5-2022

The vigilante identity and organizations

Fan Xuan CHEN
University of Illinois at Urbana-Champaign

Maja GRASO
University of Otago

Karl AQUINO
University of British Columbia

Lily LIN
Simon Fraser University

Joey T. CHENG
York University

See next page for additional authors

THE VIGILANTE IDENTITY IN ORGANIZATIONS

ABSTRACT

We test the theoretical and practical utility of the *vigilante identity*, a self-perception of being the kind of person who monitors their environment for signs of norm violations, and who punishes the perceived norm violator, without formal authority. We develop and validate a measure of the vigilante identity scale (VIS) and demonstrate the scale's incremental predictive validity above and beyond seemingly related constructs (Studies 1 – 2e). We show that the VIS predicts hypervigilance towards organizational wrongdoing (Study 2, 4), punishment intentions and behavior in and of organizations (Studies 3 and 4) as well as in the wider community (Study 1), and is activated under organizational justice failure conditions (Study 3). We maintain that vigilantes can impact organizations and society from both inside and outside organizational walls and we discuss theoretical implications for scholarship on vigilantes, as well as on morality, social norms, and third-party punishment in organizations.

Keywords: vigilantism; identity; punishment; social order; justice

Published in Organizational Behavior and Human Decision Processes, May 2022, 170, 104136, pp. 1-17. DOI: 10.1016/j.obhdp.2022.104136

Mythical tales often portray vigilantes as complex and troubled people who redirect the pain and anger they feel toward the goal of punishing wrongdoers, usually through illegal or violent means. The allure of vigilantism in the popular culture is evidenced by the public's interest in fictional characters like *Dexter,* Frank Castle (*The Punisher*), and Lisbeth Salander (*The Girl with the Dragon Tattoo*). But vigilantes are also real-life social actors who have been around since human beings began living together in communities. Examples of everyday vigilantes include a private citizen vandalizing a car for parking in a disabled space without a permit, a social media user shaming a colleague for circulating "unacceptable" thoughts in an internal company memo, or a group of community members shaming their neighbors for violating COVID-19 mandates. Vigilantes have been primarily studied by sociologists, criminologists, and political scientists (Bateson, 2020; Burrows, 1976; Jacobs et al., 2005; Johnston, 1996; Johnston, 1992; Moncada, 2017; Phillips, 2016). Although largely neglected in organizational research, a handful of scholars (Crawford & Dacin, 2020; DeCelles & Aquino, 2020; Tripp et al., 2007) have begun to theorize about their emergence and activities in organizations. In the most comprehensive theory to date, DeCelles and Aquino (2020) proposed a social control framework to speculate about the organizational and psychological conditions that could lead some employees to become vigilantes.

Research on workplace vigilantes is still in a nascent stage so it has been limited in several ways. First, it has focused primarily on vigilantes within the organization (Crawford & Dacin, 2020; Tripp et al., 2007). For instance, DeCelles and Aquino (2020) introduced the concept of the vigilante as adopting a particular kind of *work role* identity that involves monitoring and unilaterally punishing coworkers for perceived wrongdoing despite lacking the formal authority to do so. We extend this depiction of the insider vigilante by recognizing that social actors can monitor and punish organizations and employees from *outside* organizational walls. An emerging literature in the information sciences points to an

increasing use of social media to punish society members and employees, and tarnish organizations (e.g., “netilantism,” “online vigilantism,” or “diligilism”; Badaracco, 1997; Chang & Poon, 2016; Galleguillos, 2021; Solove, 2007). Moreover, outsider vigilantes are not subject to managerial oversight, influence, and accountability, making the conditions under which these vigilantes appear and operate different from those associated with insider vigilantes. Nonetheless, it is possible that common personal attributes or motives influence both insider and outsider vigilantes. We test this possibility and contribute to the literature by broadening both the definition and empirical focus of vigilantes by using the concept of a *personal* identity. A second limitation of the research on workplace vigilantes is that it consists primarily of theory-building or qualitative, deductive work. What has been missing is a study that introduces and validates a reliable measure of the proclivity to become a vigilante. To address this, in the current research we introduce and rigorously validate such a measure that assesses the strength at which an individual adopts the vigilante identity. Jointly, our new measure and conceptualization of vigilantism as a personal identity allow us to test novel predictions about why vigilantes might emerge and punish organizations or their employees.

Theoretical Foundations

What is the Vigilante Identity?

Our conception of the vigilante identity builds on the lay understanding of a vigilante as someone who routinely scans the environment for signs of crime and punishes those whom the law has failed to punish. The term *vigilante* entered the English language in the 19th century and describes a person who is a “member of a volunteer committee organized to suppress and punish crime summarily (as when the processes of law are viewed as inadequate); a self-appointed doer of justice” (*Merriam Webster Dictionary*). The etymology of the English term *vigilante* is the Spanish word of the same spelling that means

"watchman" or "guard". In their theory of the workplace vigilantism, DeCelles and Aquino (2020) defined a vigilante as *"*an employee who has taken on the self-appointed role identity of being a monitor and punisher of coworkers' deviance" (DeCelles & Aquino, 2020, p. 530). We retain core elements of DeCelles and Aquino's (2020) definition, but broaden its applicability by defining and measuring it at the level of a *personal* identity. This adjustment aligns with a basic assumption of contemporary theories (Brewer & Gardner, 1996; Brickson, 2000; Dutton et al., 2010; Skitka & Crosby, 2003) that identities can be construed at different levels of self-understanding. For example, Brewer and Gardner (1996) argued that an identity can be represented as self-knowledge at the personal, interpersonal or role-based (e.g., DeCelles & Aquino, 2020), or group level.

Our formal definition of someone who internalizes the vigilante identity at a personal level is that they hold a self-perception of being the kind of person who regularly monitors the behaviors of people around them for signs of norm violations, and who is willing to punish the perceived norm violator when they believe it is appropriate and justifiable, despite not having formal authority to do so. This definition aligns with scholarly descriptions of vigilantism as a form of autonomous, unsanctioned punishment (DeCelles & Aquino, 2020; Johnston, 1996; Rosenbaum & Sederberg, 1974). Punishing without having formal authority is an important criterion for determining whether a person is a vigilante. We use the term *unauthorized punishment* to mean that a person is not acting within a well-defined, hierarchically structured institutional role (Wenzel et al., 2008).[1] These requirements are consistent with how legal scholars (Robinson & Robinson, 2016; Robinson & Robinson, 2018) and ethicists (Dumsday, 2009) have described vigilantes' punishment activities as

[1] It is possible for people who formally occupy a disciplinary role to become "shadow vigilantes". Shadow vigilantes act within the system to administer their version of justice through ethically questionable or illegal means (Robinson & Robinson, 2016). An example would be an HR manager who intentionally ignores due process or refuses to consider exonerating evidence because they want to ensure that an employee who violated a norm is punished.

"extralegal", meaning they go "beyond the formal (albeit often idealized and permeable) boundaries of system legality" (Huggins, 1991, p. 6). Unauthorized punishment as enacted by vigilantes can range from relatively minor acts (e.g., a restaurant employee deliberately giving a rude customer an extra dose of hot sauce) to more severe and even illegal ones (e.g., damaging reputations by publicly disseminating false or slanderous information). However, this punishment would not include "authorized" punishment of deviants enacted through formal organizational processes (e.g., ethics committee disciplining a student found guilty of cheating, a supervisor disciplining an employee for committing fraud, or a regulatory agent disciplining an organization for falsely inflating profits).

Our definition stipulates that even if someone internalizes a vigilante identity, they do not indiscriminately punish. Following DeCelles and Aquino (2020), we assume they must be induced to punish by certain triggering conditions, which we discuss next.

Threats to Normative Order and Justice Failure

Vigilante scholarship suggests that two conditions often lead to the emergence of vigilantes: 1) perceptions that a group's normative order is under threat, and 2) perceptions that authorities or formal systems of social control (e.g., human resource officers, managers, security personnel, codes of conduct, and other institutional mechanisms) do not adequately punish wrongdoers. We will refer to the latter condition as a *justice failure* (e.g., Rosenbaum & Sederberg, 1974; Tenbrunsel et al., 2003; Zhu et al., 2012). Both conditions can give rise to vigilantes because they destabilize social order and diminish trust in authorities (Keizer et al., 2008; Ostrom, 2000). As result, actors other than those legitimately charged with the task of maintaining social order and dispensing justified punishments may take it upon themselves to compensate for the perceived inadequacy of various institutional control mechanisms (Burrows, 1976; Chui & Grieder, 2020; DeCelles & Aquino, 2020; Goldstein, 2003; Johnston, 1996; Johnston, 1992; Silke, 2001; Song, 2019). In these cases, the role they take

on is that of an informal punisher or third-party; someone who is not directly impacted by a norm violation but who witnesses or learns about it and decides to respond rather than remain a passive bystander (Fehr & Fischbacher, 2004).

Many vigilantes are not directly impacted by norm violations but are rather third-party punishers (Gürerk et al., 2006; Tomasello & Vaish, 2013). It is important to acknowledge that most people are unwilling to bear the risk of punishing deviants, even if the violation is severe or directed against them (Balafoutas et al., 2016). For this reason, employee bystanders or outsider third-parties might be even less likely to become vigilantes because of their greater distance from the violation. We submit that most observers will be content to let organizational authorities deal with deviance even if they are skeptical of the authorities' ability to deliver justice. Indeed, those who are cynical about the police and the criminal justice system still sometimes willingly turn to them for protection (Hagan et al., 2018). These observations raise an unanswered question that we sought to answer in our research: What kind of person can overcome the tendencies to remain a bystander or rely on authorities to deal with deviants? Our answer is that it partly depends on how strongly they have internalized a vigilante identity.

The Vigilante Identity and Its Behavioral Demands

An identity reflects a person's understanding of how their traits, beliefs, and characteristics influence and guide the way they perceive, think about, and react to different situations (Brewer & Gardner, 1996). These self-understandings are the abstract content of a vigilante identity from which more specific behavioral demands originate. Applying basic principles of identity theory, we propose that people who internalize the vigilante identity experience a strong compulsion to satisfy its behavioral demands because doing so is self-affirming (Markus & Kunda, 1986; Oyserman, 2009). In turn, satisfying these behavioral demands makes one's vigilante identity become more central to their sense of self (Burke &

Tully, 1977; McCall & L, 1978; Stryker, 1980) and more readily accessible for guiding subsequent information processing and goal pursuit (Higgins & Brendl, 1995).

Highly internalized identities possess trait-like attributes of relative stability in patterns of thought, emotion, and behavior associated with them (Markus & Kunda, 1986). However, unlike personality traits, identities are self-generated and therefore more malleable (Leavitt et al., 2016). Another property of an identity is that it can be activated by certain situational cues (e.g., witnessing a norm violation), and deactivated by others (Aquino et al., 2009; Higgins, 1989; Leavitt et al., 2016; Skitka, 2003). Based on identity activation principles, a strongly internalized vigilante identity is more likely to be readily accessible across situations (Markus & Kunda, 1986), which has implications for how frequently a person will perceive and react to events that provide opportunities to punish a presumed wrongdoer.

Monitoring as a Behavioral Demand

An essential characteristic of people who internalize the vigilante identity is a heightened state of attentiveness to detect deviations from what they consider to be expected behavior. People who internalize the vigilante identity should experience a self-generated imperative to commit mental resources to the task of monitoring the people around them to see if they are violating norms (Darley, 2009; Gray & Wegner, 2009; Haidt, 2013; Haidt & Kesebir, 2010; Turiel, 2006). Over time, this monitoring tendency can become automatic and habitual. Since detecting norm violations and judging them as unacceptable are necessary conditions for punishing a perceived wrongdoer, we assume that people who have internalized a vigilante identity will tend toward judging others' behavior as being right or wrong once it has been observed.

In addition to being attentive to norm violations, Saucier and Webster (2010) suggested that social vigilantes believe they have superior capabilities to arrive at 'correct'

moral appraisals of others' actions. Thus, DeCelles and Aquino (2020) argued that people who have adopted the vigilante identity are more likely to possess high levels of *moral certainty* which they defined as consisting of two related beliefs: 1) they know when something is self-evidently right or wrong, and 2) they are confident that they can make this determination in specific cases. The first attribute of moral certainty is akin to having strong *moral convictions*, which are inflexible and unassailable attitudes that are grounded in fundamental beliefs about right and wrong (Skitka et al., 2021; Skitka et al., 2015). The second attribute corresponds to what Haidt (2013) referred to as having a 'righteous mind.' People with a righteous mind are able to rationalize the punishment of presumed wrongdoers because they see themselves as incapable of committing moral error (Haidt, 2013). Studies support DeCelles and Aquino's (2020) argument by showing that the more a person is convinced that a norm violation is self-evidently wrong, the more strongly they believe that punishing the violator is the appropriate response (Skitka & Morgan, 2009) and the more supportive they are of vigilantism (Skitka et al., 2015).

Punishment as a Behavioral Demand

Once a person decides that a norm violation deserves moral condemnation, they can then determine how they should respond. For the would-be vigilante, punishing wrongdoers (vs. not punishing them) can be viewed as a potentially more authentic, self-expressive act that symbolizes their identity (Gollwitzer, 1986; Gollwitzer et al., 1982). Hofmann et al. (2018) provided some evidence for the self-affirming consequences of punishment by showing that a desire to punish a wrongdoer is associated with a "heightened sense of moral self-worth" and a momentary increase in well-being (p. 1702). These psychological benefits of punishment are more likely to be valued by people who have internalized the vigilante identity and therefore make punishing more appealing. However, it is possible that vigilantes might engage in other responses to a norm violation in addition to administering unauthorized

punishment (e.g., reporting the transgression to authorities, or directly communicating their displeasure to the norm violator).

Overview of the Studies

We hypothesized that people who have internalized a vigilante identity will be motivated to monitor their environment for signs of deviance and punish wrongdoers without authorization. We developed an initial vigilante identity scale (VIS; Study 1) to assess the extent to which people internalize these demands, which we subsequently modified to improve its psychometric properties (Studies 2a – 2d). We also showed how the vigilante identity is detectable to others, thus providing evidence that the vigilante is a recognizable social actor who exhibits certain patterns of behavior consistent with our definition (Studies 1 and 2e). Finally, we tested whether VIS predicts monitoring and punishing (Studies 3 – 4).

We diversified our samples using different recruiting procedures, including community-based advertising (Study 1), the MTurk platform (Studies 2a/b/d, 3, 4), a Facebook ad (Study 2c), and a market research firm (Study 2e). To ensure data quality, we imposed several inclusion criteria across these studies following the best practices recommended by Keith et al. (2017). In particular, we recruited MTurk participants via CloudResearch (Litman et al., 2017), where we limited our studies to those who passed the regular quality assessment required by CloudResearch and who had completed at least 100 tasks with an approval rate of at least 98%. Moreover, we embedded several attention check questions and excluded participants who failed any of these questions on their first attempt. Our data and analysis scripts, pre-registrations (where applicable), and study materials are available here: https://osf.io/ve7q2/?view_only=85383ed98a874aeaa84cd590427bc1cd

Study 1: The Vigilante Identity as a Predictor of COVID-19 Violation Punishment

The study was conducted in New Zealand during the first official Level 3 and 4 COVID-19 lockdown (between March – April, 2020), which lasted for a period of 7.5 weeks. During this time, New Zealand shut down all but the most essential services. It also required people to 'stay in their bubbles' (a 'bubble' was the nomenclature adopted in the official government communication to refer to units of people, such as a family or roommates, who were self-isolating together). Level 3 lockdown entailed similar restrictions but allowed for food deliveries.

We anticipated that the high collective compliance that the government demanded to try to eliminate the virus in New Zealand (Health, 2020) would make those who were noncompliant be perceived as potential threats to the well-being of their community. Since legitimate authorities (e.g., police or government health agencies) were not capable of monitoring every individual's behavior and punishing violators of virus containment policies (Cousins, 2020; Tait, 2020), we expected some private citizens to try and regulate the behavior of others by monitoring them for signs of disobeying government mandates and privately punishing people who violated safety guidelines (Osaki, 2020).

We tested the hypothesis that people who internalized the vigilante identity would be more likely to notice if others had violated self-isolation rules and deliver unauthorized punishment to rule violators. In addition, we anticipated that high VIS scorers would hold stronger punitive attitudes towards those who did not comply with mandates. We also sought to examine whether people who internalized the vigilante identity might acquire a prosocial, civic-minded, generally law-abiding reputation, rather than be seen as sadistic punishers who flaunted the rule of law for personal pleasure and satisfaction (Dumsday, 2009; Robinson & Robinson, 2016). To do so, we tested whether people who internalized the vigilante identity were perceived by others as generally good citizens. We recruited dyads and relied on self-

and other-report data to minimize the threat of social desirability bias and increase the confidence in our behaviour-based dependent variable during COVID-19 lockdowns.

Methods

Sample

Participants were a community sample based in a university town in New Zealand (33.7% men, M_{age} = 27.1, SD_{age} = 12.1). They were recruited from local outlets (e.g., student newspaper and parenting school groups) to take part in a two-wave study about their experiences during the COVID-19 government lockdown. To be eligible, participants had to be self-isolating with at least one other adult and be part of a bubble in which all participating members were willing to complete two surveys at two points of time. Bubble partners were people who were self-isolating together at that time. The primary nature of the relationships between the participating bubble partners was romantic (i.e., marital or romantic/non-marital partners; 50.4%), roommates or friends (20.5%), or family members (e.g., siblings, or an adult child and their parents; 29.1%). Not all eligible adults living together participated or completed all components of the study. Therefore, partner ratings were limited to those who officially participated in both parts. The final matched data was primarily based on dyads (there were four triads in our data, but those were reduced to dyads due to incomplete ratings).

Our final sample size was N = 170 nested within 85 dyads. Time 1 was administered towards the beginning of the first Level 4 lockdown (Weeks 2-3; March 2020) and Time 2 was administered during the Levels 3 and 4 lockdown (early April, 2020). We stopped data collection before the end date of Level 3 was announced to ensure that participants' responses were not influenced by affective changes in response to the prospects of exiting a strict lockdown. Participating bubbles received a $20 NZ gift certificate to a major local grocery store.

Procedures

At Time 1, participants completed the vigilante identity measure, reported the degree to which they noticed rule violations, and indicated their endorsement of general punitive acts towards wrongdoers. At Time 2, participants within a bubble rated each other on the extent to which their partner engaged in punishment and prosocial behavior. Participants' responses were anonymous and were linked across bubbles and time points through a series of self-generated codes (a mix of participants' street numbers, birthdays, and phone numbers, all of which were unknown to the researchers). Participants were instructed not to discuss their responses with others.

Measures[2]

VIS: Preliminary (Self-reported, Time 1). In this preliminary development of the VIS, we created 10 items reflecting the behavioral demands associated with being a vigilante as described in our conceptual definition. These items included the tendency to monitor the environment for signs of deviance, and one's willingness to rectify a social system's perceived failure in controlling the deviance by punishing norm violators while lacking the formal authority to do so. Participants were asked to indicate the extent to which each of the 10 items described their identity using the prompt of: "I am the kind of person who:" 1) Pays attention to make sure that others are doing the right thing; 2) Stays alert for signs of corruption or wrongdoing in society; 3) Makes sure other people act morally; 4) Rights wrongs whenever I see them; 5) Makes sure people are held accountable when they do something wrong; 6) Ensures that people who do something wrong get punished for it; 7) Protects people who are mistreated by punishing wrongdoers; 8) Enforces society's rules; 9)

[2] We administered additional measures to better understand New Zealanders' perceptions of the lockdown (e.g., perceptions of the government's responses and Covid-19 risks). Those were collected for exploratory purposes and are available from the corresponding author.

Makes sure people who do wrong don't get away with it; 10) Enforces justice in society. Participants provided their responses on a scale from 1 (*completely unlike me*) to 7 (*completely like me*)[3].

Detect Rule Violations (Self-reported, Time 1). We created six items to assess whether people who scored higher on the VIS would also be more likely to detect threats operationalized as rule violations within the context of COVID-19. Participants were asked to indicate: "How often do you see people breaking the self-isolation rules? It is ok to provide a general estimate." We selected all the possible contexts in which participants were able to observe such violations: 1) bubble, 2) immediate neighborhood, 3) around town, 4) in [your] community, 5) at the grocery store, and 6) as reported in the news. Participants were asked how frequently they had seen such violations on the scale from 1 (*I have never seen [this violation]*) to 8 (*constantly/two or more times every day*).

Endorsement of Punitive Practices for Violators of Self-Isolation Rules (Self-reported at Time 1). We created three items to measure the extent to which people supported punitive practices for violators of self-isolation rules in New Zealand. We selected items that were perceived as punitive and potentially controversial at the time of our data collection (April, 2020): 1) making a public list of people who do not self-isolate, 2) automatic fines ($500 or above) for people who do not self-isolate, and 3) mandatory tracking systems for those who should be self-isolating. Participants indicated their likelihood of supporting each of those measures on the scale from 1 (*extremely unlikely)* to 8 (*extremely*

[3] We conducted a separate study for scale validation. Specifically, we recruited 229 participants (65.9% men, M_{age} = 33.5, *SD* = 10.2) from Amazon's Mechanical Turk (MTurk), who completed the VIS embedded in a survey with other measures unrelated to the present study to assess its factor structure. EFA results based on a maximum likelihood extraction method showed that a one-factor solution best fit the data, with the first latent factor accounting for 64% of the total variance. Standardized factor loadings ranged from .72 to .85. The items also had high reliability (α = .95). These results provided preliminary evidence that the items have adequate psychometric properties. See SOM for our other preliminary scale validation studies and evidence for its psychometric adequacy.

likely).

Perceived Monitoring and Punishment Behavior (Partner-rated at Time 2). We created a scale for an observer (i.e., partner) to report perceived monitoring and unauthorized punitive behaviors of social control by their partner during COVID-19 lockdowns. Specifically, we used the following prompt: "The following are some behaviors that people may demonstrate when they see somebody breaking the C19 rules". Participants were asked to indicate how well these statements described their partner on a scale from 1 (*does not describe them at all*) to 5 (*describes them extremely well*). Three of the items measured monitoring behavior (e.g., "*monitors others in our immediate community in case of others' violating the C19 mandates*") and four of the items assessed perceived punishment behavior (e.g., "*reports others breaking the rules publicly (such as via Facebook, Twitter, Letter to the Editor, blogs, etc.*"). The full set of items is provided in Appendix A.[4]

Perceived Prosocial Behavior (Partner-rated at Time 2). To examine whether people who internalize a vigilante identity (based on their self-reports) acquire a prosocial reputation, we adapted an 8-item organizational citizenship scale (Lee & Allen, 2002). Participants rated the extent to which their partner engaged in prosocial behaviors during the lockdown on a 5-point scale (1 = *does not describe them at all*, 5 = *describes them extremely well*). A sample item was: "[the partner] goes out of the way to make others feel better" (see Appendix A for all items).

Results

Descriptive statistics, reliabilities, and correlations are shown in Table 1.

[4] We initially developed a total of 8 items. However, we dropped one item (i.e., "*Complains to me when others violate the rules*") as it was unclear which dimension underlies this item. If we classified this item as part of the monitoring scale, results showed that self-rated VIS did not significantly predict partner-rated monitoring behavior $\gamma = 0.17$, $SE = 0.10$, $p = .101$ (but nevertheless in the predicted direction). When this item was included as part of the punishment scale, self-rated VIS continued to predict partner-rated punishment behavior, $\gamma = 0.16$, $SE = 0.06$, $p = .006$.

Table 1

Descriptive Statistics and Correlations Among Variables in Study 1

			1	2	3	4	5	6
1	VIS (T1)		(.87)					
2	Partner-rated Monitoring Behavior (T2)		.20*	(.86)				
3	Partner-rated Punishment Behavior (T2)		19*	56**	(.80)			
4	Partner-rated Prosocial Behavior (T2)		.25**	.06	.12	(.93)		
5	Detection of Violations (T1)		.15	.05	.26**	-.06	(.81)	
6	Punishment Endorsement (T1)		.41**	.18*	.25**	.20*	.27**	(.63)
		M	4.32	2.60	1.36	3.21	2.76	4.60
		SD	0.93	1.06	0.59	0.96	1.35	1.60
		N	146	118	117	121	145	145

Notes. *$p < .05$; **$p < .01$. VIS = Vigilante Identity Scale. Values on the diagonal represent the Cronbach's α. T1 and 2 = Time 1 and 2.

Hypothesis Testing

To account for nonindependence arising due to dyads, we conducted our analysis using a hierarchical linear modeling (HLM) framework with random intercepts and standard errors clustered within dyads. Regressing noticed violations on self-rated VIS, we found no significant effects of VIS (although it was in the predicted direction and approached significance), $\gamma = 0.21$, $SE = 0.12$, 95% CI [-0.03, 0.45], $t(144.75) = 1.75$, $p = .083$. However, we observed a strong effect of self-rated VIS on self-rated punishment endorsement, $\gamma = 0.63$, $SE = 0.13$, 95% CI [0.38, 0.88], $t(142.72) = 4.91$, $p < .001$.

Next, we examined whether self-rated VIS was related to partner-rated perceived monitoring, punishment, and prosocial behaviors. Consistent with our expectations, self-rated VIS at Time 1 significantly predicted partner-rated monitoring ($\gamma = 0.22$, $SE = 0.10$, 95% CI [0.02, 0.43], $t(117) = 2.17$, $p = .032$), punishment ($\gamma = 0.12$, $SE = 0.06$, 95% CI [0.01, 0.23], $t(116) = 2.05$, $p = .043$), and prosocial ($\gamma = 0.26$, $SE = 0.09$, 95% CI [.08, .44], $t(120.89) = 2.78$, $p = .006$) behavior at Time 2.

Discussion

Our first study examined whether the vigilante identity predicts people's actual behavior in a real-world situation where norm violations are known to be occurring. Accordingly, we selected COVID-19 as a focal context. While we did not measure organizational factors specifically, the public's and employees' compliance (or lack thereof) with governmental orders to mitigate COVID-19 have significant relevance for many types of communities and organizations. Supporting our prediction, people who scored higher on the vigilante identity were reported by partners as engaging in more vigilante acts, such as they were seen as engaging in more monitoring as well as punishment behaviors in relation to COVID-19. They were also rated by partners as prosocial and helpful towards others in their communities. However, we did not find evidence that the vigilante identity predicted self-reported detection of violations (r = .15, p = .072, *ns*).

There may be two reasons why we did not observe a significant relationship between vigilante identity and self-reported detection of violations. The first reason is conceptual: detection of violations in one's community may not equate with *active* monitoring for signs of deviance. The second reason pertains to a possible psychometric inadequacy of our measure. Specifically, the majority of the items assessed the punishment rather than the monitoring behavioral demand component of the vigilante identity. This lack of *content validity* may be one of the reasons why our preliminary measure failed to reliably detect differences in behaviors that might be driven by the monitoring component.[5] For this reason, we sought to improve the measurement properties of our original vigilante identity scale before conducting further studies. This refinement of our measure to address its potential

[5] Our effort to improve our measure was inspired by a reviewer's comments and the Associate Editor's suggestions that the prevalence of punishment items in the scale may have made it difficult to conduct a strong test of the possibility that the VIS is comprised of two as opposed to a single underlying factor. We performed this comparative analysis with our new, more balanced scale and continued to find that a single-factor underlies the items.

weaknesses is in accord with Tellegen and Waller (2008) observation that scale development often involves iterative steps where the cycle of construct re-formulation, item generation, data collection, and analysis can be repeated as often as it is productive (p. 262).

We followed a multi-step process outlined by Hinkin (1998) to validate the revised VIS. In Studies 2a – d, we provided evidence of our scale's content validity (Study 2a), its psychometric adequacy (internal consistency and dimensionality; Studies 2b and c), and its nomological network (Study 2d). We conclude our validation process by showing that one's vigilante identity is observable by others (Study 2e).

Study 2a: Item Reduction and Content Validation

Our revised scale consisted of 17 items designed to assess the behavioral demands associated with our definition of being a vigilante. Ten of the items were taken from the original scale reported in Study 1 (8 punishment and 2 monitoring). We added 7 new monitoring items to increase coverage of this component of the vigilante identity. The 17 items are presented in Table 2. We then followed procedures recommended by Colquitt et al. (2019) to determine which items showed the highest degree of content validity. We examined the *definitional correspondence* and *definitional distinctiveness* of each item using approaches introduced by both Anderson and Gerbing (1991) and Hinkin and Tracey (1999). Across two samples (A and B), we calculated four indicators of content validity: 1) the proportion of substantive agreement (p_{sa}), 2) the substantive validity coefficient (c_{sv}), 3) the Hinkin-Tracey correspondence (*htc*), and 4) the Hinkin-Tracey distinctiveness (*htd*). We inspected each item's scores on these four indicators and compared them with cut-off criteria (Colquitt et al., 2019). We retained items that met at least three standards of "very strong evidence" for content validity across these validity indicators.

Sample and Procedure

Sample A

A total of 260 MTurk participants (48.7% men, M_{age} = 40.9, *SD* = 13.0) completed this study. They had an average of 18.8 years (*SD* = 12.6) of work experience, and more than 50% of them held a bachelor's degree or higher. They completed a Q-sort task, where they provided data for the p_{sa} and c_{sv}.

In a typical Q-sort task (Anderson & Gerbing, 1991), participants are provided with conceptual definitions of several constructs (including the focal construct of interest and other orbiting constructs), along with their respective items, and they are asked to place each item into the construct dimension that best represented that item. Orbiting constructs are those that are expected to correlate with the focal construct of interest. We selected three orbiting constructs that we expected to correlate with the VIS: physical aggression, kindness, and extraversion. Each of these constructs contain element(s) that one could expect in a vigilante (for they are non-mutually exclusive), as their unauthorized punishment can effectively be considered aggressive, but nevertheless prosocial in nature, and is partly driven by socially confident and outgoing underlying tendencies. Specifically, we gave participants the definitions of vigilantism (i.e., "*The act of regularly monitoring the behaviors of people for signs of wrongdoing and punishing the perceived wrongdoer even when having no formal authority to do so*"; DeCelles & Aquino, 2020), physical aggression ("*Behavior that will physically harm others*"; Buss & Perry, 1992), kindness ("*The quality of being friendly, generous, and considerate*"; Canter et al., 2017), and extraversion ("*The quality of being outgoing, socially-oriented, confident, and enthusiastic*"; Goldberg, 1993). Next, we showed them a total of 31 items, including 17 from the VIS and 14 items from all orbiting scales combined, and asked them to sort the items into the appropriate construct domain.

We then calculated the p_{sa} and c_{sv}. The p_{sa} was calculated as n_c / N, and the c_{sv} was

calculated as $(n_c - n_0) / N$, where N was the total number of participants, n_c was the number of participants who sorted the item correctly, and n_0 was the maximum number of times an item was sorted into any of the other constructs in the set. The p_{sa} reflected the proportion of participants who correctly sorted a particular item into the appropriate construct (range = 0 to 1, with 1 indicating that 100% of the participants correctly categorized an item). The c_{sv} captured how well the participants categorized a particular item to the appropriate construct compared to the other "incorrect" constructs (range = -1 to +1, where larger values equated to greater substantive validity). Our main focus was on the 17 VIS items. Based on Colquitt et al. (2019), we classified an item as exhibiting "very strong" content validity evidence if the p_{sa} and c_{sv} were larger than .91 and .81, respectively.

Sample B

Sample B consisted of 200 MTurk participants (42.0% men, M_{age} = 39.9 SD = 12.0). They completed a rating task where they provided data for the *htc* and *htd*. Specifically, participants were presented with the same definitions for *vigilantism* and *physical aggression* used in Sample A and rated how well each VIS item matched each definition (1 = *extremely bad match*, 7 = *extremely good match*). The average rating presented an index of definitional correspondence for each item, also referred to by Hinkin and Tracey (1999) as "content adequacy."

The *htc* was calculated by dividing the average definitional correspondence rating of a particular item by the number of anchors. For example, if an item received an average correspondence rating of 6 out of 7, *htc* would be calculated as 6/7 = 0.86. The *htc* statistic for a particular item would take on a perfect value of 1 when all participants selected the maximum anchor for that item. The *htd* captured definitional distinctiveness. In particular, *htd* was calculated based on the differences between the definitional correspondence rating of an item on its intended construct and the ratings on unintended, orbiting construct. This

difference was then divided by $k - 1$, where k represented the maximum number of anchors. For example, if a VIS item received a rating of 6.5 out of 7 for the definition of vigilantism, and a score of 2.5 out of 7 for the definition of physical aggression, the *htd* would be calculated as $(6.5 - 2.5) / (7 - 1) = 0.67$. Based on Colquitt et al. (2019), we classified an item as exhibiting "very strong" content validity evidence if the *htc* and *htd* were larger than .91 and .35, respectively.

Results and Discussion

We retained an item if it demonstrated at least three "very strong" content validity evidence across the four indicators guided by the cut-off values suggested by Colquitt et al. (2019). A total of 10 items out of the initial 17 met this standard. Five items assessed tendencies to monitor the environment for signs of deviance, and the other five assessed the willingness to punish norm violators. These 10 items are presented in bold in Table 2.

In our next study, we assessed the psychometric adequacy of each item by examining whether they load significantly onto the latent construct of a vigilante identity. We administered the 10-item measure to a separate online sample and subjected their responses to an exploratory factor analysis (EFA). Also, given that the definition of the vigilante identity we used consisted of two different sets of behavioral demands (i.e., monitoring and willingness to punish), we explored whether a unidimensional or multi-dimensional two-factor model provided a better fit for our data.

Study 2b: Exploratory Factor Analysis

Sample and Procedure

A total of 201 MTurk participants (53.2% men, $M_{age} = 38.8$, $SD = 12.5$) completed the current study. Subjects completed the 10-item VIS from Study 2a and basic demographic questions. For the VIS, participants were guided by the stem question: "*I am the kind of person who…*" (1 = *completely unlike me,* 7 = *completely like me*).

Results and Discussion

EFA results based on a maximum likelihood extraction method showed that a one-factor solution best fit the data, with the first latent factor accounting for 68.33% of the total variance. Factor loadings for each item are shown in Table 3. All items loaded highly on the latent factor with standardized factor loadings ranging from .68 to .89. The items also had high internal consistency (Cronbach's α = .95). These results thus provided evidence for the psychometric adequacy of each item and also suggested that a unidimensional model provided the best fit for the data, justifying their aggregation into a single composite VIS score. Our next study sought to cross-validate the results of our EFA in a new sample to increase our confidence that the use of item aggregation to produce a unidimensional scale is appropriate.

Table 2

Content Validation Results

	Items	Definitional Correspondence		Definitional Distinctiveness	
		Proportion of substantive agreement (*Psa*)	**Hinkin-Tracey Correspondence (*htc*)**	**Substantive validity coefficient (*csv*)**	**Hinkin-Tracey Distinctiveness (*htd*)**
1	**Makes sure other people act morally.**	Very Strong	Moderate	Very Strong	Very Strong
2	Rights wrongs whenever I see them.	Strong	Moderate	Strong	Strong
3	**Makes sure people are held accountable when they do something wrong.**	Very Strong	Very Strong	Very Strong	Very Strong
4	**Ensures that people who do something wrong get punished for it.**	Very Strong	Very Strong	Very Strong	Strong
5	Protects people who are mistreated by punishing wrongdoers.	Moderate	Moderate	Moderate	Strong
6	**Enforces society's rules.**	Very Strong	Moderate	Very Strong	Very Strong
7	**Makes sure people who do wrong don't get away with it.**	Very Strong	Very Strong	Strong	Very Strong
8	Enforces justice in society.	Very Strong	Moderate	Very Strong	Strong
9	Pays attention to make sure that others are doing the right thing.	Strong	Strong	Strong	Very Strong
10	**Stays alert for signs of corruption or wrongdoing.**	Very Strong	Very Strong	Very Strong	Very Strong
11	**Is vigilant for signs of injustice in society.**	Very Strong	Strong	Very Strong	Very Strong
12	Notices when others do or say something inappropriate.	Strong	Weak	Strong	Strong
13	**Actively monitors others to see if they are following society's rules.**	Very Strong	Very Strong	Very Strong	Very Strong
14	**Takes it upon themselves to monitor others for signs of wrongdoing.**	Very Strong	Very Strong	Very Strong	Very Strong
15	Notices when people are being disorderly.	Very Strong	Weak	Strong	Strong
16	**Actively detects signs of wrongdoing.**	Very Strong	Strong	Very Strong	Very Strong
17	Notices when people are doing something unethical or immoral.	Strong	Moderate	Strong	Very Strong

Notes. Items 1 – 8 measured the *punishment* component of the VIS; items 9 – 17 assessed the *monitoring* aspect of the VIS. Selected items were in bolded font.

Table 3

Factor Loadings of the 10 VIS Items

Items	Factor Loadings from the EFA (Study 2b)	Communalities (Study 2b)	Standardized Factor Loadings from the CFA (Study 2c)
1. Makes sure other people act morally.	0.86	0.74	0.70
2. Makes sure people are held accountable when they do something wrong.	0.86	0.73	0.68
3. Ensures that people who do something wrong get punished for it.	0.88	0.78	0.73
4. Enforces society's rules.	0.78	0.61	0.70
5. Makes sure people who do wrong don't get away with it.	0.89	0.79	0.74
6. Stays alert for signs of corruption or wrongdoing.	0.68	0.46	0.62
7. Is vigilant for signs of injustice in society.	0.77	0.59	0.50
8. Actively monitors others to see if they are following society's rules.	0.79	0.63	0.62
9. Takes it upon themselves to monitor others for signs of wrongdoing.	0.79	0.63	0.66
10. Actively detects signs of wrongdoing.	0.72	0.51	0.66

Notes. EFA results showed that a one-factor solution best fit the data, with the first latent factor accounting for 68.33% of the total variance. CFA results in Study 2c showed that one-factor solution fit the data as well as a two-factor solution that separated the monitoring and punishment components of the VIS.

Study 2c: Confirmatory Factor Analysis

Sample and Procedure

To diversify the sample for scale validation, we posted a paid advertisement on Facebook to recruit U.S. residents who were at least 18 years old. A total of 170 participants (24.1% men, M_{age} = 48.2, *SD* = 16.9) completed the study. Like Study 2b, participants completed the same 10-item VIS and a basic demographic questionnaire. We incentivized participation by entering participants into a draw for a $10 Amazon Gift Card. We ensured anonymity of responses by redirecting participants who completed the survey to another link, where they then entered their

email address. Every one in ten participants was randomly selected to receive the Amazon Gift Card, and we honored our agreement and issued 17 gift cards.

Results and Discussion

We conducted a CFA on the 10 items of VIS specifying a single factor model. The results showed that a one-factor solution fit the data well, $\chi^2(35) = 62.47$, $p = .003$, CFI = .96, RMSEA = .07, SMSR = .05. Standardized item loadings are shown in Table 3. We sought to ensure that our one-factor solution fit the data better than a two-factor solution by separating punishment and monitoring items into their respective latent constructs in a separate CFA model. Although the two-factor model also fit the data well, $\chi^2(34) = 62.44$, $p = .002$, CFI = .96, RMSEA = .07, SMSR = .05, it was clear that separating items by components of the VIS (i.e., monitoring and punishment) did not substantially improve model fit, $\Delta\chi$ [$\Delta df = 1$] = 0.03, $p = .855$. Favoring parsimony, we concluded that the VIS items measured a unidimensional construct and could justifiably be aggregated into a single scale.

Study 2d: Convergent, Discriminant, and Nomological Validity of the VIS

We identified measures that were conceptually related to, but that we expected to be conceptually distinct from VIS for evidence of convergent and discriminant validity. In order to assess our scale's nomological validity, we examined correlations between the VIS with other constructs that prior theory and research suggest are likely to be related to the vigilante identity. We collected this data across three time points. As we will show in Study 3 (based on Time 3 data), we used this sample to test the predictive validity of the VIS on punishment intent, while controlling for theoretically plausible competing predictors.

Convergent Validity Tests

To establish convergent validity, we examined the relationship between the VIS and

published scales that may influence people's willingness to administer unauthorized punishment to norm violators. We selected the *social vigilantism* (Saucier & Webster, 2010), *punishment orientation* (Yamamoto & Maeder, 2019), and *moral identity* (Aquino & Reed, 2002) scales for this procedure based on the following rationale. First, *social vigilantism* shares some features with our conceptualization of the vigilante identity. While they both reflect one's desire to enact corrective action, the two differ in their manifested behaviors and scope. Social vigilantes are people with tendencies to "assert their 'superior' beliefs onto others to correct others' more 'ignorant' opinions for the 'greater good'" (Saucier & Webster, 2010, p. 19), without the identity-driven behavioral demand to punish others. In contrast, punishment tendencies are central to those who adopt a vigilante identity (DeCelles & Aquino, 2020). Thus, we viewed the social vigilantism scale (SVS) as a suitable candidate for convergent validity and we expected it to be positively correlated with the VIS.

Second, we selected *punishment orientation* (PO; Yamamoto & Maeder, 2019) as another candidate for examining convergent validity. PO contains four dimensions: 1) prohibitive utilitarianism (punishment focuses on prospective benefits), 2) prohibitive retributivism (avoiding punishing innocent people), 3) permissive utilitarianism (punish to ensure public safety and deter crime), and 4) permissive retributive (punish to get back at the offender). Like SVS, however, the PO scale does not address behavioral demands inherent in vigilante identity (i.e., monitoring and punishing others). Instead, it assesses how lay people think about punishment, with the goal of capturing individual-level differences in perceptions of punishment ethics (Yamamoto & Maeder, 2019). We expected the VIS to correlate positively with all four dimensions of punishment orientation, as both scales contain punitive elements and perceptions of those who violate norms.

Finally, to provide evidence of the VIS's convergent validity, we measured people's *moral identity* as it is a well-established identity construct that has been shown to motivate third party punishment. Moral identity is defined as a self-conception organized around a set of moral traits, which although it is a trait-based concept, can be malleable and subject to social construction (Aquino & Reed, 2002, p. 1424). Aquino and Reed's (2002) model proposes two components of moral identity: 1) *internalization*, which reflects a person's subjective experience of having a moral identity and 2) *symbolization*, which reflects a person's public expression of their moral character. As some studies have shown that people with a strong moral identity are more motivated to punish deviants (Hofmann et al., 2018), it is conceivable that adoption of this identity could impose similar behavior demands as those associated with adopting the vigilante identity. Hence, we expect that the VIS and moral identity should be positively correlated, but they should still be empirically distinct. We do not make specific predictions about which facet of moral identity should be more strongly correlated with the VIS since it is not obvious what we might expect theoretically.

Although we expected positive correlations between the VIS and SVS, PO, and moral identity, we did not expect these correlations to be so high as to suggest that VIS is redundant. In other words, we also expected to find evidence of construct distinctiveness.

Nomological Network Tests

We explored the nomological network (Cronbach & Meehl, 1955) of VIS by examining whether it is related to certain personality traits and cognitive tendencies. We expected that dispositions associated with capacity and willingness to devote mental energies to detect wrongdoing should covary positively with VIS. Those dispositions include trait hypervigilance (Bernstein et al., 2015), moral attentiveness (Reynolds, 2008), and need for cognition (Cacioppo

et al., 1984).

We also explored the VIS's placement within the Big-5 personality framework (Soto & John, 2017). The Big-5 personality domains have been shown to be comprised of different sub-facets (DeYoung et al., 2007) and we used the descriptions of the behavioral tendencies associated with these sub-facets to derive predictions about how the VIS would be correlated with the Big-5. We expected the VIS to be positively related to extraversion due to extraverts' willingness to engage with the social world rather than retreat from it (Soto & John, 2017), and to be assertive in social situations (DeYoung et al., 2007). We expected the VIS to be negatively related to agreeableness since the sub-facets of agreeableness are comprised of being polite and compassionate (DeYoung et al., 2007).

We made no predictions about the relationship between the VIS and neuroticism, conscientiousness, and openness to experience. On the one hand, one facet of neuroticism is volatility which can lead people to be highly reactive to social stimuli. On the other hand, people high in neuroticism can also be socially withdrawn (DeYoung et al., 2007). We reasoned that these two conflicting tendencies can lead to indecisiveness and inaction. Similarly, it is possible that the VIS will be positively related to conscientiousness since highly conscientious people tend to be orderly (DeYoung et al., 2007; Soto & John, 2017); however, this same tendency might also lead people to refrain from becoming vigilantes because they do not want to act without authorization and break the organization's chain of command. Finally, we made no prediction about the relationship between the VIS and openness to experience as it is not obvious how being intellectually curious, imaginative, and aesthetically inclined, which are the characteristics associated with this trait (DeYoung, et al., 2007), would be related to becoming a vigilante.

We also examined the relationships between the VIS, physical aggressiveness, trait kindness, and tendency to forgive. First, punishment is a form of aggression so we expected the VIS to be related to physical aggressiveness. Second, we expected vigilantes' higher level of moral certainty (DeCelles & Aquino, 2020) would result in a self-perception of being morally praiseworthy. Thus, we expected positive relationships between the VIS and self-rated trait kindness (Canter et al., 2017). Finally, we expected VIS to be negatively related to forgiveness attitudes (Brown, 2003), since high VIS individuals' willingness to punish wrongdoers suggests that they may be less likely to excuse others' poor behavior.

Sample and Procedure

We sought to reduce participant fatigue by temporally separating our measures across three time points. Therefore, we conducted a time-lagged, three-wave online study on MTurk via CloudResearch. As stated in the preregistration, we only analyzed and retained participants who completed all three surveys and who passed our key attention checks.

A total of 318 MTurk participants (52.2% men, M_{age} = 41.7, *SD* = 12.2) completed all three waves of the study. In the Time 1 survey, they completed the VIS and various individual differences measures, which assessed the Big-5 personality traits (Soto & John, 2017), need for cognition (Cacioppo et al., 1984), trait kindness (Canter et al., 2017), tendency to forgive (Brown, 2003), and physical aggression (Buss & Perry, 1992). At Time 2, we assessed their moral identity (Aquino & Reed, 2002), punishment orientation (Yamamoto & Maeder, 2019), hypervigilance (Bernstein et al., 2015), social vigilantism (Saucier & Webster, 2010), and moral attentiveness (Reynolds, 2008). In Time 3, participants again completed the VIS and participated in an experiment where we assessed their intention to punish under different experimental conditions (i.e., justice failure vs. justice upheld). For clarity, we report the Time 3 experiment-

based results separately below as Study 3.

Measures

Unless otherwise noted, all assessments used a 7-point Likert scale (1 = *strongly disagree,* 7 = *strongly agree*). We report reliability coefficients in Table 4 provided at the end of the manuscript due to its size and cross-reference with Study 3.

VIS (Time 1 and 3). We used the same measure as in Studies 2b – 2c.

Big-5 Personality (Time 1). We used the shortened 15-item measure of the Big-5 from Soto and John (2017). Sample items for openness, conscientiousness, extraversion, agreeableness, and neuroticism included "*I am fascinated by art, music, or literature*", "*I am reliable, can always be counted on"*, "*I am dominant, act as a leader*", "*I assume the best about people*", and "*I tend to feel depressed, blue*", respectively.

Need for Cognition (Time 1). We used the scale developed by Cacioppo (1984). A sample item was "*I would prefer complex to simple problems.*"

Trait Kindness (Time 1). We used the 5-item measure developed by Canter et al. (2017). A sample item was "*I am kind to others.*"

Physical Aggression (Time 1). We used 5 items from the physical aggression dimension of the aggression scale by Buss and Perry (1992). A sample item was "*Once in a while I cannot control the urge to strike another person.*"

Tendency to Forgive (Time 1). We used the 4-item scale of tendency to forgive developed by Brown (2003). A sample item was "*I tend to get over it quickly when someone hurts my feelings.*"

Moral Identity (Time 2). Participants completed the 10-item moral identity scale (Aquino & Reed, 2002). The scale consists of two subscales assessing internalization (MI-I) and

symbolization (MI-S) of moral identity. Participants were asked to visualize a type of person who possessed certain moral traits (e.g., caring, fair, and honest), and answered five questions assessing MI-I (e.g., "*I strongly desire to have these characteristics*"), and five questions assessing MI-S (e.g., "*I often wear clothes that identify me as having these characteristics*").

Punishment Orientation (Time 2). We used the 17-item Punishment Orientation Questionnaire (POQ) to assess the four dimensions of punishment orientation (Yamamoto & Maeder, 2019, p. 1288). Sample items included: "*Punishment should be about looking forward to improve society, not backward to address the criminal's misdeeds*" (prohibitive utilitarianism), "*It is better to let 10 guilty criminals go free than to punish one innocent person*" (prohibitive retribution), "*Punishment is a necessary evil*" (permissive retributive), and "*Overly harsh punishment may be necessary to prevent crime*" (permissive utilitarian).

Hypervigilance (Time 2). We adapted the 5-item scale developed by Bernstein et al. (2015), which was originally developed to assess paranoia cognition. A sample item was "*As soon as I wake up and for the rest of the day, I am watching for signs of trouble.*"

Social Vigilantism (Time 2). We used the 15-item measure of social vigilantism from Saucier and Webster (2010). A sample item was: "*I like to imagine myself in a position of authority at work so that I could make the important decisions around my workplace*."

Moral Attentiveness (Time 2). We assessed the extent to which participants showed attentiveness to ethical dilemmas in their daily lives using a 7-item perceptive dimension subscale (Reynolds, 2008, p. 1031).[6] A sample item was: "*In a typical day, I face several ethical dilemmas*."

[6] We selected the 7-item perceptive scale due to its relevance with VIS. *Reflective* moral attentiveness, which was not measured in our study, refers to whether one ruminates on matters of morality. A sample item is: "*I like to think about ethics*".

Results and Discussion

Table 4 contains the descriptive statistics, αs, and correlations among variables in Study 2d.

Insert Table 4 about here (presented at the end due to size)

First, we assessed the nomological network of the VIS. Table 4 shows that the VIS was positively correlated with extraversion, agreeableness, need for cognition, trait kindness, physical aggression, moral identity (internalization and symbolization), hypervigilance, and moral attentiveness. Most of the above relationships were directionally consistent with our predictions. The exceptions were that the VIS was unrelated to attitudes toward forgiveness and positively related to agreeableness. One possible explanation for the former finding is that individuals high in VIS may be willing to forgive unintended mistakes but unwilling to let deliberate wrongdoings go unpunished; not specifying the source of misbehaviors in the Tendency to Forgive scale may therefore result in a null, non-significant correlation with the VIS. For the latter finding, it was possible that people high in the VIS might view themselves as generally agreeable unless provoked. This favorable perception of their interpersonal attributes is consistent with the finding that the VIS was positively related to self-perceived trait kindness. Future research is needed to determine whether the relationship between the VIS and Tendency to Forgive was indeed moderated by source of harm, and if the relationship with agreeableness can be replicated or whether our initial prediction of a negative relationship proves to be more reliable across samples.

Second, we found evidence of convergent validity. VIS was positively associated with the prohibitive utilitarianism, permissive utilitarianism, and permissive retribution dimensions of

the PO measure and negatively associated with prohibitive retribution dimension. It was also positively correlated with the SVS and moral identity. We also sought evidence of discriminant validity (i.e., construct distinctiveness). We followed the latest approach introduced by Henseler et al. (2015) where we calculated the heterotrait-monotrait (HTMT) ratio(s) between VIS and other related constructs (i.e., SVS, PO, and moral identity). According to Henseler et al. (2015), a HTMT of 0.85 or above would indicate a lack of discriminant validity. Results clearly suggest the presence of discriminant validity. HTMT criteria for each pair of constructs ranged from .152 for HTMT (VIS, prohibitive utilitarian) to .526 for HTMT (VIS, SVS) — all of which were below the suggested cut-off of 0.85.

Finally, we assessed whether the VIS exhibited test-retest reliability. The correlation between VIS scores measured at Time 1 and Time 3 administered approximately 17 days apart was high (r = .82, p < .001). This correlation was comparable to median test-retest correlations of trait constructs such as Big-5 personality traits (range r = .83 - .93), trait hostility (r = .80 - .83), trait assertiveness (r = .82 - .91), need for order (r = .80 - .90) (for a comprehensive review of reliability coefficients in personality assessments, see McCrae et al., 2010). Our results suggested that like other identities, the VIS may have trait-like stability if it is highly internalized. Collectively, we concluded that the pattern of correlations reported in Table 4 provided further evidence for the construct validity of the VIS.

Study 2e: Identity Recognition by Observers

If the internalization of a vigilante identity motivates people to enact its behavioral demands, then vigilante behavior is likely to be observed by people in the immediate social environment. The current study tests this assumption thereby providing evidence that the VIS reliably measures an identity. As reported in the SOM, we also sought to replicate our dyad-

based ratings from Study 1 using our revised scale, where we explored whether the VIS was correlated with others' perceptions that the employee displays other kinds of behaviors that are generally viewed positively (i.e., organizational citizenship, in-role performance) or negatively (i.e., counterproductive work behavior).

Sample and Procedure

We employed the services of a market research firm, Knowledge Intercept, based in India to assist with data collection. The firm recruited working adults across India who expressed a willingness to participate in online surveys for academic and market research purposes. The panel members were informed that the research team was investigating which cognitive, affective, and behavioral factors affected extra-role behaviors in the workplace. Participants were informed that the study would ask them to provide their supervisors' contact information; if they did not wish to provide this information, they were not eligible to participate.

The participating employees were asked to complete demographic questions, the VIS, and other measures (see SOM for full results). The research firm then contacted the supervisors and asked them to rate the focal employee on vigilante behavior (i.e., an adjusted VIS). Each supervisor rated only one employee. All respondents (employees and their supervisors) were paid approximately the equivalent of $10 USD for participation. All questions were presented in English, which is one of the official languages in India and is widely spoken by working professionals. Employees and their supervisors were instructed not to share their responses with each other, and they completed their surveys individually using their own computers and at different times.

The research firm provided us with the final, matched data. Our final sample included responses from 213 full time employees and their matched supervisors working across various

organizations. In the employee sample, 56.8% was male. Mean age of employees was 29.8 years (*SD* = 3.4), they had 5.0 years of work experience on average (*SD* = 2.5), and averaged 3.2 years (*SD* = 1.5) of tenure in their current organization. Most employees held a graduate degree (62.9%). In the supervisor sample, 65.3% was male. Mean age of supervisors was 35.2 years (*SD* = 4.1), they had 9.5 years of work experience on average (*SD* = 4.0), and averaged 3.9 years (*SD* = 2.4) of tenure with their current organization. Most of the supervisors (88.3%) held a graduate degree.

Measures

VIS (Self- and Supervisor-rated). Focal employees completed the same VIS used in our earlier studies. Their supervisors responded to those same items, but with the following prompt:

> Think about the employee you are being asked to evaluate. Specifically, think about the kind of person they are, the characteristics they possess, and how they conduct themselves in everyday life. Based on what you know about this employee, indicate how accurately each of the statements below describes this employee. *My employee is the kind of person who:*

The actual items were the same as in our VIS we used in Studies 2b – 2d (e.g., "*My employee is the kind of person who enforces justice in society*"). Supervisors rated their employees on a scale from 1 (*completely unlike them*) to 7 (*completely like them*).

Results and Discussion

Descriptive statistics, scale reliabilities, and correlations among all Study 2e variables are presented in Table 5.

Table 5

Descriptive Statistics and Correlations among Variables in Study 2e

		1	2	3	4
1	Employee Self-rated VIS	(.74)			
2	Supervisor-rated VIS	.53**	(.75)		
3	Employee Gender (1 = Male)	.10	.02	-	
4	Employee Age	.00	-.12	-.01	-
	M	5.02	5.09	.57	29.77
	SD	.77	.74	.50	3.40
	N	213	213	213	213

As shown in Table 5, the correlation between employee self-rated VIS and supervisor-rated VIS was large and positive, $r = .53$, $p < .001$, providing strong evidence for self-other agreement on the VIS. The results of this study thus confirm that the vigilante identity is a recognizable personal identity. Employees who more strongly internalized the vigilante identity were perceived by their supervisors as being more likely to monitor their environment for signs of violations and to punish norms violators. This result replicates a pattern we found in our dyad-based Study 1. In the next study, we go beyond an observer's perspective and directly test whether VIS can predict the actor's intention to administer unauthorized punishment to a presumed norm violator.

Study 3: Predictive Validity of the VIS

In this study, we examined the predictive validity of VIS using participants' intention to punish a norm violator as the dependent variable. This study was also designed to test one of the assumptions of DeCelles and Aquino's (2020) theory that vigilantes tend to emerge when people believe that authorities are failing to ensure justice by not punishing deviants. We hypothesized the VIS would predict punishment intent when there was evidence of previous justice failure, but not when they believed that justice had been upheld in a similar situation.

Design and Procedures

This study was conducted at Time 3 of the data collection for Study 2d, where the same group of participants were directed to complete an online experiment after they responded to the Time 1 and 2 measures (i.e., VIS) we reported in Study 2d. In this experiment, participants watched a video of what was purportedly an interaction between a university instructor and a student. The incident was set in a real classroom with actors playing different roles and following a scripted dialogue. In the video, a middle-aged, Caucasian male instructor was positioned as the putative norm violator when he changed the assignment rules for self-serving purposes and made a culturally insensitive remark to an ethnic minority female student. Specifically, the instructor informed students that he was going to change the class schedule because of his gym appointment and that they either needed to attend a make-up class on a different day or complete an onerous assignment in lieu of it. Upon hearing this announcement, a female student wearing a hijab told the instructor that she would be unable to attend the make-up class because she had made travel plans to see her family for a religious holiday. She proceeded to question the fairness of having to complete a difficult assignment for missing a class that was not on the course syllabus. The instructor asked the other students if they had any problems with attending the make-up class. When no one else objected, he turned to the student in the hijab and said:

> "As you see, you're the only one with a problem. It seems every year that there is some issue with you people. Why does this have to be so complicated? Unfortunately, the assignment is mandatory, and the issue is not up for discussion."

We manipulated perceived justice failure by informing participants what had happened to another instructor who behaved in a similar way at this university. Participants in the *justice failure* condition read that even though the university administration was made aware of the prior incident involving another instructor, they did not acknowledge the wrongdoing. That instructor did not face any repercussions and was able to continue teaching at the university. In the *justice*

upheld condition, the university had issued a public statement condemning the instructor's behavior as inappropriate and required him to issue an apology to the student and attend a 3-week-long diversity training before he could teach a class again.[7]

Measures

Manipulation Check

As a manipulation check, all participants rated the perceived fairness of the university's response to the previous norm violation by indicating the extent to which they agreed with the statement "*The university's response failed to deliver justice to the parties involved in this incident.*" on a 7-item scale (1= s*trongly disagree*, 7 = *strongly agree*).

Intent to Punish the Instructor

We operationalized the intention to punish the instructor in the video by asking participants to indicate the likelihood that they would post the video on social media if they had it in their possession. This measure corresponded to what information systems researchers refer to as *netilantism* or *digilantism* (Chang & Poon, 2016; Galleguillos, 2021). Participants indicated their likelihood of posting the video on social media using both a continuous, 9-point scale (1 = *not likely at all*, 9 = *very likely*) as well as a binary choice option (0 = *not post*, 1 = *post*). Finally, they answered demographic questions and were given a debriefing form indicating that the video was created strictly for research purposes only, and that all characters (the students and the instructor) were trained actors.

[7] We also included an extreme punishment condition for exploratory reasons. Here, the university fired the instructor, and the instructor was unable to find a new job and experienced psychological problems as a result. Results related to this condition are reported in the SOM.

Results and Discussion

Manipulation Check

Subjects in the justice-upheld condition (M = 3.83, SD = 1.69) perceived a significantly lower level of justice failure by the university compared to those in the justice failure condition (M = 6.12, SD = 1.18), $t(208) = -11.34$, $p < .001$, indicating that our manipulation was effective.

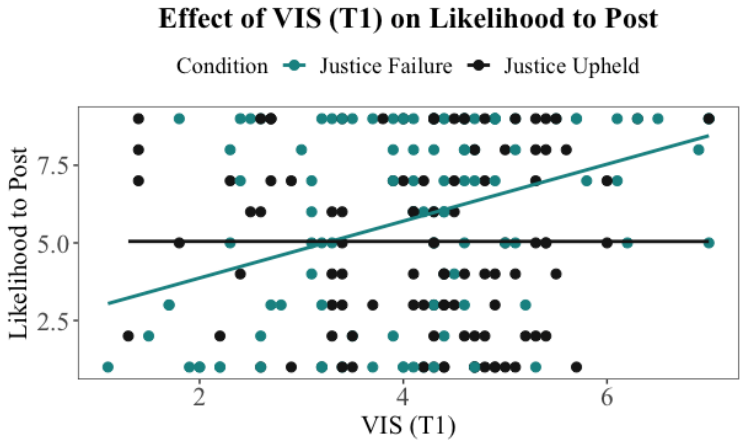
Hypothesis Test

We used the 9-point scale measure of likelihood to post the video as our main DV and VIS measured at Time 1 as our IV. Supporting our hypothesis, the VIS at Time 1 positively predicted likelihood to post the video, b = .92, SE = .22, $p < .001$. Moreover, we found the expected moderation effect by justice failure condition, b = -.92, SE = .34, p = .007. Simple slopes analysis revealed that the VIS had a positive effect only in the justice failure condition, b = .92, SE = .22, $t(102) = 4.18$, $p < .001$, but had no effect in the justice-upheld condition, b = -.01, SE = .25, $t(104) = .00$, p = .998 (see Figure 1). As a robustness check, we subsequently included SVS, moral identity, and all four dimensions of the punishment orientation in the model as covariates. The VIS at Time 1 remained a significant predictor of likelihood to post the video, b = .90, SE = .23, $p < .001$, and the same interaction effect with justice failure condition also held, b = -.71, SE = .34, p = .036. Finally, our results remained the same and our conclusions remained unchanged when we repeated all the analyses above using the VIS measured at Time 3 as the IV.[8]

[8] The 9-point scale reported above significantly correlated with the binary measure we additionally included, r = .89, suggesting that they are empirically indistinguishable. When we conducted a logistic regression analysis by using the binary response as the dependent variable (1 = *post*, 0 = *do not post*), results were consistent with our results above using the continuous measure, such that the VIS at Time 1 positively predicted the binary intention to post the video, *Odds Ratio (OR)* = 1.79, z = 3.19, p = .001. We again found a moderation effect by justice failure condition, b = -.70, p = .005. Simple slope analysis revealed that the VIS had a positive effect only in the justice failure condition, OR = 1.79, z = 3.19, p = .001, but had no effect in the justice-upheld condition, OR = .89, z = 0.70, p = .487. Also, consistent with results above based on the continuous DV, VIS at Time 1 continued to predict this binary outcome even after adding covariates (SVS, punishment orientation, moral identity), OR = 1.77, z = 2.79, p = .005. The same interaction effect with justice failure condition also continued to hold, b = -.61, p = .024.

Figure 1

The Effect of VIS (Time 1) on Likelihood to Post by Justice Failure Condition



Study 3 provided evidence for the predictive validity of the VIS. It also supported our hypothesis that the VIS would only predict intentions to punish when authorities failed to enact justice for a similar violation that occurred in the institution on a prior similar occasion. Next, we moved beyond examining punishment intention and tested the predictive validity for monitoring.

Study 4:

The VIS as a Predictor of Vigilant Monitoring

In Study 4, we sought evidence that the VIS can reliably predict monitoring among vigilantes who are organizational outsiders (i.e., customers). We hypothesized that the VIS would be positively related to the intention to use social media to monitor the activities of the organization in which a presumed norm violation occurred.

Sample and Procedures

A total of 136 participants (50.0% men, M_{age} = 41.2, *SD* = 11.9) recruited from MTurk completed the current study. They had an average of 18.8 years (*SD* = 12.7) of work experience. This study utilized a correlational, non-experimental design. Participants read a short news-like article that described a female customer who was expelled from a restaurant by the owner after

she verbally confronted another customer who was wearing a T-shirt that allegedly displayed a racist symbol. In efforts to extend the real-world relevance of our study, we based our descriptions on an actual incident (see the case of Kachka restaurant in Portland; Herron, 2018). Specifically, they read the following description:

> "A woman in Idaho recently became the topic of conversation on social media over an encounter she had with another customer at the local diner, Ricky's Kitchen. Beverly Tanners, a frequent patron of the diner, said that she was asked by the owner to leave when she confronted another customer who she says was wearing a T-shirt with the word "Luftwaffe" prominently displayed. Luftwaffe is the current name of the German air force, but it was also the name of the air force for the Nazi party during World War II.
>
> Tanners said that she was asked by the owner of Ricky's Kitchen to leave after she confronted the customer and expressed her concerns because, according to the owner of Ricky's Kitchen, she had caused a disturbance in the restaurant and was upsetting and threatening the customer.
>
> "We want all our customers to feel safe in our restaurant," the owner said. The customer wearing the T-shirt was not asked to leave."

After reading the vignette above, participants were asked to rate the restaurant. Ratings were on a 5-point scale in 0.5 intervals where higher scores meant a more positive review for the restaurant.[9] Then, we told participants that the incident quickly became one of the widely discussed topics across various social media platforms. Subjects subsequently indicated their likelihood (1 = *very unlikely*, 7 = *very likely*) of 'following' the relevant discussions across different social media platforms, such as: 1) the relevant #hashtag (e.g., #justiceforBeverly on Twitter), 2) Reddit posts that openly discussed other wrongdoings by the restaurant, 3) Facebook groups created specifically for revealing other wrongdoings by this restaurant, and 4) Google

[9] As stated in the preregistration, we initially treated this variable as an indicator of willingness to punish. However, during the review process questions arose as to whether this measure fully represented unauthorized punishment as we conceptualized in the introduction. For this reason, we decided to exclude this variable as part of our key dependent variable and instead used it as a control variable to assess whether the VIS continued to predict monitoring after punishment intent has been statistically accounted for. Nevertheless, the VIS was negatively related to ratings for the restaurant, providing some evidence that the VIS was a valid predictor of punishment sentiments.

groups that shared negative reviews about this restaurant. These four items were used to create a composite of *willingness to monitor* the restaurant (α = .90). Subjects then completed the same measure of VIS (α = .95), followed by basic demographic questions.[10]

Results

Table 6 shows the descriptive statistics, scale reliabilities, and correlations among variables in Study 4. Consistent with our expectations, the VIS was positively related to willingness to monitor the restaurant's misbehaviors on social media platforms, r = .42, p <.001. As a robustness check, when we controlled ratings for the restaurant on the relationship between the VIS and willingness to monitor the restaurant, the partial correlation remained significant, r = .37, p < .001. Finally, we additionally explored whether the effect of VIS on vigilant monitoring may be shaped by certain demographic characteristics (e.g., age, conservatism, religiosity, race), but found no evidence of moderation (p-values ranged from .095 to .961).

Table 6

Correlations Among Variables in Study 4

		1	2	3	4	5	6	7
1	VIS	(.95)						
2	Ratings for the Restaurant	-.22**						
3	Willingness to Monitor	.42**	-.45**	(.90)				
4	Age	-.15	-.17	-.08				
5	Gender (1 = Male)	-.04	-.09	-.15	-.20*			
6	Conservatism	-.02	-.29**	-.19*	.08	.13		
7	Religiosity	.16	-.11	.03	.06	-.26*	.32**	
	M	3.39	2.89	2.46	41.17	1.50	4.05	4.51
	SD	1.40	1.42	1.67	11.88	.50	2.39	2.12
	N	136	136	136	136	136	136	94

Notes. *p <.05; ** p <.01.

[10] We presented participants with additional information after they completed the VIS for exploratory purposes. Specifically, participants were informed that the business owners were actually Jewish and that their decision to expel the 'vigilante-like' customer was motivated by prosocial motives to protect the customer wearing the shirt. We then assessed participants' moral self-appraisal (e.g., "*My ratings of the restaurant provided earlier would not change; 5 items in total*") after reading this new information. We reported this result in the SOM.

Discussion

Study 4 extended the results of Study 3 and showed that the VIS predicted willingness to devote greater attention to monitoring the organization for any future deviance - a relationship that continued to hold even after we controlled for punishment sentiments. Combined with the findings in Study 3, these results support the notion that the vigilante identity contains both punitive and monitoring demands.

General Discussion

We introduced the *vigilante identity* as an individual difference to help explain why organizational insiders and outsiders might take it upon themselves to punish employees or organizations that they believe have violated a social norm. We provide empirical evidence that this identity has two behavioral demands: the motivation to *monitor* the environment for signs of wrongdoing and the willingness to *punish* norm violators when they deem it appropriate. We defined vigilantes as self-appointed monitors and punishers, meaning they have not been assigned formal jurisdiction by institutional or state authorities to control the actions of their targets through punishment. Since they are not granted legitimacy to perform either of these tasks, employee vigilantes can be subjected to disciplinary action by superiors. However, outsider vigilantes are beyond the reach of managerial control and are generally unaccountable to any authorities, especially if they act in anonymity. By departing from the primary focus of current scholarship on vigilantes as organizational insiders and presenting evidence for *vigilante identity* as a construct that can explain the behavior of outsider vigilantes, we expand the opportunities for theory development and hypothesis testing. For instance, there might be different consequences for organizations when it comes to managing vigilantes who are

employees versus outsiders, and there could also be different antecedents that predict when vigilantes emerge to punish organizations, depending on their relationship to the organization.

In an age of digital archiving, the permanence of social media posts, and the ability of company surveillance technology and personal cell phones to record the most minute and intimate actions of employees, nearly everyone is vulnerable to having their foibles exposed to the prying eyes of the public. Sufficiently motivated vigilantes thus have an unprecedented ability to access and disseminate such information to a large, global audience for the purpose of inflicting social or even material harm upon their intended targets. To date, the organizational literature has not sufficiently theorized about or empirically investigated what might motivate people who are unconnected to the organization to intervene in the lives of its employees for the purpose of controlling or punishing them. The concept of the vigilante identity that we introduce in this paper provides one possible answer to this question.

Theoretical and Practical Contributions

Our work makes several contributions to established lines of research. First, it contributes to the literature on organizational punishment. There is a small but emerging literature in organizational studies exploring topics of violence and informal punishment (Costas & Grey, 2018; DeJordy, 2010; Gill & Burrow, 2017; Martí & Fernández, 2013; Remington Abramson & Senyshyn, 2010), including those labeled as vigilantes (Crawford & Dacin, 2020; de Rond et al., 2021; DeCelles & Aquino, 2020). Our emphasis is primarily on better understanding *who* might enact such informal punishment, and we trace our answer to organizational insiders and outsiders who may self-identify as a vigilante. In turn, their vigilante identity motivates their attempts to control others' behavior by monitoring and punishing them for norm violations. We examine this relationship across a diverse set of norm violations, including others' non-compliance with

government COVID-19 regulations (Study 1), cultural insensitivity towards a student by an instructor (Study 3), and the presumed toleration of a reviled symbol of racism by a restaurant (Study 4), highlighting the many potential ways that vigilante behavior affects organizations.

Of course, not all people who notice a violation of social norm will become a vigilante and punish the violator. Thus, another contribution of our research is to show that vigilantes are not gratuitous punishers but are likely to emerge only under certain circumstances, most notably where they believe formal authorities have failed at proper enactment of organizational justice. This finding suggests that most people are probably reluctant to become vigilantes, but a few may do so if they believe that systems of social control are inadequate or fail to punish wrongdoers (DeCelles & Aquino, 2020; Dumsday, 2009; Robinson & Robinson, 2016; Tripp et al., 2007; Weisburd, 1988).

For this reason, our research also contributes to our emerging understanding of how justice enactment in organizations can affect the behavior of individuals, both employees and organizational outsiders. The recent justice enactment literature focuses on the extent to which agents (e.g., supervisors, decision makers, or others with the power to treat others fairly) embrace or violate justice rules (Graso et al., 2020; Huang et al., 2017; Koopman et al., 2015). DeCelles and Aquino (2020) argued that one of the ways that managers can try to discourage employees from becoming vigilantes is by effectively enacting justice, which includes not only giving employees what they are due, which includes punishing them for legitimate violations of organizational norms, but also following basic principles of procedural fairness. Our research shows that when people perceive authorities will not enact justice, they may be motivated to take matters into their own hands. What is worth noting here is that vigilantes who believe they are enacting justice to compensate for presumed justice failure by authorities could themselves be

viewed by observers as committing a graver injustice by acting without either full knowledge of the circumstances or without granting the alleged perpetrator the basic right of the presumption of innocence. Our studies show that people who internalize the vigilante identity are more willing to accept the possibility of punishing an innocent person (Study 2d), perhaps because they prioritize the goals of retribution or inducing conformity over defending the presumption of innocence.

A third contribution of our research is to advance the testing of theories and hypotheses about workplace vigilantism by introducing a rigorously validated measure of the vigilante identity. Theoretical, inductive, and qualitative scholarship has far outpaced deductive hypothesis (Crawford & Dacin, 2020; de Rond et al., 2021; DeCelles & Aquino, 2020) in the vigilante literature. To advance deductive theory testing, we introduced a new measure of the vigilante identity (i.e., the VIS) and showed that it predicts monitoring one's environment for signs of deviance, intent to punish, and actual punitive behavior directed at norm violators. We observed these relationships cross-culturally (New Zealand, USA, and India) and in different contexts (e.g., a community during COVID-19 in Study 1, organizational insiders in Study 2e, and organizational outsiders in Studies 3 and 4), and using different methodological paradigms (self-report surveys, an experiment, and two dyad-based surveys). We conclude from these patterns and other findings that we report in this paper, that the VIS meets widely accepted standards for establishing construct validity and we believe researchers should be confident about using it to test novel hypotheses.

Practical Implications

We do not take a normative stance about whether vigilantes have a net positive or negative influence on organizations, as the appraisal of their activities is highly dependent on the

context, the severity of wrongdoing, and how vigilantes achieve their goals. Nonetheless, our research suggests that organizations need to be prepared to deal with the possibility of vigilante behavior from both insiders and outsiders to limit the damages they might cause. For example, the scenario we used in Study 4 was based on an actual case in which a rush to judgment after seeing a customer's Facebook post tarnished the reputation of the business and its owners (Herron, 2018). The potential damage to employees and organizational reputations that vigilantes might inflict on digital platforms (e.g., posting incriminating videos or damaging a business' reputation online) will likely become even more difficult to manage as information systems become increasingly vulnerable to attack. While one can argue that the informal punishment of norm violating employees may be helpful for discouraging future transgressors, those tools are also potentially dangerous when untampered by deliberate efforts to determine whether allegations are true (Crockett, 2017). Thus, organizations need to plan for the potential for vigilantes to inflict unauthorized punishment and how they might best be able to prevent the harms they cause, some of which might be irreparable and irreversible.

Limitations and Future Directions

Our research has limitations that should be acknowledged. First, we cannot conclude whether the VIS predicts repeated unauthorized punishment by a single individual across situations. Although a theoretical advantage of the vigilante identity as an explanatory construct is that it allows us to test predictions about why some people may administer unauthorized punishment more frequently than others over time, future research based on longitudinal data is needed to document this more convincingly. Second, we examined only one contextual factor – justice failure – that could predict vigilante emergence. Future research should examine whether

other situational conditions (e.g., threats to normative order or ethical infrastructure) might lead to a rise in vigilantism and whether those acts differ for insider versus outsider vigilantes.

Third, we used the term unauthorized punishment to distinguish a vigilante's actions from that of people whom societies generally recognize as having a legitimate right to punish (e.g., judges, parents, teachers, or bosses) within well-defined, hierarchically structured institutional roles. However, it is possible for people even within such well-defined roles to meet our definition of a vigilante if, for example, they stretch the boundaries of their roles or ignore its formal requirements to achieve an outcome they desire. These are referred to as "shadow vigilantes" (Robinson & Robinson, 2016) and we did not directly examine whether the vigilante identity would also predict excessive or unacceptable punishment by those who are trusted to deliver it justly.

Fourth, our research did not compare the predictive power of the vigilante identity relative to group-based social identities. Recognizing that the self is comprised of multi-faceted identities (Brewer & Gardner, 1996; Dutton et al., 2010; Sluss & Ashforth, 2007) opens the opportunity to further study psychological consequences of being a vigilante. For example, scholars have suggested that moral judgments and behaviors can be used to construct or reinforce people's social identities (Ellemers, 2018; Ellemers et al., 2013). If so, we might hypothesize that people who have internalized this identity will be more likely to participate in collective acts of vigilantism even when they have no stake in the grievance that mobilized the group into being. When they do so, it may allow them to feel a sense of belongingness within a valued in-group (Ellemers et al., 2013), which can explain the allure of mob violence directed against an alleged wrongdoer (Klatetzki, 2017).

Finally, we did not examine how the vigilante identity can serve an identity-expressive function (Erikson, 1964) that contributes to the maintenance of positive self-regard (DeCelles & Aquino, 2020). It is conceivable that the vigilante punisher will experience a positive, self-enhancing experience or *schadenfreude* from knowing they gave a norm violator their "just deserts" (Darley et al., 2000). Future research can use VIS and examine this possibility by looking at whether people who have internalized the vigilante identity experience self-relevant outcomes like esteem, meaning, or power when punishing others.

Conclusion

Vigilantes have always been present in societies. In an increasingly democratized and electronically connected world, would-be vigilantes have many means at their disposal to punish – either individually or collectively – those who they deem has having trespassed the norms of society. The current research examines what might motivate such self-appointed doers of justice to direct their activities towards employees from within and outside of organizations. In doing so, it contributes to the broader investigation of how people use informal punishments to control the behavior of others and impose their vision of justice upon the world.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). Grant was awarded to Prof. Karl Aquino (Grant # F17 – 04376). This research is partially supported by the Ministry of Education, Singapore, under its 2/2019 Academic Research Fund Tier 2 Grant Call (MOE2019-T2-1-192).

Table 4

Correlations Among Variables in Studies 2d and 3

		1	2	3	4	5	6	7	8	
1	VIS	(.92)								
2	Openness	.20**	(.76)							
3	Conscientiousness	.14*	.20**	(.80)						
4	Extraversion	.40**	.28**	.41**	(.72)					
5	Agreeableness	.18**	.32**	.25**	.19**	(.69)				
6	Neuroticism	-.10	-.10	-.55**	-.45**	-.20**	(.84)			
7	Forgiveness Attitude	.00	.14*	.04	.05	.44**	-.10	(.73)		
8	Need for Cognition	.25**	.55**	.16**	.32**	.28**	-.15**	.10	(.94)	
9	Trait Kindness	.23**	.35**	.21**	.19**	.66**	-.16**	.39**	.34**	(
10	Physical Aggression	.16**	-.02	-.19**	.14*	-.39**	.05	-.28**	-.04	-.
11	MI-I	.17**	.29**	.14*	.08	.50**	-.07	.37**	.25**	.
12	MI-S	.43**	.31**	.22**	.39**	.41**	-.18**	.18**	.21**	.
13	Prohibitive-U	.14*	.30**	-.03	.08	.25**	.01	.18**	.33**	.
14	Prohibitive-R	-.15**	.12*	-.08	-.11	.09	.09	.13*	.13*	
15	Permissive-U	.23**	-.09	.09	.20**	-.06	-.09	-.07	-.13*	-
16	Permissive-R	.15**	-.19**	.11*	.13*	-.08	-.10	-.03	-.21**	-
17	Hypervigilance	.26**	.07	-.09	.08	-.19**	.22**	-.21**	-.01	-
18	SVS	.48**	.19**	-.05	.23**	-.10	.10	-.12*	.19**	-
19	Moral Attentiveness	.31**	.11*	-.04	.24**	.10	.14*	.15**	.21**	
20	VIS (Time 3)	.82**	.26**	.14*	.38**	.25**	-.06	.12*	.27**	.
21	Likelihood to Post	.20**	.04	-.08	0	.09	.09	-.10	.00	
22	Age	-.19**	-.10	.23**	.09	.10	-.16**	.15**	-.04	
23	Male	.00	-.06	-.10	.02	-.21**	-.13*	-.05	.12*	-.
24	Conservatism	-.02	-.23**	.16**	.14*	-.07	-.20**	.10	-.12*	-
	Mean	4.18	5.04	5.48	3.51	5.04	3.34	4.80	4.48	5
	SD	1.24	1.42	1.37	1.40	1.24	1.67	0.95	1.53	0
	N	318	318	318	318	318	318	318	318	3

Notes. * $p < .05$; ** $p < .01$

Table 4 (Cont.)

		12	13	14	15	16	17	18	19	20
12	MI-S	(.92)								
13	Prohibitive-U	.22**	(.93)							
14	Prohibitive-R	-0.07	.38**	(.85)						
15	Permissive-U	.16**	-.29**	-.55**	(.91)					
16	Permissive-R	.11*	-.51**	-.47**	.70**	(.88)				
17	Hypervigilance	.11*	-0.1	-.24**	.31**	.31**	(.85)			
18	SVS	.19**	.16**	-0.01	.19**	.14*	.41**	(.88)		
19	Moral Attentiveness	.31**	.19**	0.06	0.06	0.06	.27**	.34**	(.96)	
20	VIS (Time 3)	.45**	.23**	-.14*	.24**	.13*	.27**	.47**	.33**	(.96)
21	Likelihood to Post	.13*	.20**	.14*	-0.06	-.17**	0.04	.12*	0.05	.20**
22	Age	-0.08	-.14*	-0.02	-0.02	.11*	-.16**	-.18**	-.13*	-.14*
23	Male	-.18**	0.1	.13*	-0.03	-0.09	-0.04	.21**	0.02	-0.01
24	Conservatism	-0.01	-.40**	-.35**	.43**	.52**	.11*	-0.06	-0.03	-0.06
	Mean	3.79	4.46	4.98	3.60	4.54	2.51	4.00	3.10	3.86
	SD	1.52	1.49	1.44	1.57	1.44	0.92	0.96	1.54	1.40
	N	318	318	318	318	318	318	318	318	318

Notes. * $p < .05$; ** $p < .01$

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology, 76*(5), 732-740. https://doi.org/10.1037/0021-9010.76.5.732

Aquino, K., Freeman, D., Reed, A., Felps, W., & Lim, V. K. G. (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), 123-141.

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423-1440. https://doi.org/10.1037//0022-3514.83.6.1423

Badaracco, J. L. (1997). The internet, intel and the vigilante stakeholder. *Business Ethics: A European Review, 6*(1), 18-29. https://doi.org/10.1111/1467-8608.00044

Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2016). Altruistic punishment does not increase with the severity of norm violations in the field. *Nature Communications, 7*(1), 13327. https://doi.org/10.1038/ncomms13327

Bateson, R. (2020). The politics of vigilantism. *Comparative Political Studies, 54*(6), 923-955. https://doi.org/10.1177/0010414020957692

Bernstein, R. E., Delker, B. C., Knight, J. A., & Freyd, J. J. (2015). Hypervigilance in college students: Associations with betrayal and dissociation and psychometric properties in a brief hypervigilance scale. *Psychological Trauma: Theory, Research, Practice, and Policy, 7*(5), 448-455. https://doi.org/10.1037/tra0000070

Brewer, M. B., & Gardner, W. (1996). Who is this "we"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71*(1), 83-93. https://doi.org/10.1037/0022-3514.71.1.83

Brickson, S. (2000). The impact of identity orientation on individual and organizational outcomes in demographically diverse settings. *The Academy of Management Review, 25*(1), 82-101. https://doi.org/10.2307/259264

Brown, R. P. (2003). Measuring individual differences in the tendency to forgive: Construct validity and links with depression. *Personality and Social Psychology Bulletin, 29*(6), 759-771. https://doi.org/10.1177/0146167203029006008

Burke, P. J., & Tully, J. C. (1977). The measurement of role identity. *Social Forces, 55*(4), 881-897. https://doi.org/10.1093/sf/55.4.881

Burrows, W. E. (1976). *Vigilante*. Harcourt Brace Jovanovich.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*(3), 452-459. https://doi.org/10.1037/0022-3514.63.3.452

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*(3), 306-307. https://doi.org/10.1207/s15327752jpa4803_13

Canter, D., Youngs, D., & Yaneva, M. (2017). Towards a measure of kindness: An exploration of a neglected interpersonal trait. *Personality and Individual Differences, 106*, 15-20. https://doi.org/10.1016/j.paid.2016.10.019

Chang, L. Y. C., & Poon, R. (2016). Internet vigilantism: Attitudes and experiences of university students toward cyber crowdsourcing in Hong Kong. *International Journal of Offender*

Therapy and Comparative Criminology, 61(16), 1912-1932. https://doi.org/10.1177/0306624X16639037

Chui, C., & Grieder, M. (2020). The effects of investigative sanctioning systems on wrongdoing, reporting, and helping: A multiparty perspective. *Organization Science, 31*(5), 1090-1114. https://doi.org/10.1287/orsc.2019.1340

Colquitt, J. A., Sabey, T. B., Rodell, J. B., & Hill, E. T. (2019). Content validation guidelines: Evaluation criteria for definitional correspondence and definitional distinctiveness. *Journal of Applied Psychology, 104*(10), 1243-1265. https://doi.org/10.1037/apl0000406

Costas, J., & Grey, C. (2018). Violence and organization studies. *Organization Studies, 40*(10), 1573-1586. https://doi.org/10.1177/0170840618782282

Cousins, S. (2020). New Zealand eliminates Covid-19. *The Lancet, 395*(10235), 1474. https://doi.org/10.1016/S0140-6736(20)31097-7

Crawford, B., & Dacin, M. T. (2020). Policing work: Emotions and violence in institutional work. *Organization Studies*, 1-22. https://doi.org/10.1177/0170840620941614

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. https://doi.org/10.1037/h0040957

Darley, J. M. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science, 5*(1), 1-23. https://doi.org/10.1146/annurev.lawsocsci.4.110707.172335

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior, 24*(6), 659-683. https://doi.org/10.1023/A:1005552203727

de Rond, M., Lok, J., & Marrison, A. (2021). To catch a predator: The lived experience of extreme practices. *Academy of Management Journal*. https://doi.org/10.5465/amj.2020.1492

DeCelles, K., & Aquino, K. (2020). The dark knight. *Academy of Management Review, 45*(3), 528-548. https://doi.org/10.5465/amr.2017.0300

DeJordy, R. (2010). *Institutional guardianship: The role of agency in preserving threatened institutional arrangements* (Publication Number 3413608) [Ph.D., Boston College]. ProQuest One Academic. Ann Arbor. https://www.proquest.com/dissertations-theses/institutional-guardianship-role-agency-preserving/docview/750076477/se-2?accountid=14700

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880-896. https://doi.org/10.1037/0022-3514.93.5.880

Dumsday, T. (2009). On cheering Charles Bronson: The ethics of vigilantism. *The Southern Journal of Philosophy, 47*(1), 49-67. https://doi.org/10.1111/j.2041-6962.2009.tb00131.x

Dutton, J. E., Roberts, L. M., & Bednar, J. (2010). Pathways for positive identity construction at work: Four types of positive identity and the building of social resources. *Academy of Management Review, 35*(2), 265-293. https://doi.org/10.5465/amr.2010.48463334

Ellemers, N. (2018). Morality and social identity. In *The Oxford Handbook of the Human Essence.* (pp. 147-158). Oxford University Press.

Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology, 24*(1), 160-193. https://doi.org/10.1080/10463283.2013.841490

Erikson, E. H. (1964). *Insight and responsibility: Lectures on the ethical implications of psychoanalytic insight*. W W Norton & Co.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63-87. https://doi.org/10.1016/s1090-5138(04)00005-4

Galleguillos, S. (2021). Digilantism, discrimination, and punitive attitudes: A digital vigilantism model. *Crime, Media, Culture*, 17416590211017937. https://doi.org/10.1177/17416590211017937

Gill, M. J., & Burrow, R. (2017). The function of fear in institutional maintenance: Feeling frightened as an essential ingredient in haute cuisine. *Organization Studies, 39*(4), 445-465. https://doi.org/10.1177/0170840617709306

Goldstein, D. M. (2003). "In our own hands": Lynching, justice, and the law in Bolivia. *American Ethnologist, 30*(1), 22-43. https://doi.org/10.1525/ae.2003.30.1.22

Gollwitzer, P. M. (1986). Striving for specific identities: The social reality of self-symbolizing. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 143-159). Springer New York. https://doi.org/10.1007/978-1-4613-9564-5_7

Gollwitzer, P. M., Wicklund, R. A., & Hilton, J. L. (1982). Admission of failure and symbolic self-completion: Extending Lewinian theory. *Journal of Personality and Social Psychology, 43*(2), 358-371. https://doi.org/10.1037/0022-3514.43.2.358

Graso, M., Camps, J., Strah, N., & Brebels, L. (2020). Organizational justice enactment: An agent-focused review and path forward. *Journal of Vocational Behavior, 116*(B). https://doi.org/10.1016/j.jvb.2019.03.007

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96*, 505-520. https://doi.org/10.1037/a0013748

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science, 312*(5770), 108. https://doi.org/10.1126/science.1123633

Hagan, J., Mccarthy, B., Herda, D., & Chandrasekher, A. C. (2018). Dual-process theory of racial isolation, legal cynicism, and reported crime. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 7190 - 7199.

Haidt, J. (2013). *The righteous mind: Why good people are divided by politics and religion*. Vintage Books.

Haidt, J., & Kesebir, S. (2010). Morality. In *Handbook of social psychology, vol. 2, 5th ed.* (pp. 797-832). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002022

Health, Ministry of Health. (2020). *Covid-19: Elimination strategy for Aotearoa New Zealand*. Ministry of Health Manatu Hauora. Retrieved September 20 from https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-response-planning/covid-19-elimination-strategy-aotearoa-new-zealand

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Herron, E. (2018). *A woman claimed she was confronting Nazi sympathizers at a Portland restaurant, and now it's a whole thing*. Willamette Week. Retrieved July 14 from

https://www.wweek.com/culture/2018/03/15/a-woman-claimed-she-was-confronting-nazi-sympathizers-at-a-portland-restaurant-and-now-its-a-whole-thing/

Higgins, E. T. (1989). Self-discrepancy theory: What patterns of self-beliefs cause people to suffer? In *Advances in experimental social psychology, vol. 22.* (pp. 93-136). Academic Press. https://doi.org/10.1016/S0065-2601(08)60306-8

Higgins, E. T., & Brendl, C. M. (1995). Accessibility and applicability: Some "activation rules" influencing judgment. *Journal of Experimental Social Psychology, 31*(3), 218-243. https://doi.org/10.1006/jesp.1995.1011

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104-121. https://doi.org/10.1177/109442819800100106

Hinkin, T. R., & Tracey, J. B. (1999). An analysis of variance approach to content validation. *Organizational Research Methods, 2*(2), 175-186. https://doi.org/10.1177/109442819922004

Hofmann, W., Brandt, M. J., Wisneski, D. C., Rockenbach, B., & Skitka, L. (2018). *Moral punishment in everyday life* (Vol. 44). https://doi.org/10.1177/0146167218775075

Huang, J. L., Cropanzano, R., Li, A., Shao, P., Zhang, X., et al. (2017). Employee conscientiousness, agreeableness, and supervisor justice rule compliance: A three-study investigation. *Journal of Applied Psychology, 102*(11), 1564-1589. https://doi.org/10.1037/apl0000248

Huggins, M. K. (1991). *Vigilantism and the state in modern Latin America: Essays on extralegal violence*. Praeger.

Jacobs, D., Carmichael, J. T., & Kent, S. L. (2005). Vigilantism, current racial threat, and death sentences. *American Sociological Review, 70*(4), 656-677. https://doi.org/10.1177/000312240507000406

Johnston, L. (1996). What is vigilantism? *British Journal of Criminology, 36*(2), 220-236. https://doi.org/10.1093/oxfordjournals.bjc.a014083

Johnston, L. E. S. (1992). The politics of private policing. *The Political Quarterly, 63*(3), 341-349. https://doi.org/10.1111/j.1467-923x.1992.tb00905.x

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in Psychology, 8*(1359). https://doi.org/10.3389/fpsyg.2017.01359

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science, 322*(5908), 1681. https://doi.org/10.1126/science.1161405

Klatetzki, T. (2017). Organizing vigilante lynching: A neo-institutional perspective In G. Krücken, C. Mazza, R. E. Meyer, & P. Walgenbach (Eds.), *New themes in institutional analysis: Topics and issues from European research* (pp. 345–376). Edward Elgar. https://doi.org/10.4337/9781784716875.00017

Koopman, J., Matta, F. K., Scott, B. A., & Conlon, D. E. (2015). Ingratiation and popularity as antecedents of justice: A social exchange and social capital perspective. *Organizational Behavior and Human Decision Processes, 131*, 132-148. https://doi.org/10.1016/j.obhdp.2015.09.001

Leavitt, K., Zhu, L., & Aquino, K. (2016). Good without knowing it: Subtle contextual cues can activate moral identity and reshape moral intuition. *Journal of Business Ethics, 137*(4), 785-800. https://doi.org/10.1007/s10551-015-2746-6

Lee, K., & Allen, N. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *Journal of Applied Psychology, 87*(1), 131-142. https://doi.org/10.1037/0021-9010.87.1.131

Litman, L., Robinson, J., & Abberbock, T. (2017). Turkprime.Com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433-442. https://doi.org/10.3758/s13428-016-0727-z

Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology, 51*(4), 858-866. https://doi.org/10.1037/0022-3514.51.4.858

Martí, I., & Fernández, P. (2013). The institutional work of oppression and resistance: Learning from the holocaust. *Organization Studies, 34*(8), 1195-1223. https://doi.org/10.1177/0170840613492078

McCall, G. J., & L, S. J. (1978). *Identities and interactions*. The Free Press.

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2010). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28-50. https://doi.org/10.1177/1088868310366253

Moncada, E. (2017). Varieties of vigilantism: Conceptual discord, meaning and strategies. *Global Crime, 18*, 1-21. https://doi.org/10.1080/17440572.2017.1374183

Osaki, T. (2020). *Japan's 'virus vigilantes' take on rule-breakers and invaders*. Retrieved January 20 from https://www.japantimes.co.jp/news/2020/05/13/national/coronavirus-vigilantes-japan/

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives, 14*(3), 137-158. https://doi.org/10.1257/jep.14.3.137

Oyserman, D. (2009). Identity-based motivation and consumer behavior. *Journal of Consumer Psychology, 19*(3), 276-279. https://doi.org/10.1016/j.jcps.2009.06.001

Phillips, B. J. (2016). Inequality and the emergence of vigilante organizations: The case of mexican autodefensas. *Comparative Political Studies, 50*(10), 1358-1389. https://doi.org/10.1177/0010414016666863

Remington Abramson, N., & Senyshyn, Y. (2010). Effective punishment through forgiveness: Rediscovering Kierkegaard's knight of faith in the Abraham story. *Organization Studies, 31*(5), 555-581. https://doi.org/10.1177/0170840610372202

Reynolds, S. J. (2008). Moral attentiveness: Who pays attention to the moral aspects of life? *Journal of Applied Psychology, 93*(5), 1027-1041. https://doi.org/10.1037/0021-9010.93.5.1027

Robinson, P. H., & Robinson, S. M. (2016). Shadow vigilante officials manipulate and distort to force justice from an apparently reluctant system. *Faculty Scholarship at Penn Law, 1662*.

Robinson, P. H., & Robinson, S. M. (2018). *Shadow vigilantes: How distrust in the justice system breeds a new kind of lawlesness*. Prometheus.

Rosenbaum, H. J., & Sederberg, P. C. (1974). Vigilantism: An analysis of establishment violence. *Comparative Politics, 6*(4), 541. https://doi.org/10.2307/421337

Saucier, D. A., & Webster, R. J. (2010). Social vigilantism: Measuring individual differences in belief superiority and resistance to persuasion. *Personality and Social Psychology Bulletin, 36*(1), 19-32. https://doi.org/10.1177/0146167209346170

Silke, A. (2001). Dealing with vigilantism: Issues and lessons for the police. *The Police Journal: Theory, Practice and Principles, 74*(2), 120-133. https://doi.org/10.1177/0032258x0107400204

Skitka, L. J. (2003). Of different minds: An accessible identity model of justice reasoning. *Personality and Social Psychology Review, 7*(4), 286-297. https://doi.org/10.1207/s15327957pspr0704_02

Skitka, L. J., & Crosby, F. J. (2003). Trends in the social psychological study of justice. *Personality and Social Psychology Review, 7*(4), 282-285. https://doi.org/10.1207/S15327957PSPR0704_01

Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The psychology of moral conviction. *Annual Review of Psychology, 72*(1), null. https://doi.org/10.1146/annurev-psych-063020-030612

Skitka, L. J., & Morgan, G. S. (2009). The double-edged sword of a moral state of mind. In *Personality, identity, and character: Explorations in moral psychology.* (pp. 355-374). Cambridge University Press. https://doi.org/10.1017/CBO9780511627125.017

Skitka, L. J., Washburn, A. N., & Carsel, T. S. (2015). The psychological foundations and consequences of moral conviction. *Current Opinion in Psychology, 6*, 41-44. https://doi.org/10.1016/j.copsyc.2015.03.025

Sluss, D. M., & Ashforth, B. E. (2007). Relational identity and identification: Defining ourselves through work relationships. *The Academy of Management Review, 32*(1), 9-32. https://doi.org/10.2307/20159278

Solove, D. J. (2007). I've got nothing to hide and other misunderstandings of privacy *San Diego Law Review 44*(4), 745-772.

Song, E. Y. (2019). Protect to damage? Institutional work, unintended consequences and institutional dynamics. *Organization Studies, 42*(3), 495-517. https://doi.org/10.1177/0170840618814564

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five inventory–2: The bfi-2-s and bfi-2-xs. *Journal of Research in Personality, 68*, 69-81. https://doi.org/10.1016/j.jrp.2017.02.004

Stryker, S. (1980). *Symbolic interactionism: A social structural version* Benjamin-Cummings Publishing.

Tait, A. (2020). *Pandemic shaming: Is it helping us keep our distance?* . The Guardian. Retrieved May 1 from https://www.theguardian.com/science/2020/apr/04/pandemic-shaming-is-it-helping-us-keep-our-distance

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the multidimensional personality questionnaire. In *The Sage Handbook of Personality Theory and Assessment, Vol 2: Personality Measurement and Testing.* (pp. 261-292). Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n13

Tenbrunsel, A. E., Smith-Crowe, K., & Umphress, E. E. (2003). Building houses on rocks: The role of the ethical infrastructure in organizations. *Social Justice Research, 16*(3), 285-307. https://doi.org/10.1023/A:1025992813613

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology, 64*(1), 231-255. https://doi.org/10.1146/annurev-psych-113011-143812

Tripp, T. M., Bies, R. J., & Aquino, K. (2007). A vigilante model of justice: Revenge, reconciliation, forgiveness, and avoidance. *Social Justice Research, 20*(1), 10-34. https://doi.org/10.1007/s11211-007-0030-3

Turiel, E. (2006). The development of morality. In *Handbook of child psychology: Social, emotional, and personality development, vol. 3, 6th ed.* (pp. 789-857). John Wiley & Sons, Inc.

Weisburd, D. (1988). Vigilantism as community social control: Developing a quantitative criminological model. *Journal of Quantitative Criminology, 4*(2), 137-153. http://doi.org/10.1007/BF01062870

Wenzel, M., Okimoto, T. G., Feather, N. T., & Platow, M. J. (2008). Retributive and restorative justice. *Law and Human Behaviour, 32*, 375-389.

Yamamoto, S., & Maeder, E. M. (2019). Creating the punishment orientation questionnaire: An item response theory approach. *Personality and Social Psychology Bulletin, 45*(8), 1283-1294. https://doi.org/10.1177/0146167218818485

Zhu, L., Martens, J. P., & Aquino, K. (2012). Third party responses to justice failure. *Organizational Psychology Review, 2*(2), 129-151. https://doi.org/10.1177/2041386611434655

APPENDIX A

Study 1: Partner-rated Monitoring and Punishment Behavior
The following are some behaviors that people may demonstrate when they see somebody break the C19 rules. Consider how [your partner] usually behaves. How well do these statements describe [your partner]? 1 = *does not describe them at all* to 5 = *describes them extremely well*

Monitoring Behavior

1. Monitors others in our immediate community in case of others' violating the C19 mandates
2. Monitors others in case they break the rules
3. Notices when others break the rules

Punishment Behavior

1. Complains to me when others violate the rules
2. Makes formal complaints about people breaking the self-isolation rules (such as through a NZ C19 hotline, police, or human resources)
3. Reprimands others for breaking the rules
4. Reports others breaking the rules behavior publicly (such as via Facebook, Twitter, letter to the editor, blogs, etc.)
5. Makes an effort to punish others for breaking the rules

Study 1: Citizenship behavior (adapted from Lee & Allen, 2002)

The following statements refer to your partner's behavior during this time. Consider how well these statements describe [your partner]. 1 = *does not describe them at all* to 5 = *describes them extremely well*

1. Seeks opportunities to do good in our community
2. Goes out of their way to help those who need it
3. Willingly gives their time to help others
4. Adjusts their schedule to accommodate others
5. Goes out of the way to make others feel better
6. Shows genuine concern and courtesy toward others, even under the most challenging situations
7. Gives up time to help others who are experiencing various problems.
8. Seeks opportunities to help those who are struggling right now