Something About Bad Cops. *The Atlantic*. Retrieved from [http://www.theatlantic.com/politics/archive/2015/06/its-time-for-good-cops-to-do-something-about-bad-cops/396890/{\#}disqus{\\\_}thread](http://www.theatlantic.com/politics/archive/2015/06/its-time-for-good-cops-to-do-something-about-bad-cops/396890/{\#}disqus{\_}thread)
- Friedersdorf, C. (2017a, dec). Footage of a police shooting that jurors chose not to punish. *The Atlantic*. Retrieved from <https://www.theatlantic.com/politics/archive/2017/12/footage-of-a-police-killing-that-jurors-didnt-punish/547868/{\#}article-comments>
- Friedersdorf, C. (2017b). *Footage of the Police Shooting of Daniel Shaver - The Atlantic*. Retrieved 2017-12-10, from <https://www.theatlantic.com/politics/archive/2017/12/footage-of-a-police-killing-that-jurors-didnt-punish/547868/>
- Friedersdorf, C. (2017c, dec). A police killing without a hint of racism. *The Atlantic*. Retrieved from <https://www.theatlantic.com/politics/archive/2017/12/a-police-killing-without-a-hint-of-racism/546983/>

- Friedman, B. (2017). *Unwarranted : policing without permission*. New York: Farrar, Straus and Giroux.
- Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender Differences in Responses to Moral Dilemmas: A Process Dissociation Analysis. *Personality and Social Psychology Bulletin*, *41*(5), 696–713. Retrieved from <http://psp.sagepub.com/cgi/doi/10.1177/0146167215575731> doi: 10.1177/0146167215575731
- Fryer Jr., R. G. (2016). *An empirical analysis of racial differences in police use of force*.
- Garner, J. H., Schade, T., Hepburn, J., & Buchanan, J. (1995). Measuring the continuum of force used by and against the police. *Criminal Justice Review*, *20*(2), 183–205. Retrieved from <http://hjb.sagepub.com.proxy.lib.umich.edu/content/9/2/183.full.pdf+html> doi: 10.1177/07399863870092005
- Gerber, M. M., & Jackson, J. (2017). Justifying Violence: Legitimacy, Ideology and Public Support for Police Use of Force. *Psychology, Crime and Law*, *23*(1), 79–95.
- Gorner, J., & Dardick, H. (2016, oct). *Citing beating of officer, Chicago's top cop says police are 'second-guessing themselves'*. Chicago. Retrieved from <http://www.chicagotribune.com/news/local/breaking/ct-citing-beating-of-officer-chicago-s-top-cop-says-police-are-second-guessing-themselves-20161006-story.html>
- Graham, J., Haidt, J., & Nosek, B. a. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, *96*(5), 1029–1046.
- Graham v. Connor*, 490 U.S. 3. (1989).
- Gray, K., Schein, C., & Ward, A. F. (2014). The Myth of Harmless Wrongs in Moral Cognition: Automatic Dyadic Completion From Sin to Suffering. *Journal of experimental psychology. General*, *143*(4), 1600–1615.
- Gray, K., & Wegner, D. M. (2009). Moral Typecasting: Divergent Perceptions of Moral Agents and Moral Patients. *Journal of Personality and Social Psychology*, *96*(3),

505–520. doi: 10.1037/a0013748

- Gureckis, T. M., Martin, J., McDonnell, J., Alexander, R. S., Markant, D. B., Coenen, A., ... Chan, P. (2015). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavioral Research Methods*, 1–15. doi: 10.3758/s13428-015-0642-8
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Knopf Doubleday Publishing Group.
- Hall, A. V., Hall, E. V., Perry, J. L., & Hall, A. V. (2016). Black and Blue : Exploring Racial Bias and Law Enforcement in the Killings of Unarmed Black Male Civilians. *Journal of Applied Social Psychology*, 46(3), 175–186.
- Hamm, J. A., Trinkner, R., & Carr, J. D. (2017). Fair Process, Trust, and Cooperation: Moving Toward an Integrated Framework of Police Legitimacy. *Criminal Justice and Behavior*, XX(X), 009385481771005. doi: 10.1177/0093854817710058
- Handelman, S. (2014). *Ferguson Leads TCR's Top News Stories of 2014*. Retrieved from <http://www.thecrimereport.org/news/inside-criminal-justice/2014-12-ferguson-leads-tcrs-top-news-stories-of-2014>
- Harris, S., & Loury, G. C. (2016). *Racism and Violence in America*. Retrieved from <https://www.samharris.org/podcast/item/racism-and-violence-in-america>
- Hickman, M. J., Atherley, L. T., Lowery, P. G., & Alpert, G. P. (2015). Reliability of the Force Factor Method in Police Use-of-Force Research. *Police Quarterly*, 1098611115586175–. doi: 10.1177/1098611115586175
- Hogan, S. (2017, may). *Tulsa officer acquitted in fatal shooting of unarmed black man*. Retrieved from [https://www.washingtonpost.com/news/morning-mix/wp/2017/05/18/tulsa-officer-acquitted-in-fatal-shooting-of-unarmed-black-man/?utm{\\\_}term=.c0dcff495ece](https://www.washingtonpost.com/news/morning-mix/wp/2017/05/18/tulsa-officer-acquitted-in-fatal-shooting-of-unarmed-black-man/?utm{\_}term=.c0dcff495ece)
- Hume, D. (2003). *A treatise of human nature*. Dover Publications.

- Hyland, S., Langton, L., & Davis, E. (2015). *Police Use of Nonfatal Force, 2002–11* (Tech. Rep. No. November). U.S. Department of Justice Bureau of Justice Statistics.
- Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, 7(8).
- Jackson, J., Huq, A. Z., Bradford, B., & Tyler, T. R. (2013). Monopolizing force? Police legitimacy and public attitudes toward the acceptability of violence. *Psychology, Public Policy, and Law*, 19(4), 479–497. doi: 10.1037/a0033852
- Jefferis, E., Butcher, F., & Hanley, D. (2011). Measuring perceptions of police use of force. *Police Practice and Research*, 12(1), 81–96. doi: 10.1080/15614263.2010.497656
- Jenkinson, E., Neeson, C., & Bleetman, A. (2006). The relative risk of police use-of-force options: Evaluating the potential for deployment of electronic weaponry. *Journal of Clinical Forensic Medicine*, 13(5), 229–241. doi: 10.1016/j.jcfm.2005.11.006
- Johnson, D., & Kuhns, J. B. (2009). Striking Out: Race and Support for Police Use of Force. *Justice Quarterly*, 26(3), 592–623. doi: 10.1080/07418820802427825
- Johnson, J. A., Hogan, R., & Zonderman, A. B. (1981). Moral judgment, personality, and attitudes toward authority. *Journal of Personality and Social Psychology*, 40(2), 370–373. doi: 10.1037/0022-3514.40.2.370
- Kahan, D. M., Hoffman, D. A., & Braman, D. (2009). Whose eyes are you going to believe? Scott V. Harris and the perils of cognitive illiberalism. *Harvard Law Review*, 122(3), 837–906. doi: 10.2307/40379726
- King, S. (2016). *Shaun King Facebook Post and Commentary RE Dallas Shooting*. Retrieved 2017-01-01, from <https://m.facebook.com/shaunking/posts/1084886048216979>
- Klinger, D. A. (1995). The micro-structure of nonlethal force: baseline data from an observational study. *Criminal Justice Review*, 20(2), 169–186.

- Klinger, D. A., & Slocum, L. A. (2017). Critical Assessment of an Analysis of a Journalistic Compendium of Citizens Killed by Police Gunfire. *Criminology & Public Policy*, *16*(1), 1–14. doi: 10.1111/1745-9133.12283
- Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and STAN* (2nd ed.). Academic Press.
- Kruschke, J. K., & Liddell, T. M. (2018a). Bayesian analysis for newcomers. *Psychonomic Bulletin & Review*, *25*(1), 155–177. doi: 10.3758/s13423-017-1272-1
- Kruschke, J. K., & Liddell, T. M. (2018b). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. doi: 10.3758/s13423-016-1221-4
- Lemoine, P. (2017). *Police Violence against Black Men Is Rare: Here's What the Data Actually Say*. Retrieved from <http://www.nationalreview.com/article/451466/police-violence-against-black-men-rare-heres-what-data-actually-say>
- Leovy, J. (2015). *Ghettoside: A True Story of Murder in America*.
- Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a “common is moral” heuristic in the stability and change of moral norms. *Journal of Experimental Psychology General*, *147*(September), 228–242. doi: 10.1037/xge0000365
- Loury, G. C. (2015). *Ferguson Won't Change Anything. What Will?* Retrieved 2016-11-16, from <http://bostonreview.net/forum/glenn-c-loury-ferguson-wont-change-anything-what-will>
- Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. R. (2013). Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice. *Criminology*, *51*(1), 33–63. doi: 10.1111/j.1745-9125.2012.00289.x
- McGraw, A. P., & Tetlock, P. E. (2005). Taboo Trade-Offs, Relational Framing, and the Acceptability of Exchanges. *Journal of Consumer Psychology*, *15*(1), 2–15. doi: 10.1207/s15327663jcp1501\_2

- McWhorter, J. (2016, jul). Police kill too many people—white and black. *Time*. Retrieved from <http://time.com/4404987/police-violence/>
- Morin, R., Parker, K., Stepler, R., & Mercer, A. (2017). Behind the Badge. *Pew Research*, 11.
- Nix, J., Campbell, B. A., Byers, E. H., & Alpert, G. P. (2017). A Bird's Eye View of Civilians Killed by Police in 2015: Further Evidence of Implicit Bias. *Criminology and Public Policy*, 16(1), 309–340. doi: 10.1111/1745-9133.12269
- Paolacci, G., & Chandler, J. (2014). Inside the Turk. *Current Directions in Psychological Science*, 23(3), 184–188. doi: 10.1177/0963721414531598
- Pew. (2014). Pew: Sharp Racial Divisions in Reactions to Brown, Garner Decisions. Retrieved from <http://www.people-press.org/files/2014/12/12-8-14-Police-Race-release.pdf>
- Pinker, S. (2011). *The Better Angels of Our Nature: Why Violence Has Declined*. Penguin Publishing Group.
- Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*.
- Reaves, B. A. (2011). *Census of State and Local Law Enforcement Agencies, 2008* (Vol. Census of; Tech. Rep. No. July). Bureau of Justice Statistics. doi: 10.3886/ICPSR27681
- Reaves, B. A. (2012). *Federal Law Enforcement Officers, 2008* (Tech. Rep. No. June). Bureau of Justice Statistics.
- Salinger, T. (2016, sep). *Video released in fatal police shooting of unarmed Oklahoma man Terence Crutcher*. Retrieved from <https://www.nydailynews.com/news/crime/video-released-fatal-police-shooting-unarmed-oklahoma-man-article-1.2798435>
- Schein, C., & Gray, K. (2017). The Theory of Dyadic Morality : Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, 1–39.

doi: 10.1177/1088868317698288

- Schwitzgebel, E., & Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind and Language*, 27(2), 135–153.
- Selby, N. (2016). *Guest post: The 'low-hanging fruit' of police reform - The Washington Post*. Retrieved from [https://www.washingtonpost.com/news/the-watch/wp/2016/06/20/guest-post-the-low-hanging-fruit-of-police-reform/?utm{\\\_}term=.85743965ca24](https://www.washingtonpost.com/news/the-watch/wp/2016/06/20/guest-post-the-low-hanging-fruit-of-police-reform/?utm{\_}term=.85743965ca24)
- Shaffer, L. (1983). Toward Pepitone's vision of a normative social psychology: What is a social norm? *Journal of Mind and Behavior*, 4(2).
- Silver, J. R., & Pickett, J. T. (2015). Toward a better understanding of politicized policing attitudes: Conflicted conservatism and support for police use of force. *Criminology*, 53(4), 650–676. doi: 10.1111/1745-9125.12092
- Skinner, A. L., & Haas, I. J. (2016). Perceived threat associated with police officers and black men predicts support for policing policy reform. *Frontiers in Psychology*, 7(JUL), 1–17. doi: 10.3389/fpsyg.2016.01057
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: another contributor to attitude strength or something more? *Journal of personality and social psychology*, 88(6), 895–917.
- Stan Development Team. (2016). *RStan: the R interface to Stan*. Retrieved from <http://mc-stan.org>
- Stetser, M. (2001). *The use of force in police control of violence: Incidents resulting in assaults on officers*. El Paso: LFB Scholarly Publishing LLC.
- Stewart, E. (2018). *Florida shooting: multiple armed officers waited outside during massacre*. Retrieved 2018-06-23, from <https://www.vox.com/policy-and-politics/2018/2/24/17048720/florida-shooting-law-enforcement-gun>

- Sullivan, A. (2017, sep). NFL Protests Obscure the Facts on Race and Policing. *New York Magazine*. Retrieved from <http://nymag.com/daily/intelligencer/2017/09/nfl-protests-obscure-the-facts-on-race-and-policing.html>
- Sunshine, J., & Tyler, T. R. (2003). The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing. *Law & Society Review*, 37(3), 513–548.
- Team, R. (2016). *RStudio: Integrated Development Environment for R*. Retrieved from <http://www.rstudio.com/>
- Tennessee v. Garner*, 471 U.S. 1. (1985).
- Terrill, W., Alpert, G. P., Dunham, R. G., & Smith, M. R. (2003). A management tool for evaluating police use of force: an application of the force factor. *Police Quarterly*, 6(2), 150–171. doi: 10.1177/1098611102250491
- Terrill, W., & Mastrofski, S. D. (2002). Situational and officer-based determinants of police coercion. *Justice Quarterly*, 19(2), 215–248. doi: 10.1080/07418820200095221
- Terrill, W., & Paoline, E. A. (2013). Examining Less Lethal Force Policy and the Force Continuum: Results From a National Use-of-Force Study. *Police Quarterly*, 16(1), 38–65. doi: 10.1177/1098611112451262
- Terrill, W., Paoline, III, E. A., & Ingram, J. (2011). *Final Technical Report Draft: Assessing Police Use of Force Policy and Outcomes* (Tech. Rep.). National Institute of Justice. Retrieved from <https://www.nij.gov/publications/pages/publication-detail.aspx?ncjnumber=237794> doi: NCJNumber:237794
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of personality and social psychology*, 78(5), 853–870.
- Thompson, B. L., & Daniel Lee, J. (2004). Who Cares If Police Become Violent? Explaining Approval of Police Use of Force Using a National Sample. *Sociological Inquiry*, 74(3), 381–410. doi: 10.1111/j.1475-682X.2004.00097.x

- Tyler, T. R. (2004). Enhancing Police Legitimacy. *The annals of the American academy of political and social science*, 593(1), 84–99.
- Tyler, T. R. (2006). *Why People Obey the Law*. New Jersey: Princeton University Press.
- Tyler, T. R., Goff, P. A., & MacCoun, R. J. (2015). *The Impact of Psychological Science on Policing in the United States: Procedural Justice, Legitimacy, and Effective Law Enforcement* (Vol. 16) (No. 3). doi: 10.1177/1529100615617791
- Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116(1), 87–100.
- Waxman, O. B. (2017). *The History of Police in America and the First Force*. Retrieved 2019-04-03, from <http://time.com/4779112/police-history-origins/>
- Weitzer, R. (2005). Determinants of Public Satisfaction with the Police. *Police Quarterly*, 8(3), 279–297.
- Wikipedia Contributors. (2017). *2016 shooting of Dallas police officers*. Retrieved 2017-01-01, from <https://en.wikipedia.org/w/index.php?title=2016 shooting of Dallas police officers&oldid=792017195>
- Wikipedia Contributors. (2018). *Stoneman Douglas High School shooting*. Retrieved from [https://en.wikipedia.org/wiki/Stoneman\\_Douglas\\_High\\_School\\_shooting](https://en.wikipedia.org/wiki/Stoneman_Douglas_High_School_shooting)
- Worden, R. E. (2015). The causes of police brutality: The "causes" of police brutality: Theory and evidence on police use of force. In E. R. Maguire & D. E. Duffee (Eds.), *Criminal justice theory: Explaining the nature and behavior of criminal justice* (2nd ed., pp. 149–200). Routledge.

Curriculum Vitae / Résumé

Education

- May 2019 **Ph.D. in Psychology**, *Indiana University*, Bloomington, Indiana.  
Dissertation: *Moral evaluations of police and civilian use of force*  
Advisor: Dr. John K. Kruschke  
Committee: Dr. John K. Kruschke, Dr. Eliot Smith, Dr. Edward Hirt, Dr. Robert J. Rydell
- May 2012 **Bachelor of Arts in Psychology**, *Bethel College*, North Newton, Kansas.  
BA thesis: *Implicit sexual identity bias as a function of religiosity: an event-related potential study*  
Advisor: Dr. Dwight Krehbiel
- May 2012 **Certificate in Neuroscience**, *Bethel College*, North Newton, Kansas.

Honors, Awards, & Scholarships

- 2019 Gordon Kato Fellowship, Indiana University Social Psychology Faculty  
2019 Sharon Stephens Brehm Travel Award, Indiana University  
2019 Society for Personality and Social Psychology Graduate Travel Award  
2017 Center for Innovative Teaching and Learning Grant for Developing Statistical Models of Course Evaluation Data (*Statistical Analyst: Brad Celestin, PI: Dr. Richard Hullinger*)  
2017 Indiana University Graduate and Professional Student Organization Travel Award  
2017 Indiana University College of Arts and Sciences Travel Award  
2016 Indiana University Grant-in-Aid of Doctoral Research  
2014 National Science Foundation Graduate Research Fellowship Honorable Mention  
2014 Richard McFall Summer Research Fellowship, Indiana University  
2013 National Science Foundation Graduate Research Fellowship Honorable Mention  
2012 Thresher Award for Outstanding Academic Achievement in Psychology and Neuroscience, Bethel College  
2008 Silver (life-saving) Award, Kansas Association of Chiefs of Police  
2007 Medal of Merit, City of Newton  
2007 Certificate of Honor for Life-Saving Actions, City of Newton  
2002-2013 Commendations for Outstanding Service, Newton Police Department  
2003 Top Marksmanship Award, Kansas Law Enforcement Training Center  
2001 Music Scholarship in Vocal and Instrumental Performance, Bethel College

## Publications & Manuscripts

Celestin, B.D. & Kruschke, J.K. (in press). *Lay evaluations of police and civilian use of force: Action severity scales*, Law and Human Behavior <https://osf.io/jth52/>

Celestin, B.D. & Kruschke, J.K. (2019). *Lay evaluations of police and civilian use of force: Moral and physical magnitude ratings of officer and civilian actions*, Manuscript submitted for publication, <https://osf.io/gdnb4/>

Celestin, B.D. & Kruschke, J.K. (2019). *Disparate police and public expectations: Lay evaluators intuitively choose less severe and more normative police actions relative to the severity of civilian resistance*, Manuscript in preparation

Anderson, A. & Celestin, B.D. (2013) *Presenting Undergraduate Research at National Events*, Council on Undergraduate Research Quarterly

## Research Presentations

- 2019 Symposia Talk: *Lay expectations for appropriate police use of force*, Coauthor: Dr. John K. Kruschke, American Psychology-Law Society, Portland, OR
- 2019 Poster: *The effects of role-normativity on moral evaluations of police use of force*, Coauthor: Dr. John K. Kruschke, Society for Personality and Social Psychology, Portland, OR
- 2019 Talk: *The effects of perceived action normativity on moral evaluations of police use of force*, Social Psychology and Law: An International Perspective Preconference, Coauthor: Dr. John K. Kruschke, Society for Personality and Social Psychology, Portland, OR
- 2018 Poster: *Lay evaluations of police and civilian use of force*, Coauthor: Dr. John K. Kruschke, Society for Personality and Social Psychology, Atlanta, GA
- 2018 Talk: *Lay evaluations of police and civilian use of force*, Social Psychology and Law: In Court and Beyond Preconference, Coauthor: Dr. John K. Kruschke, Society for Personality and Social Psychology, Atlanta, GA
- 2018 Talk: *Lay evaluations of police and civilian use of force*, Coauthor: Dr. John K. Kruschke, Indiana University Social Psychology Seminar, Bloomington, IN
- 2017 Poster: *Moral evaluations of police and civilian use of force*, Coauthor: Dr. John K. Kruschke, Society for Personality and Social Psychology, Austin, TX
- 2017 Poster: *Moral evaluations of police and civilian use of force*, Coauthor: Dr. John K. Kruschke, Justice and Morality Preconference, Society for Personality and Social Psychology, Austin, TX
- 2016 Talk: *Lay moral evaluations of police and civilian use of force*, Coauthor: Dr. John K. Kruschke, Indiana University Social Psychology Seminar, Bloomington, IN
- 2014 Poster: *Differential amygdala responses for cultural vs. racial emotion faces*, Coauthor: Dr. Anne Krendl, Indiana University Annual Research Symposium and Poster Session, Bloomington, IN

- 2013 Talk: *Unraveling morality: the differential effects of stigmatized groups on moral decision-making*, Coauthor: Dr. Anne Krendl, Indiana University Social Psychology Seminar, Bloomington, IN
- 2012 Talk: *Sexual identity bias as a function of religiosity*, Coauthor: Dr. Dwight Krehbiel, Bethel College URICA Symposium, North Newton, KS
- 2012 Talk: *Sexual identity bias as a function of religiosity*, Coauthor: Dr. Dwight Krehbiel, STEM Community Meeting, Bethel College, North Newton, KS
- 2012 Poster: *Neural correlates of implicit sexual identity bias as a function of religiosity*, Coauthor: Dr. Dwight Krehbiel, Council on Undergraduate Research Posters on the Hill, US House of Representatives, Rayburn Office Building, Washington, DC
- 2011 Poster: *Neural correlates of implicit sexual identity bias as a function of religiosity*, Coauthor: Dr. Dwight Krehbiel, Society for Neuroscience Faculty for Undergraduate Research, Washington, DC
- 2011 Poster: *Brain electrical responses when remembering French-English vocabulary pairs*, Coauthor: Dr. Dwight Krehbiel, Bethel College URICA Symposium, North Newton, KS
- 2010 Talk: *Physiological changes during the recall of an offense*, Coauthors: Kelsie Wilson & Sonia Barrera, Bethel College URICA Symposium, North Newton, KS

#### Teaching Experience

- Spring 2019 Assistant Instructor – PY533 Bayesian Statistics (graduate level), Indiana University
- Fall 2018 Assistant Instructor – P553 Advanced Statistics in Psychology (graduate level)
- Summer 2018 Assistant Instructor – P304 Social Psychology, Indiana University
- Spring 2018 Invited Guest Lecturer – Q400 Cognitive Psychology Senior Seminar
- Spring 2018 Assistant Instructor – PY533 Bayesian Statistics (graduate level), Indiana University
- Fall 2017 Assistant Instructor/Guest Lecturer – P553 Advanced Statistics in Psychology (graduate level), Indiana University
- Summer 2017 Faculty Member – Bethel College Summer Science Institute (Understanding and Measuring Attitudes)
- Summer 2017 Guest Lecturer – P304 Social Psychology, Indiana University
- Spring 2017 Invited Guest Lecturer – Q400 Cognitive Psychology Senior Seminar
- Spring 2016 Assistant Instructor – K300 Statistics, Indiana University
- Fall 2016 Assistant Instructor/Guest Lecturer – P553 Advanced Statistics in Psychology (graduate level), Indiana University
- Summer 2016 Invited Guest Lecturer – PY211 Research Methods in Psychology, Indiana University
- Summer 2016 Assistant Instructor/Guest Lecturer – PY304 Social Psychology, Indiana University
- Spring 2016 Associate Instructor – PY211 Research Methods in Psychology (2 sections), Indiana University
- Spring 2016 Invited Guest Lecturer – PY304 Social Psychology, Indiana University
- Fall 2015 Assistant Instructor/Guest Lecturer – PY440 Cognitive psychology: Thinking and Reasoning, Indiana University

- Summer 2015 Assistant Instructor/Guest Lecturer – PY304 Social Psychology, Indiana University
- Summer 2015 Faculty Member – Bethel College Summer Science Institute (Understanding and Measuring Implicit and Explicit Attitudes)
- Spring 2015 Assistant Instructor/Guest Lecturer – PY457 The Science of Morality, Indiana University
- Fall 2014 Assistant Instructor – PY472 Lab in Brain Electrical Activity, Indiana University
- Spring 2014 Assistant Instructor – PY421 Research Methods in Social Psychology, Indiana University
- Fall 2013 Assistant Instructor – PY421 Research Methods in Social Psychology, Indiana University
- Summer 2013 Faculty Member – Bethel College Summer Science Institute (Understanding and Measuring Implicit and Explicit Attitudes)
- Spring 2013 Adjunct Instructor for Cognitive Psychology Lab, Bethel College

██████████ Professional Training & Education ( $\geq 40$  hours per course)

- Kansas Law Enforcement Training Center
- 2002-2003 172nd Basic Police Training Class – *University of Kansas*
- Federal Bureau of Investigation
- Investigative Interviewing of Children
- Conducting Undercover Cyber Investigations
- Advanced Microsoft Computer Forensics
- Accessdata Forensic Toolkit Bootcamp
- Annual Safeguard Assessment
- Backstopping/Covert Investigations
- Cyber Task Force Basic Training
- National White Collar Crime Center
- Identifying and Seizing Electronic Evidence
- Secure Techniques for On-site Preview
- Basic Data Recovery and Acquisition
- Intermediate Data Recovery and Analysis
- Forensic Computer Construction
- Windows NT Operating System Forensics
- Advanced Internet Trace Evidence
- Child Abuse Investigations
- Finding Words (Cornerhouse: RATAAC Interviewing Protocol)
- Beyond Finding Words: Advanced Forensic Interviewing
- Via Christi St. Joseph Hospital SANE/SART Medical Exam Training
- 2009-2010 Kansas Internet Crimes Against Children Conference
- 2011-2012 United States Attorney's Protect Our Children Conference

Kansas Internet Crimes Against Children Internship

*(See appendix for additional professional education)*

██████████ Employment History

2019-Present **Assistant Professor of Psychology**, *Bethel College*, 300 East 27th Street, North Newton, Kansas 67117.

2013-2019 **Associate Instructor, Assistant Instructor, or Research Fellow**, *Department of Psychological and Brain Sciences*, Indiana University, Bloomington, Indiana.

Primary Duties:

- o Generate course content (e.g., assignments, exams, lecture slides, reading lists, syllabi, etc.)
- o Give lectures & lead discussions
- o Run lab sessions including technical (e.g., EEG, statistical) and writing labs
- o Hold office hours
- o Meet individually with students
- o Grade and provide feedback to students
- o Generate documents and interact with the Indiana University Institutional Review Board (IRB) to obtain formal approval to run experiments on human subjects
- o Design, program, and implement scientific experiments
- o Develop programs to clean and organize data
- o Formulate mathematical models to describe data
- o Program, debug, and interpret statistical analyses
- o Write technical papers collaboratively
- o Present research results
- o Apply for grants (institutional and government) individually and collaboratively

2011-2013 **Task Force Agent**, *Federal Bureau of Investigation, Cyber Division*, Kansas City Field Office, Wichita Resident Agency.

Primary Duties:

- o Conduct investigations into cyber intrusion, fraud, technology-facilitated child abuse, and the production, collection, and distribution of child sexual abuse media
- o Conduct physical and electronic surveillance
- o Prepare and execute search and arrest warrants
- o Conduct victim/witness/suspect interviews & interrogations
- o Execute formal forensic examinations of digital media (e.g., computer hard drives, cellular phones, etc.)
- o Provide technical and legal expertise and assistance to other agents & agencies throughout Kansas and surrounding states
- o Provide expert testimony in the United States Federal Court and the various District Courts of the State of Kansas

- o Work collaboratively with numerous agencies including the U.S. Attorney's Office, the U.S. Marshall's Service, the Drug Enforcement Administration, and the Department of Homeland Security

2008-2013 **Detective**, *Newton Police Department*, City of Newton, Kansas, Investigations Division.

Primary Duties:

- o Conduct investigations into homicide, robbery, aggravated battery, child sexual and physical abuse, fraud, burglary, technology-facilitated child abuse, and the production, collection, and distribution of child sexual abuse media
- o Conduct physical and electronic surveillance
- o Prepare and execute search and arrest warrants
- o Conduct victim/witness/subject interviews & interrogations
- o Execute forensic examinations on digital media
- o Provide technical expertise and assistance to other officers & agencies
- o Provide expert testimony in federal, district, & municipal courts
- o Work collaboratively with the various elected District Attorneys of Kansas.

2007-2008 **USD 373 Elementary School Resource Officer**, *Newton Police Department*, City of Newton, Kansas, Support Division.

Primary Duties:

- o Act as liaison between the Newton Police Department and the five elementary schools of USD373
- o Develop positive relationships with K-5 students, parents, & teachers
- o Collaborate with school officials and social workers to ensure school safety & develop emergency safety plans
- o Enforce all municipal, state, & federal laws, including school-relevant laws (e.g., weapons violations, truancy, etc.)
- o Generate and provide law/police-related public talks to students, parents, and staff
- o Investigate all cases of child abuse and neglect involving school students
- o Research, develop, & engage in community policing strategies

2002-2007 **Patrol/Master Police Officer**, *Newton Police Department*, City of Newton, Kansas, Operations Division.

Primary Duties:

- o Serve as an active first-responder to all relevant emergency situations
- o Enforce local, state, and federal laws
- o Train newly recruited police officers in the field (FTO)

## Selected Examples of Vocational Teaching and Presentations

2012 Overview of FBI Innocent Images program and cyber division, *Hutchinson, KS Rotary Club*

- 2011 Digital Forensics Procedure and Best Practice (FBI Cyber Division), *Kansas Institute of Internal Auditors*
- 2011 Forensic Interviewing & Child Abuse, *Assessment of Human Systems Class, Bethel College Social Work Department*
- 2009 Forensic Interviewing & Child Abuse, *Assessment of Human Systems Class, Bethel College Social Work Department*
- 2007 Taser User Certification Instructor, *Kansas Bureau of Investigation SWAT Team*
- 2004-2013 Classes taught to *Harvey County Joint Law Enforcement Agencies (i.e., Police Departments of Newton, Hesston, Halstead, Burrton, North Newton, Walton, Sedgwick, and the Harvey County Sheriff's Department)*
  - o Use of Force Case Law
  - o Ground and Standing Combatives
  - o Collapsible Baton Deployment
  - o Digital Forensics Evidence Collection Procedures in the Field
  - o Ethics & Diversity Training
  - o Interview & Interrogation
  - o Internet Crimes Against Children
  - o Child Abuse Investigations
  - o Peer to Peer Networks
  - o Cell Phone Forensics & Case Law
  - o Taser User Certification
  - o Handcuffing
  - o Weapon Retention
  - o Lateral Vascular Restraint
- 2009-2013 Classes taught to *Newton Citizen's Police Academy*
  - o Child Abuse
  - o Technology-facilitated Criminal Investigations
  - o Interview & Interrogation Strategy
  - o Crime Scene Processing

■ Additional Professional Roles and Organizational Memberships

- 2019-Present Member, American Psychology-Law Society
- 2014-Present Member, Society of Personality & Social Psychology
- 2011-2012 Member, Society for Neuroscience
- 2011-2013 Board Member, Board of Directors, Finding Words Kansas
- 2011-2013 Child Forensic Interviewer, Federal Bureau of Investigation
- 2007-2013 Child Forensic Interviewer and Multidisciplinary Team Member, Heart to Heart Child Advocacy Center
- 2011-2013 Digital Forensic Examiner, Federal Bureau of Investigation
- 2008-2013 Digital Forensic Examiner, Newton Police Department

- 2011-2013 Collaborator with the Kansas Equality Coalition (introducing LGBT issues to law enforcement through diversity training)
- 2008-2013 Hostage Negotiator, Harvey County Emergency Response Team (SWAT)
- 2004-2008 Field Training Officer, Newton Police Department
- 2009-2013 Team Leader, Harvey County On-line Sex-Crime Apprehension and Removal Task Force
- 2012-2013 Law Enforcement Ethics Instructor, Harvey County Joint Law Enforcement Agencies
- 2008-2013 Interview and Interrogation Instructor, Harvey County Joint Law Enforcement Agencies
- 2004-2013 Taser Instructor, Harvey County Joint Law Enforcement Agencies
- 2004-2013 Use of Force Instructor & Internal Investigation Team Member, Harvey County Joint Law Enforcement Agencies
- 2008-2013 Member, Kansas Internet Crimes Against Children Task Force
- 2002-2013 Member, Kansas Peace Officers Association
- 2010-2013 Committee Member, City of Newton Technology Committee

### Technical Skills

- o Bayesian statistical methods
- o JAGS
- o Javascript, HTML, CSS
- o R/RStudio
- o RMarkdown
- o jsPsych
- o Amazon Mechanical Turk API
- o Forensic ToolKit (AccessData)
- o Sleuthkit & Other Linux Forensics Suites
- o Python
- o Canvas
- o Apache
- o Matlab
- o Virtualbox
- o Frequentist statistical methods (p values)
- o Stan
- o L<sup>A</sup>T<sub>E</sub>X
- o Knitr
- o Linux/Mac/Windows
- o Qualtrics
- o psiTurk
- o EnCase
- o SQL
- o Git
- o SONA
- o PHP
- o EEGlab/ERPLab
- o Moodle

### References

Available upon request.