Honor among Students: The Effects of Punishment Severity on Whistleblowing

Sophie Clayton
Charlottesville, Virginia

A predissertation research project presented to the
graduate faculty of the University of Virginia
in candidacy for the Degree of
Master of Arts

Department of Psychology
University of Virginia
August 2024

Primary advisor: Amrisha Vaish, Ph.D.
Second reader: Tobias Grossmann, Ph.D.

Honor among Students: The Effects of Punishment Severity on Whistleblowing

Sophie Clayton[1]†, Stefen Beeler-Duden[1]†, Udeyvir Cheema[1,2], Amrisha Vaish[1]

† These authors contributed equally to this work

[1]Early Social Development Laboratory, University of Virginia, Developmental Psychology, Charlottesville, Virginia, United States

[2]Duke University, School of Medicine, Durham, North Carolina, United States

***Correspondence:**
Sophie Clayton
sophieclayton@virginia.edu

Keywords: whistleblowing, punishment, moral conflict, moral concerns.

Abstract

The current study used a classroom cheating paradigm to investigate how severity of punishment affected the rate at which bystander students reported a confederate's blatant cheating behavior to an authority. Participants were asked to read and memorize a passage of literature before taking a memory and comprehension quiz about its contents. The instructions included one of three conditions: low, medium, and severe levels of punishment for cheating during the quiz. During the memorization and testing time, a confederate student cheated by using their phone. The rate of reporting was found to be significantly different across the punishment conditions; participants reported cheating more often in the lowest punishment condition than in the highest severity condition. These findings suggest that more severe punishments for transgressions may discourage whistleblowing behavior from peers; future work should further explore motivations at play, elaborating on the link between punishments and reports.

Honor among Students: The Effects of Punishment Severity on Whistleblowing

Humans regularly cooperate with others, often even with strangers and often even at a cost to themselves (Sober & Wilson, 1998). However, cooperation can result in a greater loss for co-operators (who invest resources such as time, energy, or material goods) than for non-co-operators or free riders (who benefit from the outcomes of the cooperation without investing any resources; Fehr & Fischbacher, 2004). This raises the question of how such cooperation could evolve and be maintained. One effective solution to this puzzle is that those who break the norms of cooperation are held accountable in some way, which induces norm violators to cooperate more in future interactions and thus helps to maintain cooperation in the group (Nowak, 2006; Boyd & Richerson, 2009).

Norm enforcement can take several forms (Marshall & McAuliffe, 2022). Perhaps the best understood form is punishment, i.e., imposing costs or harm on a transgressor (Fehr & Fischbacher, 2004; Raihani et al., 2012). This may include direct punishment such as physically or verbally confronting the transgressor, or indirect punishment such as gossiping or social exclusion (Henrich et al., 2006; Molho et al., 2020). Such punishment has been extensively studied and has indeed been shown to help maintain cooperation in groups (Gächter et al., 2008; Boyd et al., 2010; Balliet et al., 2011).

There are other forms of norm enforcement as well. One important form is whistleblowing, or reporting transgressors to those who have the authority to punish them (Waytz et al., 2013). Reporting may result in the transgressor being punished, but the norm enforcer faces a lower risk than if they engaged in direct punishment, which puts them at risk of retaliation (Nikiforakis, 2008; Balafoutas et al., 2014). Thus, to understand the phenomenon of

norm enforcement more fully, it is vital to understand the mechanisms underlying whistleblowing and, in particular, when people do or do not choose to blow the whistle.

Whistleblowing can occur in a range of contexts, both structured, such as professional or academic systems, as well as unstructured or interpersonal interactions (Bocchiaro, Zimbardo, & Van Lange, 2012; Gino & Bazerman, 2009; Waytz et al., 2013). Even young children have been shown to blow the whistle. For instance, by 2-3 years of age, children tattle to authority figures about witnessed transgressions, including third-party transgressions that do not impact them in any way (Dunn & Munn, 1985; Vaish et al., 2011; Yucel & Vaish, 2018). By early school age, children show greater complexity in their decisions to report. For instance, between 6-11 years, children judged reporting to be appropriate only for major transgressions, whereas younger children endorsed reporting for both major and minor violations (Loke et al., 2011). Whistleblowing is thus an early-emerging and fairly common form of norm enforcement.

Yet not everyone blows the whistle when they witness a transgression. A recent systematic review of factors that affect whistleblowing found eight significant dimensions, including personal factors, organizational factors, costs and benefits, outcome expectancies, aspects of the offense, the mechanisms of reporting, characteristics of the wrongdoer, and social factors (Nicholls et al., 2021). The key findings highlight several concerns external to the witness, such as situational context, financial incentives, perceived protection from retaliation, knowledge regarding whistleblowing pathways, and the severity of the transgression (Oh & Teo, 2010; Latan et al., 2023; Dungan et al., 2019; Near & Micelli, 1985; Richardson et al., 2012; Alleyne et al., 2013; Cassematis & Wortley, 2013; Dungan et al., 2015). Personal factors such as closeness with the transgressor, affiliation with group interests, and individual endorsements of

moral concerns like fairness and loyalty were also found to be significant (Trevino & Victor, 1992; Waytz et al., 2013).

Individuals' moral concerns also play a role, as whistleblowing presents a conflict between competing moral concerns. In particular, the reporter may need to consider their loyalty to the transgressor against their concerns of justice, fairness, and empathy for those hurt by the wrongdoing. A body of recent research has demonstrated that tradeoffs between such competing moral concerns predict people's likelihood of blowing the whistle. In particular, higher concerns for fairness predict greater likelihood of whistleblowing, whereas higher concerns for loyalty decrease the likelihood (Waytz et al., 2013; Dungan et al., 2019). Relatedly, research with children revealed that for severe, but not mild, transgressions, 5-year-olds were less willing to report their ingroup members than outgroup members (Misch et al. 2018). The authors proposed that the ingroup difference between mild and severe transgressions may have resulted from children's assumption that the punishment for more severe transgressions would be harsher and from their greater loyalty and desire to protect ingroup members more than outgroup members from the harsh punishment.

This line of work on the competing moral concerns surrounding whistleblowing raises an interesting question about whether the severity of the punishment faced by the transgressor may also impact the likelihood of whistleblowing. It is possible, for instance, that in addition to their loyalty to the transgressor, potential whistleblowers' concern that the punishment for a transgression is too severe may also reduce their likelihood to blow the whistle. Prior work suggests that whistleblowing is motivated by both vindictiveness (the desire to have the transgressor punished for their wrongdoing) and the desire to help the wrongdoer by reporting their behavior and seeking support (Rennie & Crosby, 2002). Thus, potential whistleblowers are

attuned to the consequences of their reporting and may be seeking what they deem as an appropriate punishment.

Whistleblowing decisions may indeed be influenced by concern for the transgressor. Weidman et al. (2020)'s model of the relationship between closeness and whistleblowing to a police officer revealed a mediating factor of concern for harm to the perpetrator. When participants were more concerned about the harm a transgressor faced, they reported decreased intentions to blow the whistle. On a self-report survey, fear of hurting a colleague was the fourth most reported reason why participants did not report a violation after fears of retribution, dismissal, and disruption (Moore & McAuliffe, 2010).

Although punishment's effect on whistleblowing has not, to our knowledge, been directly tested in prior work, the limited research on related questions has painted a mixed picture. Some recent work suggests that harsh punishments are viewed negatively, and peer punishment is met with disapproval from group members, especially when punishments are severe (Henrich et al., 2006; Nelissen, 2008; Eriksson et al., 2016). On the other hand, Krügel and Uhl (2023) found that mild and unreliable punishment resulted in decreased whistleblowing, whereas harsh and consistent punishment increased it. However, as the authors noted in their discussion, their results could be due to differences in consistency rather than severity of punishment, since the two factors were not separated. It thus remains an open question whether the severity of punishment for transgressions impacts the likelihood of observers reporting the transgressions. This was the focal question of the current research. We addressed this question in the context of peer reporting of academic misconduct.

Whistleblowing about academic misconduct

Although academic misconduct garners a great deal of attention from researchers and administrators, the role of whistleblowing in this context is still unclear. Colleges and universities often implement honor systems to discourage academic and personal dishonesty. These traditionally include a written pledge, a judicial structure with active student involvement, and, critically, an expectation that students report other students' dishonest behavior (Melendez, 1985; Arnold et al., 2007). Indeed, an atmosphere of whistleblowing may directly impact cheating instances; in one study, increased perception of peer-reporting was associated with a reduction in both instances and severity of cheating (Burrus et al., 2013).

Yet although students consistently state that cheating is wrong (Davis et al., 1992; Miller et al., 2011; Waltzer et al., 2021; Waltzer & Dahl, 2022), they are nonetheless quite reluctant to report their cheating peers (Davis et al., 1992; Jendrek, 1992; Rennie & Crosby, 2002; Stone et al., 2009; J. Scrimpshire et al., 2017; Yachison et al., 2017; Waltzer et al., 2024). What, then, holds students back from reporting? One explanation, echoed in the general whistleblowing literature described above, is that reporting others' violations can have adverse social consequences for the reporter, including lower likeability ratings from peers, social anxiety surrounding retaliation, and exclusion (Friman et al., 2004; Greenberger et al., 1987; Nora & Zhang, 2010; Rennie & Crosby, 2002; Waltzer et al., 2024). Additionally, age, gender, area of study, certain personality traits, group membership, and clarity on what is considered a violation are also associated with changes in a students' willingness to report (Dungan et al., 2015; Misch et al., 2018; Pupovac et al., 2019; Stevenson et al., 2023; Stone et al., 2012). Jenkel and Haen (2012) found that group scoring criteria (e.g. curved exams) resulted in increased peer reporting of cheating compared to individual test scoring; students were more likely to report cheating when it would directly affect their own scores. Rennie & Crosby (2002) found that student

support for reporting decreased with year level: Medical students in their first year were most likely to support peer-reporting compared to those in later years of the program. One possible explanation for these age-related trends is the emergence of a social norm that encourages students to protect others by refraining from reporting. As students spend more time together in a group or program, such social conventions have time to grow in normative strength and exert increasing force (Przepiorka et al., 2022).

Students' self-reported reasons for not reporting also reflect competing moral concerns documented in whistleblowing more generally. For instance, students report the desire to maintain camaraderie, concerns about the cheater's welfare, social conformity, a desire to protect other students from punishment, and even a belief that their report could ruin the cheater's life (Nitsch et al., 2005; Pupovac et al., 2019; Rennie & Crosby, 2002; Waytz et al., 2013). In an experimental paradigm to explore the impact of these self-reported reasons, students' decisions to report a hypothetical peer were affected by the ambiguity around the cheating instance, severity of the violation, grades and punishment consequences, and welfare concerns, among other factors (Waltzer et al., 2022).

Importantly, Waltzer et al. found that students who were less willing to report their peers were also more likely to state incorrect beliefs about punishments for cheating, including overestimating the severity of the punishment. This suggests that the severity of punishment for academic dishonesty may indeed play an important role in shaping whether or not students report observed academic dishonesty. However, because this work relied on hypothetical scenarios, it did not directly assess students' experiences of observing cheating and deciding whether to report. In previous whistleblowing studies, researchers have found a tendency to overpredict

one's own willingness to report; it is essential to examine students' actual behaviors and decisions in real life scenarios (Bocchiaro et al., 2012).

Together, these findings indicate that as with whistleblowing in general, whether students do or do not report a peer's cheating is shaped by the moral ambivalence and competing concerns that would-be reporters face: Reporting a peer may appeal to a student's moral concerns of fairness, honesty, or integrity, yet their compassion and empathy for the cheater, who faces punishment, as well as their sense of the fairness or proportionality of the punishment, could potentially counteract their sense of obligation to report. This leads to the possibility that when the punishment for cheating is severe, students may be less likely to report a peer's cheating than when the punishment is light or moderate. This was the hypothesis tested in the present study.

The current study

In this experimental study, we investigated how the severity of punishment for cheating impacts the likelihood of students reporting clear cases of observed cheating in an academic setting. Four undergraduate student participants were administered a literary memory quiz in the same room; a fifth confederate student used their phone to cheat on the quiz. Prior to taking the quiz, students were informed about the punishment for cheating, which was manipulated to be either low, medium, or severe. We predicted that students would be less likely to report observed cheating when the associated punishment for the cheater was severe rather than low or moderate.

To capture all levels of willingness to report and to encourage consideration of the choice to report, students were given three increasingly explicit opportunities to report the confederate for cheating. This was based on prior findings that most students did not spontaneously report a peer's cheating or confront the cheater, but they did report the cheating when asked directly by an examiner, and more students reported peer cheating when directly

asked a second time (Vaughn et al., 2009; Yachison et al., 2017). Thus, repeated opportunities for reporting may allow for greater consideration and higher rates of reporting. In our paradigm, three opportunities to report (spontaneous reporting, indirect questions on a post-survey, and direct questioning by the experimenter) allowed for a binary measure of reporting (yes/no) and an ordinal measure of time and amount of prompting needed to report. We did not have specific hypotheses regarding the number of prompts students would need to report the cheater across conditions.

Finally, as a first, exploratory step towards identifying potential individual variation in reporting behavior, participants also completed two questionnaires: The Moral Foundations Questionnaire (MFQ) is based on a theoretical model of universally available moral intuitions (Graham et al., 2011), and the Buss-Perry Aggression Questionnaire (BPAQ) measures aggression through four scales: physical aggression, verbal aggression, anger, and hostility (Buss & Perry, 1992). We tentatively expected that individuals with higher ratings of fairness, submission to authority, and purity may have higher rates of reporting in our study, and that individuals with higher loyalty and harm/care ratings may have lower rates of reporting (Waytz et al., 2013; Weidman et al., 2020; Nicholls et al., 2021). Further, as previous work has shown that individuals who score higher on trait aggression tend to show less disapproval of punishment, we also tentatively predicted that those with higher trait aggression scores would be more likely to blow the whistle (Eriksson & Andersson, 2015).

Method

Participants

Participants were recruited from the host institution's departmental research recruitment system through which undergraduate students receive course credit for participation.

Because the tasks and surveys in this study were written in English, we excluded any non-English speakers if they indicated that they had trouble understanding the instructions. A total of 116 students were recruited, 8 of whom were excluded due to deviations from the protocol during their session or denial of media release and data usage after the debriefing. The final sample included 108 students (36 per condition; 76 female; 103 aged 18-21 years, 4 aged 22-23, and 1 aged 25 or above). Of the 107 participants who provided their race and ethnicity, 58 identified as White, 28 as Asian or Pacific Islander, four as Hispanic or Latino, 6 as African American, and 11 identified as multi-racial or selected more than one race.

Although we did not conduct an a priori power analysis, a post-hoc sensitivity analysis conducted in G*Power 3.1 revealed that the sample size of 108 would provide 80% power to detect an effect size of $d = 0.27$ (two-tailed, $\alpha = .05$; Faul et al., 2009). Note that our final sample of 108 is also near the middle of the range of samples included in comparable prior studies (44, 82, and 184 students in Jenkel & Haen, 2012, Vaughn et al., 2009, and Yachison et al., 2017, respectively).

Although our study paradigm did not explicitly invoke the host institution's honor system, it is worth noting that at the time of data collection, all participants were students under the institution's honor system. Under this honor system, at the time that the study was conducted, the punishments for students found guilty included either expulsion or a one-year leave of absence if guilt was admitted prior to a trial [citation blinded for review]. In 2019, only 18% of reported cases at the institution came from peer reports, and a 2018 audit found that two major explanations students gave for not engaging in whistleblowing were "uneasiness about the possibility of the student being dismissed from the university" and not believing that "the offense

is serious enough" [citation blinded for review]. This context informs our understanding of these participants in their academic honor environment.

Materials and Procedure

Participants were scheduled in groups of four (for a total of 10-11 groups per condition) in addition to one confederate (a trained research assistant acting as an additional participant). In the case of no-shows, the study proceeded as long as two participants were present in addition to the confederate (i.e., at least a group of 3). In our final sample, 5 groups had two participants (2 in the medium condition, 3 in the severe condition), 10 groups had three participants (4 low, 4 medium, 2 severe), and the remaining 17 groups had the planned four participants (6 low, 5 medium, 6 severe). One session with two participants was excluded due to deviations from the protocol. Six participants were excluded because they did not provide post-debrief consent.

The participants and the confederate arrived at the same time to the lab room and were all invited to sit around a table. An experimenter acting as a proctor consented the participants in-person. To reliably investigate the mechanisms of reporting academic dishonesty in a realistic situation, it was necessary to deceive participants into believing that they were participating in an unrelated task (not concerned with academic dishonesty or whistleblowing per se). We thus told participants that they were completing a literary memory retention test: They would be asked to read and memorize a passage of literature for 5 minutes before taking a 3-minute memory and comprehension quiz about its contents.

Instructions with punishment manipulation. The experimenter handed out an instruction sheet to each participant. These were handed out face down so that the experimenter would remain blind to condition. The sheet informed participants that they would be presented with a passage and given a set time (5 minutes) to memorize the passage before the passage was

collected and the memory and comprehension quiz was given out. The instruction sheet stated that participants would receive a set testing time (3 minutes) to complete the memory and comprehension quiz, and if they achieved a perfect score on the quiz, they would earn $10 in addition to their course credit. To ensure uniform testing difficulty for all participants, the quiz included one impossible question so that no perfect scores could be achieved.

The instruction sheet also stated that the participants should not speak to one another, any cheating or use of external aids should be reported to the proctor so that the offender may be punished, and if an individual was found to be committing misconduct, they would be punished according to one of three conditions. Each group of participants was randomly assigned to read one of three punishment conditions (for a total of 3 groups, or 36 participants per condition):

1. Low punishment: cheater would receive only $9 (instead of $10) for a perfect score.
2. Medium punishment: cheater would receive $0 if they achieve a perfect score.
3. Severe punishment: cheater would receive $0 if they achieve a perfect score, would be reported to the administrator in charge of the participant database for academic credit, and may no longer be eligible to participate in experiments that semester.

Importantly, to keep the confederate blind to condition, they received an instruction sheet with no information about the severity of punishment.

The instruction sheet contained six free-response comprehension questions to ensure that each participant understood the instructions and the punishment for rule violations. After reading through the instruction sheet, participants were asked to sign at the bottom to indicate that they had read and understood the study guidelines (to simulate the signing of an academic honor pledge). The proctor then collected the completed instruction sheets and signed consent forms from the participants and handed out one passage selected randomly from three possible

passages to each participant. The proctor reminded the participants not to speak to one another, began a 5-minute timer visible to all participants, and left the room.

Memorization and quiz. Participants then proceeded to read the literary passage. During the 5-minute memorization stage, a confederate pulled their phone out of their pocket and, with the ringer on full volume, took a picture of the passage. When the shutter sounded**,** the confederate scrambled to turn off the ringer switch and put the phone back in their pocket. Following the memorization time, the proctor re-entered the room, collected the passage from each participant, handed them the memory quiz, reminded them of the testing period length, began the timer again, and left the room. The confederate then pulled out their phone, placed it on the table, and pulled up the picture they had taken of the passage. The confederate proceeded to quite clearly look back and forth between the passage and their quiz. With approximately 30 seconds left on the timer, the confederate placed their phone back in their pocket.

After the testing period ended, the proctor returned to the room and collected the quizzes. The proctor then asked each student to enter a separate, smaller room (each attached to the main study room) to privately fill out a study survey and directed them to open the door once finished with the survey.

Survey and interview. In the separate room, participants were given multiple opportunities to report the confederate's cheating behavior:

1. First opportunity (Least explicit): The post-quiz written survey participants filled out included the question: "Is there anything you'd like to tell us about the study?"
2. Second opportunity (More explicit): After the participant completed the survey and questionnaires and opened the door, the proctor entered the participant's room, closed the door, collected the paper documents, and asked the participant several interview

questions. The first few questions asked participants about how the quiz was written and how well the participant thought they had done on the quiz. Finally, they asked "One of the participants did really well, I've never seen anyone get such a high score before! Any idea how they could have done so well?"

3. Third opportunity (Most explicit): If the participant did not report any cheating at the second opportunity, the proctor then immediately followed it by asking: "Did any participant cheat during the study?"

After providing these opportunities to report cheating, the proctor then asked the participant what they thought the study was about, how harsh they thought the punishment was (on a scale of 1 to 10, with 10 being the harshest; manipulation check), and if they could recall the punishment (comprehension check).

Questionnaires. After the survey and interview, the proctor handed the participant three questionnaires to complete: a demographic form, the Moral Foundations Questionnaire, and the Buss-Perry Aggression Questionnaire.

The Moral Foundations Questionnaire (MFQ) is based on a theoretical model of universally available moral intuitions (Graham et al., 2011). It is a 30-item questionnaire in which participants indicate on a 6-point Likert scale how relevant certain considerations are in their thinking about morality. The five official subscales, with six items each, are harm/care, fairness/reciprocity, in-group/loyalty, authority/respect, and purity/sanctity.

We calculated unique subscores for fairness and loyalty, each based on the three most relevant items determined by Waytz et al. (2013). Fairness included relevance ratings for the items "Whether or not someone acted unfairly" and "Whether or not someone was denied his or her rights" as well as agreement with the statement "Justice is the most important requirement

for a society." Loyalty included relevance ratings for "Whether or not someone did something to betray his or her group" and "Whether or not someone showed a lack of loyalty" and agreement with "People should be loyal to their family members, even when they have done something wrong." These subscores excluded statements about loyalty to one's country and government lawmaking.

Additionally, following Waytz et al. (2013)'s approach to fairness and loyalty, we also calculated unique authority, purity, and harm/care subscores based on the three most relevant items. Authority included relevance ratings of "Whether or not someone showed a lack of respect for authority and "Whether or not an action caused chaos or disorder" and agreement with "Respect for authority is something all children need to learn." The purity subscore included relevance ratings of "Whether or not someone violated standards of purity and decency" and "Whether or not someone did something disgusting" and agreement with "People should not do things that are disgusting, even if no one is harmed." This subscore excluded statements on conformity to gender roles, traditions, and commanding officer orders. The harm/care subscore included relevance ratings of "Whether or not someone suffered emotionally" and "Whether or not someone cared for someone weak or vulnerable" and agreement with "Compassion for those who are suffering is the most crucial virtue." This subscore excluded statements on cruelty, murder, and animal harm.

The Buss-Perry Aggression Questionnaire (BPAQ) measures aggression through four scales: physical aggression, verbal aggression, anger, and hostility (Buss & Perry, 1992). It is a 29-item questionnaire in which participants indicate on a 5-point Likert scale how much the statements are characteristic of themselves (e.g. "I have trouble controlling my temper"). For our

measure of trait aggression, we super-scored all four scales to create a single measure of aggression.

Once all participants completed their study survey, interview, and questionnaires, the proctor gathered all participants in the main study room and handed out the debriefing form, materials release form, and post-debrief consent. Participants were then debriefed to inform them of the purpose of the study. We also assured them that no part of the study would affect their academic standing, and any cheating or reactions to cheating would not be reported to any authorities, to minimize any stress the students may have experienced due to having cheated themselves or having reported or not reported the observed cheating. Due to the deception involved in the study, participants completed a post-debrief consent and materials release form to indicate whether they agreed to continue their participation and allow their data to be used after the true purpose of the study was revealed.

Results

Of the 108 total students who witnessed cheating, 61 (56%) reported the cheating through at least one avenue. One student confronted the cheater directly during the quiz period. This student was recorded as reporting at the earliest time point, the post-quiz survey, since after confronting, they did also report the cheating on the survey. There were no significant trends in reporting based on group size (varying from 2 to 4 participants in the session); 6 participants (60%) reported in the groups of 2, 8 (27%) reported in groups of 3, and 33 (49%) reported in groups of 4.

Manipulation check

During the post-quiz interview, participants were asked to rate the severity of their condition's punishment on a scale of 1-10. The punishment conditions followed the expected

trend, with perception of severity increasing from the low (M = 1.00, SD = 1.46) to medium (M = 3.59, SD = 1.86) to high severity conditions (M = 6.17, SD = 1.56). Thus, even though participants were only exposed to one level of punishment severity and the severity was not described as 'low,' 'medium,' or 'high,' they did nonetheless accurately perceive the manipulated level of severity according to condition.

Comprehension check

During the post-quiz interview, participants were asked to recall the punishment for their condition. Out of 108 students, 11 (10%) did not remember or incorrectly recalled the punishment. These students are not excluded from the reported analyses, as the same findings hold when excluded.

Reporting behavior

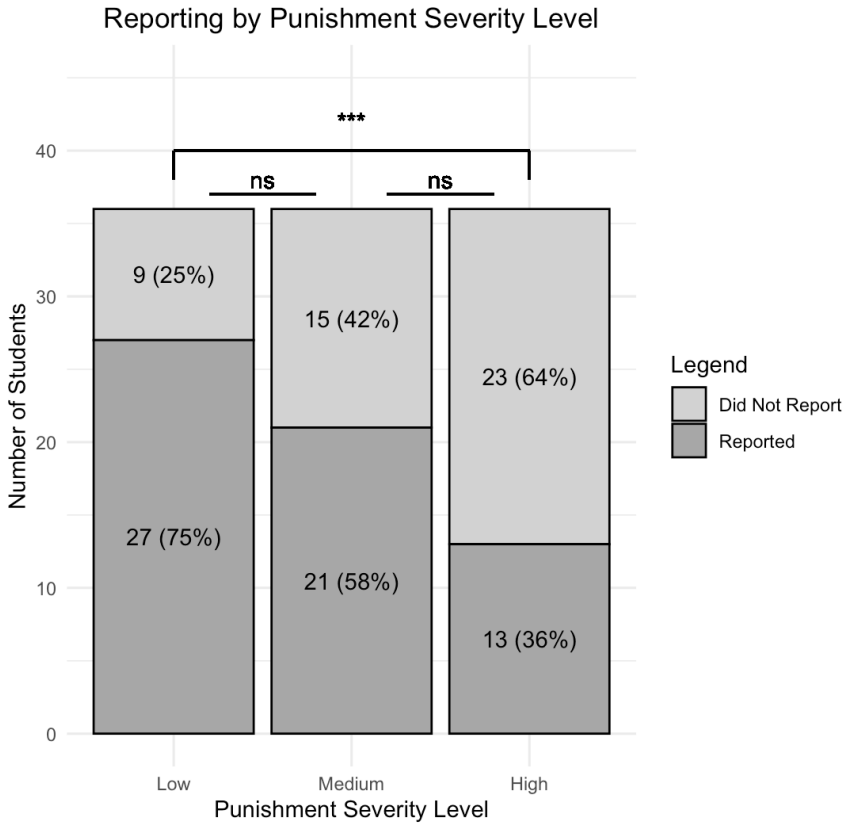
As hypothesized, fewer students reported cheating as severity increased across the three conditions: Out of the 36 students in each punishment condition, 27 (75%) reported cheating in the Low severity condition, 21 (58%) reported in the Medium severity condition, and 13 (36%) reported in the High severity condition (Figure 1). A chi-square test of independence was performed to examine the association between punishment severity and whether students reported cheating. This association was significant, X^2 (2, N = 108) = 11.15, p = .004, V = .29. The effect size of .29, calculated using Cramer's V, is considered moderate.

Post-hoc pairwise chi-square tests with Bonferroni corrections revealed that significantly more participants reported the cheating in the Low than the High severity condition (p = 0.002, adjusted p = 0.006). Somewhat more participants also reported in the Medium than the High severity condition, though this difference was not significant (p = 0.098, adjusted p =

0.20). No significant differences were found between the Low and Medium severity conditions ($p = 0.21$, adjusted $p = 0.21$).

A logistic regression was also conducted to estimate the strength and direction of the association between severity of punishment and likelihood of reporting cheating. The model revealed that the intercept was significant ($\beta = 1.099$, $p = 0.004$), indicating significant likelihood of reporting cheating in the Low severity condition. Using the Low severity condition as the reference level, the Medium severity condition was not significantly different ($\beta = -0.76$, $p = 0.14$), but the High severity condition ($\beta = -1.67$, $p = 0.001$) was associated with a significant decrease in the likelihood of reporting cheating.

Fig. 1.



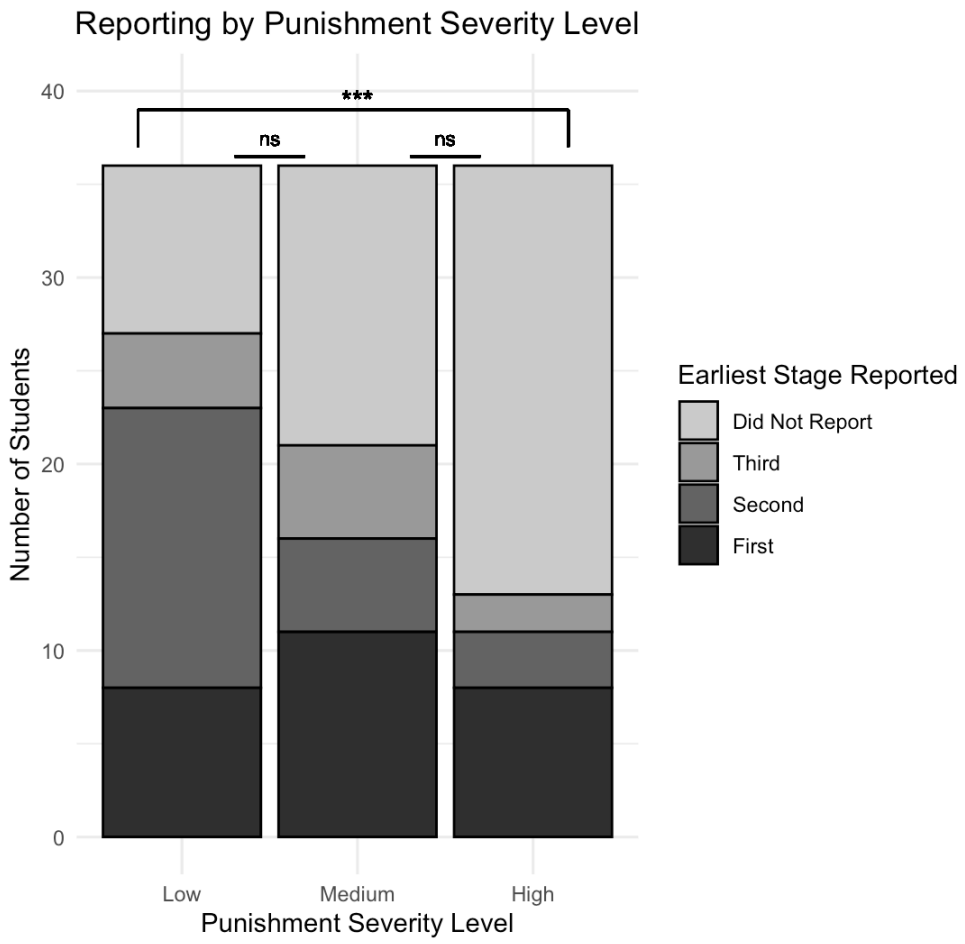
Exploratory analyses

Timepoint of reporting

Students were offered three opportunities to report cheating that sequentially increased how explicitly they were asked about observed cheating. We first conducted an exploratory analysis to examine whether the 61 students who reported did so at different timepoints based on condition (Figure 2). A Kruskal-Wallis test revealed no significant association between punishment severity level and the earliest stage at which students reported, X^2 (2, N = 61) = 2.16, p = .34.

This first analysis excluded students who did not report at all, lowering the total sample size and resulting in unequal group sizes. We therefore conducted another analysis that included a fourth timepoint to represent participants who never reported cheating (thus including the full sample). A Kruskal-Wallis test comparing conditions was significant, X^2 (6, N = 108) = 6.07, p = .048 (see Figure 2). Pairwise Wilcoxon rank-sum tests revealed a significant difference only between the Low and High severity conditions (p = 0.043). The Low and Medium conditions, as well as the Medium and High conditions, were not significantly different from one another, both ps > .337. This pattern replicates – and likely results from – the pattern reported above for whether or not students chose to report. We conclude that the sizable portion of students who did not report at all in the High severity condition may be driving this significant finding on earliest timepoint of reporting; therefore, we do not draw any strong conclusions regarding the trends in time at which students reported by condition.

Fig. 2. Number of students in each condition whose earliest reports were at the first, second, or third opportunities to report or who did not report at all.



Questionnaire responses

Our tentative hypothesis was that fairness, authority, purity, and aggression scores may be positively associated with reporting behavior, whereas loyalty and harm may be negatively associated with reporting behavior. Students who reported cheating (M = 3.35, SD = .72) had significantly higher authority concern scores than non-reporters (M = 3.02, SD =.69; $t(34)$ = 1.94, p = .020). We examined authority score within punishment severity conditions, and none had significantly different authority scores between reporters and non-reporters. This discrepancy may be due to a significant difference in authority scores between the severe condition (M = 2.96) and both the medium (M = 3.31; $t(70)$ = 2.03, p = .045, d = .48) and low (M

= 3.36; $t(70) = 2.53$, $p = .014$, $d = .60$) conditions. No other questionnaire scores were significantly different between reporters and non-reporters, all ps > .166.

In a correlation analysis, only authority scores were significantly correlated with whistleblowing, such that higher authority scores were associated with choosing to blow the whistle, $r(108) = .22$, $p = .020$. However, when controlling for level of severity, this correlation disappears.

We also explored the four subscales of the aggression questionnaire separately, but none were significantly correlated with reporting behavior.

Table 1. Zero-order (Pearson) correlations and partial correlations (controlling for punishment severity level) between reporting behavior and the MFQ 3-item subscales of authority, loyalty, fairness, purity, and harm, and the BPAQ score of aggression.

Correlations

Control Variables			Reported (Y/N)	Loyalty	Fairness	Authority	Purity	Harm	Aggression	Severity Level
-none-[a]	Reported (Y/N)	Correlation	1.000	.067	.009	.224*	.120	.042	.134	-.320**
	Loyalty	Correlation	.067	1.000	.052	.252**	.322**	.028	.225*	.086
	Fairness	Correlation	.009	.052	1.000	.192*	.278**	.235*	-.052	.081
	Authority	Correlation	.224*	.252**	.192*	1.000	.494**	.047	-.026	-.229*
	Purity	Correlation	.120	.322**	.278**	.494**	1.000	.187	.158	-.233*
	Harm	Correlation	.042	.028	.235*	.047	.187	1.000	-.188	-.056
	Aggression	Correlation	.134	.225*	-.052	-.026	.158	-.188	1.000	.043
	Severity Level	Correlation	-.320**	.086	.081	-.229*	-.233*	-.056	.043	1.000
Severity Level	Reported (Y/N)	Correlation	1.000	.100	.037	.164	.049	.026	.156	
	Loyalty	Correlation	.100	1.000	.045	.280**	.353**	.033	.222*	
	Fairness	Correlation	.037	.045	1.000	.217*	.306**	.241*	-.056	
	Authority	Correlation	.164	.280**	.217*	1.000	.465**	.035	-.016	
	Purity	Correlation	.049	.353**	.306**	.465**	1.000	.179	.173	
	Harm	Correlation	.026	.033	.241*	.035	.179	1.000	-.186	
	Aggression	Correlation	.156	.222*	-.056	-.016	.173	-.186	1.000	

*. Correlation is significant at 0.05 level

**. Correlation is significant at 0.01 level

a. Cells contain zero-order (Pearson) correlations.

Discussion

Blowing the whistle is a complicated decision for a witness; we investigated if the decision to report is affected by the severity of punishment that the transgressor will face. In this

cheating paradigm, we found that severity of punishment for cheating was significantly associated with whistleblowing, such that students reported cheaters more under the threat of a low punishment than a severe punishment. These results suggest that not only do witnesses consider consequences for the transgressor in their whistleblowing decisions, but they also demonstrate protective behavior in not reporting them when the punishment is severe.

No prior whistleblowing studies, to our knowledge, experimentally manipulate punishment severity for the transgressor; ours is the first to introduce witnesses' consideration of consequences beyond a fear for themselves (e.g., retaliation). These results do align with current literature on retaliation; similar to our finding, witnesses are less likely to blow the whistle under a more severe threat of retaliation (Khan et al., 2022). When consequences were severe, our participants also hesitated to report even though the punishment was directed towards the transgressor and not themselves. These similar trends suggest alternative explanations for past findings; measurements of punishment fears may capture whistleblowers' fears for the transgressor along with their fear for themselves. In Ogungbamila et al. (2024), participants with higher scores on the Punishment Anxiety Scale had lesser intentions to blow the whistle, interpreted as fear of retaliation. However, in line with their discussion of internal versus external dimensions of punishment anxiety, those with higher punishment anxiety could view transgressors' punishments similarly as their own, and thus are less likely to blow the whistle on others than those who have lower punishment anxiety. The introduction of our factor, consequences for the transgressor, may require more specific measures to understand which aspect of punishment is motivating for a whistleblower. Measures of empathy may allow us to tease apart considerations for the self and for others.

Our finding that a weak punishment for transgressors resulted in more whistleblowing than a severe punishment seems to contradict Krügel and Uhl (2023), who found that weak and inconsistent punishments resulted in almost no whistleblowing. However, our range of punishments likely sit on the higher end of their severity scale. Our "weak" punishment was still consistent, material, and supported by an honor-code environment, which implies reliable enforcement. Krüger and Uhl (2023) suggest that proper punishments can help witnesses feel as though their reports will be followed by meaningful action to resolve the violation. While the whistleblower may hope for a just punishment and may not report if the punishment is too weak or unreliable, our findings introduce the possibility that an overly severe punishment will dissuade peers from reporting transgressions.

Systematic reviews have identified many factors that affect whistleblowing decisions (e.g., closeness to the transgressor, trust in the organization); however, it is unclear how consciously individuals consider these factors in their decision-making process. Our manipulation check helps to address this for punishment severity; overall, participants accurately assessed their punishment condition as low, medium, or severe on a scale of 1-10 without any comparison conditions or descriptors provided by the proctor. Importantly, these harshness ratings were made after witnessing the cheating instance; participants provided ratings informed by the specific transgression in the study. This measure supports the conclusion that potential whistleblowers consider and assess punishment severity as appropriate or inappropriate for the specific transgression they witness. This also supports our divergence from Krüger and Uhl (2023), as punishments are not necessarily perceived as appropriate or inappropriate based on their severity alone; witnesses may consider the associated violation in their judgement of the punishment and ultimately in their decision to blow the whistle.

Exploratory analyses revealed no significant relationship between the time points at which students reported their peers and the punishment severity. One explanation is that punishment severity affects only the decision to report or stay silent and has no effect on other aspects of whistleblowing like when, to whom, and how to report. However, our paradigm cannot identify the significance, or non-significance, of punishments in determining when a person reports.

We found no significant correlations between whistleblowing and concerns of fairness, loyalty, authority, purity, harm/care, or aggression scores when controlling for punishment severity level. Without this control, we did find a relationship between authority concerns and whistleblowing, such that greater endorsements of authority in moral decisions were associated with choosing to report a cheating peer. Whistleblowing requires a witness to report a wrongdoing to an authority, leading to a natural conclusion that greater respect for authority can be associated with whistleblowing.

Past studies have found a positive relationship between the emotion of anger and whistleblowing (Gundlach et al., 2008; Jones et al., 2014), but no prior work references aggression. Considering the self-reported motivation of vindictiveness from Rennie and Crosby (2002)'s interviews, aggression could potentially affect desire to seek punishment. Especially when punishments are severe, individual differences in aggression may drive the decision to seek out harsh punishment.

Prior research suggests that whistleblowing is associated with concerns of fairness and loyalty, (Waytz et al., 2013; Dungan et al., 2019); however, we did not find significant results for either concern. One possible explanation is that these moral concerns may be in conflict not just with each other but with themselves. What aspect of the situation appeals to a witness's fairness

concerns, the transgression or the punishment? Participants may choose not to report because they think the violation is unfair (e.g., free riding) or because they view the punishment for cheating as unduly harsh. Loyalty also has two applications in whistleblowing cases; to whom is loyalty stronger, the authority or the transgressor? The negative relationship between closeness with a transgressor and whistleblowing is well-established (Alleyne et al., 2013; Olesen et al., 2019), and loyalty seems to drive protective behaviors for close targets (Waytz et al., 2013; Weidman et al., 2020). However, trust in and commitment to one's organization may motivate whistleblowing (Taylor, 2018; Latan et al., 2018). Loyalty to one's company may drive different behavior than loyalty to one's peer groups, especially when they are pit against each other.

Overall, it is possible that current fairness, loyalty, authority, purity, harm, and aggression measures may not be sensitive enough to determine the driving motivation for reporting behavior, especially in circumstances that introduce conflict within those moral concerns. Additionally, larger sample sizes within conditions and less variation between conditions may help to reveal any present trends in these scores. More work, including experimental manipulations of moral concerns, is needed to understand the impact of these concerns on whistleblowing behavior.

Implications

Our results suggest that introducing more severe punishments to deter violations may inadvertently reduce instances of whistleblowing. For any organization that depends on peer whistleblowing to identify and address violations, punishment severity level may decrease the number of transgressors reported. Furthermore, an opposing effect may be possible; if increased severity of punishments decreases whistleblowing, violators may engage in more transgressions as they become aware of their immunity among potential whistleblowers. Recent work also

suggests that frequent, small enforcements may reduce violations more than infrequent, severe enforcements (Teodorescu et al., 2021). In tandem with violation frequency studies, our work suggests that consistent but not overly harsh punishments have the potential to both decrease transgressions and increase whistleblowing for the remaining violations.

Measuring efficacy in reducing transgressions is complicated; due to the deceptive nature of most transgressions, it is difficult to discover if a violation occurred unless it is reported or causes obvious, immediate consequences. If an organization considers number of reports as a proxy measure of total transgressions, factors that only reduce a witness's willingness to blow the whistle may wrongly suggest a reduction in actual violations. We suggest that a measure of punishment effectiveness should come from its impact on the middle man, the reporter.

Our work has special implications for the academic context, especially regarding honor code systems. The cheating paradigm took place in under ten minutes but revealed significantly different results depending only on brief instructions given to students. This environment closely matches classroom conditions for tests and quizzes. Our findings can aid educators who have the freedom to tailor assessment instructions and administrators who can alter overall policies; overly severe punishments may reduce peer reporting of cheaters. In large classes where teachers may depend on peer-reports to curtail cheating, punishment severity may be especially attention-worthy. Additionally, our participants came from a university with a strong emphasis on the honor code, which explicitly states that students are expected to report cheating, and they simulated an honor code commitment by signing their names on our instruction sheets. Even in an environment with strong expectations of conduct, we found differences in rule-abiding behavior (i.e., reporting peers) based on condition-specific punishments. This has implications for administrative systems that depend on general codes-of-conduct; the existence of a conduct

code with an expectation of whistleblowing does not always result in consistent reporting behavior. Aspects of the situation, including punishment and transgression severity, influence witnesses' decisions to report cheating and abide by their honor code.

Where consequences can be made clear, such as in professional, academic, and legal systems, authorities may consider selecting severity of their punishments carefully to avoid reductive effects on whistleblowing.

Limitations and Future Research

While the experimental nature of our study allowed for deeper analysis of the research question, it also had practical limitations. The varying number of participants in each group, ranging from 2-4, is not ideal, and future work could benefit from ensuring an equal number of test-takers in each session. The absence of an observing proctor during the quiz and the nature of the cheating (taking a photo with a cell phone) may not mirror real-world instances of cheating in the classroom.

Our generalizations are limited by the academic focus of the study; participants faced unique motivational factors. Our participants engaged in the study as course credit, implying that their college's honor code applied. This was emphasized with their signatures on the study instructions. More than half of the students (56%) reported the cheating behavior, suggesting that this paradigm offers sufficient opportunity to notice, react to, and report cheating to an authority figure, but our reporting numbers may be inflated from real behavior due to the demands of the study. Whistleblowing was easy; participants were given multiple indirect and direct opportunities that were anonymized from other participants since they were in another room. These conditions of clear instruction, scaffolding, and anonymity may not generalize to other

environments where expectations and pathways for reporting are non-existent, unclear, or un-enforced.

Relationships between real-world whistleblowers and transgressors vary; our classroom paradigm examined only an equal, peer relationship between the test-takers. Manipulating the whistleblower-transgressor relationship through closeness or power dynamics could reveal important moderation effects. Weidman et al. (2020)'s participants perceived that greater harm would come to transgressors close to the participant than to distant transgressors if the participant reported their violation to a police officer. Perceived harm increased with closeness even when no punishment information was provided; a punishment manipulation may expand on this relationship between closeness, punishment severity, perceived harm, and whistleblowing. Additionally, developmental work could build upon Misch et al. (2018)'s finding of in-group bias based on transgression severity; one of their explanations was that the children believed more severe transgressions would provoke greater punishments and thus protected their in-group from greater harm. As they did not manipulate punishment severity, future work should address the questions: Do children report their peers less when the punishment is severe? Is an in-group bias present as well?

While our measures of moral concerns showed limited findings, this line of research may benefit from exploring moral emotions. In the moral emotions literature, measures of anticipated shame and guilt have been linked to prosocial and antisocial behavior, in that guilt seems to drive prosociality while shame inhibits antisocial actions (Olthof, 2011). When considering whether to report a peer, the witness's judgement of how they may feel in the future could inform their behavior in the moment. The prospect of subjecting a peer to more severe punishment may elicit greater feelings of guilt, which could motivate a witness to act prosocially towards the

transgressor, protecting them from punishment. Research on these moral emotions, in tandem with moral concerns, may explain the decision-making process for whistleblowing on an affective level.

If witnesses are more likely to report transgressions when the violator faces lesser punishment, other aspects of the punishment system may also affect their willingness to report. While punishment severity level was not associated with reporting cheating at an earlier opportunity, this measure of students' eagerness to report violations deserves further exploration. The time points offered to report cheating in this study were relatively close in time; future research could examine extended periods of time to report violations to represent real-world scenarios more accurately. Additionally, the prompts and method through which students could report could be standardized across the opportunities to control for other factors like directness, anonymity, and demand characteristics.

Future work should also explore motivations driving the difference between the low and high severity conditions. This study's high severity condition escalated the reported cheating to a higher authority, introducing social consequences absent in the lower severity categories. Social rewards are often reported as more motivating than monetary rewards (Wang et al., 2017), and some evidence shows social rewards activate different neural pathways than nonsocial rewards (Rademacher et al., 2010). Social punishments (i.e. shame) may also be viewed as more motivating than non-social punishments; because our conditions combined them, further work is needed to untangle the two.

Our social system relies on cooperation, and violations of social norms can result in disarray. To address violations, courses of action include directly punishing transgressors as well as reporting their behavior to an authority in charge of punishment. However, our findings

suggest that individuals may be unwilling to enforce a norm by blowing the whistle if they perceive the authority's punishment for the transgressor as too severe. When acting as an "accomplice" in punishing another person, whistleblowers are sensitive to the consequences faced by the transgressor; punishments that are overly severe for the transgression may induce moral conflict for potential whistleblowers. Empathy and concern for others are often assumed to enhance cooperation, but our findings show that punishment conditions that appeal to one's concern for a transgressor may instead decrease norm enforcement. Cooperation suffers when partners hesitate to enforce norms; one transgression could snowball into multiple norm violations when norm-enforcement could subject transgressors to overly harsh consequences.

References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Alleyne, P., Weekes-Marshall, D., & Arthur, R. (2013). Exploring Factors Influencing Whistle-blowing Intentions among Accountants in Barbados. *Journal of Eastern Caribbean Studies*, *38*(1/2), 35–62.

Arnold, R., Martin, B. N., & Bigby, L. (2007). Is There a Relationship Between Honor Codes and Academic Dishonesty? *Journal of College and Character*, *8*(2). https://doi.org/10.2202/1940-1639.1164

Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*, *111*(45), 15924–15927. https://doi.org/10.1073/pnas.1413170111

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*(4), 594–615. https://doi.org/10.1037/a0023489

Basu, K., Basu, K., & Cordella, T. (2016). Asymmetric Punishment as an Instrument of Corruption Control. *Journal of Public Economic Theory*, *18*(6), 831–856. https://doi.org/10.1111/jpet.12212

Bernardi, R. A., Landry, A. C., Landry, E. E., Buonafede, M. R., & Berardi, M. E. (2016). What actions can be taken to increase whistle-blowing in the classroom? *Accounting Education*, *25*(1), 88–106. https://doi.org/10.1080/09639284.2015.1107496

Bocchiaro, P., Zimbardo, P. G., & Van Lange, P. A. M. (2012). To defy or not to defy: An experimental study of the dynamics of disobedience and whistle-blowing. *Social Influence*, *7*(1), 35–50. https://doi.org/10.1080/15534510.2011.648421

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science*, *328*(5978), 617–620. https://doi.org/10.1126/science.1183665

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1533), 3281–3288. https://doi.org/10.1098/rstb.2009.0134

Bricci, L., Fragata, A., & Antunes, J. (2015). The Effects of Trust, Commitment and Satisfaction on Customer Loyalty in the Distribution Sector. *Journal of Business Economics and Management*, *4*. https://doi.org/10.7763/JOEBM.2016.V4.386

Burrus, R. T., Jones, A. T., Sackley, B., & Walker, M. (2013). It's The Students, Stupid: How Perceptions of Student Reporting Impact Cheating. *The American Economist*, *58*(1), 51–59. https://doi.org/10.1177/056943451305800106

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

Cassematis, P. G., & Wortley, R. (2013). Prediction of Whistleblowing or Non-reporting Observation: The Role of Personal and Situational Factors. *Journal of Business Ethics*, *117*(3), 615–634. https://doi.org/10.1007/s10551-012-1548-3

Chen, S., Xu, K., & Yao, X. (2022). Empirical study of employee loyalty and satisfaction in the mining industry using structural equation modeling. *Scientific Reports*, *12*, 1158. https://doi.org/10.1038/s41598-022-05182-2

Dahl, A., & Waltzer, T. (2023). A Canary Alive: What Cheating Reveals about Morality and Its Development. *Human Development*, *68*(1), 6–25. https://doi.org/10.1159/000534638

Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic Dishonesty: Prevalence, Determinants, Techniques, and Punishments. *Teaching of Psychology*, *19*(1), 16–20. https://doi.org/10.1207/s15328023top1901_3

DeZoort, F. T. (2023). The Experimental Effects of a Student Whistleblower Policy on Reporting Classmate Misconduct. *Issues in Accounting Education*, *38*(4), 61–79. https://doi.org/10.2308/ISSUES-2021-062

Dungan, J. A., Young, L., & Waytz, A. (2019). The power of moral concerns in predicting whistleblowing decisions. *Journal of Experimental Social Psychology*, *85*, 103848. https://doi.org/10.1016/j.jesp.2019.103848

Dungan, J., Waytz, A., & Young, L. (2015). The psychology of whistleblowing. *Current Opinion in Psychology*, *6*, 129–133. https://doi.org/10.1016/j.copsyc.2015.07.005

Dunn, J. F., & Munn, P. (1985). Becoming a family member: Family conflict and the development of social understanding in the second year. *Child Development*, *56*(2), 480–492. https://doi.org/10.2307/1129735

Eriksson, K., Andersson, P. A., & Strimling, P. (2016). Moderators of the disapproval of peer punishment. *Group Processes & Intergroup Relations*, *19*(2), 152–168. https://doi.org/10.1177/1368430215583519

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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

Friman, P., Woods, D., Freeman, K., Gilman, R., Short, M., Davis, A., & Handwerk, M. (2004). Relationships Between Tattling, Likeability, and Social Classification: A Preliminary Investigation of Adolescents in Residential Care. *Behavior Modification*, *28*, 331–348. https://doi.org/10.1177/0145445503258985

Gächter, S., Renner, E., & Sefton, M. (2008). The Long-Run Benefits of Punishment. *Science*, *322*(5907), 1510–1510. https://doi.org/10.1126/science.1164744

Gino, F., Moore, D. A., & Bazerman, M. H. (2010). See no evil: When we overlook other people's unethical behavior. In *Social decision making: Social dilemmas, social values, and ethical judgments* (pp. 241–263). Routledge/Taylor & Francis Group.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*, 366–385. https://doi.org/10.1037/a0021847

Greenberger, D. B., Miceli, M. P., & Cohen, D. J. (1987). Oppositionists and Group Norms: The Reciprocal Influence of Whistle-Blowers and Co-Workers. *Journal of Business Ethics*, *6*(7), 527–542. https://doi.org/10.1007/bf00383744

Gundlach, M. J., Martinko, M. J., & Douglas, S. C. (2008). A New Approach to Examining Whistle-Blowing: The Influence of Cognitions and Anger. *SAM Advanced Management Journal (07497075)*, *73*(4), 40–50.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science (New York, N.Y.)*, *312*(5781), 1767–1770. https://doi.org/10.1126/science.1127333

Huang, I.-C., Du, P.-L., Lin, L.-S., Liu, T.-Y., Lin, T.-F., & Huang, W.-C. (2021). The Effect of Perceived Value, Trust, and Commitment on Patient Loyalty in Taiwan. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, *58*, 00469580211007217. https://doi.org/10.1177/00469580211007217

J. Scrimpshire, A., Stone, T. H., Kisamore, J. L., & Jawahar, I. M. (2017). Do Birds of a Feather Cheat Together? How Personality and Relationships Affect Student Cheating. *Journal of Academic Ethics*, *15*(1), 1–22. https://doi.org/10.1007/s10805-016-9267-5

Jeffreys, H. (1961). *The Theory of Probability* (Third). Oxford University Press.

Jendrek, M. P. (1992). Students' reactions to academic dishonesty. *Journal of College Student Development*, *33*(3), 260–273.

Jenkel, I., & Haen, J. J. (2012). Influences on Students' Decisions to Report Cheating: A Laboratory Experiment. *Journal of Academic Ethics*, *10*(2), 123–136. https://doi.org/10.1007/s10805-012-9154-7

Jones, J. C., Spraakman, G., & Sánchez-Rodríguez, C. (2014). What's in it for Me? An Examination of Accounting Students' Likelihood to Report Faculty Misconduct. *Journal of Business Ethics*, *123*(4), 645–667. https://doi.org/10.1007/s10551-013-2015-5

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences*, *111*(35), 12710–12715. https://doi.org/10.1073/pnas.1402280111

Khan, J., Saeed, I., Zada, M., Ali, A., Contreras-Barraza, N., Salazar-Sepúlveda, G., & Vega-Muñoz, A. (2022). Examining Whistleblowing Intention: The Influence of Rationalization on Wrongdoing and Threat of Retaliation. *International Journal of Environmental Research and Public Health*, *19*(3), 1752. https://doi.org/10.3390/ijerph19031752

Khodaie, E., Moghadamzadeh, A., & Salehi, K. (2011). Factors Affecting the Probability of Academic Cheating School Students in Tehran. *Procedia - Social and Behavioral Sciences*, *29*, 1587–1595. https://doi.org/10.1016/j.sbspro.2011.11.401

Kisamore, J. L., Stone, T. H., & Jawahar, I. M. (2007). Academic Integrity: The Relationship between Individual and Situational Factors on Misconduct Contemplations. *Journal of Business Ethics*, *75*(4), 381–394. https://doi.org/10.1007/s10551-006-9260-9

Kohls, G., Perino, M. T., Taylor, J. M., Madva, E. N., Cayless, S. J., Troiani, V., Price, E., Faja, S., Herrington, J. D., & Schultz, R. T. (2013). The nucleus accumbens is involved in both the pursuit of social reward and the avoidance of social punishment. *Neuropsychologia*, *51*(11), 2062–2069. https://doi.org/10.1016/j.neuropsychologia.2013.07.020

Krügel, S., & Uhl, M. (2023). Internal whistleblowing systems without proper sanctions may backfire. *Journal of Business Economics*, *93*(8), 1355–1383. https://doi.org/10.1007/s11573-023-01144-w

LaSalle, R. E. (2009). The Perception of Detection, Severity of Punishment and the Probability of Cheating: Journal of Forensic Studies in Accounting & Business. *Journal of Forensic Studies in Accounting & Business*, *1*(2), 93–112.

Latan, H., Chiappetta Jabbour, C. J., Ali, M., Lopes de Sousa Jabbour, A. B., & Vo-Thanh, T. (2023). What Makes You a Whistleblower? A Multi-Country Field Study on the Determinants of the Intention to Report Wrongdoing. *Journal of Business Ethics*, *183*(3), 885–905. https://doi.org/10.1007/s10551-022-05089-y

Latan, H., Ringle, C. M., & Jabbour, C. J. C. (2018). Whistleblowing Intentions Among Public Accountants in Indonesia: Testing for the Moderation Effects. *Journal of Business Ethics*, *152*(2), 573–588. https://doi.org/10.1007/s10551-016-3318-0

Lim, V. K. G., & See, S. K. B. (2001). Attitudes Toward, and Intentions to Report, Academic Cheating Among Students in Singapore. *Ethics & Behavior*, *11*(3), 261–274. https://doi.org/10.1207/S15327019EB1103_5

Loke, I. C., Heyman, G. D., Forgie, J., McCarthy, A., & Lee, K. (2011). Children's Moral Evaluations of Reporting the Transgressions of Peers: Age Differences in Evaluations of Tattling. *Developmental Psychology*, *47*(6), 1757–1762. https://doi.org/10.1037/a0025357

Malesky, A., Grist, C., Poovey, K., & Dennis, N. (2021). The Effects of Peer Influence, Honor Codes, and Personality Traits on Cheating Behavior in a University Setting. *ETHICS & BEHAVIOR*. https://doi.org/10.1080/10508422.2020.1869006

Marshall, J., & Mcauliffe, K. (2022). Children as assessors and agents of third-party punishment. *Nature Reviews Psychology*, *1*, 1–11. https://doi.org/10.1038/s44159-022-00046-y

McCabe, D. L., Butterfield, K. D., & Treviño, L. K. (2012). *Cheating in College: Why Students Do It and What Educators Can Do about It*. Johns Hopkins University Press. http://ebookcentral.proquest.com/lib/uva/detail.action?docID=3318632

McCabe, D. L., & Trevino, L. K. (1993). Academic Dishonesty: Honor Codes and Other Contextual Influences. *The Journal of Higher Education*, *64*(5), 522–538. JSTOR. https://doi.org/10.2307/2959991

McCabe, D. L., & Trevino, L. K. (2016). Academic Dishonesty. *The Journal of Higher Education*. https://www.tandfonline.com/doi/abs/10.1080/00221546.1993.11778446

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2002). Honor Codes and Other Contextual Influences on Academic Integrity: A Replication and Extension to Modified Honor Code Settings. *Research in Higher Education*, *43*(3), 357–378. JSTOR.

Melendez, B. (1985). *Honor Code Study*. Harvard University.

Miceli, M. P., & Near, J. P. (1985). Characteristics of Organizational Climate and Perceived Wrongdoing Associated with Whistle-Blowing Decisions. *Personnel Psychology*, *38*(3), 525–544. https://doi.org/10.1111/j.1744-6570.1985.tb00558.x

Miller, A., Shoptaugh, C., & Wooldridge, J. (2011). Reasons Not to Cheat, Academic-Integrity Responsibility, and Frequency of Cheating. *The Journal of Experimental Education*, *79*(2), 169–184. https://doi.org/10.1080/00220970903567830

Misch, A., Over, H., & Carpenter, M. (2018). The whistleblower's dilemma in young children: When loyalty trumps other moral concerns. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00250

Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications*, *11*(1), 3432. https://doi.org/10.1038/s41467-020-17286-2

Moore, L., & McAuliffe, E. (2010). Is inadequate response to whistleblowing perpetuating a culture of silence in hospitals? *Clinical Governance: An International Journal*, *15*(3), 166–178. https://doi.org/10.1108/14777271011063805

Near, J. P., & Miceli, M. P. (1985). Organizational dissidence: The case of whistle-blowing. *Journal of Business Ethics*, *4*(1), 1–16. https://doi.org/10.1007/BF00382668

Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, *29*, 242–248. https://doi.org/10.1016/j.evolhumbehav.2008.01.001

Nicholls, A. R., Fairs, L. R. W., Toner, J., Jones, L., Mantis, C., Barkoukis, V., Perry, J. L., Micle, A. V., Theodorou, N. C., Shakhverdieva, S., Stoicescu, M., Vesic, M. V., Dikic, N., Andjelkovic, M., Grimau, E. G., Amigo, J. A., & Schomöller, A. (2021). Snitches Get

Stitches and End Up in Ditches: A Systematic Review of the Factors Associated With Whistleblowing Intentions. *Frontiers in Psychology*, *12*, 631538. https://doi.org/10.3389/fpsyg.2021.631538

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*(1), 91–112. https://doi.org/10.1016/j.jpubeco.2007.04.008

Nitsch, D., Baetz, M., & Hughes, J. C. (2005). Why Code of Conduct Violations go Unreported: A Conceptual Framework to Guide Intervention and Future Research. *Journal of Business Ethics*, *57*(4), 327–341. https://doi.org/10.1007/s10551-004-8203-6

Nora, W. L. Y., & Zhang, K. C. (2010). Motives of cheating among secondary students: The role of self-efficacy and peer influence. *Asia Pacific Education Review*, *11*(4), 573–584. https://doi.org/10.1007/s12564-010-9104-2

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science (New York, N.y.)*, *314*(5805), 1560–1563. https://doi.org/10.1126/science.1133755

Ogungbamila, B., Osamika, B. E., & Job, E. D. (2024). Whistleblowing and Corrupt Tendencies among Selected Employees in Three Public Organizations: Roles of Corruption Tolerance, Punishment Anxiety, and Neutralization. *Journal of Management Studies and Development*, *3*(02), Article 02. https://doi.org/10.56741/jmsd.v3i02.535

Oh, L.-B., & Teo, H. (2010). To Blow or not to Blow: An Experimental Study on the Intention to Whistleblow on Software Piracy. *J. Org. Computing and E. Commerce*, *20*, 347–369. https://doi.org/10.1080/10919392.2010.516639

Olesen, A. P., Amin, L., Mahadi, Z., & Ibrahim, M. (2019). Whistle blowing and research integrity: Potential remedy for research misconduct in Malaysian institutions of higher

education. *Accountability in Research*, *26*(1), 17–32. https://doi.org/10.1080/08989621.2018.1554444

Przepiorka, W., Szekely, A., Andrighetto, G., Diekmann, A., & Tummolini, L. (2022). How Norms Emerge from Conventions (and Change). *Socius*, *8*, 23780231221124556. https://doi.org/10.1177/23780231221124556

Pupovac, V., Popović, S., & Blažina, V. (2019). What Prevents Students from Reporting Academic Misconduct? A Survey of Croatian Students. *Journal of Academic Ethics*, *17*(4), 389–400. https://doi.org/10.1007/s10805-019-09341-5

Rademacher, L., Krach, S., Kohls, G., Irmak, A., Gründer, G., & Spreckelmeyer, K. N. (2010). Dissociation of neural networks for anticipation and consumption of monetary and social rewards. *NeuroImage*, *49*(4), 3276–3285. https://doi.org/10.1016/j.neuroimage.2009.10.089

Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. *Trends in Ecology & Evolution*, *27*(5), 288–295. https://doi.org/10.1016/j.tree.2011.12.004

Rennie, S. C., & Crosby, J. R. (2002). Students' perceptions of whistle blowing: Implications for self-regulation. A questionnaire and focus group survey. *Medical Education*, *36*(2), 173–179. https://doi.org/10.1046/j.1365-2923.2002.01137.x

Richardson, B. K., Wang, Z., & Hall, C. A. (2012). Blowing the Whistle Against Greek Hazing: The Theory of Reasoned Action as a Framework for Reporting Intentions. *Communication Studies*, *63*(2), 172–193. https://doi.org/10.1080/10510974.2011.624396

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior* (p. 394). Harvard University Press.

Stevenson, S. M., Flannigan, K., Willey, A., & Kaur, T. (2023). Exploring Factors That Contribute to Nursing Students' Willingness to Report Peer Academic Integrity Violations.

Nursing Education Perspectives, *44*(3), 140–146. https://doi.org/10.1097/01.NEP.0000000000001090

Stone, T. H., Jawahar, I. M., & Kisamore, J. L. (2009). Using the theory of planned behavior and cheating justifications to predict academic misconduct. *Career Development International*, *14*(3), 221–241. https://doi.org/10.1108/13620430910966415

Stone, T., Kisamore, J., Kluemper, D., & Jawahar, I. (2012). *Whistle-Blowing in the Classroom? 12*, 11–26.

Taylor, J. (2018). Internal Whistle-Blowing in the Public Service: A Matter of Trust. *Public Administration Review*, *78*(5), 717–726. https://doi.org/10.1111/puar.12946

Teodorescu, K., Plonsky, O., Ayal, S., & Barkan, R. (2021). Frequency of enforcement is more important than the severity of punishment in reducing violation behaviors. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(42), e2108507118. https://doi.org/10.1073/pnas.2108507118

Trevino, L. K., & Victor, B. (1992). Peer reporting of unethical behavior: A social context perspective. *Academy of Management Journal*, *35*(1), 38–64. https://doi.org/10.2307/256472

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology, 29*, 124–130. https://doi.org/10.1348/026151010X532888

Vaughn, A. J., Firmin, M. W., & Chi-En Hwang. (2009). Efficacy of Request Presentation on Compliance. *Social Behavior & Personality: An International Journal*, *37*(4), 441–449. https://doi.org/10.2224/sbp.2009.37.4.441

Waltzer, T., Cox, R. L., Moser, C. F., & Heyman, G. D. (2024). Don't be a rat: An investigation of the taboo against reporting other students for cheating. *Journal of Experimental Child Psychology*, *242*, 105894. https://doi.org/10.1016/j.jecp.2024.105894

Waltzer, T., & Dahl, A. (2021). Students' perceptions and evaluations of plagiarism: Effects of text and context. *Journal of Moral Education*, *50*(4), 436–451. https://doi.org/10.1080/03057240.2020.1787961

Waltzer, T., Samuelson, A., & Dahl, A. (2022). Students' Reasoning About Whether to Report When Others Cheat: Conflict, Confusion, and Consequences. *Journal of Academic Ethics*, *20*(2), 265–287. https://doi.org/10.1007/s10805-021-09414-4

Wang, D., Liu, T., & Shi, J. (2017). Development of Monetary and Social Reward Processes. *Scientific Reports*, *7*(1), 11128. https://doi.org/10.1038/s41598-017-11558-6

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027–1033. https://doi.org/10.1016/j.jesp.2013.07.002

Weidman, A. C., Sowden, W. J., Berg, M. K., & Kross, E. (2020). Punish or Protect? How Close Relationships Shape Responses to Moral Violations. *Personality and Social Psychology Bulletin*, *46*(5), 693–708. https://doi.org/10.1177/0146167219873485

Yachison, S., Okoshken, J., & Talwar, V. (2017). Students' Reactions to a Peer's Cheating Behavior. *Journal of Educational Psychology*, *110*. https://doi.org/10.1037/edu0000227

Yucel, M. & Vaish, A. (2018). Young Children Tattle to Enforce Moral Norms. *Social Development, 27*(4). https://doi.org/10.1111/sode.12290

Zakharin, M., & Bates, T. C. (2023). Moral Foundations Theory: Validation and replication of the MFQ-2. *Personality and Individual Differences*, *214*, 112339. https://doi.org/10.1016/j.paid.2023.112339

Zhu, W., Zeng, R., Li, X., Zhu, Y., & Zhang, Z. (2017). Managerial drivers of Chinese labour loyalty in international construction projects. *Journal of Civil Engineering and Management*, *23*(8), 1109–1122. https://doi.org/10.3846/13923730.2017.1381644