

ABSTRACT

Title of Dissertation: HERO OR VICTIM: THE CONSEQUENCES
OF MORAL STIGMA ON NECESSARY EVIL
PRACTITIONERS

Virginia Choi, Doctor of Philosophy, 2025

Dissertation directed by: Professor Michele Gelfand
Department of Psychology

Necessary evil practitioners (NEPs) perform tasks that cause harm for a greater societal good, placing them in morally complex roles. Across three studies, this research examined how NEPs apply moral accounting to justify their harm-doing, and how this shapes their hero versus victim self-identification. Results showed moral accounting predicted greater hero identification, which promoted prosocial outcomes such as restitution support for their targets of harm and greater work transparency. Under elevated levels of perceived moral stigma, the pathway from hero to victim identification counteracted these prosocial effects, rendering them statistically non-significant. These findings contribute to both the theoretical and practical understanding of enacting necessary evil tasks, as they relate to occupational stigma and relevant work outcomes.

HERO OR VICTIM: THE CONSEQUENCES OF MORAL STIGMA ON
NECESSARY EVIL PRACTITIONERS

by

Virginia Choi

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

Advisory Committee:
Professor Michele Gelfand, Chair
Professor Paul Hanges
Professor Rellie Rozin
Professor Jennifer Wessel
Professor Linda Zou

© Copyright by
Virginia Choi
2025

Acknowledgements

I am deeply grateful to my parents, who supported me on this academic path, even when its twists and turns seemed unfamiliar and strange. Although the world of academia was not something you always fully understood, your steadfast encouragement and love never wavered. To my advisor, thank you for your belief in me from the moment I stumbled into your office as an eager acolyte of cross-cultural research to these last few months of finalizing the dissertation. You have lifted me up, during countless moments when I *truly* struggled to believe in myself, and you reminded me of all the exciting and amazing wonderment that comes with our work. Your guidance, support, enthusiasm, and encouragement gave me the freedom to follow my research passions ("anything goes!"); and you have forever instilled in me the importance of gratitude and celebrating life. I am especially thankful to my fellow graduate researchers, i.e., my academic foxhole companions. We survived the trenches of this process together and I could not have made it through without your company, solidarity, advice, and camaraderie. Although our collaboration and friendship will continue far past these PhD years, I will deeply miss these formative times when we were growing and learning together as emerging scholars—and I owe special thanks to my cohort companion Josh, who plodded through this adventure with me from day one. To my extended family and friends—my aunts, uncles, in-laws, collaborators, and the amazingly kind and generous staff who keep our department running smoothly—thank you so much for your love and support, especially during those times when this path felt unruly and protracted. To my partner, who had front-row seats to the uncut version of this entire PhD journey—you saw the stress, the

challenges, the tears, and the frustrations that came with it all. Your daily hugs were (and continue to be) my recharge station, giving me the strength to return to the work again and again. I could not have done this without you. And finally, to my Spoticus, my furry and ferociously loyal companion—the consummate research assistant—thank you for keeping me company through endless drafts and all those late nights.

Table of Contents

Acknowledgements.....	ii
Introduction.....	1
Theoretical Background.....	2
Summary of Hypotheses.....	19
Overview of Present Research.....	20
Study 1.....	21
Study 2.....	30
Study 3.....	39
Conclusion and General Discussion.....	50
Appendices.....	58
Appendix A: Study Materials.....	58
Appendix B: Pilot Studies and Vignette Development.....	63
Tables and Figures.....	73
References.....	92

Introduction

A necessary evil (NE) refers to a work task wherein employees must knowingly conduct harm to others in the service of a greater good (Molinsky & Margolis, 2005). These harms include, for example, managers conducting layoffs, eviction officers carrying out notices, or social workers removing children from neglectful homes. In each case, the harm is sanctioned as part of the practitioner's job role, and it is typically embedded into their occupational mandate. Practitioners who perform these tasks—whom will be referred to in this paper as necessary evil practitioners (NEPs)—often express prosocial motivations for assuming their roles (Margolis & Molinsky, 2008). However, the enactment of harm, even, when necessary, presents a unique moral conflict. This research seeks to better understand how NEPs make sense of their moral identities as they process these NE tasks, and the consequences of when practitioners must face public scrutiny towards their profession.

In addition to NEPs coping with the internal ethical burden of their work, they also face significant social pressures (Patil & Lebel, 2019). Many NE-relevant job roles are public facing, exposing these individuals to moral critique when their NE tasks are misunderstood, misrepresented, or condemned. The moral ambiguity of the tasks found in these jobs can be seen by some as a helpful service, and by others as mere cruelty. When such polarized social perceptions exist, negative views toward one's work may produce serious consequences for NEPs' moral self-concept, psychological states, and professional behaviors. The present research proposes the following processes and theoretical framework: At the task level, practitioners who engage in moral reasoning to justify the harm of NE tasks via a greater-good lens (i.e., moral accounting) are more

likely to identify themselves as heroic at the person-level. This hero self-identification, in turn, is predicted to promote prosocial attitudes and behaviors (Aquino & Reed, 2002; Goranson et al., 2022). However, drawing on symbolic interactionism, this study examines NEPs as socially situated individuals whose self-regard is informed by how others perceive them (Becker, 1971; Cooley, 1902; Mead, 1934). When NEPs are faced with moral stigmatization towards their profession from the greater public, this hero identification can be undermined. NEPs may, then, identify themselves as victims, a self-perception that can prompt reduced prosocial responses, including greater detachment from the relational and public-facing demands that are often involved in performing NE tasks (Margolis & Molinsky, 2008). This research not only addresses how NEPs process their own job-related ethical uncertainties, but also how they reconcile their moral self-image when threatened by a stigmatized social image. This dual examination aims to explain the decline in prosocial responses of NEPs, which can become occupational barriers that adversely impact their relationships with the public and compromise their professional conduct.

Theoretical Background

Drawing on Molinsky and Margolis (2005)'s foundational conceptual work on this topic, necessary evil tasks have three defining characteristics: *(i.)* a valued purpose and general agreement of its benefits, which renders them necessary; *(ii.)* they involve the infliction of harm; and *(iii.)* they constitute a mandated aspect of one's job role. To better understand what NEs are and what they are not, the following section details two neighboring constructs: unethical pro-organizational behaviors and moral incongruence,

as experienced in work settings. Both phenomena provide insights into NEs through some of their overlapping characteristics as well as the sources of their divergence.

Unethical Pro-Organizational Behaviors. First, it is valuable to define unethical pro-organizational behaviors (UPBs), which is a well-researched topic among business ethics and corporate social responsibility scholarship. UPBs describe the actions of employees who behave unethically in the service of helping their organizations (Mishra et al., 2021; Umphress et al., 2010). These unethical actions involve violating societal standards and ethicality norms or “hypernorms” (Umphress & Bingham, 2011; Donaldson & Dunfee, 1994). Similar to NE practitioners (NEPs), employees who perpetrate UPBs engage in unethical acts out of prosocial motivations, i.e., helping their organizations versus their self-interest. However, the “harm” found in NE tasks are more directly associated with its societal utility and, most critically, they are explicitly part of one’s job directives. UPBs typically involve morally deviant behaviors as employees operate outside their job requirements such as lying or hiding information from stakeholders, in order to protect the reputation of their organization.

Workplace Moral Incongruence. NEPs, when tasked with harm-doing, can experience moments of moral incongruence. However, moral incongruence does not fully nor accurately capture the NEP’s experience, especially given how this research mostly focuses on person-role conflicts. For example, Shukla and colleagues (2024)’s study on workplace moral incongruence focused on situations when highly ethical employees were mandated to complete unethical tasks due to their organization’s immoral operations. Hence, the source of moral incongruence and discomfort arose from personal disagreement with the moral principles that underpinned one’s occupational roles.

Another possible combination in the study of moral incongruence has been an unethical employee who cannot live up to the high moral standards expected by their organization. In this sense, moral incongruence is akin to *role conflict* as employees' preferred ideals and behaviors are misaligned with their job roles (House & Rizzo, 1972). These conflicts often result in employees leaving their roles and organizations. On the other hand, the current understanding of work involving NEs assumes a self-selective process, wherein workers knowingly assume these occupational roles and value their societal contributions (Coldwell et al., 2008; Dierdorff & Morgeson, 2013; Wilk & Sachett, 1996). NEPs are expected, at least to a certain extent, to have taken on these job roles because they agree with the purpose or the “necessary” goals undergirding NE tasks, despite or including the means of achieving them (Molinsky & Margolis, 2005; Margolis & Molinsky, 2008).

Nevertheless, the harms committed by NEPs can still be psychologically taxing, despite a practitioner's personal recognition and agreement with the merit of completing NE tasks. This contrasts with unethical pro-organizational behaviors (UPBs), in which employees are driven by their organizational loyalty to violate societal norms and bring benefits to their organizations (Umphress & Bingham, 2011). In contrast, NEPs are not engaging in unlawful misconduct, but in morally complex service work. Moreover, unlike workers experiencing moral incongruence, due to a misalignment between personal values and job requirements (Shukla et al., 2024), NEPs typically choose their roles and endorse the societal goals of their work (Coldwell et al., 2008; Dierdorff & Morgeson, 2013).

Moral Accounting and Hero Self-Identification

Cognitive dissonance theory suggests that people are motivated to maintain consistency between their values and actions (Festinger, 1957). Ethical dissonance, which is the psychological discomfort that arises from perceived contradictions between one's actions and one's moral self-image, is especially relevant for NEPs (Barkan et al., 2012; Voisin & Fointiat, 2012). In order to stay gainfully employed, avoiding or not doing the NE task is often not an option for NEPs. Instead, practitioners may rely on justification-based coping strategies. One such strategy, herein referred to as *moral accounting*, involves post-hoc justification processes that account greater weight towards the task's potential benefits. This is an expected sense-making response of NEPs (Ashforth & Kreiner, 2013). Moral judgments that justify harm by focusing on their benefits align with a consequentialist framework of ethical judgements (Shaw, 2006; Singer, 1991; Sinnott-Armstrong, 2008) and akin to utilitarian moral judgements (Conway & Gawronski, 2012). Indeed, this cost-benefit rationalization process is the foundation in which NE tasks have become institutionally approved, as enough pockets of society considered these tasks as putatively necessary to be performed and professionalized (Ashforth & Kreiner, 2013). Moral accounting may enable NEPs to resolve their ethical dissonance towards performing necessary evil tasks (Molinsky & Margolis, 2005; Shaw, 2006; Singer, 1991).

Moral Reasoning Processes. In the present research, I draw upon the extensive body of psychological research dedicated to understanding how individuals maintain a favorable view of themselves, specifically the moral cognitive processing that is involved in facilitating a desired positive self-view (Aquino & Reed, 2002; Crocker & Knight,

2005). Notably, the perception of oneself as a moral actor is found to be a key contributor to individuals' overall positive self-worth.

Cognitive dissonance theory, a cornerstone of socio-cognitive theories, describes people's desire to maintain a stable positive self-image and how individuals diffuse threats that jeopardize this equilibrium (Andiappan & Dufour, 2017; Aronson, 1992 ; Festinger, 1957). Most individuals strive to maintain or secure a positive self-image (Crocker & Park, 2004). Accordingly, people who experience significant discomfort when they face discrepancies in their positive self-image are, then, driven to take actions that reduce this dissonance (Voisin & Fointiat, 2012). Ethical dissonance is a subset of cognitive dissonance experiences, which is specifically associated with threats to one's moral self-concept (Barkan et al., 2012). Of the many favorable social evaluations possible (e.g., intelligence and friendliness), people seem to prioritize being viewed as moral in the eyes of others (Aquino & Reed, 2002; Dunning, 2007). In the context of NEPs, the enactment of harm can lead to ethical dissonance, which can impair their moral self-image and positive views of themselves. In accordance with cognitive dissonance theories, NEPs are likely to instantly embark on introspective cognitive processing that aims at coping with this ethical dissonance.

Among several dissonance reduction strategies that individuals exhibit, there are two overarching categories that these coping styles can be divided into: active coping and avoidance coping (Snow et al., 2003). Avoidance coping involves avoidance of future dissonance-causing occurrences, which includes reactions like psychological and emotional distancing (Nandkeolyar et al., 2014). Unlike active coping strategies, however, avoidance methods do not directly help people reduce their internal dissonance

(Abraham, 1999). Thus, these passive avoidance coping strategies are not optimal. Festinger (1957) summarized three ways people *actively* embark on resolving the problem of cognitive dissonance: (1) to change the dissonance-causing behavior, i.e., stop the behavior; (2) to minimize the effect of one of the variables causing the value-behavior discrepancy such as trivializing the harm enacted; and (3) to focus on justifications, viewed as a cognitive restructuring or supplanting the dissonance with new cognitions that re-establishes self-alignment. In NE jobs, since harm enactment is mandated as a job requirement, the first option to neglect the behavior altogether may not be feasible without quitting the job. Additionally, trivialization may be difficult to maintain as a strategy if enactors are directly involved in the harm-doing, thereby, confronted by the undeniability of their actions' aversive consequences (Clair & Dufresne, 2004; Schultz & Darley, 2014). Therefore, the proposed research focuses on NEPs' post-hoc justification processes.

Justification is an active coping strategy that has been well-researched in helping people override their dissonance and maintain self-esteem (Barkan et al., 2012; Taylor & Brown, 1988; Tesser, 2000). This coping strategy of justification can be accomplished in many ways such as reinterpreting one's actions as normative ("everyone does it"), or attributing greater external responsibility, etc. (Andiappan & Dufour, 2017; Harvey & Weary, 1984; Margolis & Molinsky, 2008; Matza & Sykes, 1957). Through justifications, people can actively neutralize their ethical dissonance and protect their self-image (Bandura, 1999; Tsang, 2002). Philosopher Alfred Mele (1987) describes this motivated reasoning process as a "desire-influenced manipulation of data" (p. 126). Baumeister and Newman (1994) categorized these self-serving cognitive reasoning

processes into two thinking modes: (i.) *the intuitive scientist*, and (ii.) *the intuitive lawyer*. The intuitive scientist strives to make impartial and objectively accurate conclusions—with the inevitable susceptibility to biases and human error notwithstanding. The intuitive lawyer selectively finds justifications that reinforce their predetermined stance or preferred conclusions. Research on motivated moral reasoning suggests most people come to the defense of their moral self-image as intuitive lawyers (Haidt, 2007).

Furthermore, once individuals meaningfully attach their NE tasks as helpful, given that moral accounting involves seeing one’s actions as having societal or community gains, it is possible that NEPs may even identify themselves as heroic actors. Goranson and colleagues (2022) suggest that individuals classify themselves and others into certain moral archetypes: heroes (help-doers), villains (harm-doers), victims (harm-receivers), and beneficiaries (help-receivers). When NEPs morally account for their actions, which links them to a broader altruistic purpose, they may be more likely to identify themselves as heroes, i.e., morally upright actors who endure hardships for others’ benefit (Kinsella et al., 2015; Walker, 2016). While the content of narratives featuring heroism has changed over time—for instance, in comparison to epics written by the ancient Greeks—conceptualization of the contemporary hero has retained its moral quality of helpfulness and selflessness (Allison & Goethals, 2011). Additionally, the modern hero serves others at great personal risk (Franco et al., 2011, 2018; Kinsella et al., 2015). As individuals reflect on a NE task’s overall positive gains to society, it is possible that performers of NEs are more likely to view themselves as benevolent actors—even heroes.

Hypothesis 1: Moral accounting is positively associated with NEPs' hero self-identification.

As part of an exploratory component of this research, I will aim to further understand the mechanisms that could explain the possible link between moral accounting and hero identification (see Tetlock & Levi, 1982). For instance, moral accounting may enable a more favorable view of oneself as a moral actor (Aquino & Reed, 2002; Crocker & Knight, 2005), as indicated by high self-esteem (Tetlock & Levi, 1982). Vogel and Willems (2020) conducted an intervention-based study with public servants, which demonstrated that external reminders on the prosocial significance of their work led to improved well-being and stronger job retention intentions. Alternatively, as individuals undergo moral accounting, this process may be viewed as an ethical development exercise that generates moral self-efficacy. In accordance with social-cognitive theory (Bandura, 1986; Schwarzer & Luszczynska, 2008), self-efficacy beliefs are enhanced through enactive mastery and direct experiences of overcoming a challenging task. In this regard, NEPs' moral accounting may serve as an opportunity for meaningful exposure and effective deliberation of difficult ethical situations. Moral accounting, in this sense, represents an impactful learning opportunity that could increase one's self-perceived moral efficacy (Smith & Kouchaki, 2021); thus, this could encourage greater alignment with a heroic self-concept. Lastly, when moral accounting is seen as a reparatory action taken by NEPs to regain a positive moral self-image, perhaps there is an amplified or compensatory effect that translates into moral superiority (Dong et al., 2019). Viewed through this lens, hero identification may be representative of one's

self-enhancement efforts. These three possible explanations—to regain self-esteem, obtain self-expansion via moral self-efficacy, or enhance one’s self-regard—will be explored in the current research.

Prosocial Responses. Heroes, across time and cultures, are presented as agentic individuals who actively participate in behaviors that align with their reputation as doers of good (Stanley & Kay, 2024; Stanley et al., 2023). This is in accordance with self-consistency and moral self-regulation theories (Mullen & Monin, 2016), which would suggest greater hero identification is likely to promote attitudes and behaviors that align with this self-image. For practitioners of NE tasks, who identify their professional selves as heroic, they may be more likely to pursue their work in a prosocial fashion. Past studies of NEPs have observed how individual NEPs disassociate from their work roles in response to the emotional and cognitive demands of enacting harm (Clair & Dufresne, 2004; Margolis & Molinsky, 2008). Yet, no empirical research to date has examined the extent to which moral accounting and hero self-identification helps to resolve this dissonance and keep practitioners positively engaged with their work.

An important expression of prosocial conduct that is of utmost importance to NEPs involve their ability to carry out NE tasks with interpersonal sensitivity towards their targets of harm (Margolis & Molinsky, 2008; Molinsky & Margolis, 2005). The ideal NEP is described as an individual who can humanely and compassionately enact harm. This would require committing harms, whilst maintaining sympathy towards one’s targets of harm as a matter of course. Humane NEPs, therefore, must endure a “sufficient” state of distress, rather than suppress their negative emotions altogether (Molinsky & Margolis, 2005, p. 256). However, since most people are motivated to

maintain value-behavior congruence (Aquino & Reed, 2002; Dunning, 2007; Monin & Jordan, 2009; Verplanken & Holland, 2002), some coping efforts mobilized to reduce one's felt distress may lead to psychological distancing (Margolis & Molinsky, 2008) or greater external attribution of responsibility (Andiappan & Duffour, 2017). This would undermine the interpersonal sensitivity that practitioners are expected to showcase towards their harm recipients. Self-identification with the moral archetype of a hero, on the other hand, may help practitioners better navigate this intrapsychic difficulty associated with NE tasks.

By viewing themselves as heroes, this may affirm the prosocial goals of a NEP's work, perhaps even reinforcing their moral commitment to helping others (Fagin-Jones, 2024; Goranson et al., 2022). Accordingly, hero self-identification should increase NEPs' prosocial motivations and reinforce their commitment to helping others through their work, which is predicted to extend to targets of harm as well. At the interpersonal level, this would be demonstrated by practitioners exhibiting a stronger willingness to express compassion towards those they have harmed, as a holistic expression of professionalism. In addition, when NEPs self-image is aligned with being a hero, they may be more willing to facilitate restorative efforts to aid their harm recipients. Further, these prosocial responses by NEPs may also be expressed at a broader level. Since being of service to a broader community is integral to the heroic self-concept (Fagin-Jones, 2024; Franco et al., 2018), this self-schema may strengthen one's commitment to building and maintaining public trust. A dedication to fostering positive public relations and promoting accountability would be evidenced by NEPs' favorable attitudes toward the public and support for transparent work practices (Lee et al., 2024). Altogether, greater

hero self-identification, as elicited by the moral accounting of NE tasks, is predicted to increase practitioners' prosocial responses. See Figure 1.

***Hypothesis 2:** Hero self-identification is positively associated with NEPs' prosocial responses, including prosocial motivation continuance, professional compassion, restitution support, public esteem, and work transparency.*

***Hypothesis 3:** Moral accounting predicts greater prosocial responses through its effect on greater hero self-identification.*

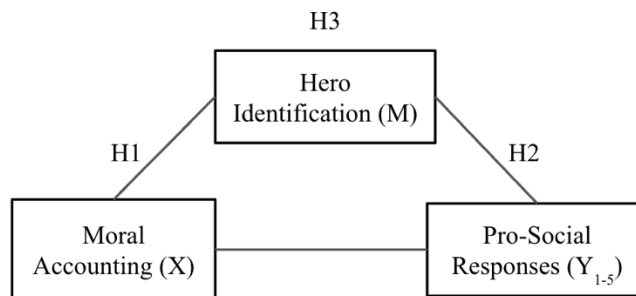


Figure 1. Model depicting H3's proposed relationship between moral accounting and prosocial responses, through the indirect effect of hero identification.

Moral Stigma and Victim Identification

Most research pertaining to NEPs provide detailed accounts regarding the coping responses of having to carry out morally compromising tasks (Molinsky & Margolis, 2005; Margolis & Molinsky, 2008; Shukla et al., 2024). Little attention has been paid to how NEPs must also navigate the morally charged public scrutiny that can emerge from their jobs' association with NE tasks. While these jobs can involve a variety of tasks, the NEs linked to a job's professional requirements may sway public opinion against the

entire occupation. When the actual necessity of NE tasks is brought to question, public criticism can erode the reputation of NEPs at the occupational level and paint their jobs as more “evil than necessary” (Ashforth & Kreiner, 2013, p. 84). Perceived moral stigmatization of their work can damage practitioners’ moral self-image and contribute to reductions in their prosocial behaviors (Patil, 2019; Patil & Lebel, 2019).

Moral Occupational Stigma. Although the direct literature on NEs rarely acknowledges the effect of public disapproval on practitioners’ experiences, research on “dirty work” has examined stigmatized occupations more generally. Public stigma toward occupations can stem from the job’s affiliation with physical uncleanliness, discredited social groups, or immoral activities (Ashforth & Kreiner, 1999; Goffman, 1963; Hughes, 1962). Ashforth and Kreiner (2013) theorized *morally* stigmatized jobs are likely to face the most public disdain. By contrast, jobs that are stigmatized due to their physical uncleanliness are expected to be more indisputably viewed as necessary for the effective functioning of society. This does not exempt them from stigmatization, but certainly far less public outrage. On the other hand, the public’s doubts on the societal usefulness of morally stigmatized occupations may lead them to perceive these jobs as mostly immoral (Ashforth & Kreiner, 2013). Morally stigmatized jobs that Ashforth and Kreiner (2013) classify as befitting the “big E,” i.e., highly evil, are often jobs that challenge society’s longstanding moral standards such as sex workers or black-market vendors. The jobs that connote a “little E” are morally stigmatized because the public may view the occupation as primarily associated with activities involving dishonesty, deceit, or self-gain purposes (e.g., psychics or ambulance chasers).

The dirty work literature examines occupation-wide moral stigma as proceeding from a post hoc assumption that society already views these jobs as disreputable due to their association with morally questionable tasks or goals (Kreiner et al., 2022). On the contrary, the study of NEs has mainly focused its research lens at the task-level to describe a subset of morally compromising responsibilities and experiences found within a job (Molinsky & Margolis, 2005). However, since NEs are more broadly represented in certain jobs, research on NEs do often rely on a sample of subjects that are employed in specific occupations (e.g., human resource managers, police officers, judges, soldiers). In short, the jobs studied to investigate NE-related experiences are not already presumed to be stigmatized. Another important difference is that unlike jobs stigmatized especially with the “big E,” the “necessary” stipulation of NEs implies there is some social agreement about the value of performing these tasks, in terms of fulfilling societal needs (Molinsky & Margolis, 2005).

Nevertheless, it is the case that NE tasks can still invite public scrutiny and result in occupational stigma, especially given the evolving normative constraints and social construction of moral standards (Fineman, 2006). Constancy in the social approval of NE tasks is likely tenuous (Ashforth & Kreiner, 2013). As public evaluation over the “necessity” of NEs inevitably change, growing public disapproval over a NE task can lead to the moral stigmatization of the profession that is most affiliated with conducting these tasks. Due to the inconsistency and instability of socio-cultural norms surrounding ethics and moral standards, jobs involving NEs are likely to attract “pointed and persistent derogation from at least certain segments of society” (Ashforth & Kreiner, 2013, p. 93). For instance, the action of soldiers killing enemy combatants can be viewed

as more unnecessarily cruel at different points in time across different social audiences. Such differing views over this NE task can result in public attitudes that either heroize or morally stigmatize the military profession.

Numerous studies have explored how stigmatized workers cope with public disapproval, often examining these dynamics through a group-level lens that is founded on a shared stigmatized identity (Ashforth & Kreiner, 1999; Bentein et al., 2017; Kreiner et al., 2022). A rich source of ethnographies as well as both theoretical and empirical work on occupational stigma have observed ways in which stigmatized occupational groups enhance their collective self-worth. For example, Ashforth and Kreiner (2013) found that morally stigmatized professional groups are especially likely to adopt an occupational ideology that frames their work as a vital service to society. This ideological stance closely resembles heroic self-identification. Alternatively, the present research examines if a hero narrative, which has been positioned by the occupational stigma literature as a group-level coping mechanism in response to social stigma, may independently emerge from individual sense-making processes as NEPs undergo moral accounting. Even absent social stigma information or a strongly held occupational identification, individual NEPs may self-identify as heroes in the process of constructing a positive self-image while making sense of their jobs at the task level.

While NEPs engage in private deliberations to make sense of their work tasks in a way that enhances their positive self-view, especially as ethical actors, stigma has the power to interfere with their self-definitions (Parmar, 2014; Wyatt et al., 2021). As discussed earlier, since NEs involve enacting harm, NEPs' jobs can become subject to moral occupational stigma. When these individuals encounter salient displays of public

disapproval, this experience can trigger heightened awareness of a stigmatized self (Eriksson, 2019; Major & O'Brien, 2005). Stigmas are context-specific and socially experienced, thus, can be viewed as a phenomenon that presents itself through disruptions in one's daily preoccupations, e.g., "I come to think about my minority status the moment I hear some racial slur dropped by a stranger I pass on the street" (Eriksson, 2019, p. 397). Thus, a reprobation by an unfamiliar entity about one's ethical inconsistencies or moral failures initiates stigma awareness and furthers an individual's sense-making processes (see Ashforth et al., 2008; Parmar, 2014; Schaubroeck et al., 2018). In general, social mistreatment leads people to try to understand why they are being mistreated (Andersson & Pearson, 1999); and formulate an interpretation of this negative social interaction (Debono et al., 2020). This research on NEPs, unlike many past studies on stigma management, aims to account for the introspective processes of workers that precede these salient stigma consciousness experiences. When NEPs face morally-charged hostility from others, their prior *self*-considerations and *other*-expectations will likely inform their processing of these social criticisms.

This negative feedback, which instantiates a NEP's moral stigma awareness, would certainly contend with their self-identification as a hero. Both individuals and collective groups across multiple studies have reported a higher preference for being viewed as moral over other positive attributes such as being seen as competent or sociable (Ellemers & de Gilder, 2023; Leach et al., 2007; Tappin & McKay, 2017; Pagliaro et al., 2016). A favorable moral self-image is an essential component of an individual's self-worth (Dunning, 2007). Moral condemnation, therefore, provokes an especially damaging threat to a person's sense of positive self-image and social worth

(Barkan et al., 2012; Gausel & Leach, 2011; Wenzel et al., 2020). As NEPs are ultimately socially embedded actors (Becker, 1971), their self-identification as heroes will likely be influenced by their meta-perceptions and expectations concerning how others perceive them (Cooley, 1902; Mead, 1934). For NEPs experiencing these instances of moral stigma, the maintenance of their hero-based identities may be susceptible to disruption (Chopova & Ellemers, 2023). Negative outside feedback communicates to NEPs that the harms they perform are not justifiable, but rather, condemnable. Indeed, the public's reprobation in the form of moral stigma is often provocatively communicated by dramatizing the harms committed (Shulman et al., 2021; Snow & Benford, 1988), which clearly sends the message that one is viewed by members of the public as villains. An important finding from work by Goranson and colleagues (2022) illustrates that people would rather see themselves as victims over being seen as villains. Therefore, NEPs' self-identification as a victim may arise due to people's strong resistance to being villainized.

***Hypothesis 4:** Hero self-identification will interact with perceived moral stigma to predict greater victim identification, such that NEPs with strong hero identification will report greater victim identification under high (vs. low) moral stigma.*

This identification as a victim, in avoidance of viewing oneself as a villain, can lead to a psychological state that emphasizes one's own personal suffering (Gray & Kubin, 2024). In correspondence with this increased focus on one's own suffering, victim self-identification has been linked with decreased recognition and awareness of one's

role in causing harm. As NEPs self-identify as victims, this may prompt lower prosociality, as reflected by greater antisocial responses found in their work conduct and how they interface with the public. In contrast to hero self-identification, victim self-identification with its heightened focus on the mistreated self may reduce the prosocial motivations of NE practitioners (Chopova & Ellemers, 2022; Gray & Kubin, 2024; Patil, 2019). Rather than expressing compassion towards targets of harm, practitioners may view compassion as unprofessional and reduce their restitution support for their harm recipients (Ashforth & Kreiner, 2013; Gray & Kubin, 2024). As prior studies on antisocial responses to social threats have shown, this dynamic can lead to greater denigration towards the public and prompt less commitment to transparent work practices (Gausel et al., 2012). This failure to appropriately balance performing harm with prosocial intent can lead to social reprobation, public outcry, and intensify occupational stigma (Patil, 2019; Patil & Lebel, 2019).

***Hypothesis 5:** Victim identification is associated with less prosocial responses, including diminished prosocial motivation continuance, less compassion and restitution support towards targets of harm, greater public denigration, and less transparency regarding one's work practices.*

***Hypothesis 6:** The indirect effect of moral accounting on prosocial responses via hero and victim identification will be conditional on moral stigma. Specifically, when perceived stigma is high, moral accounting will indirectly predict reduced prosocial*

responses through the sequential path from hero to victim identification. This model is illustrated in Figure 2.

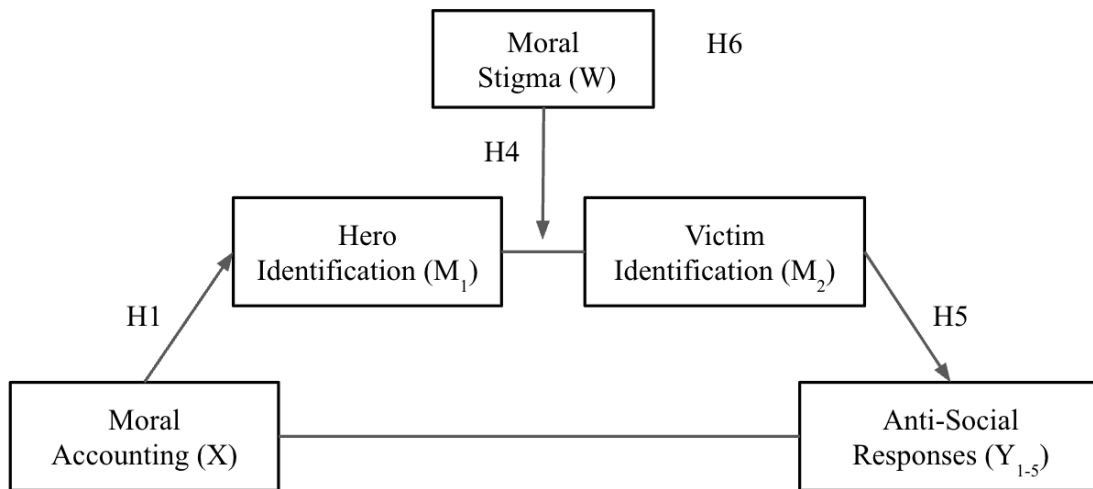


Figure 2. Conceptual framework of conditional serial mediation model. This model depicts the predicted effects of moral accounting on NEPs’ anti-social responses via hero and victim identification, as moderated by perceived moral stigma (H6).

Summary of Hypotheses

Hypothesis 1: *Moral accounting is positively associated with NEPs' hero self-identification.*

Hypothesis 2: *Hero self-identification is positively associated with NEPs' prosocial responses, including professional compassion, restitution support, public esteem, work transparency, and prosocial motivation continuance.*

Hypothesis 3: Moral accounting predicts greater prosocial responses through its effect on greater hero self-identification.

Hypothesis 4: Hero self-identification will interact with perceived moral stigma to predict greater victim identification, such that NEPs with strong hero identification will report greater victim identification under high (vs. low) moral stigma.

Hypothesis 5: Victim identification is associated with less prosocial responses, including diminished prosocial motivation continuance, less compassion and restitution support towards targets of harm, greater public denigration, and less transparency regarding one's work practices.

Hypothesis 6: The indirect effect of moral accounting on prosocial responses via hero and victim identification will be conditional on moral stigma. Specifically, when perceived stigma is high, moral accounting will indirectly predict reduced prosocial responses through the sequential path from hero to victim identification.

Overview of Present Research

The present research investigated the proposed hypotheses across three studies, which aimed to capture the consequences of NEPs' moral accounting processes, as they may relate to practitioners' self-identifications as heroes or victims. Studies 1 and 2 employed experimental vignette designs in which participants imagined themselves as social workers engaged in a NE task. Study 1 manipulated the opportunity for moral

accounting to test its association with hero identification and explored potential mediators that might lead to this moral self-evaluation as a hero. Study 2 examined whether hero identification promoted greater prosocial responses and whether perceived moral stigma moderated the process by which hero self-identification gave rise to victim self-identification, ultimately contributing to practitioner's reduced prosocial responses. Study 3 recruited a sample of real NEPs from a broad spectrum of occupations to test all the proposed hypotheses, in hopes of enhancing the ecological validity of this research as well as strengthen its generalizability. All three studies were pre-registered through AsPredicted.org.

Study 1

The purpose of Study 1 was to test the hypothesis that following the enactment of a NE task, moral accounting would be positively associated with an individual's moral self-identification as a hero (H1). With the use of an experimental vignette, participants were instructed to mentally place themselves in the identity of a particular social worker, who has been tasked to remove a child from a neglectful home. Both the occupation and vignette presented in this study previously underwent piloting, which included a battery of construct validity checks and feedback on its perceived realism (see Appendix B). All participants immersed themselves in this NE task. Afterwards, all participants were provided with an opportunity to morally account for their actions, while others were not. For exploratory purposes, the indirect influence of state self-esteem, moral self-efficacy, and moral superiority was assessed to gain further insights into the potential relationship between moral accounting and hero identification.

Method

Participants. Participants were invited to take part in this study from the online research recruitment platform Prolific. For both Study 1 and Study 2, the eligibility criteria for the recruited participants were employed U.S. residents over the age of 18, fluent in English, and with no prior or current work experience as social workers. For Study 1, of the three hundred participants invited to participate in the study, eight participants were excluded for failing attention checks. As part of another exclusion criteria, the conclusion of this study incorporated Stanley et al. (2019)'s self-assessment that asks participants how focused they felt while taking the study. To promote truthful reporting, this measure began with a message assuring subjects that they will not be penalized for their responses. None of the participants reported feeling distracted during Study 1. Thus, the final sample consisted of 292 participants ($M_{\text{age}} = 38.18$, $SD = 11.55$; 56.16% female; 75% White).

Procedure. This study applied a one-factor (moral accounting vs. no moral accounting), between-subjects experimental design. All participants read a vignette, which instructed them to imagine themselves as social workers and immerse themselves in a NE-relevant work situation. In this work situation, they were tasked to authorize separating a child from a neglectful parent. Following this authorization, participants were randomly assigned to one of two conditions: the individuals in the moral accounting condition were asked to reflect on the beneficial consequences of their actions as the hypothetical social worker before proceeding with the rest of the study ($n = 144$), while the other half of the study's sample advanced directly to the other survey measures without receiving this reflection prompt ($n = 148$).

Necessary evil scenario. Participants engaged in a simulated work scenario that involved a NE task relevant to the responsibility of certain social workers. The complete transcript of this vignette can be found in Appendix A1. In this scenario, participants were placed in the position of a social worker, who were assigned to investigate a reported child neglect case and conduct a follow-up home visit. Upon witnessing signs of neglect during their visit, such as lack of running water in the house, participants encountered a decision checkpoint—to proceed or to not proceed with taking protective custody of the neglected child. If participants selected to not proceed with this action (14.73%), they were shown a message stating they could take more time to reflect before proceeding with the necessary child removal order. This structured decision flow was included to stress participants’ direct involvement and evaluate their discomfort with the execution of this task. After participants proceeded with the removal of the child, the vignette ended with an excerpt about witnessing the distress of both parent and child as the child was being taken away by law enforcement. Appendix B details the pilot studies and vignette development procedure.

Moral accounting. After participants completed this scenario, subjects in the control condition bypassed a moral accounting prompt and proceeded to the next set of measures in the study. The participants assigned to the moral accounting condition, alternatively, were asked to write a short essay given the following prompt-

Despite its challenges, what are some benefits (immediate or more general) that may come out of this situation? Explain why your response to this situation was the right thing to do within your professional capacity.

As a manipulation check, all participants rated the extent to which they felt their actions as the social worker in the vignette would produce a positive impact (i.e., “I believe my actions in this case may have a positive effect with beneficial outcomes”; 1 = *strongly disagree*, 7 = *strongly agree*; $M = 5.42$; $SD = 1.21$).

Measures

Hero identification. To measure hero identification, a 5-item scale was adapted from Kinsella et al. (2015)’s work identifying lay perceptions of a hero. Using deductive and inductive approaches, Kinsella and colleagues determined multiple functions associated with the public’s prototypical idea of a hero. Five of these functions (“to help,” “to save,” “to protect,” “to make the world better,” and “to do what no one else will do”) were found to be most attributable to heroes, in comparison to other related figures such as leaders or role models. These functions served as the foundation for the items found in the present study’s hero identification measure. In particular, the measure’s stem (“In the role of this social worker, I would identify myself as someone who...”) was accompanied by the following five statements: “...helps others,” “...saves others in need,” “...protects others in need,” “...makes the world a better place,” and “...steps up and does what no one else will do.” Participants answered their agreement to these statements with a scale that ranged from 1 (*not at all*) to 7 (*very much so*) ($\alpha = 0.92$; $M = 5.89$; $SD = 1.04$).

State self-esteem. The state version of participants’ self-esteem was measured with a 4-item scale adapted from Heatherton and Polivy (1991). Each item followed the stem, “As the social worker in this story...” The four items were, “I feel like the kind of person who is respected by others;” “I feel like the kind of person who is admired by

others;” “I feel good about myself;” and “I feel like a worthy person.” Participants rated each item on a 7-point scale (1 = *not at all*, 7 = *very much so*; $\alpha = 0.92$; $M = 4.34$; $SD = 1.43$).

Moral self-efficacy. Three items measured participants’ moral self-efficacy (Owens et al., 2019; see also Spreitzer, 1995). They followed the stem, “Imagining myself in the role of this social worker...” The statements were “I feel like someone who is confident in my ability to perform this work in a way that meets the profession's ethical standards,” “I feel like someone who is self-assured about my capabilities to perform this work in an ethical manner,” and “I feel like someone who is an expert on the ethical regulations and skills necessary to do this job.” Participants rated their agreement with these three statements using a 7-point scale (1 = *not at all*, 7 = *very much so*; $\alpha = 0.91$; $M = 5.75$; $SD = 1.14$).

Moral superiority. This 4-item measure was informed by Dong et al. (2019)’s conceptualization of moral superiority as a self-evaluation of being better than average others, specifically with regards to one’s morals and ethics. In accordance with this conceptual approach, each statement for this study’s measure of moral superiority followed the stem, “Compared to the average working adult, I believe my experiences as this social worker would prepare me to be...” The specific items, adapted from Bandura (1990)’s self-efficacy scale but modified to reflect a focus on moral judgements and behaviors, included: “More willing to take the toll of dealing with a morally challenging situation,” “more helpful in a morally challenging situation,” “more competent at handling a morally challenging situation in the best manner possible,” and “more capable

of making difficult calls regarding ethical dilemmas.” Responses ranged from 1 (*strongly disagree*) to 7 (*strongly agree*) ($\alpha = 0.92$; $M = 5.63$; $SD = 1.13$).

Control variables. The study’s control variables included: age, gender (1 = female, 0 = male), race (1 = White, 0 = Not White), and political orientation (1 = *very liberal*, 7 = *very conservative*; $M = 3.55$, $SD = 1.83$). Past research has demonstrated an important link between these variables and moral reasoning processes (Graham et al., 2009; Haidt et al., 1993).

In addition, the study incorporated two more control variables—*deontic orientation* and *moral normlessness*—to measure participant’s moral standards. These variables were included as they could potentially be primary drivers of an individual’s heroic self-identification, as opposed to the experimental manipulation. Accounting for these dispositional factors would help to ensure any observed effects were not primarily attributable to pre-existing moral orientations. Originally, to measure deontic orientation, the study included four items from a measure by Radtke (2008) on individuals’ adherence to a more deontological moral framework, such as categorically prohibiting harm towards others (e.g., “A person should always avoid harming others in any way”). Respondents answered their agreement with statements using a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*; $M = 6.05$; $SD = 0.73$). However, this measure, due to its poor reliability ($\alpha = 0.62$), was omitted from the present research to maintain the interpretive integrity of the study’s results. In its place, deontic orientation was derived from the hesitance exhibited by participants before proceeding with the child removal order. While some participants chose to immediately take action at this decision checkpoint, others selected to initially not proceed. For the latter set of participants, the

following screen told them to take their time until they were able to execute the recommended action. Participants' discomfort with carrying out the NE task was applied as an indicator of their deontic orientation, specifically, the duration of their hesitance in seconds ($M = 48.69$, $SD = 137.43$). This determination is consistent with prior research that has applied participants' reluctance toward causing harm as a measure of their congruence with a deontic moral orientation in similar decision-making paradigms (Bartels & Medin, 2007; Uhlmann et al., 2009).

Lastly, a three-item measure for *moral normlessness* by Kohn (1976) was included as a covariate. This is a measure that evaluates the degree to which individuals personally condone socially deviant actions in pursuit of self-serving goals. Sample items included: "It's all right to get around the law as long as you don't actually break it" and "If something works, it doesn't matter whether it's right or wrong." This scale was incorporated due to its significant negative relationship found with the Self-Importance of Moral Identity Scale (SIMIS; Crocker et al., 2003; Aquino & Reed, 2002), which measures the extent to which individuals value being identified as a moral person. Moral normlessness served to control for the self-importance of one's moral identity, whilst addressing the potential social desirability bias that has been observed with skewed responses to the SIMIS (Aquino & Reed, 2002; Hall & Derryberry, 2010). The moral normlessness measure was rated on a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*; $\alpha = 0.81$; $M = 2.75$; $SD = 1.14$).

Results

Table 1 provides the correlations among all variables in this study.

As a manipulation check, participants rated the extent to which they believed their handling of the case as the social worker in the vignette would yield beneficial effects. Indeed, participants' perceptions of having a positive impact significantly differed by condition, $b = 0.34$, $SE = 0.14$, $t(287) = -2.41$, $p < 0.05$, $d = 0.28$. In the moral accounting condition, perceptions of one's positive impact were higher ($M = 5.59$, $SD = 1.13$) than in the control condition where participants were not prompted to moral account for their action in the given scenario ($M = 5.25$, $SD = 1.28$). This indicated that the moral accounting reflection prompt was an effective manipulation.

Next, a regression analysis testing Hypothesis 1 examined the effect of moral accounting (1, no moral accounting = 0) on hero identification. As predicted, individuals who underwent moral accounting expressed greater self-identification as heroes ($M = 6.04$, $SD = 0.90$), relative to the referent group that did not engage in moral accounting ($M = 5.74$, $SD = 1.14$; $b = 0.30$; $SE = 0.12$, $t(278) = -2.48$, $p < 0.05$, $d = 0.29$). Hero identification was then regressed onto moral accounting, with the inclusion of the following control variables: age, gender, race, political orientation, deontic orientation, and moral normlessness. Results from this analysis indicated that the effect of moral accounting on hero identification remained statistically significant, even after accounting for these critical covariates, $b = 0.27$; $SE = 0.12$, $p < 0.05$, $d = 0.27$. A full summary of these results can be found in Table 2.

Exploratory analyses. An exploratory set of mediation analyses was subsequently conducted to investigate the psychological mechanisms that could explain the positive relationship between moral accounting and hero identification. The three mediators of interest were state self-esteem, moral self-efficacy, and moral superiority.

To explore these indirect effects, mediation models were estimated (see Hayes, 2013; PROCESS Model 4), which used maximum likelihood with bootstrapped standard errors (5,000 samples). All mediation analyses included the following covariates across all regression paths: age, gender, race, politics, deontic orientation, and moral normlessness. The results of these analyses showed a significant indirect effect of moral accounting on hero identification via state self-esteem, $\beta = 0.07$; $SE = 0.05$, $p < 0.01$, 95% CI [0.05, 0.25]. Moral accounting significantly predicted state self-esteem ($\beta = 0.18$, $p < 0.01$), and state self-esteem significantly predicted hero self-identification ($\beta = .39$, $p < 0.001$). However, while the total effect was significant ($p < 0.05$), the direct effect of moral accounting on hero identification was not statistically significant ($p = 0.26$), suggesting this link between moral accounting and hero self-identification was fully mediated by state self-esteem. None of the other exploratory mediators (moral self-efficacy and moral superiority) yielded a significant indirect effect on hero identification.

Discussion

The findings of this first study showed that prompting individuals, following a NE task, to engage in moral accounting heightened their moral self-identification as heroes. Supporting Hypothesis 1, subjects in the moral accounting condition reported significantly higher levels of hero identification compared to their counterparts in the referent group. The effect remained significant after adjusting for demographic covariates and controls regarding participants' moral self-perceptions.

Exploratory mediation analyses further suggested that moral accounting's effect on hero identification may be partially explained by enhanced state self-esteem. Specifically, engaging in moral accounting improved individuals' feelings of self-worth,

which in turn predicted their stronger moral identification with a heroic self-image of themselves. Neither moral self-efficacy nor moral superiority emerged as significant mediators, indicating that the observed effect may be more closely tied to a positive self-regard than perceived gains in ethical competence or moral exceptionalism. These results provide initial support that moral accounting serves as a meaningful cognitive tool, which can enable NE practitioners to reconcile their morally difficult tasks with a more positive self-evaluation to procure an overall heroic moral identity.

Study 2

This next study employed the same vignette in Study 1, whilst extending the study procedure to include a moral stigma manipulation, as well as measures that assessed individuals' prosocial responses. Study 2 aimed to replicate the positive association found between moral accounting and greater hero self-identification in Study 1. Moreover, the present study examined whether moral accounting predicted NEPs' prosocial attitudes and behaviors, via greater hero self-identification. It also included measures of victimhood and prosocial behaviors to test the theorized framework delineated by Hypotheses 4, 5, and 6.

Method

Participants. Study 2's participants were recruited online through Prolific. Using the same parameters for participant recruitment as Study 1, the study invited 400 subjects to complete the study, excluding 26 participants who failed attention checks and one participant who indicated they felt distracted while taking the survey (see Stanley et al., 2019). The remaining final sample for Study 2 totaled 373 participants ($M_{\text{age}} = 40.70$, $SD = 12.01$; 58.44% female; 71.58% White).

Procedure. Participants were randomly assigned to a 2 (moral accounting, no moral accounting) X 2 (stigma-present, stigma-absent) between-subjects experimental study. As with Study 1, all participants first engaged with the same child neglect scenario in the role of a social worker. Then, subjects in the moral accounting condition ($n = 187$) responded to the essay prompt that was used in the first study, while those in the control condition ($n = 186$) bypassed this prompt. All participants next answered the same state self-esteem scale ($M = 5.33$; $SD = 1.30$) and manipulation check question from the first study—rating the extent to which they felt their actions as the social worker in the vignette would produce a positive impact (1 = *strongly disagree*, 7 = *strongly agree*; $M = 5.33$; $SD = 1.30$).

Next, participants were exposed to the moral stigma manipulation (see Appendix A2). While continuing to imagine themselves as the social worker in the vignette, participants read about coming across a flyer displayed at their workplace and a public opinion poll they came across. In the stigma condition ($n = 187$), both the flyer and poll conveyed the message that the public held unfavorable and distrustful views toward social workers. In contrast, the control condition ($n = 186$) featured neutral content, including a flyer that detailed a parking lot construction notice and a poll about America's least favorite car colors. The manipulation checks associated with Study 2's stigma manipulation can be found in the Results section. After completing this portion of the study, all participants responded to the rest of the study's measures and provided their demographic data.

Measures

Hero and victim identification. Study 2 applied the same 5-item measure for hero identification ($\alpha = 0.96$; $M = 5.98$; $SD = 1.05$) from the first study. The victim identification measure was adapted from Thai et al. (2023)'s single-item measure of perceived victimhood. Participants rated the following statement: "In the role of this social worker, I would identify myself as someone who is a *victim* of unfair treatment." Ratings ranged from 1 (*not at all*) to 7 (*very much so*) ($M = 4.13$; $SD = 1.62$).

Moral stigma. Moral stigma was measured using Shnabel and Nadler (2008)'s public moral image scale. To better align this measure with the present study, the stem was revised to specifically reflect public opinion towards social workers ("The public seems to generally perceive social workers as..."). This revised stem was preceded by the following three descriptives: "immoral," "unethical," and "behaving inappropriately." These items were rated on a 7-point scale (1 = *not at all*, 7 = *very much so*; $\alpha = 0.96$; $M = 4.15$; $SD = 1.74$), which was measured in this study as a manipulation check.

State self-esteem. This was measured with the 4-item scale applied in Study 1 ($\alpha = 0.93$; $M = 4.44$; $SD = 1.49$).

Professional compassion. This 4-item scale was reverse-coded from the measure of professional distancing was informed by Molinsky et al. (2012)'s study that measured employees' withholding of compassion in work situations that may call for interpersonal sensitivity—with the excuse to maintain professionalism. Specifically, they measured participants' perceptions of unprofessionalism with the idea of writing a compassionate letter to an employee that the participant had recently fired. This scenario relates well with the present research's focus on NEPs' maintenance of interpersonal sensitivity when executing necessary evil tasks. Thus, Study 2 applied the same scenario to measure

professional distancing, and representing a reduction in interpersonal sensitivity towards one's target of harm. Participants in the present study rated their support for writing an empathetic letter to the parent of the child from the social worker vignette ("How would you feel about sending a personal message to [the parent] that would express compassion for her situation?"). Participants read the following stem, "As the social worker in this story, I believe sending [the parent] a personal message like that would...", preceded by these statements: "make me look foolish," "be seen as unprofessional," "reflect poorly on me," and "make me appear unpolished." Ratings for these items were presented on a scale that ranged from 1 (*not at all*) to 7 (*very much so*) ($\alpha = 0.95$; $M = 3.52$; $SD = 1.97$). Professional distancing, which is a measure of one of the hypothesized anti-social responses of NEPs in this research, was reverse-coded to represent the opposing prosocial response of professional compassion.

Restitution support. As a promising indicator of behavioral intent, restitution support was assessed by participants' willingness to direct financial aid toward the parent in the social worker vignette, i.e., their target of harm ($M = 2822.25$; $SD = 2704.85$). Participants were instructed to remain in the mindset of the social worker and read the following prompt: "Your agency has \$10,000 remaining from last quarter's discretionary expenses. Given the following choices, how would you recommend these extra funds be allocated?" Their allocation options for the \$10,000 were (*i.*) donations to help [the parent] in the social worker scenario or (*ii.*) funds providing more resources for employee(s) at their social work agency. Participants could recommend dividing up these funds or fully donating the sum of money to just one of the two funds. This presentation of options (displayed in random order) was selected to ensure this measure of restitution

support reflected participants' willingness and their prioritization of offering financial help towards their target of harm, over the chance of supporting other positive initiatives.

Public esteem. Public denigration was adapted from Cho and Fast (2012)'s 8-item measure of denigration. The measure was modified to shift the target of denigration to one's felt esteem towards the general public. Participants rated the degree to which they believed most members of the public were competent, intelligent, capable of good judgment, likable, incompetent (reverse-scored), simple-minded (reverse-scored), unlikable (reverse-scored), and lacking in good judgment (reverse-scored). Ratings applied a 7-point scale (1 = *not at all*, 7 = *very much so*). These items were averaged to create the public esteem measure ($\alpha = 0.94$; $M = 3.76$; $SD = 1.14$).

Work transparency. Endorsement of greater work transparency, in relation to the public, was measured using one item informed by Gausel et al. (2012)'s cover-up measure on anti-social reactions to social threat. Using a 7-point rating system (1 = *not at all*, 7 = *very much so*; $M = 6.10$; $SD = 1.24$), participants answered their agreement with the statement, "As the social worker in this story, I would support total public transparency about my job's work practices."

Prosocial motivation continuance. This 3-item measure, which came toward the end of the study, aimed to capture participants' motivation to take prosocial actions after taking part in the present study's role-play scenario. Adapted from Aquino et al. (2011)'s measure, participants were prompted with the statement, "As a result of putting yourself 'in the shoes' of the social worker in this story, indicate how much more or less you feel motivated to engage in the following behaviors right now." The action items that participants rated on a scale from 1 (*not at all*) to 7 (*very much so*) were: "I feel like

doing something good for others,” “I feel like helping someone in need,” and “I feel like caring for people less fortunate than me” ($\alpha = 0.97$; $M = 5.47$; $SD = 1.41$).

Control variables. The following demographic measures were included as covariates: age, gender (1 = female, 0 = male), race (1 = White, 0 = Not White), political orientation (1 = *very liberal*, 7 = *very conservative*; $M = 3.56$, $SD = 1.91$), and religiosity (1 = *not at all religious*, 7 = *very religious*; $M = 3.46$, $SD = 2.28$). Prior literature has demonstrated that each of these demographic variables share a meaningful association with human moral processing (Côté et al., 2013; Graham et al., 2009; Haidt et al., 1993; Skitka et al., 2009). In addition, participants’ deontic moral orientation was measured ($M = 86.10$, $SD = 55.35$), based on the duration of their hesitancy (in seconds) during the study’s decision checkpoint. As in Study 1, deontic orientation was computed based on the recorded deliberation time of participants who hesitated before proceeding with the child removal order, while participants who immediately chose to complete the NE task were assigned a value of 0 seconds.

Results

Manipulation checks. In evaluating their handling of the child neglect case, participants in the moral accounting condition ($M = 5.66$, $SD = 1.13$) reported greater perceived positive impact than those in the control condition ($M = 4.99$, $SD = 1.33$; $b = 0.67$, $SE = 0.13$, $t(364) = -5.14$, $p < 0.001$, $d = 0.53$). For the moral stigma manipulation, the study incorporated two manipulation checks. First, right after participants viewed the flyers and polls depicting public opinion of social workers (or car colors for the stigma-absent condition), participants were asked to rate the extent to which the public seemed to view social workers as trustworthy (1 = *not at all*, 7 = *very much so*). We found

support for the effectiveness of the stigma manipulation, $b = -2.37$, $SE = 0.12$, $t(267) = 20.39$, $p < 0.001$, $d = 2.11$). Results showed significant group differences between participants who were in the stigma-present condition ($M = 1.98$, $SD = 0.70$) versus the stigma-absent condition ($M = 4.35$, $SD = 1.42$) in their ratings of perceived trustworthiness by the public towards social workers. To confirm the persistence of the stigma manipulation, participants completed another manipulation check with the moral stigma measure about the extent to which the public morally stigmatizes social workers (see Measures section). Significant differences were found between the stigma-present condition ($M = 5.06$, $SD = 1.32$) and the stigma-absent condition ($M = 3.24$, $SD = 1.64$; $b = 1.82$, $SE = 0.15$, $t(353) = -11.83$, $p < 0.001$, $d = 1.23$).

Hypotheses testing. Table 3 provides the correlations among all main variables Study 2. As a first step, the positive relationship between moral accounting and hero identification (HI), which was observed in Study 1, was not replicated in Study 2, as indicated by a non-significant t -test ($p = 0.12$). However, results from a mediation analysis as with Study 1, found a significant indirect effect of moral accounting predicting greater hero self-identification via higher state self-esteem, $\beta = 0.15$; $SE = 0.07$, $p < 0.001$, 95% CI [0.20, 0.45]. This model controlled for all covariates mentioned in the Measures section. The direct effect of moral accounting predicting greater hero self-identification was marginally significant ($p = 0.054$), but state self-esteem was significantly predictive of hero self-identification ($\beta = 0.27$; $SE = 0.05$, $p < 0.001$). The total effect was not significant, $\beta = 0.06$; $SE = 0.11$, $p = 0.21$, 95% CI [-0.10, 0.34].

Without empirical support for Hypotheses 1, there was no statistical justification for conducting the planned mediation analyses ($H3$). However, as supplementary

analyses relevant to Hypothesis 2, OLS regression analyses were carried out to assess positive links between hero identification and the prosocial responses of interest (i.e., professional compassion, restitution support, public esteem, work transparency, and prosocial motivation continuance). All these regression models controlled for age, gender, race, political orientation, religiosity, and deontic orientation. It is worth noting that interpretation of the following results should take into account that moral accounting is not a significant driver of hero self-identification in these analyses.

Results revealed hero identification significantly predicted both higher support for public transparency with regards to work practices ($b = 0.14$, $SE = 0.06$, $p < 0.05$) and stronger prosocial motivation continuance ($b = 0.37$, $SE = 0.06$, $p < 0.001$). However, hero identification was found to be unrelated to professional compassion ($p = 0.24$) or greater esteem felt toward the public ($p = 0.12$). Contrary to expectations, a significant negative relationship emerged between hero identification and restitution support ($b = -444.94$, $SE = 131.94$, $p < 0.001$), as hero-identifying participants were less likely to recommend providing financial assistance towards the parent in the social worker scenario. Instead, participants with greater hero identification favored directing their funds toward their hypothetical work colleagues, over their target of harm.

Next, results of a moderation analysis showed no empirical support for the hypothesized two-way interaction between hero identification and moral stigma as a predictor of greater victim identification ($p = 0.50$). As this interaction was not established (H4), no further tests of the theoretical model were conducted.

Discussion

Study 2 did not yield evidence supporting the hypothesized relationships under investigation. This lack of empirical support likely reflects several limitations with the study's experimental materials, among other possible contributing factors. Although the first study demonstrated significant differences in hero identification as a function of moral accounting—an effect that was central to the theoretical framework of study's hypotheses—this pattern failed to emerge in Study 2. Given the smaller sample size of Study 1, it was not likely a problem with Study 2's power. The inability to reproduce this prior finding using the same manipulation materials underscores the value of replication efforts, particularly with research that attempts to experimentally induce complex and nuanced experiences such as those of NEPs.

Another clear and readily identifiable weakness of the present study is the inherent limitations of conducting experimental role-based studies via online platforms. While participants were asked to imagine themselves as a social worker, it is unlikely that all were able to fully get themselves into the mindset of completing this emotionally distressing NE task through a hypothetical, text-based medium. In addition, unlike field studies or lab-based role plays, online environments may not sufficiently facilitate the psychological realism or accountability checks needed to elicit the intended emotional and cognitive engagement required of this research.

Furthermore, even though manipulation checks confirmed that participants in the moral justification group (vs. the control group) were more likely to report that they felt their actions left a positive impact, this manipulation check question may not have been indicative of the psychological conditions that equate moral accounting. Indeed, participants may all naturally engage in some degree of self-justification, without needing

a prompt that instructs them to do so. That is, given people's general motivation to maintain a positive self-concept (Aquino & Reed, 2002), participants may have spontaneously engaged in moral reasoning to rationalize their actions, thereby, attenuating the intended contrast between the control and experimental conditions. If most individuals are inclined to naturally engage in some level of moral accounting, the extent to which they internalize their justifications can likely vary—suggesting that this study may have been better suited to assess degree, rather than the mere presence versus absence of moral accounting. In short, this deeper personalization of one's moral accounting may have been better captured by a continuous rating scale, rather than a two-cell experimental manipulation. In addition, the exploratory analyses from Study 1 provided preliminary evidence that the association between moral accounting and hero identification may be fully mediated by state self-esteem. This was also found to be case in Study 2.

Study 3

Study 3 shifted away from the vignette-based experimental approach found in the two prior studies to an observational approach. The primary objective of incorporating this third study was to strengthen the ecological validity of this research through a cross-sectional survey of real-world practitioners who perform NE tasks. In capturing the self-reports of employees with direct exposure to NEs, the psychological and moral processes central to this research could be evaluated with reduced dependence on one's capacity for imagination. Alongside this methodological shift, Study 3 was able to fully test the hypothesized paths concerning the moral accounting processes of NEPs and their possible outcomes, illustrated in the proposed theoretical model (Figures 1 and 2).

Method

Participants and Procedure. Participants in Study 3 were working adults who indicated that they perform NEs in their current jobs. They were recruited online through a screening process via Prolific, wherein prospective survey-takers initially responded to the following pre-programmed question: “In your current job, are there times when you are tasked to do something that causes emotional or physical harm/pain to others—but it’s something you must do as part of your work responsibilities and it’s necessary in some beneficial way?” (1 = *never*, 2 = *rarely*, 3 = *sometimes*, 4 = *frequently*, 5 = *very often*). The wording of this question, which describes necessary evils without explicit use of its terminology, was directly adapted from Margolis and Molinsky (2008)’s qualitative study exploring the construct space of necessary evils through interviews with NEPs. The Prolific users who answered 3 or above to this screener question and met Study 3’s other eligibility criteria (i.e., over 18 years of age, employed, resided in the U.S., and fluent in English), were provided access to participate in this study. Through this screening process, 302 participants took part in Study 3. Four participants were excluded for answering that they felt distracted while taking the study via Stanley et al. (2019)’s self-assessment questionnaire. In addition, participants’ responses to an open-ended question asking them to describe the NE tasks they performed at work were content-coded for relevance and comprehension. Thirteen responses were excluded as they raised concerns about screener comprehension and attention; for example, responses describing work conditions that led to personal harm rather than instances in which they caused harm to others.

As a result, the final sample was composed of 285 participants ($M_{\text{age}} = 40.54$, $SD = 10.49$; 48.42% female; 69.12% White). The study followed a straightforward survey procedure, which included measures related to NEs, perceived public opinion toward participants' actual occupations, as well as questions capturing their demographic information. Participants also completed the same measures of prosocial responses used in Study 2, with survey items that inquired them to consider how they would act in the context of their current job roles.

Measures

Many of the same measures from the previous study were incorporated in Study 3, with minor modifications to anchor responses to participants' own perceptions and real-life occupational roles. For moral accounting, we asked participants to rate their agreement on two statements ("I believe I am serving a greater good and/or greater purpose," "I believe I am doing what is necessary to do") on a 7-point scale (1 = *Strongly Disagree*, 7 = *Strongly Agree*). These ratings were averaged together for the moral accounting measure, $\alpha = 0.81$, $M = 5.97$; $SD = 1.17$. Both hero identification ($\alpha = 0.89$; $M = 5.45$; $SD = 1.25$) and victim identification ($M = 2.67$; $SD = 1.65$) were measured in Study 3 with the same items found in Study 2. The stem for both measures was slightly revised to begin with "To what extent in your job role, do you feel like someone who..." State self-esteem was measured with the same scale found in the prior studies ($\alpha = 0.93$, $M = 5.47$; $SD = 1.26$). To measure moral stigma, the survey included Shnabel and Nadler (2008)'s public moral image scale with the stem, "The public generally perceives my job as ..." followed by the items, "completely moral," "100% appropriate in what I do," and

“ethical.” This was rated on a 7-point scale (1 = *not at all*, 7 = *very much so*) and reverse-coded, $\alpha = 0.94$; $M = 2.32$; $SD = 1.28$.

Participants of this study also completed the following outcome-related measures that was applied in Study 2: public denigration or public esteem, reverse-ordered ($\alpha = 0.91$; $M = 3.21$; $SD = 1.93$); work secrecy or work transparency, reverse-ordered ($M = 5.96$; $SD = 1.30$), professional distancing or compassion, reverse ordered ($\alpha = 0.95$; $M = 3.21$; $SD = 1.93$); prosocial motivation continuance ($\alpha = 0.94$; $M = 5.85$; $SD = 1.14$); and restitution support ($M = 0.75$; $SD = 0.66$). For restitution support, participants were asked how they might earmark \$10,000 of leftover discretionary funds, if given the opportunity, from their current workplace budget. They were provided two options for allocating these funds: (i.) “Donations to assist the individual(s) you have harmed in the course of doing your job;” and (ii.) “Resources and bonuses for deserving colleague(s) at your workplace.” Specifically, restitution support was coded based on how much of the available funds participants allocated towards their target(s) of harm, which was standardized for analysis.

For Study 3’s control variables, consistent with the prior studies in the present research, the following demographic data was collected: age, gender, race, political orientation ($M = 4.04$, $SD = 1.98$), and religiosity ($M = 4.15$, $SD = 2.23$). In recognition of the broad variation present in this sample’s job backgrounds and NE task features, this study employed an extensive set of control variables to reduce noise and account for alternative sources of potential variance. For instance, several theoretically grounded covariates were included such as *job tenure* and *NE recall*, i.e., the last time participants recalled having performed a NE task. Both variables have demonstrated predictive

importance in prior research involving job perceptions and memory (Aquino et al., 2011). Participants reported their job tenure by selecting from predefined time intervals (1 = *Less than 1 year*, 2 = *1-2 years*, 3 = *2-3 years*, 4 = *3-4 years*, 5 = *4-5 years*, 6 = *More than 5 years*; $M = 4.96$; $SD = 1.48$). For NE recall, participants were asked when they last had to perform a necessary evil task (1 = *Less than one week ago*, 2 = *two weeks ago*, 3 = *one month ago*, 4 = *two months ago*, 5 = *three months ago*, 6 = *more than three months ago*; $M = 2.06$; $SD = 1.33$). Another two covariates were included to reflect how participants generally viewed their targets of harm, with regards to the *closeness* of their relationship with them (1 = *not at all*, 7 = *very much so*; $M = 3.93$; $SD = 1.95$) and how *blameworthy* they felt targets of harm were for “bringing [these situations] upon themselves” (1 = *not at all*, 7 = *very much so*; $M = 4.65$; $SD = 1.98$). Target closeness has previously been shown to impact an offending party’s prosocial intentions following the offense (Schumann, 2014) and target blame is a well-established indicator of moral disengagement (Andiappan & Dufour, 2017; Moore et al., 2012). In addition, to account for variation in NE task characteristics, a measure of *job complexity*—a central feature of distinguishing NE tasks—was incorporated (Margolis & Molinsky, 2008). This was measured with a 4-item scale from Chen et al. (2024), which begins with the stem, “Professionals in my line of work...,” followed by the items, “provide services that are actually very complicated,” “provide expertise that is very complex,” “need to take a lot of time to complete their services,” and “do not have it easy, in terms of having the knowledge and skills required by this profession.” Participants rated their agreement on these job complexity statements using a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*; $\alpha = 0.82$; $M = 5.27$; $SD = 1.18$). Another NE-relevant task dimension, the

perceived centrality of NEs in participants' jobs, was included as a covariate. To measure *NE centrality*, participants answered whether NEs were a major or minor aspect of their jobs (1 = *very minor*, 2 = *somewhat minor*, 3 = *moderate*, 4 = *somewhat major*, 5 = *very major*; $M = 3.10$; $SD = 1.10$). Finally, given this research's focus on the moderating effect of moral and occupational stigma, *job status* was included as a control variable ("To what extent do others view your occupation as a low-status or high-status job?"). Ashforth and Kreiner (1999)'s *status shield* theory posits that higher-prestige job roles buffer individuals from the psychological effects of experiencing stigma. Job status was rated with a scale that ranged from 1 (*very low status*) to 7 (*very high status*) ($M = 5.01$; $SD = 1.36$).

Results

Table 4 summarizes the pairwise correlations among the focal variables of this study. As a first step, Hypothesis 1 was tested with an OLS regression analysis. Results found moral accounting significantly predicted greater hero self-identification ($b = 0.31$, $p < 0.001$). Table 5 provides a summary of this output, which was conducted with the inclusion of all 12 covariates: age, race, gender, political orientation, religiosity, tenure, NE recall, NE centrality, job complexity, job status, target closeness, and target blame. This positive association between moral accounting and hero self-identification was significant without controls as well ($p < 0.001$).

For all subsequent analyses, to minimize overfitting the hypothesized path models, a preliminary series of regressions were conducted to identify which control variables meaningfully contributed to the explained variance of each outcome variable. Only the covariates that emerged as significant predictors (threshold, $p < .05$) for each

given dependent variable were retained for the corresponding models. As displayed in Table 6, several control variables emerged as predictors. Predictors of restitution support towards harm recipients were job complexity ($p < .01$) target closeness ($p < .01$), and target blame ($p < .001$). In the model examining predictors of public denigration, i.e., public esteem when reverse-coded, the significant control variables were religiosity ($p < .05$), job status ($p < .001$), and target closeness ($p < .05$). Gender ($p < .05$) and necessary evil centrality ($p < .05$) emerged as significant predictors of work secrecy (i.e., work transparency, when reverse-coded). In the model predicting prosocial motivation continuance, the following control variables reached significance: gender ($p < .05$), religiosity ($p < .001$), and job status ($p < .001$). Notably, no control variables reached significance in the model predicting professional distancing (i.e., professional compassion, reverse-coded).

Mediation. Table 7 provides the summary of the mediation analyses conducted using 5,000 bootstrapped resamples to estimate confidence intervals for indirect effects. Each model tested whether hero self-identification mediated the relationship between moral accounting and the five prosocial outcomes of interest (H1, H2, H3). They each included the specified covariates that are significantly associated with each outcome variable found in Table 6. Across all models, the results from these mediation analyses found support for Hypothesis 1 with moral accounting significantly predicting greater hero self-identification ($ps < .001$). Hypothesis 2 was supported in four of the five models: hero self-identification positively predicted higher restitution support toward targets of harm ($\beta = .18, p < .01$); greater public esteem ($\beta = .24, p < .001$); more support for public transparency regarding one's work practices ($\beta = .33, p < .001$); and increased

prosocial motivation continuance ($\beta = .42, p < .001$); but was only marginally significantly related to greater professional compassion ($\beta = .12, p = 0.10$).

Significant indirect effects, in support of Hypothesis 3, were found for restitution support ($\beta = .07, 95\% \text{ CI } [0.01, 0.08]$), public esteem ($\beta = .08, 95\% \text{ CI } [0.03, 0.14]$), work transparency ($\beta = .13, 95\% \text{ CI } [0.06, 0.24]$), and prosocial motivation continuance ($\beta = .15, 95\% \text{ CI } [0.07, 0.23]$). However, the indirect effect for professional compassion was not significant ($\beta = .05, 95\% \text{ CI } [-0.02, 0.11]$). Direct effects of moral accounting on prosocial outcomes were non-significant in models predicting public esteem, work transparency, and prosocial motivation continuance, indicating full mediation. The exception was the model predicting restitution support, as illustrated in Figure 3, in which the direct effect was significant ($\beta = -0.13, 95\% \text{ CI } [-0.29, -0.00]$), but showing the opposite direction. This means moral accounting on its own generally will lead to a significant reduction in supporting one's target of harm, however, hero self-identification counteracts this result and increases restitution support. Collectively, the results of these models provided support for the hypothesized indirect path linking moral accounting to prosocial outcomes through hero self-identification, except for the outcome variable professional compassion.

Conditional serial mediation effects. To evaluate the theoretical framework corresponding to Hypotheses 1, 4, 5, and 6 (as depicted in Figure 2), conditional serial mediation analyses were conducted as outlined in Hayes (2017; PROCESS Model 91). This is a type of conditional process analysis, which integrates both serial mediation and moderation within a regression-based path analytic framework. Moral accounting served as the predictor in these models. Both hero self-identification and victim self-

identification served as mediators, in this respective order, to assess the relationship between moral accounting and NEPs' reduced prosocial response.

To test Hypotheses 4-6, the first mediator, hero self-identification, was configured as a predictor of the second mediator, victim identification, when interacting with the moderator of moral stigma (high vs. low, computed with ± 1 SD). Both hero identification and moral stigma were mean-centered before being inputted as interaction terms. A conditional serial mediation model was conducted with bootstrapped standard errors (5,000 samples) per outcome variable, i.e., the five different anti-social responses. The outcome variables assessed included: greater professional distancing (model 1), less support for aiding targets of harm (model 2), higher public denigration (model 3), more secrecy regarding one's work practices (model 4), and lower prosocial motivation continuance (model 5).

Across all models, Hypothesis 1 was consistently supported: moral accounting positively predicted hero self-identification in each model ($ps < 0.001$). Hypothesis 4 predicted that hero self-identification would interact with moral stigma to predict greater victim identification, such that individuals would be more likely to see themselves as victims when perceived moral stigma was high. Results showed that the interaction between hero identification and stigma predicted victim identification across most of the outcome-based models. Specifically, this predicted path was found significant in models predicting denigrating attitudes toward the public ($\beta = 0.125, SE = 0.042, p < .05$); greater work-related secrecy ($\beta = 0.121, SE = 0.041, p < .05$); and prosocial motivation continuance ($\beta = 0.133, SE = 0.042, p < .05$); while only marginally significant in the model predicting professional distancing ($\beta = 0.102, SE = 0.043, p = .09$). However, this

hypothesized interaction effect was not observed for the model predicting assistance towards targets ($p = .20$). Figure 4 provides a visualization of this interaction.

Hypothesis 5, which posited that victim identification would predict less prosocial responses, received mixed support depending on the outcome variable. Victim identification significantly predicted less public esteem ($\beta = -.17$, $SE = 0.04$, 95% CI [0.032, 0.177]), and less professional compassion ($\beta = -.13$, $SE = 0.07$, 95% CI [0.015, 0.295]), but it was not significant for less work transparency ($\beta = -.02$, $SE = 0.04$, 95% CI [-0.066, 0.100]) or prosocial motivation continuance ($\beta = -0.10$, $SE = 0.04$, 95% CI [-0.124, 0.010]). Unexpectedly, a higher level of victim identification was significantly associated with increased support for restitution towards harm recipients, ($\beta = 0.13$, $SE = 0.03$, 95% CI [0.003, 0.105]). All direct effects are presented in Figures 5-9.

The conditional indirect effect (H6) of moral accounting on anti-social responses through hero and victim identification, as moderated by moral stigma, was significant for less public esteem at high levels of stigma ($\beta = -0.02$, $SE = 0.01$, 95% CI [0.004, 0.036]); and marginally significant for less professional compassion at high levels of stigma ($\beta = -0.02$, $SE = 0.02$, 95% CI [0.003, 0.066]). A summary of the indirect effects and conditional serial mediation results from these models can be found in Table 8.

Discussion

The findings from Study 3 provide support for the proposed pathway linking moral accounting to prosocial responses via greater hero self-identification (H1, H2, and H3). Supporting Hypothesis 3, hero self-identification significantly mediated the relationship between moral accounting and four of the five prosocial responses—restitution support, public esteem, transparency support, and motivation continuance.

These relationships were accompanied by significant indirect effects, indicating that moral accounting promotes prosocial behaviors through the internalization of a heroic identity. Importantly, the direct effects of moral accounting on these outcomes were non-significant in most cases, implying full mediation via hero identification. The one exception was the model predicting restitution support, where a negative direct effect was found, suggesting moral accounting may reduce perceived responsibility unless reframed through a heroic self-concept.

Results from the conditional serial mediation models demonstrated consistent support for the relationships between moral accounting and hero identification (H1). Overall, across most models, the findings indicated that greater hero identification led to greater victim identification when stigma was high (H4), suggesting that the relationship between heroism and victimhood can be shaped by moral stigma, as when individuals' hero identifications are threatened.

However, the results of these conditional serial mediation analyses found more selective support for the predicted downstream behavioral effects related to NEPs' reduced prosocial response. The path analyses that correspond to Hypotheses 5 showed victim identification significantly led to greater public denigration and was marginally significant in predicting greater professional distancing. Surprisingly, victim identification also predicted greater support for restitution toward harm recipients, counter to what had been hypothesized. This unexpected finding may reflect a public-facing moral self-repair strategy (see Barkan et al., 2012), whereby those who see themselves as victims still seek to restore their public moral image. This would mean they are likely to double-down on behaviors that showcase their concern for the harmed

party, when they feel negatively judged to seek reputational repair. Victim identification did not predict secrecy or lower prosocial motivation continuance; on the other hand, hero identification did significantly predict greater promotion of work transparency and prosocial motivation continuance.

One potential reason for the mixed or null effects observed in the conditional serial mediation analyses may lie in the distributional characteristics of the moral stigma measure. Specifically, the measure exhibited positive skew, with most participants in Study 3 reporting low levels of perceived stigma. As a result, the “high stigma” condition—operationalized using scores one standard deviation above the mean—may not have captured a sufficiently meaningful range of stigma as experienced by NE practitioners in vivo. This restricted range and relative lack of variability could have limited statistical power to detect moderated effects, particularly in the upper tail of the distribution where key conditional effects were expected to emerge. From a theoretical standpoint, the core mechanism proposed by the model—that hero identification may backfire in stigmatizing environments by prompting a shift to victim identity—presumes that individuals are experiencing at least moderate levels of perceived moral stigma. If such experiences were rare in this sample, the full psychological dynamics underlying this pathway may not have been activated, thereby attenuating the conditional effects.

Conclusion and General Discussion

The present research sought to understand how necessary evil practitioners (NEPs) psychologically navigate their morally complex roles, particularly with respects to the interplay between their internal ethical conflicts and exposure to external moral stigma. Across three studies, findings shed light on the cognitive and identity-driven

processes that NEPs undergo from performing tasks involving harm, as mandated by their occupational roles.

The most compelling evidence emerged from Study 3, which sampled real-world NEPs across various professions. Supporting Hypothesis 1, moral accounting was positively associated with hero self-identification. This finding aligns with prior work on individuals' desire to retain a positive self-image, particularly as moral actors (Aquino & Reed, 2002; Festinger, 1957). This moral accounting process, then, predicted individual practitioners' self-identification as heroes. Furthermore, moral accounting significantly predicted greater prosocial responses, including prosocial motivation continuance, restitution support, public esteem, and support for work transparency (H2) via hero identification (H3). In addition, Study 3 demonstrated that when NEPs perceived high levels of moral stigma from the public towards their jobs, their identification as heroes prompted a shift towards victim identification (H4). This hero-to-victim identity shift was expected to attenuate NEPs' prosocial responses (H5). Supporting this hypothesis to an extent, victim identification failed to significantly predict some of the prosocial outcomes observed under hero identification such as prosocial motivation continuance and work transparency. Victim identification, in fact, led to significantly less public esteem and professional compassion. As proposed in Hypothesis 6, these effects were part of a conditional serial mediation process. These findings suggest that while NEPs can leverage moral accounting to affirm a positive, heroic self-concept, perceived public disapproval may disrupt this process and no longer yield significant prosocial effects comparable to those seen with hero self-identification. Taken together, the results underscore the importance of both individual-level moral processing and socially

constructed perceptions in shaping NEPs' self-concept and downstream work behaviors. The findings highlight a critical psychological tension: moral accounting of NE tasks fosters heroic prosociality among individual practitioners, but not under external criticism—revealing the fragility of this prosocial pathway, which can be easily disrupted by salient experiences of moral stigma.

Exploratory mediation analyses from Studies 1 and 2 suggest that the link between moral accounting and hero identification operates through the restoration of a positive self-image via state self-esteem, rather than self-enhancement or heightened competence mechanisms. This finding does illuminate the effectiveness of moral accounting as a reparative cognitive dissonance reduction strategy, by which individuals may view themselves as heroes and remain committed to prosocial engagement in their work. Although this was not the primary focus of the present investigation, future research may aim to further explore other complementary processes such as self-affirmation processes, which involve affirming one's positive characteristics in domains outside the dissonance-causing event (Sherman & Cohen, 2006; Steele, 1988). Past studies have found this affirmation strategy can lead to some reductions in defensiveness among transgressors (Wenzel et al., 2020), but findings are mixed in terms of its ability to increase prosociality (Sherman & Cohen, 2006). Moral accounting, alternatively, focuses on affirming the self by direct reinterpretations of the dissonance-causing event.

Of course, the present research is not without its limitations, some of which reflect deliberate choices made to prioritize other methodological strengths. Some of these limitations involve the experimental manipulations found in Studies 1 and 2. Although both studies applied the same experimental procedure, moral accounting failed

to predict hero self-identification in Study 2. The lack of consistent support for Hypothesis 1 across both studies suggests potential issues with both the experimental materials and/or its delivery, perhaps calling into question the suitability of using an online paradigm to simulate the NEP experience. Although the results of the pilot study on the social worker vignette (see Appendix B) indicated that participants found completing the NE task difficult, participants also reported experiencing low levels of guilt and negative affect on average. Festinger (1957) proposed that cognitive dissonance and its subsequent dissonance-reducing actions would occur only when the discomfort is strong enough to motivate corrective action. It is possible that all participants, regardless of whether they were experimentally prompted to engage in moral accounting, did not experience a sufficient level of dissonance-related discomfort for the manipulation to produce discernible effects across conditions. Another important research area in need of further attention is revisiting the concept of moral accounting to ensure its reliable measurement and manipulation. In Studies 1 and 2, the moral accounting condition was designed to prompt participants to reflect on the positive outcomes associated with their NE task. However, this approach may have captured only a partial representation of the moral accounting process. Specifically, the manipulation may have constrained participants' responses by directing attention exclusively to beneficial consequences, thereby omitting the full evaluative nature of moral accounting. A more comprehensive approach might involve asking participants to consider both the costs and benefits of the NE task, followed by a reflective judgment of this tradeoff—free from valence-based cues. Such a design may yield a more ecologically valid and psychologically authentic measure of moral accounting. In addition, the moral accounting manipulation used in

Studies 1 and 2 involved a reflective writing task in the experimental condition, whilst the control group skipped this procedure altogether. It is possible that the experimental exercise may have simply introduced cognitive processing differences. Participants in the moral accounting condition may have spent more time engaged in introspective thought and slower cognitive processing (Evans, 2008), which could have confounded effects beyond the intended manipulation. Future studies should ensure greater parity across conditions by including control prompts of similar length and complexity, thereby better isolating the specific effect of moral accounting. In addition, the binary stigma manipulation (stigma vs. control conditions) may not have sufficiently represented NEPs' experiences with public feedback. Study 2's stigma manipulation may have benefited from including not only a stigma-directed threat to hero identification, but also a condition in which one's hero identification is affirmed by the public. This would have provided a wider range of social feedback as they impact practitioners' moral identities.

Furthermore, while Study 3 provided greater ecological validity by surveying real NEPs, its sampling and cross-sectional design limits causal inferences. Notably, findings from Study 3 involving perceived moral stigma were complicated by the positively skewed distribution of this measure. Most participants in Study 3 reported low levels of perceived stigma toward their professions, meaning that the sample contained relatively few respondents who fell into the "high stigma" condition. As a result, conditional effects—particularly those requiring variance at the higher end of the stigma scale—may have been underpowered. Future research should seek to oversample NEPs who experience higher levels of public criticism, or employ experimental manipulations of stigma salience, to better test the robustness of these conditional effects. In addition,

future longitudinal or experimental field studies are needed to establish causal pathways as well as temporal order (Maxwell & Cole, 2007). Tetlock and Levi (1982) maintain that self-schemas are shaped over time, rather than by a single event. Accordingly, hero self-identification would be expected to evolve through multiple cycles of moral reflection and reinforcement of feedback. Future research should examine within-person variation in moral accounting of NEs as it could illuminate whether repeated interpretations of NEs contribute to the emergence of a heroic self-concept. Both experimental and cross-sectional studies employing such an experience sampling methodology (ESM) to better understand the psychological strain of NEs would offer multiple time points, allowing for a more detailed examination of the proposed pathways.

Importantly, the ESM approach would allow for a clearer disentanglement of the influence of social stigma and its potential role in shaping other relationships within the proposed model (H6), including currently untested pathways such as the link between moral accounting and hero identification. It is plausible that stigma, functioning as a form of social feedback, exerts a meaningful influence on the transition from moral accounting to an individual adopting a heroic identity. However, the present research focused on examining how practitioners engage in moral reasoning following harm-doing as a self-directed moral regulatory process—a process capable of directly informing one’s moral identity development. This focus is grounded in the view that individuals are strongly motivated to preserve a coherent and positive self-concept (Aronson, 1992). In the case of NEPs, the execution of NE tasks represents immediate, routine threats to their self-concept, thereby, prompting moral and self-identification processes more prominently than occasional encounters with harsh outside judgement (Eriksson, 2019). The

theoretical model advanced in this research prioritized privately held moral reasoning processes and professional identity formation at the micro-level. Nevertheless, broader social perceptions and dominant professional narratives do ultimately shape how individuals define their professional identities (Caza & Creary, 2016; Reid, 2015; Wyatt et al., 2021). Self-definitions may not simply be altered by salient experiences of stigma but dynamically constructed with these social attitudes in mind. In either case, it remains incumbent upon the individual to navigate these various influences and ideally arrive at a state of positive self-regard. Further research is needed to clarify these interactions.

Ultimately, the findings from this research underscore the delicate interplay between the moral processing of NE tasks and greater societal pressures in shaping the professional behaviors and attitudes of NE practitioners. The findings reveal that NEPs are not simply passive executors of morally fraught tasks, but active moral agents who work to psychologically reconcile their harm-enactment, which further influences beliefs about who they are and who they ought to be in their job roles. Moral accounting emerged as a powerful tool in this process—capable of fostering a heroic self-concept that fuels prosocial responses. Yet, this same identity can be destabilized by moral stigma, with implications for NEPs reduced compassion and positive engagement with the public. The latter can negatively impact the professional viability of jobs involving NEs. If NEPs do not effectively manage these public impressions, their NE tasks may be deemed “more evil than necessary” (Ashforth & Kreiner, 2013, p. 84). Based on the results from Study 3, NEPs’ moral accounting is not merely a private affair, its downstream consequences are embedded in a broader social context that either legitimizes or undermines practitioners’ moral narratives from heroes to victims. This

shift—while self-protective—can yield mixed behavioral outcomes, some anti-social and some restorative, underscoring the complexity of how stigmatization is internalized and manifests in one’s work conduct. From a practical standpoint, these insights are especially timely for organizations whose workers carry out NE tasks that are often under public scrutiny (e.g., law enforcement, healthcare, social services). The encouragement of moral accounting may foster prosocial engagement and mitigate burnout among NEPs. At the same time, when public criticism towards these professions are high and moral accounting tendencies are strong, this combination can undermine workers’ prosocial intentions. In sum, by identifying conditions under which moral accounting enhances prosociality—and when they may inadvertently backfire—this research advances understanding of how NEPs regulate their moral identities in the face of morally challenging tasks and societal judgments. It highlights both the promise and fragility of heroic self-concepts in morally complex work, with implications for professional support structures, public messaging, and drivers of continued prosocial engagement.

Appendices

Appendix A: Study Materials

Appendix A1: Social Worker Vignette

Imagine you are a social worker with over 9 years of experience.

Today, you are following up on a household that you last visited 3 months ago—initially based on reports of child neglect.

When you get to the house, Jamie (the mother) opens the door. She appears nervous as she lets you into the house. Compared to your last visit, you notice the house has fallen into greater disrepair. Jamie is a single mom, who has not been able to hold down a stable job. It is clear that she loves her child. But despite past warnings, the child still appears malnourished and has missed numerous days of school. You also find out that the house has no running water.

- Decision Checkpoint Screen Part 1 -

The assessment during your visit suggests you should take the following action -

Place Jamie's child under alternative housing and care.

* Are you ready to proceed with this course of action?

Yes

No

- Decision Checkpoint Screen Part 2 - If “NO” is Selected -

*The assessment during your visit suggests you should
take the following action -*

**Place Jamie's child under alternative housing and
care.**

Please take your time and select "Yes" when you are ready to proceed.

Yes

Without extended family members who can help with childcare, you must coordinate with law enforcement and foster care to remove the child from the home. Jamie cries and pleads with you. She begins sobbing. "Please don't take my child away from me." While being taken away, the child looks terrified and confused.

Later, you reflect on what happened with Jamie and her child as you finalize your case report.

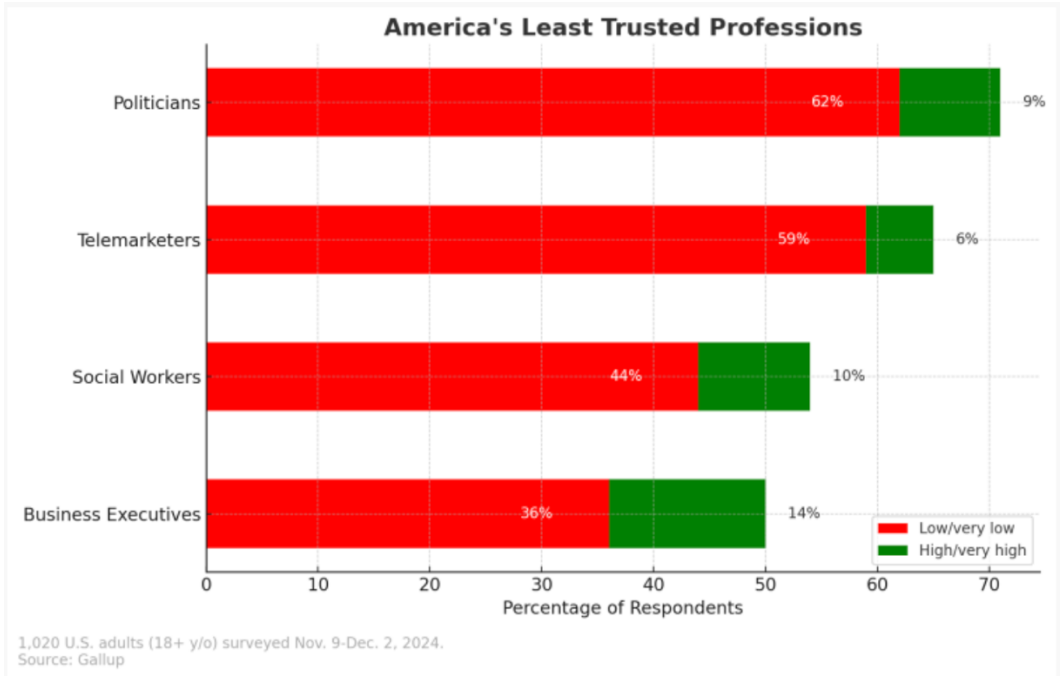
Appendix A2: Moral Stigma Manipulation Materials

-Experimental Group-

The next day, you return to work and see a flyer pinned to the front door of your office building...



Later, you go online to learn more about the flyer, when you come across this public poll about social workers:

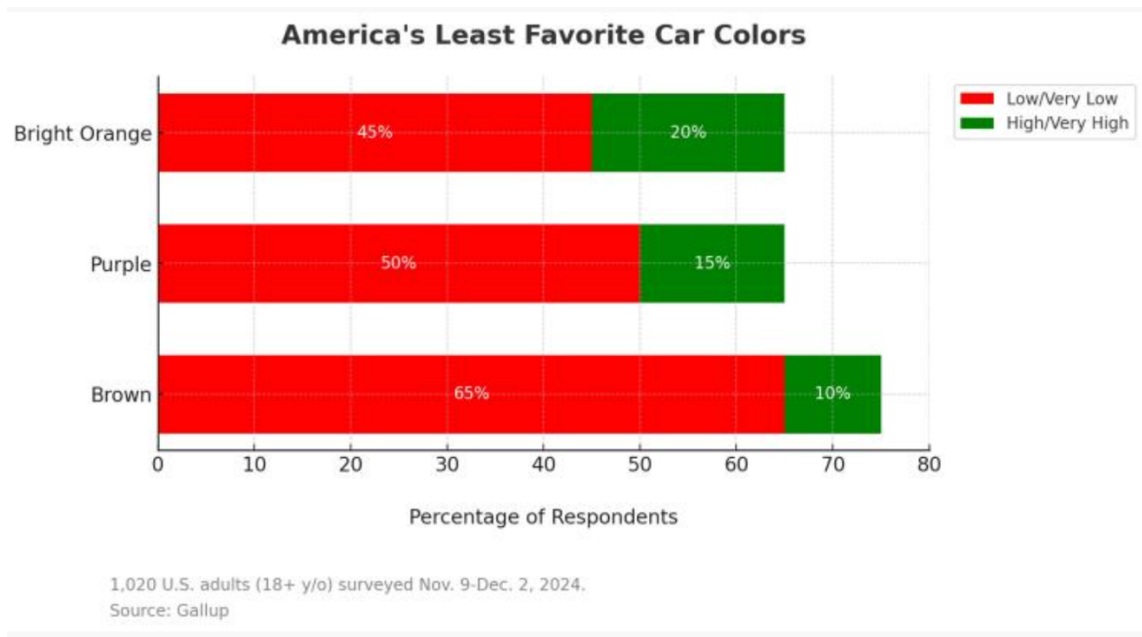


-Control Group-

The next day, you return to work and see a flyer pinned to the front door of your office building...



Later, you go online to check your work emails, when you come across this public poll about car colors:



Appendix B: Pilot Studies and Vignette Development

Two pilot studies were carried out in developing the experimental vignette used in Studies 1 and 2. First, qualitative data was collected from employees about tasks they experienced at work that fit the description of a necessary evil (Pilot 1). This collection of stories informed the creation of three vignettes, which were next pre-tested by a sample of subjects who responded to a variety of construct validity questions and felt realism regarding each vignette (Pilot 2). The pilot studies identified the social worker vignette (see Appendix A1) as the most appropriate scenario for capturing important facets of a necessary evil task and addressing the objectives of this research.

Pilot 1

The first pilot study involved a survey that collected stories from Prolific users with occupational backgrounds in law enforcement and medicine/social work, which were also professions targeted by Margolis and Molinsky (2008) in their qualitative study of necessary evil practitioners (NEPs). Twenty participants completed this study. One participant said they had never had to perform a NE task in their job, leaving a total of 19 participants who provided rich descriptions of their NE experiences. When asked for their specific job titles, the sample of answers included: detention officer ($n = 1$), government employee ($n = 1$), law enforcement ($n = 3$), detective ($n = 1$), security services supervisor ($n = 1$), veterinary technician ($n = 1$), social worker ($n = 4$), program manager ($n = 2$), personal support worker ($n = 1$), medical coordinator ($n = 1$), case manager ($n = 1$), office manager ($n = 1$), and mental health technician ($n = 1$).

In this survey, participants were prompted to list the NE tasks they performed in their jobs as well as write about a specific work experience in-detail that involved “doing

something that has negative consequences—specifically, it [caused] emotional or physical harm/pain to others,” but it was necessary for them to do in their jobs. We elicited critical incident stories from participants by applying a similar procedure from Lees, Young, and Waytz (2022). First, participants read: “Please take a few moments to think about a particularly vivid work experience you had that involved performing one of these tasks that inflicted harm on someone, but it was necessary to do.” This was followed by the prompt, “What took place? Please be as specific as possible. Provide a chronological account. This account should enable someone to recreate the scene in their mind.” After participants wrote an answer to this question, they were asked for more details: “As this experience unfolded, what were you experiencing personally? What thoughts went through your mind? What emotions did you feel? Please share this experience as you would to a close friend or partner.”

I content-coded these responses and the most common themes that arose were as follows: (1) Experiencing emotional or moral distress while doing the task, 52.6%; (2) describing the task as a job-imposed necessity, such as mentioning words like “had to,” “was required,” “as part of my job,” “due to policy,” 42.1%; (3) having to physically restrain or take forceful control of another person, 42.1%; (4) conflicting feelings regarding the compassion they felt toward their targets and how their job task was requiring them to harm these targets, 26.3%; (5) safety concerns regarding themselves, others, or the target’s—if they had not conducted the NE task, 26.3%; (6) intervening in situations that involved individuals with mental health challenges, 21.1%; and (7) a task that involved delivering bad news to someone, 10.5%.

Themes 3, 6, and 7 were related to the content of a NE task, which aligns with the jobs of the NEPs who made up this sample. Other themes (1, 2, 4), however, were notably about NEPs' psychological states as they experienced a particularly "vivid" NE task. For example, Theme 1 encompassed half the participants, who described feeling some form of negative emotion or moral distress while conducting their NE task ("I personally felt very bad about having to tell the patient this news..."; "I felt guilty and did not want to do it"). In accordance with Theme 4, over a quarter of the participants expressed feeling caught between fulfilling a task mandated by their jobs, due to the empathy they felt towards their target's situation (e.g., "I had to send staff to evict a person from a shelter... I was told that the rules did not allow them to stay after they missed too many nights. It didn't feel right..."; "Reprimanding staff for mistakes... I didn't want to be harsh but had to document it formally"). Drawing from the themes that emerged from these personal narratives that were shared by participants, three vignettes were created to serve as potential NE scenarios for use as experimental stimuli.

Pilot 2

Following the first pilot, three vignettes were developed to immerse participants in a hypothetical necessary evil task from the viewpoint of its central character, i.e., the employee carrying out the NE task. These vignettes were piloted with online surveys via Prolific ($N = 104$). Participants were randomly selected to read and assess one of three vignettes. Each vignette was of similar length and followed a comparable sequence of events. However, each story was framed from the lens of a different profession: a social worker ($n = 34$), a bankruptcy trustee administrator ($n = 36$), and a law enforcement officer ($n = 34$). The pilot study assessed vignette quality, occupational bias, and

construct validity. Mean differences informed the selection of the most suitable vignette for Study 1.

Experimental Vignettes

In each vignette, the storyline structure began with prompting readers to imagine themselves as the story's central worker, who was about to carry out a routine job task. Next, to set up the situation for performing the NE task, the reader was provided details about why the target of harm was under scrutiny. The target had committed a wrongdoing, and the reader, given their job role, had to address this situation. All the vignettes intended to establish a level of irrefutable wrongdoing on the part of the target of harm to signal to the reader that their NE task had become procedurally necessary. This also reduced ambiguity surrounding whether a misdemeanor had occurred or not, which would otherwise engage a cognitive reasoning process that is outside the scope of this study (Bartels et al., 2015). Moreover, NEPs are generally more psychologically affected by the act of inflicting harm itself than by their adjudication regarding the target's innocence or guilt (Margolis & Molinsky, 2008). To evoke emotional discomfort and heighten the reader's feelings of internal tension, each vignette portrayed the target as culpable yet worthy of compassion by providing details that humanized the target. After introducing this situation, the reader was guided to a critical decision point, in which they were nudged to carry out the harmful action. After confirming to "proceed" in completion of the NE task, readers were provided with a brief description of the immediate aftermath, as they witnessed the distress felt by the target, to further elicit emotional unease and emphasize the human impact of this NE task.

Social worker scenario. See Appendix A1.

Bankruptcy trustee scenario. This vignette began with this introduction: “Imagine you are a bankruptcy trustee with over 9 years of experience. Today, you are meeting with a small business owner as the court-appointed accountant reviewing his finances and liabilities. This is your second meeting with Jamie the business owner). During your initial consultation, he provided you with his financial records. He explained his restaurant never recovered after the pandemic. Jamie took out large loans to buy time and revitalize his restaurant, which deepened his debt. Unfortunately, the entire surrounding corridor, once thriving, has experienced sustained decline in customers. Jamie hopes to relocate the business. But due to his outstanding debts, lenders refuse to offer more funding.”

At the decision checkpoint, it is suggested to the reader that they proceed in “[initiating] Chapter 7 liquidation of Jamie's business.” Once the reader has confirmed to do this task, they read the following epilogue: “Now, seated across from Jamie, you prepare to deliver the news. He begs, ‘Is there any way I can get a bit more time? Please don’t do this!’ His weary eyes meet yours as you explain that his debts far exceed his assets, and liquidation is necessary to prevent aggressive creditor actions. Jamie slumps into his chair. His hands begin trembling. ‘It’s not just a restaurant,’ he says quietly. ‘My father helped me build it. My kids grew up there. This restaurant is my whole life.’ Later, you reflect on what happened with Jamie as you finalize your case report for the judge.”

Police officer scenario. This scenario opens with: “Imagine you are a law enforcement officer with over 9 years of experience. Today, you must follow up on a store’s report of a shoplifting incident. When you arrive, the store manager appears frustrated. They show you video surveillance of an elderly man slipping items into his

coat pocket. The manager also pulls up footage of another time the same suspect stole items from their store. Jamie (the suspect) is being held by the security guard in the manager's office. Jamie's elderly wife is also there. She says Jamie has been experiencing cognitive decline. When you sit down to question Jamie, he appears confused. However, he does not dispute that he is the man caught stealing in the video footage."

After reading this first part of the vignette, readers were given a choice to proceed with "[bringing] Jamie to the police station for processing." Once the reader committed to this action, they read the following conclusion: "You prepare to detain Jamie. As you read Jamie his rights, his wife begs, "He's sick. He's not a criminal. Please don't do this!" You tell her the incident will be sorted out at the station, and she can make her case there. You escort Jamie to the back of your patrol car. His wife is trembling and begins to tear up. As you drive out of the parking lot, you can hear her sobbing. Later, you reflect on what happened with Jamie as you finalize your case report."

Measures

The survey questions that followed the reading of this vignette evaluated the following: the narrative quality of the vignette, possible confounding biases related to the occupational role itself and the likability of the NEP protagonist, as well as the scenario's construct validity with regards to a necessary evil task.

Vignette quality was measured with questions regarding the ease in which it was to follow along, how realistic the situation felt to participants, and to what extent participants were able to imagine themselves as the NEP in the story. The 5-point scale that was used for responding to these three questions ranged from 1 (*Not at all*) to 5

(*Very much so*). They were averaged together into one vignette quality score. To assess the emotional reactivity to reading this vignette, the survey applied Thompson (2007)'s shortened version of the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). This shortened version included ratings of positive affect (inspired, alert, attentive, active, determined) and negative affect (afraid, upset, nervous, ashamed, hostile) using a 5-point response scale (1 - *very slightly or not at all*, 2 - *a little*, 3 - *moderately*, 4 - *quite a bit*, 5 - *extremely*).

To assess occupational biases, participants were asked baseline questions about their opinion on the general public's view of these occupations (1 – *very negative*, 5 – *very positive*) and to what extent most people would consider the NEP's job in the story to be a low-status or a high-status job (1 – *very low-status*, 5 – *very high-status*). We also asked participants to rate on a 5-point scale how they felt about the NEP's likability, specifically, if they found them unlikable-likable, incompetent-competent, and unprofessional-professional. These items were combined into one likability score.

For ensuring each vignette reflected the task characteristics of a NE, as Molinsky and Margolis (2005) specified, we asked participants several construct validity questions. The first of these questions asked just how necessary the NE task felt to participants (“The [worker's] action taken in this case reflected a necessary part of their job responsibilities”; 1 – *Strongly Disagree*, 5 – *Strongly Agree*). In terms of the harm and occurrence of harm, we asked participants to rate whether “the story made clear that [the target] experienced distress,” given their actions in the role of the NEP, which was rated from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). The survey included a question on the agency of the NEP, i.e., “In the story, the [worker's] response to this case seemed guided

by their own judgement and not by other people's input," which used a 5-point rating scale (1 – *Strongly Disagree*, 5 – *Strongly Agree*). Adapted from Hanselmann and Tanner (2008)'s study, we also had participants immediately after their reading of the vignette rate what it was like for them to proceed with an action that harmed the target, from 1 (*very easy*) to 5 (*very difficult*). To further assess the psychological state of participants, we asked how guilty they felt (1 - *very slightly or not at all*, 5 – *extremely*).

Results

Table A1 presents the mean scores of the piloted measures across the three vignette conditions: social worker (SW), police officer (PO), and bankruptcy trustee (BT). To determine which vignette was most suitable for the experimental study, mean comparisons were examined, and one-way ANOVAs were conducted with Tukey HSD post-hoc tests. Overall, all vignettes received high mean ratings on vignette quality ($M_{SW} = 4.75$, $M_{PO} = 4.65$, $M_{BT} = 4.34$). However, both the social worker ($p < 0.01$) and police officer ($p < 0.05$) vignettes, in comparison to the bankruptcy trustee vignette, were rated significantly higher on vignette quality. PANAS scores indicated moderate positive affect and low negative affect across conditions, suggesting none of the vignettes strongly induced either positive or negative emotions.

The perceived public bias toward these occupations were not significantly different between the three job roles ($p = 0.053$), but the job of bankruptcy trustee was viewed most negatively ($M_{BT} = 2.75$), followed by a police officer ($M_{PO} = 3$). Social workers were viewed most favorably ($M_{SW} = 3.35$), but there were no statistically significant group differences ($p = 0.07$). Interestingly, the perceived status of a social worker was viewed as significantly lower, in comparison to both police officers ($p < 0.05$) and

bankruptcy trustees ($p < 0.01$). All NEPs, in terms of their likability as characters in the vignettes, were rated highly on likability, but the police officer received the lowest likability; however, there were no significant group differences ($M_{SW} = 4.08$, $M_{PO} = 3.80$, $M_{BT} = 4.12$).

When examining which vignette appeared to provide the most construct validity, the social worker scenario's NE task was viewed as significantly more necessary to enact than the police officer's NE task ($p < 0.01$). In addition, participants rated to what extent the vignette made clear how distressed the target felt, in response to their enactment of harm. The police officer scenario was rated the lowest on this measure, in comparison to the bankruptcy trustee ($p < .001$) and the social worker ($p < 0.01$) scenarios. Thus, people in the police officer condition did not think the vignette effectively demonstrated how their actions as the NEP led to distress for the target of harm. Although there were no significant group differences on these following measures, participants in the social worker condition exhibited the highest average scores in having felt agency ($M_{SW} = 3.35$) while performing the task, as well as finding the situation difficult to do ($M_{SW} = 3.35$).

Given that the social worker scenario received the highest ratings for vignette quality and outperformed other conditions on several construct validity indicators, it was selected as the experimental vignette for use in the main studies.

Table A1. Mean Ratings on Piloted Measures by Vignette Condition

	Bankruptcy Trustee	Police Officer	Social Worker
Vignette Quality	4.34	4.65	4.75
PANAS - Positive Affect	3.18	3.02	3.62
PANAS - Negative Affect	2.34	2.05	2.28
Perceived Job Bias	2.75	3.00	3.35
Perceived Job Status	3.42	3.26	2.88
NEP Likability	4.12	3.8	4.08
Necessity of Task	4.5	4.09	4.71
Harm Check	4.08	2.97	3.82
Agency of NEP	3.64	3.91	4.06
Task Difficulty	3.61	3.76	3.94
Guilt	2.78	2.82	2.53

Tables and Figures

Table 1. Correlations between Study 1 Variables

	1	2	3	4	5	6	7	8	9	10	11
1. Moral Accounting (0, 1)	-										
2. Hero Identification	0.14*	-									
3. State Self-Esteem	0.17**	0.43***	-								
4. Moral Self-Efficacy	0.07	0.58***	0.42***	-							
5. Moral Superiority	0.04	0.54***	0.33***	0.47***	-						
6. Age	-0.04	-0.04	0.07	0.11	-0.05	-					
7. Gender	0.00	0.12*	0.17**	0.05	0.01	0.15*	-				
8. Race	0.05	-0.12*	-0.07	-0.05	0.00	0.21***	-0.06	-			
9. Politics	0.02	0.15*	0.11	0.06	0.00	0.08	-0.07	0.04	-		
10. Deontic Orientation	0.05	0.03	0.08	0.04	0.03	0.02	0.04	-0.09	0.05	-	

	1	2	3	4	5	6	7	8	9	10	11
11. Moral Normlessness	-0.11	-0.03	-0.02	0.00	0.01	-0.16**	-0.18**	-0.12*	-0.02	-0.01	-

Note:

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 2. Study 1's OLS Results with Covariates

	Hero Identification
Moral Accounting	0.265** (0.118)
Age	-0.005 (0.005)
Gender	0.264** (0.122)
Race	-0.264* (0.140)
Politics	0.087*** (0.032)
Deontic Orientation	-0.00003 (0.0004)
Moral Normlessness	-0.009 (0.053)
Constant	5.707*** (0.316)
Observations	290
R ²	0.071
Adjusted R ²	0.048
Residual Std. Error	0.992 (df = 282)
F Statistic	3.060*** (df = 7; 282)

Note: *** $p < .001$, ** $p < .01$, * $p < .05$

Table 3. Correlations between Study 2's Main Variables

	1	2	3	4	5	6	7	8	9	10
1. Moral Accounting (0,1)	-									
2. Moral Stigma (0, 1)	-0.02	-								
3. Hero Identification	0.08	0.07	-							
4. Victim Identification	-0.03	0.34***	0.30***	-						
5. State Self-Esteem	0.28***	0.06	0.54***	0.06	-					
6. Prosocial Motivation Continuance	0.01	-0.05	0.31***	0.10	0.21***	-				
7. Professional Compassion	0.10	-0.07	0.08	-0.07	0.19***	0.17***	-			
8. Restitution Support	-0.09	-0.09	-0.14**	0.01	-0.16**	0.24***	0.11*	-		
9. Public Esteem	0.02	-0.10*	0.10	-0.13*	0.16**	0.39***	0.16**	0.08	-	

Table 3. Correlations between Study 2's Main Variables

	1	2	3	4	5	6	7	8	9	10
10. Work Transparency	0.00	0.06	0.12*	0.04	0.10	0.20***	0.10	0.01	0.17***	-

Note:

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 4. Correlations between Study 3's Main Variables

	1	2	3	4	5	6	7	8	9
1. Moral Accounting	-								
2. Hero Identification	0.43***	-							
3. Victim Identification	-0.27***	-0.06	-						
4. Moral Stigma	-0.43***	-0.49***	0.45***	-					
5. Professional Compassion	0.11	0.15*	-0.15*	-0.20***	-				
6. Restitution Support	-0.07	0.13*	0.12*	0.02	0.22***	-			
7. Public Esteem	0.22***	0.35***	-0.19**	-0.28***	0.24***	0.21***	-		
8. Work Transparency	0.15**	0.35***	-0.03	-0.22***	0.13*	0.09	0.25***	-	
9. Prosocial Motivation Continuance	0.35***	0.54***	-0.14*	-0.40***	0.21***	0.21***	0.38***	0.41***	-

Table 4. Correlations between Study 3's Main Variables

	1	2	3	4	5	6	7	8	9
--	---	---	---	---	---	---	---	---	---

Note: *** $p < .001$, ** $p < .01$, * $p < .05$

Table 5. Study 3's OLS Results Testing Hypothesis 1

	Hero Identification
Moral Accounting	0.307*** (0.060)
Age	-0.00000 (0.007)
Gender	0.428*** (0.129)
Race	0.022 (0.146)
Political Orientation	0.089** (0.038)
Religiosity	0.026 (0.033)
Job Tenure	-0.012 (0.049)
Necessary Evils Recall	0.078 (0.051)
Necessary Evils Centrality	0.095 (0.065)
Job Complexity	0.168*** (0.057)
Job Status	0.175*** (0.050)
Target Closeness	0.067* (0.034)
Target Blame	0.0003 (0.034)
Constant	0.507 (0.570)
R ² , Adjusted R ²	0.317, 0.284
Residual Std. Error	1.063 (df = 269)
F Statistic	9.609*** (df = 13; 269)

Note: *** $p < .001$, ** $p < .01$, * $p < .05$

Table 6. Regression Results for Selection of Covariates in Study 3's Path Models

	(1) Professional Compassion	(2) Restitution Support	(3) Public Esteem	(4) Work Transparency	(5) Prosocial Motivation Cont.
Age	-0.02 (0.01)	-0.001 (0.004)	0.004 (0.01)	-0.001 (0.01)	0.001 (0.01)
Gender	-0.01 (0.23)	0.05 (0.08)	0.16 (0.12)	0.34* (0.16)	0.32* (0.13)
Race	0.33 (0.26)	0.04 (0.09)	-0.04 (0.14)	0.23 (0.18)	-0.02 (0.15)
Political Orientation	-0.01 (0.07)	-0.03 (0.02)	0.03 (0.04)	-0.06 (0.05)	-0.004 (0.04)
Religiosity	-0.01 (0.06)	0.03 (0.02)	0.07* (0.03)	0.01 (0.04)	0.11*** (0.03)
Job Tenure	0.05 (0.09)	0.03 (0.03)	0.06 (0.05)	0.03 (0.06)	0.06 (0.05)
Necessary Evils Recall	-0.13 (0.09)	-0.003 (0.03)	0.08 (0.05)	0.08 (0.06)	0.05 (0.05)
Necessary Evils Centrality	-0.17 (0.12)	0.05 (0.04)	0.0002 (0.06)	0.20* (0.08)	0.10 (0.06)
Job Complexity	-0.03 (0.10)	-0.09** (0.03)	-0.04 (0.05)	0.02 (0.07)	0.05 (0.06)
Job Status	0.12 (0.09)	0.05 (0.03)	0.16*** (0.05)	0.06 (0.06)	0.17*** (0.05)
Target Closeness	0.09 (0.06)	0.05** (0.02)	0.06* (0.03)	0.05 (0.04)	0.06 (0.03)
Target Blame	-0.09 (0.06)	-0.10*** (0.02)	-0.04 (0.03)	-0.004 (0.04)	-0.04 (0.03)
Constant	5.43*** (0.99)	0.87** (0.32)	3.09*** (0.51)	4.35*** (0.66)	3.34*** (0.54)

Observations	283	283	283	283	283
R ²	0.05	0.15	0.16	0.06	0.18
Adjusted R ²	0.01	0.11	0.12	0.02	0.14
Residual Std. Error (df = 270)	1.93	0.62	1.00	1.29	1.06
F Statistic (df = 12; 270)	1.16	3.89***	4.20***	1.39	4.90***

Note:

* ** p*** p<0.001

Table 7. Study 3's Mediation Models Linking Moral Accounting (MA) to Prosocial Outcomes via Hero Identification

DVs	MA → Hero	Hero → DV	Direct Effect	95% CI (c')	Indirect Effect	95% CI (ab)	Total Effect	95% CI (total)
Professional Compassion	.43*** (.07)	.12† (.11)	.10 (.12)	[-.13, .33]	.09† (.06)	[-.02, .21]	.18† (.11)	[-.02, .39]
Restitution Support	.40*** (.07)	.18** (.03)	-.13* (.04)	[-.14, .00]	.07* (.02)	[.01, .08]	-.05 (.04)	[-.09, .04]
Public Esteem	.35*** (.07)	.24*** (.06)	.04 (.06)	[-.08, .14]	.08** (.03)	[.03, .14]	.12* (.05)	[.00, .22]
Work Transparency	0.41*** (.07)	.33*** (.10)	-.002 (.09)	[-.15, .19]	.13** (.05)	[.06, .24]	.13† (.08)	[.01, .33]
Prosocial Motivation Continuance	0.35*** (.07)	.42*** (.07)	.12† (.07)	[-.02, .27]	.15*** (.04)	[.07, .23]	.27*** (.07)	[.14, .41]

Note: Estimates are standardized. Each model included covariates specified by outcome variable (see Table 6).

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Table 8. Defined Parameters from Conditional Serial Mediation Models

	<u>Indirect Effects</u>		<u>Serial Indirect Effect</u>	<u>Total Indirect Effect</u>	<u>Conditional Serial Mediation</u>	
	a_1*b_1	a_2*b_2	$a_1*d_1*b_2$		<i>Stigma-High</i>	<i>Stigma-Low</i>
Model 1: Professional Compassion	0.073 (0.108)	0.024 (0.025)	-0.014† (0.012)	0.083 (0.106)	-0.021† (0.016)	-0.006 (0.011)
Model 2: Restitution Support	-0.012 (0.039)	-0.026 (0.010)	0.012† (0.004)	-0.026 (0.037)	0.018 (0.005)	0.007 (0.004)
Model 3: Public Esteem	0.075 (0.057)	0.035† (0.017)	-0.013* (0.006)	0.097 (0.055)	-0.022* (0.008)	-0.004 (0.005)
Model 4: Public Transparency	0.126† (0.082)	0.004 (0.013)	-0.002 (0.006)	0.129† (0.081)	-0.003 (0.008)	-0.001 (0.004)
Model 5: Prosocial Motivation Continuance	0.216* (0.070)	0.015 (0.011)	-0.006 (0.004)	0.225* (0.070)	-0.010 (0.006)	-0.001 (0.003)

Notes: Coefficient estimates are standardized. Results for the conditional serial mediation paths are presented at high and low levels of the moderator, i.e., moral stigma (± 1 SD).

† $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

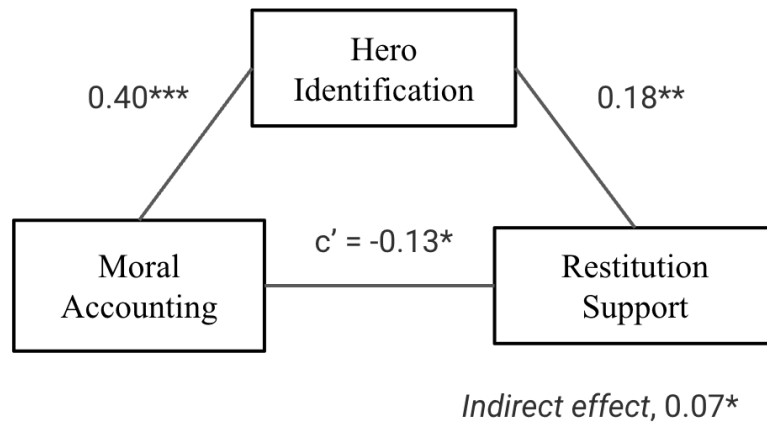


Figure 3. Study 3's Mediation Results for Outcome Variable Restitution Support.

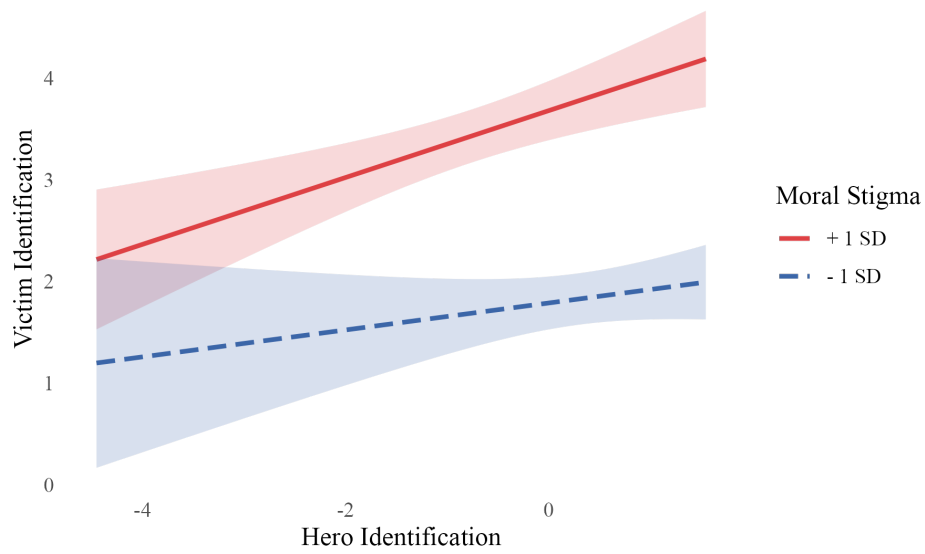


Figure 4. Simple Slopes for Hero Identification’s Effect on Victim Identification at High and Low Levels of Perceived Moral Stigma. This figure visualizes the d_3 path tested in Hypothesis 4, showing how moral stigma influences the relationship between hero identification and victim identification, including covariates. Lines represent ± 1 SD from the mean of stigma with 95% confidence intervals.

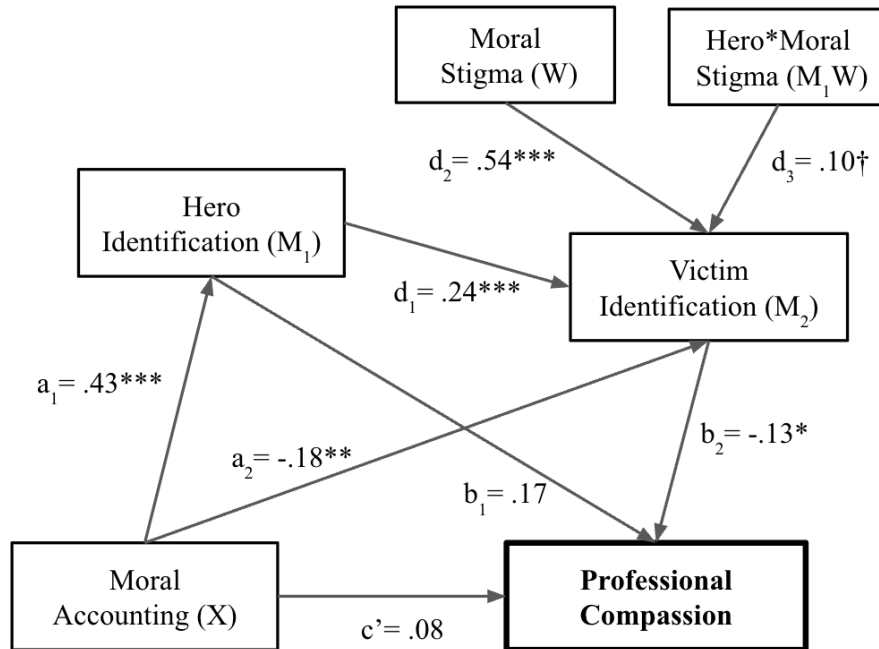


Figure 5. Direct Effects from Study 3's Conditional Serial Mediation Model Predicting Professional Compassion.

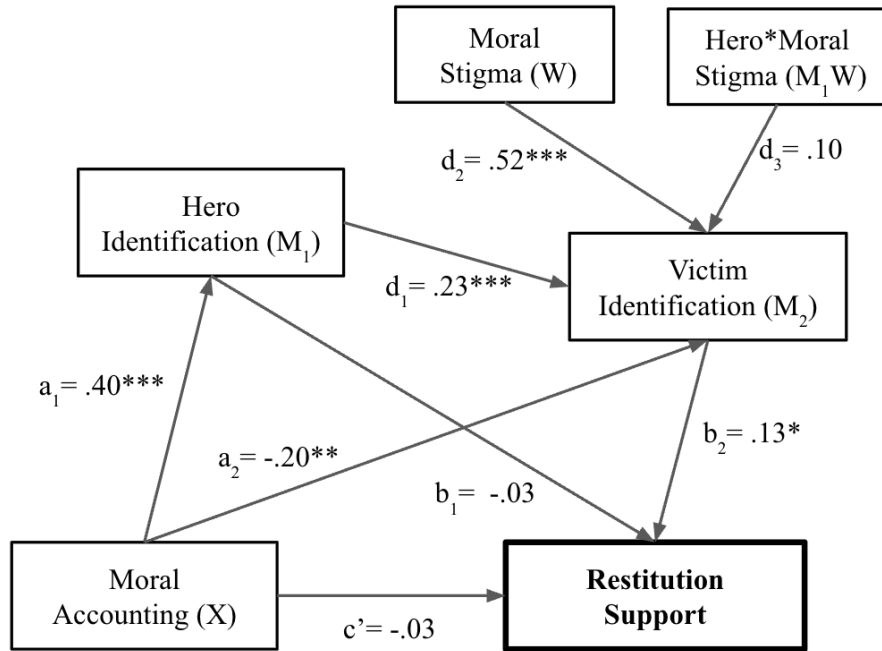


Figure 6. Direct Effects from Study 3's Conditional Serial Mediation Model Predicting Restitution Support.

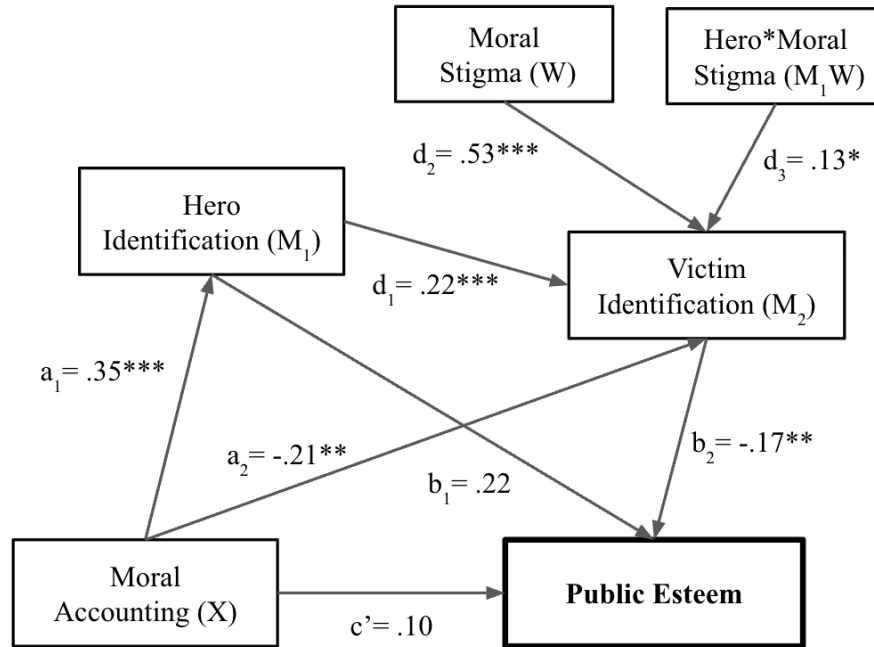


Figure 7. Direct Effects from Study 3's Conditional Serial Mediation Model Predicting Public Esteem.

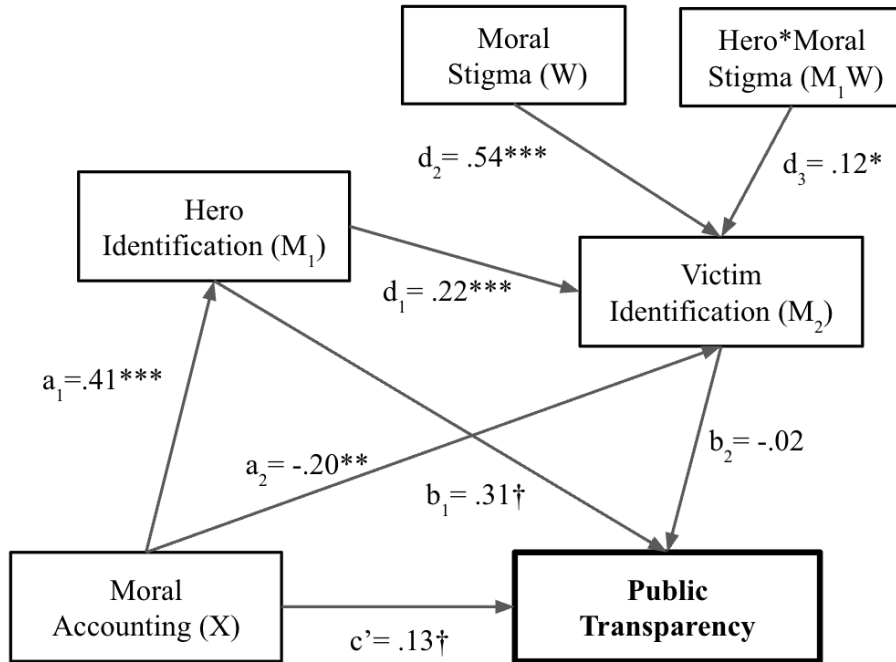


Figure 8. Direct Effects from Study 3's Conditional Serial Mediation Model Predicting Public Transparency.

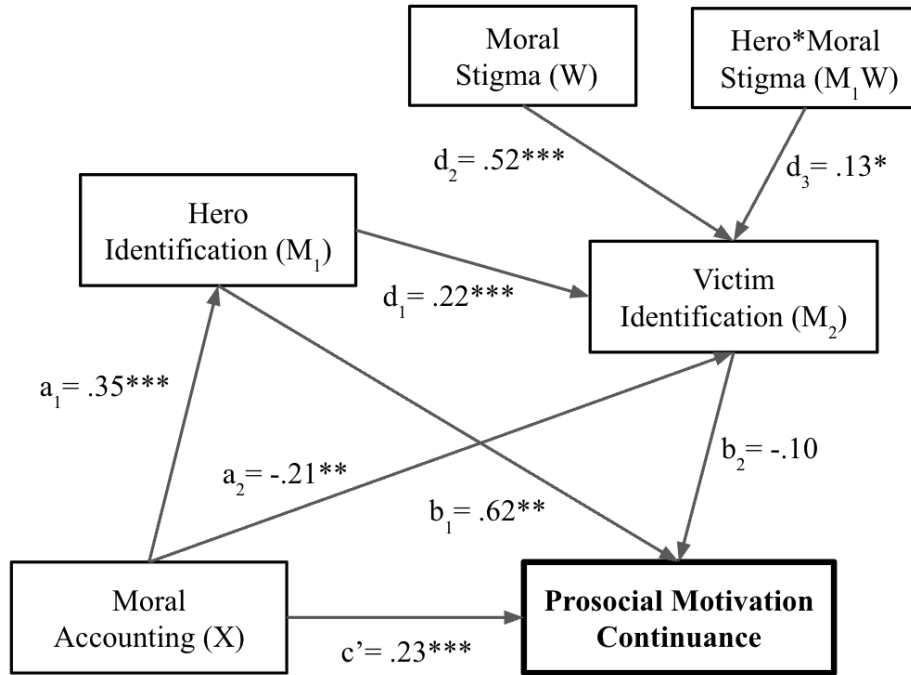


Figure 9. Direct Effects from Study 3's Conditional Serial Mediation Model Predicting Prosocial Motivation Continuance.

References

- Abraham, R. (1999). Emotional dissonance in organizations: conceptualizing the roles of self-esteem and job-induced tension. *Leadership & Organization Development Journal*, 20(1), 18-25.
- Allison, S. T., & Goethals, G. R. (2010). *Heroes: What they do and why we need them*. Oxford University Press.
- Andersson, L. M., & Pearson, C. M. (1999). Tit for tat? The spiraling effect of incivility in the workplace. *Academy of management review*, 24(3), 452-471.
- Andiappan, M., & Dufour, L. (2017). A difficult burden to bear: the managerial process of dissonance resolution in the face of mandated harm-doing. *Journal of business ethics*, 141, 71-86.
- Aquino, K., Freeman, D., Reed II, A., Lim, V. K., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: the interactive influence of situations and moral identity centrality. *Journal of personality and social psychology*, 97(1), 123.
- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of personality and social psychology*, 83(6), 1423.
- Aquino, K., McFerran, B., & Laven, M. (2011). Moral identity and the experience of moral elevation in response to acts of uncommon goodness. *Journal of personality and social psychology*, 100(4), 703.
- Aronson, E. (1992). The return of the repressed: Dissonance theory makes a comeback. *Psychological inquiry*, 3(4), 303-311.
- Ashforth, B. E., Harrison, S. H., & Corley, K. G. (2008). Identification in organizations: An examination of four fundamental questions. *Journal of management*, 34(3), 325-374.
- Ashforth, B. E., & Kreiner, G. E. (1999). "How can you do it?": Dirty work and the challenge of constructing a positive identity. *Academy of management Review*, 24(3), 413-434.
- Ashforth, B. E., & Kreiner, G. E. (2013). Profane or profound? Finding meaning in dirty work.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1-26.
- Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ*, 1986(23-28), 2.

- Barraza, J. A., Alexander, V., Beavin, L. E., Terris, E. T., & Zak, P. J. (2015). The heart of the story: Peripheral physiology during narrative exposure predicts charitable giving. *Biological psychology*, *105*, 138-143.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, *141*(4), 757.
- Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices?. *Psychological Science*, *18*(1), 24-28.
- Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral judgment and decision making. *The Wiley Blackwell handbook of judgment and decision making*, *2*, 478-515.
- Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices?. *Psychological Science*, *18*(1), 24-28.
- Baumeister, R. F., & Newman, L. S. (1994). How stories make sense of personal experiences: Motives that shape autobiographical narratives. *Personality and Social Psychology Bulletin*, *20*(6), 676-690.
- Becker, H. S. (1971) *Sociological Work: Method and Substance*, London: Allen Lane.
- Bentein, K., Garcia, A., Guerrero, S., & Herrbach, O. (2017). How does social isolation in a context of dirty work increase emotional exhaustion and inhibit work engagement? A process model. *Personnel Review*, *46*(8), 1620-1634.
- Caza, B. B., & Creary, S. (2016). The construction of professional identity. In *Perspectives on contemporary professional work* (pp. 259-285). Edward Elgar Publishing.
- Chen, H. C., Lin, S. Y., & Chen, I. H. (2024). Image discrepancy around professional work matters: a moderated mediation study. *Current Psychology*, *43*(8), 6840-6859.
- Cho, Y., & Fast, N. J. (2012). Power, defensive denigration, and the assuaging effect of gratitude expression. *Journal of Experimental Social Psychology*, *48*(3), 778-782.
- Chopova, T., & Ellemers, N. (2023). The importance of morality for collective self-esteem and motivation to engage in socially responsible behavior at work among professionals in the finance industry. *Business Ethics, the Environment & Responsibility*, *32*(1), 401-414.

- Clair, J. A., & Dufresne, R. L. (2004). Playing the grim reaper: How employees experience carrying out a downsizing. *Human relations*, 57(12), 1597-1625.
- Coldwell, D. A., Billsberry, J., Van Meurs, N., & Marsh, P. J. (2008). The effects of person–organization ethical fit on employee attraction and retention: Towards a testable explanatory model. *Journal of business ethics*, 78, 611-622.
- Cooley, C. H. (1902). The looking-glass self. *The production of reality: Essays and readings on social interaction*, 6(1902), 126-28.
- Côté, S., Piff, P. K., & Willer, R. (2013). For whom do the ends justify the means? Social class and utilitarian moral judgment. *Journal of personality and social psychology*, 104(3), 490.
- Crocker, J., & Knight, K. M. (2005). Contingencies of self-worth. *Current directions in psychological science*, 14(4), 200-203.
- Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003). Contingencies of self-worth in college students: theory and measurement. *Journal of personality and social psychology*, 85(5), 894.
- Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological bulletin*, 130(3), 392.
- Debono, A., Corley, N., & Muraven, M. (2020). Why am I left out? Interpretations of exclusion affect antisocial and prosocial behaviors. *The American Journal of Psychology*, 133(1), 63-79.
- Dierdorff, E. C., & Morgeson, F. P. (2013). Getting what the occupation gives: Exploring multilevel links between work design and occupational values. *Personnel Psychology*, 66(3), 687-721.
- Donaldson, T., & Dunfee, T. W. (2002). Ties that bind in business ethics: Social contracts and why they matter. *Journal of Banking & Finance*, 26(9), 1853-1865.
- Dong, M., van Prooijen, J. W., & van Lange, P. A. (2019). Self-enhancement in moral hypocrisy: Moral superiority and moral identity are about better appearances. *PloS one*, 14(7), e0219382.
- Dunning, D. (2007). Self-image motives and consumer behavior: How sacrosanct self-beliefs sway preferences in the marketplace. *Journal of consumer psychology*, 17(4), 237-249.
- Ellemers, N., & de Gilder, D. (2024). Overcoming unintended consequences of social

- impact accountability: How to avoid counterproductive responses of individuals and groups in organizations. *Social Issues and Policy Review*, 18(1), 194-219.
- Eriksson, K. (2019). Self-stigma, bad faith and the experiential self. *Human Studies*, 42(3), 391-405.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59(1), 255-278.
- Fagin-Jones, S. (2024). Prosocialization of Heroism. In *Encyclopedia of Heroism Studies* (pp. 1645-1649). Cham: Springer International Publishing.
- Festinger, L. (1957). Social comparison theory. *Selective Exposure Theory*, 16(401), 3.
- Fineman, S. (2006). On being positive: Concerns and counterpoints. *Academy of management review*, 31(2), 270-291.
- Franco, Z. E., Allison, S. T., Kinsella, E. L., Kohen, A., Langdon, M., & Zimbardo, P. G. (2018). Heroism research: A review of theories, methods, challenges, and trends. *Journal of humanistic psychology*, 58(4), 382-396.
- Franco, Z. E., Blau, K., & Zimbardo, P. G. (2011). Heroism: A conceptual analysis and differentiation between heroic action and altruism. *Review of general psychology*, 15(2), 99-113.
- Frisch, D., & Baron, J. (1988). Ambiguity and rationality. *Journal of Behavioral Decision Making*, 1(3), 149-157.
- Gausel, N., & Leach, C. W. (2011). Concern for self-image and social image in the management of moral failure: Rethinking shame. *European Journal of Social Psychology*, 41(4), 468-478.
- Gausel, N., Leach, C. W., Vignoles, V. L., & Brown, R. (2012). Defend or repair? Explaining responses to in-group moral failure by disentangling feelings of shame, rejection, and inferiority. *Journal of personality and social psychology*, 102(5), 941.
- Goffman, E. (1963). Embarrassment and Social Organization. In N. J. Smelser & W. T. Smelser (Eds.), *Personality and social systems* (pp. 541–548). John Wiley & Sons, Inc.
- Goranson, A., O’Fallon, C., & Gray, K. (2022). The moral identity picture scale (MIPS): Measuring the full scope of moral identity. *Self and Identity*, 21(6), 609-643.
- 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.
- 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.
- Gray, K., & Kubin, E. (2024). Victimhood: The most powerful force in morality and politics. *Advances in Experimental Social Psychology*.
- Haidt, J. (2007). The new synthesis in moral psychology. *science*, 316(5827), 998-1002.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog?. *Journal of personality and social psychology*, 65(4), 613.
- Hall, B., & Derryberry, W. P. (2010). Are aversive racists distinguishable from those with high explicit racial prejudice. *Beliefs & Values*, 2(2), 138-153.
- Harvey, J. H., & Weary, G. (1984). Current issues in attribution theory and research. *Annual review of psychology*, 35(1), 427-459.
- Hayes, A. F. (2013). *Methodology in the social sciences. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. London: Guilford Publications..
- Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social psychology*, 60(6), 895.
- House, R. J., & Rizzo, J. R. (1972). Toward the measurement of organizational practices: Scale development and validation. *Journal of Applied Psychology*, 56(5), 388.
- Hughes, E. C. (1962). Good people and dirty work. *Social problems*, 10(1), 3-11.
- Kelley, H. H., & Thibaut, J. (1978). *Interpersonal relations: A theory of interdependence*. New York Wiley.
- Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, 342(6156), 377-380.
- Kinsella, E. L., Ritchie, T. D., & Igou, E. R. (2015). Lay perspectives on the social and psychological functions of heroes. *Frontiers in psychology*, 6, 126986.

- Kohn, M. L. (1976). Occupational structure and alienation. *American Journal of Sociology*, 82(1), 111-130.
- Korman, A. K. (1970). Toward a hypothesis of work behavior. *Journal of Applied psychology*, 54(1p1), 31.
- Kreiner, G., Mihelcic, C. A., & Mikolon, S. (2022). Stigmatized work and stigmatized workers. *Annual Review of Organizational Psychology and Organizational Behavior*, 9(1), 95-120.
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of personality and social psychology*, 93(2), 234.
- Lee, E., Lewis-Liu, T., Khurana, S., & Lu, M. (2024). A systematic review of the link between public service motivation and ethical outcomes. *Asia Pacific Journal of Public Administration*, 46(3), 236-263.
- Lees, J., Young, L., & Waytz, A. (2022). Morally questionable actors' meta-perceptions are accurate but overly positive. *Journal of Experimental Social Psychology*, 102, 104371.
- Major, B., & O'Brien, L. T. (2005). The social psychology of stigma. *Annu. Rev. Psychol.*, 56, 393-421.
- Margolis, J. D., & Molinsky, A. (2008). Navigating the bind of necessary evils: Psychological engagement and the production of interpersonally sensitive behavior. *Academy of management journal*, 51(5), 847-872.
- Matza, D., & Sykes, G. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review*, 22(6), 664-670.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological methods*, 12(1), 23.
- Mead, G. H. (1934). *Mind, self & society*. University of Chicago press.
- Mele, A. (1987). Recent work on self-deception. *American Philosophical Quarterly*, 24(1), 1-17.
- Mishra, M., Ghosh, K., & Sharma, D. (2021). Unethical pro-organizational behavior: A systematic review and future research agenda. *Journal of Business Ethics*, 1-25.
- Molinsky, A., & Margolis, J. (2005). Necessary evils and interpersonal sensitivity in organizations. *Academy of Management Review*, 30(2), 245-268.

- Molinsky, A. L., Grant, A. M., & Margolis, J. D. (2012). The bedside manner of homo economicus: How and why priming an economic schema reduces compassion. *Organizational behavior and human decision processes*, 119(1), 27-37.
- Monin, B., & Jordan, A. H. (2009). The Dynamic Moral Self: A Social Psychological Perspective.
- Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual review of psychology*, 67(1), 363-385.
- Nandkeolyar, A. K., Shaffer, J. A., Li, A., Ekkirala, S., & Bagger, J. (2014). Surviving an abusive supervisor: The joint roles of conscientiousness and coping strategies. *Journal of Applied Psychology*, 99(1), 138.
- Owens, B. P., Yam, K. C., Bednar, J. S., Mao, J., & Hart, D. W. (2019). The impact of leader moral humility on follower moral self-efficacy and behavior. *Journal of Applied Psychology*, 104(1), 146.
- Pagliaro, S., Ellemers, N., Barreto, M., & Di Cesare, C. (2016). Once dishonest, always dishonest? The impact of perceived pervasiveness of moral evaluations of the self on motivation to restore a moral reputation. *Frontiers in psychology*, 7, 586.
- Parmar, B. (2014). From intrapsychic moral awareness to the role of social disruptions, labeling, and actions in the emergence of moral issues. *Organization Studies*, 35(8), 1101-1126.
- Patil, S. V. (2019). "The public doesn't understand": The self-reinforcing interplay of image discrepancies and political ideologies in law enforcement. *Administrative Science Quarterly*, 64(3), 737-769.
- Patil, S. V., & Lebel, R. D. (2019). "I want to serve but the public does not understand:" Prosocial motivation, image discrepancies, and proactivity in public safety. *Organizational Behavior and Human Decision Processes*, 154, 34-48.
- Puleston, J. (2013). Gamification of market research. *Social media, sociality, and survey research*, 253-293.
- Radtke, R. R. (2008). Role morality in the accounting profession—how do we compare to physicians and attorneys?. *Journal of Business Ethics*, 79, 279-297.
- Reid, E. (2015). Embracing, passing, revealing, and the ideal worker image: How people navigate expected and experienced professional identities. *Organization science*, 26(4), 997-1017.
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological science*, 20(4), 523-528.

- Schaubroeck, J. M., Lam, L. W., Lai, J. Y., Lennard, A. C., Peng, A. C., & Chan, K. W. (2018). Changing experiences of work dirtiness, occupational disidentification, and employee withdrawal. *Journal of Applied Psychology, 103*(10), 1086.
- Schwarzer, R., & Luszczynska, A. (2008). Self efficacy. *Handbook of positive psychology assessment, 2*(0), 7-217.
- Shaw, W. (2006). The consequentialist perspective. *Contemporary debates in moral theory, 5-20*.
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in experimental social psychology, 38*, 183-242.
- Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of personality and social psychology, 94*(1), 116.
- Shukla, J., Stein, C., Bush, J. T., & Janardhanan, N. S. (2024). Should I do this? Incongruence in the face of conflicting moral and role expectations. *Human Resource Management Review, 34*(2), 101010.
- Shulman, D., Shnitzer-Akuka, M., & Reifen-Tagar, M. (2021). The cost of attributing moral blame: Defensiveness and resistance to change when raising awareness to animal suffering in factory farming. *Plos one, 16*(8), e0254375.
- Shultz, T. R., & Darley, J. M. (2014). An information-processing model of retributive moral judgments based on “legal reasoning”. In *Handbook of moral behavior and development* (pp. 247-278). Psychology Press.
- Singer, P. (1991). A refutation of ordinary morality. *Ethics 101*(3), 625-633.
- Sinnott-Armstrong, W. (2008). Is moral phenomenology unified?. *Phenomenology and the Cognitive Sciences, 7*, 85-97.
- Skitka, L. J., Bauman, C. W., & Lytle, B. L. (2009). Limits on legitimacy: moral and religious convictions as constraints on deference to authority. *Journal of personality and social psychology, 97*(4), 567.
- Smith, I. H., & Kouchaki, M. (2021). Ethical learning: The workplace as a moral laboratory for character development. *Social Issues and Policy Review, 15*(1), 277-322.
- Snow, D. A., & Benford, R. D. (1988). Ideology, frame resonance, and participant mobilization. *International social movement research, 1*(1), 197-217.

- Snow, D. L., Swan, S. C., Raghavan, C., Connell, C. M., & Klein, I. (2003). The relationship of work stressors, coping and social support to psychological symptoms among female secretarial employees. *Work & Stress, 17*(3), 241-263.
- Spreitzer, G. M. (1995). Psychological empowerment in the workplace: Dimensions, measurement, and validation. *Academy of management Journal, 38*(5), 1442-1465.
- Stanley, M. L., & Kay, A. C. (2024). The consequences of heroization for exploitation. *Journal of Personality and Social Psychology, 126*(1), 5.
- Stanley, M. L., Shepherd, S., & Kay, A. C. (2023). Heroization and ironic funneling effects. *Journal of Personality and Social Psychology, 125*(1), 29.
- Stanley, M. L., Yin, S., & Sinnott-Armstrong, W. (2019). A reason-based explanation for moral dumbfounding. *Judgment and Decision making, 14*(2), 120-129.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In *Advances in experimental social psychology* (Vol. 21, pp. 261-302). Academic Press.
- Tappin, B. M., & McKay, R. T. (2017). The illusion of moral superiority. *Social psychological and personality science, 8*(6), 623-631.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin, 103*(2), 193.
- Tesser, A. (2000). On the confluence of self-esteem maintenance mechanisms. *Personality and Social Psychology Review, 4*(4), 290-299.
- 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.
- Tetlock, P. E., & Levi, A. (1982). Attribution bias: On the inconclusiveness of the cognition-motivation debate. *Journal of Experimental Social Psychology, 18*(1), 68-88.
- Thai, M., Wenzel, M., & Okimoto, T. G. (2023). Turning tables: Offenders feel like “victims” when victims withhold forgiveness. *Personality and social psychology bulletin, 49*(2), 233-250.
- Tsang, A. H. (2002). Strategic dimensions of maintenance management. *Journal of Quality in Maintenance Engineering, 8*(1), 7-39.

- Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision making*, 4(6), 479-491.
- Umphress, E. E., Bingham, J. B., & Mitchell, M. S. (2010). Unethical behavior in the name of the company: the moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior. *Journal of applied psychology*, 95(4), 769.
- Umphress, E. E., & Bingham, J. B. (2011). When employees do bad things for good reasons: Examining unethical pro-organizational behaviors. *Organization science*, 22(3), 621-640.
- Van Lange, P. A. M., & Balliet, D. (2015). Interdependence theory. In M. Mikulincer, P. R. Shaver, J. A. Simpson, & J. F. Dovidio (Eds.), *APA handbook of personality and social psychology, Vol. 3. Interpersonal relations* (pp. 65–92). American Psychological Association.
- Verplanken, B., & Holland, R. W. (2002). Motivated decision making: effects of activation and self-centrality of values on choices and behavior. *Journal of personality and social psychology*, 82(3), 434.
- Vogel, D., & Willems, J. (2020). The effects of making public service employees aware of their prosocial and societal impact: A microintervention. *Journal of Public Administration Research and Theory*, 30(3), 485-503.
- Voisin, D., & Fointiat, V. (2013). Reduction in cognitive dissonance according to normative standards in the induced compliance paradigm. *Social psychology*.
- Walker, L. J. (2016). The moral character of heroes. In *Handbook of heroism and heroic leadership* (pp. 121-141). Routledge.
- Wenzel, M., Woodyatt, L., & McLean, B. (2020). The effects of moral/social identity threats and affirmations on psychological defensiveness following wrongdoing. *British Journal of Social Psychology*, 59(4), 1062-1081.
- West, C., & Zhong, C. B. (2015). Moral cleansing. *Current Opinion in Psychology*, 6, 221-225.
- Wilk, S. L., & Sackett, P. R. (1996). Longitudinal analysis of ability–job complexity fit and job change. *Personnel Psychology*, 49(4), 937–967.
- Wyatt, T. R., Rockich-Winston, N., White, D., & Taylor, T. R. (2021). “Changing the narrative”: a study on professional identity formation among Black/African American physicians in the US. *Advances in Health Sciences Education*, 26(1), 183-198.