RUNNING HEAD: Whistleblowing

The Whistleblower's Dilemma and the Fairness-Loyalty Tradeoff

Adam Waytz[1], James Dungan[2], Liane Young[2]

1. Northwestern University
2. Boston College

[**UNCORRECTED PROOF, in press at *JESP***]

Address correspondence to:

Adam Waytz
Northwestern University
2001 Sheridan Rd
Evanston, IL 60208
United States
Email: a-waytz@kellogg.northwestern.edu

Word count: 5,000

Abstract

Whistleblowing – reporting another person's unethical behavior to a third party – often constitutes a conflict between competing moral concerns. Whistleblowing promotes justice and fairness but can also appear disloyal. Five studies demonstrate that a fairness-loyalty tradeoff predicts people's willingness to blow the whistle. Study 1 demonstrates that individual differences in valuing fairness over loyalty predict willingness to report unethical behavior. Studies 2a and 2b demonstrate that experimentally manipulating endorsement of fairness versus loyalty increases willingness to report unethical behavior. Study 3 demonstrates that people recall their decisions to report unethical behavior as driven by valuation of fairness, whereas people recall decisions *not* to report unethical behavior as driven by valuation of loyalty. Study 4 demonstrates that experimentally manipulating the endorsement of fairness versus loyalty increases whistleblowing in an online marketplace. These findings reveal the psychological determinants of whistleblowing and shed light on factors that encourage or discourage this practice.

Keywords: ethics, morality, loyalty, fairness, whistleblowing

The decision to report another person's unethical behavior to a third party — to engage in whistleblowing — presents a dilemma. Although some whistleblowers receive heroic acclaim (Johnson, 2003), other whistleblowers face revenge from their community (Dyck, Adair, & Zingales, 2010). For example, a recent editorial reprimanding National Security Agency whistleblower, Edward Snowden, stated that Snowden "faced a moral dilemma" and ultimately "betrayed his employers," contributing to "the fraying of social fabric" (Brooks, 2013).

What then drives whistleblowing decisions? Previous research has investigated structural and organizational factors that influence whistleblowing, including the professional status of whistleblowers, organizational support for whistleblowing (Near & Miceli, 1985; Dozier & Miceli, 1985; Vadera, Vadera, & Caza, 2009), and the type of behavior that people deem unethical and therefore reportable (Gino & Bazerman, 2009). Existing research has not, however, investigated the psychological determinants of whistleblowing. Here, we investigate the cognitive processes underlying people's decision to blow the whistle or not. Specifically, we propose that differences in people's valuation of moral norms, fairness versus loyalty, contribute to whistleblowing decisions.

Fairness and loyalty alike represent basic moral values, as reflected in developmental and evolutionary approaches to moral cognition. Infants endorse distributive and retributive justice – before age two, children expect resources to be divided fairly among individuals according to each individual's effortful contribution in a group task (Sloane, Baillargeon, & Premack, 2012; Kanngiesser & Warneken, 2012). Furthermore, 8-month-olds prefer to reward helpful, prosocial behavior and punish selfish, antisocial behavior (Hamlin, Wynn, Bloom, & Mahajan, 2011). At the same time, young children's adherence to fairness and justice norms are powerfully modified by group membership – children share disproportionate resources with family and friends over

strangers (Olson & Spelke, 2008) and often choose to act loyally versus fairly, especially when expectations for friendship are made salient (Smetana, Killen, & Turiel, 1991). Research on third-party judgments shows that infants also prefer those who harm dissimilar others and help similar others (Hamlin, Mahajan, Liberman, & Wynn, 2013), and toddlers prefer those who behave loyally (i.e., who reciprocate) to those who behave fairly in certain competitive contexts (Shaw, DeScioli, & Olson, 2012). Finally, whereas toddlers consider tattling in some cases to be a just, prosocial act (Ingram & Bering, 2010), adolescents, who place a premium on group membership, respond far more negatively to peer tattlers (Friman, et al., 2004). Notably, precursors to both fairness (e.g., Brosnan, Schiff, & de Waal, 2005) and loyalty have been observed in our primate ancestors as well (Mahajan, et al., 2011), revealing the fundamental nature of both moral norms.

Although fairness and loyalty represent basic moral values (Haidt, 2007; Walker & Hennig, 2004), they do, at times, conflict. At their core, norms for fairness and justice demand that all persons and groups be treated equally. By contrast, loyalty norms dictate preferential treatment, a responsibility to favor one's own group over other groups. Studies have shown that fairness norms typically dominate behavior but may be overwritten in contexts that pit fairness against loyalty. For example, factors such as psychological closeness (Batson, Klein, Highberger, & Shaw, 1995), national culture (Miller & Bersoff, 1992), residential mobility (Lun, Oishi, & Tenney, 2012), perceived duty (Baron, Ritov, & Greene, 2011), and relationship type (Rai & Fiske, 2011) modulate people's preference for loyalty versus fairness (see also Shaw, DeScioli, & Olson, 2012). Because of the fundamental tension between these norms, the present research assesses the loyalty-fairness tradeoff rather than assessing each in isolation.

We propose that fairness and loyalty norms clash during whistleblowing decisions. Our definition of whistleblowing corresponds to organizational definitions of this behavior as well as with definitions of "tattling" from the social cognitive development perspective (Ingram & Bering, 2010) and "snitching" from a legal perspective (Natapoff, 2004). We take whistleblowing to involve two key components: (1) reporting unethical behavior (2) to a third party (e.g., an authority figure).

On one hand, whistleblowers may act in the service of fairness and justice when exposing corporate wrongdoing (Near & Miceli, 1985; Miceli & Near, 1992), neighborhood crime (Natapoff, 2004), or scientific fraud (Vogel, 2011; Yong, 2012). On the other hand, whistleblowing may constitute an act of disloyalty, depending also on the relationship between the offender and the whistleblower. Indeed, the vast majority of corporate whistleblowers face negative outcomes as a result of their actions: revenge, reassignment, firing, and personal distress (Dyck et al., 2010) and such "moral rebels" are often ostracized (e.g., Monin, Sawyer, & Marquez, 2008; Parks & Stone, 2010; Minson & Monin, 2012). Would-be whistleblowers are thus faced with the dilemma of choosing between competing demands. Whereas fairness norms typically require that people report and punish wrongdoing, loyalty norms—even in the abstract—indicate that reporting another person to a third party may constitute an act of betrayal, associated with potential repercussions as detailed above.

We propose that whistleblowing behavior constitutes a tradeoff between fairness and loyalty. A direct prediction of this proposal is that the endorsement of fairness versus loyalty tracks subsequent decisions to blow the whistle. Evidence from five studies supports this prediction. First, individual differences in the endorsement of fairness versus loyalty correspond to decisions to blow the whistle (Study 1). Second, experimentally manipulating concern for

fairness versus loyalty predicts willingness to blow the whistle (Studies 2a and 2b). Third, people describe real-life decisions to blow the whistle as motivated by concerns for fairness more than loyalty, whereas they describe decisions to *not* blow the whistle as motivated by concerns for loyalty more than fairness (Study 3). Finally, experimental inductions of fairness versus loyalty predict real-life whistleblowing in an online marketplace (Study 4).

Study 1: Individual Differences

Study 1 assessed individual differences in valuation of fairness versus loyalty and the relation to whistleblowing.[1]

Method

Eighty-three individuals (M_{age}=35.72, SD_{age}=13.93; 65% female) participated via Amazon.com's Mechanical Turk in exchange for a small payment; all subsequent studies used the same methodological approach. Participants completed three measures to assess their valuation of fairness versus loyalty. In this study and subsequent studies, we included only participants who completed all measures.

The first measure consisted of six-point Likert scale items from the Moral Foundations Questionnaire (MFQ; Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011), assessing valuation of fairness and loyalty. Two fairness items and two loyalty items assessed the relevance of various considerations for judgments of right and wrong (e.g., "Whether or not someone acted

---

[1] Following Simmons, Nelson, and Simonsohn (2012), we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures. Sample sizes were determined separately for each study based on prior similar studies in the literature.

unfairly," "Whether or not someone was denied his or her rights" and "Whether or not someone did something to betray his or her group," "Whether or not someone showed a lack of loyalty"); one fairness item and one loyalty item asked about agreement with moral statements ("Justice is the most important requirement for a society" and "People should be loyal to their family members, even when they have done something wrong"). Following Graham et al. (2011), we averaged the three loyalty and three fairness items separately and subtracted the loyalty score from the fairness score to produce a composite *values* score (these items were embedded amongst three other MFQ items irrelevant to the current hypothesis).

The second measure asked, "Objectively speaking, who do you think is the more morally good person?" with a forced-choice response option: "Someone who is fair and just, impartial and unprejudiced" (fairness; coded 1) or "Someone who is loyal and faithful, devoted and dependable" (loyalty; coded 0). This constituted participants' *judgment* score.

The third measure asked, "Who would you rather be friends with?" with a forced-choice response option: "Someone who is fair and just to others, who is impartial and unprejudiced regardless of how it affects their family and friends" (fairness; coded 1) or "Someone who is loyal and faithful to their family and friends, who is devoted and dependable regardless of how it affects outsiders" (loyalty; coded 0). This constituted participants' *friendship* score.

We standardized and averaged the three scores (α=.64) to compute a composite fairness-versus-loyalty score. Higher values indicate a preference for fairness over loyalty, whereas lower values indicate a preference for loyalty over fairness.

To measure attitudes toward whistleblowing, we asked participants about seven

violations ranging in severity[2]:

1. Stealing $1 from a restaurant's tip jar.
2. Embezzling $1000 from their work place.
3. Robbing a woman of her cell phone and wallet.
4. Cheating on their final exam in college.
5. Spraying rude graffiti on the side of a local store.
6. Using and selling marijuana to other adults.
7. Fatally stabbing a convenience store owner.

For each scenario, participants indicated (1=*Very unlikely*; 7=*Very likely*) how likely they would be to blow the whistle if the perpetrator were:

1. A total stranger you've never met.
2. An acquaintance you see occasionally
3. A close friend you've known for years
4. A family member you're very close to

Specifically, for each offense, participants read a version of the following, "Imagine that you witness someone stealing $1 from a restaurant's tip jar. How likely would you be to report the perpetrator of this incident if this perpetrator were… a stranger/acquaintance/close friend/family member?" each accompanied by the 7-point scale. We computed an overall whistleblowing score, by averaging all 28 responses (four levels of closeness for each of seven scenarios) (α=.93). We also computed separate whistleblowing scores for each target (i.e., stranger, acquaintance, friend, and family member) by averaging responses over all seven scenarios (all αs>.74). For exploratory purposes, we also computed separate whistleblowing scores for each offense by averaging over all four levels of closeness (all αs>.77).

Results and Discussion

Our primary hypothesis was that fairness-versus-loyalty valuation would be associated

---

[2] Unrelated pilot items for a separate study were randomly presented before or after the whistleblowing scenarios and did not affect results.

with positive whistleblowing decisions.[3] As predicted, participants' fairness-versus-loyalty score was correlated with whistleblowing overall, $r(81)=.28$, $p=.011$.

For exploratory purposes, we also explored relations between fairness-versus-loyalty and whistleblowing for specific targets (i.e., levels of relationship closeness) and specific offenses. Although willingness to blow the whistle decreased as relationship closeness increased (*stranger:* $M=5.32$, $SD=1.10$, *acquaintance:* $M=5.02$, $SD=1.11$, *friend:* $M=4.48$, $SD=1.25$, and *family member:* $M=4.27$, $SD=1.35$; paired-samples t-tests demonstrated all differences were significant, $ps<.005$), fairness-versus-loyalty scores correlated with whistleblowing at each level of closeness (stranger: $r(81)=.22$, $p=.051$, acquaintance: $r(81)=.28$, $p=.01$, friend: $r(81)=.26$, $p=.023$, and family member: $r(81)=.25$, $p=.022$). See Table 1 for correlations for specific offenses. Overall, individual differences in the endorsement of fairness over loyalty predicted willingness to blow the whistle on a range of perpetrators committing crimes of varying severity.

Study 2: Inducing Fairness and Loyalty

Study 1 tested whether individual differences in moral valuation could predict whistleblowing decisions. Study 2 tested whether experimentally manipulating subjective valuation of fairness/justice versus loyalty could produce the same pattern of differences. Study 2a provides an initial test, while Study 2b provides a replication of the effect.

---

[3] The forced choice nature of the judgment and friendship items precludes us from regressing whistleblowing separately on the dimensions of fairness and loyalty. However, separate values scores for fairness ($p=.98$) and loyalty ($p=.14$) did not predict whistleblowing overall.

Study 2a

Method

One hundred forty-two MTurk workers (M_{age}=33.19, SD_{age}=12.01; 51% female) participated as in Study 1. Participants were first randomly assigned to write three short essays on the importance of fairness/justice or the importance of loyalty, designed to induce participants to prioritize the target norm. For example, in the fairness/justice condition, participants were asked, "Please write a few sentences about why it is more important to be just than to be loyal" whereas in the loyalty condition, participants were asked, "Please write a few sentences about why it is more important to be loyal than to be just" (see supplementary materials for full instructions). Participants then completed the same measures of whistleblowing from Study 1; however, we specified the relevant authority to whom the offense would be reported (see supplementary materials).

Results and Discussion

Independent-samples *t*-tests revealed that overall willingness to blow the whistle (α=.94) was higher in the fairness condition (M=4.98, SD=1.10) than in the loyalty condition (M=4.47, SD=1.06), $t(140)=2.78$, p=.006, d=.47. This finding replicates the pattern of Study 1 using an experimental manipulation of fairness-versus-loyalty.

We then conducted a 2 (condition: fairness/justice vs. loyalty) x 4 (target: stranger, acquaintance, friend, family member) mixed-effects ANOVA on participants' willingness to blow the whistle averaged across the seven scenarios (all αs>.69 for composite scores for each target). We observed a main effect of target (multivariate: $F(3,138)=52.606$, p<.001, η^2=.53; univariate: $F(3, 420)=131.893$, p<.001, η^2=.49): the closer the relationship between participant and violator, the less willing participants were to blow the whistle (*stranger:* M=5.42, SD=1.01;

acquaintance: M=5.16, SD=1.09; *friend:* M=4.31, SD=1.39; *family member:* M=4.00, SD=1.54). We also observed a significant interaction between condition and target (multivariate: $F(3,138)=3.735$, $p=.013$, $\eta^2=.075$; univariate: $F(3, 420)=8.581$, $p<.001$, $\eta^2=.058$): the effect of the loyalty versus fairness prime was more pronounced for close versus distant targets.

Independent-samples t-tests on dependent measures showed that condition predicted whistleblowing for *family member* (M=4.44, SD=1.50 vs. M=3.56, SD=1.46; p=.001), *friend* (M=4.65, SD=1.34 vs. M=3.96, SD=1.35; p=.003), marginally for *acquaintance* (M=5.33, SD=1.01 vs. M=4.99, SD=1.14; p=.062), and non-significantly for *stranger* (M=5.48, SD=1.05 vs. M=5.37, SD=0.98; p>.50). See Table 2 for whistleblowing scores by condition for each offense for this study and Study 2b.

Participants reported more willingness to blow the whistle when primed to consider fairness versus loyalty. This result further supports our hypothesis that endorsement of fairness versus loyalty norms tracks with subsequent decisions to blow the whistle.

Study 2b

Method

One hundred fifty-one MTurk workers (M_{age}=33.74, SD_{age}=12.77; 45% female) participated as in Study 1. Study 2b was identical to Study 2a, with two exceptions. (1) To alleviate task demands, we told participants at the start of the assignment that they would be completing *two* studies: a first study consisting of a writing task on values and a second study involving answering questions about social decision-making. The writing section, which involved the experimental manipulation, was clearly labeled, “STUDY 1: VALUES,” and the section that included the whistleblowing measures was clearly labeled, “STUDY 2: SOCIAL

DECISION MAKING." Second, we included additional questions at the study's end asking whether participants had previously seen any of the study materials as well as a series of suspicion checks to assess potential effects of task demands (Inbar, Pizarro, Gilovich, & Ariely, 2013). First, we asked whether participants thought there was anything "strange, confusing or suspicious about these studies" (Yes/No). If participants answered, "yes," they saw an item asking whether they thought the first and second study were related at all. If participants answered, "yes," they were asked to write about how they were related. Only three people answered, "yes" to any question; excluding these individuals did not alter results.

Results and Discussion

Independent-samples *t*-tests again revealed that overall willingness to blow the whistle (α=.93) was higher in the fairness condition (M=4.67, SD=1.07) than in the loyalty condition (M=4.30, SD=1.04), $t(149)=2.78$, p=.029, d=.36.

We also conducted the same 2 (condition: fairness/justice vs. loyalty) x 4 (target: stranger, acquaintance, friend, family member) mixed-effects ANOVA on participants' willingness to blow the whistle averaged across the seven scenarios (all αs>.72 for composite scores for each target). A main effect emerged for target (multivariate: $F(3,147)=55.76$, $p<.0001$, $\eta^2=.53$; univariate, Greenhouse-Geisser correction: $F(1.344, 200.31)=141.77$, $p<.0001$, $\eta^2=.49$): the closer the relationship between participant and violator, the less willing participants were to blow the whistle (*stranger:* M=5.23, SD=1.05; *acquaintance:* M=4.89, SD=1.07; *friend:* M=4.09, SD=1.30; *family member:* M=3.74, SD=1.46). Sphericity was violated ($p<.0001$), and the condition x target interaction was nonsignificant using the multivariate approach (F=1.51, p=.21) but was marginally significant using the univariate approach (Greenhouse-Geisser correction, $F(1.344, 200.31)=3.40$, p=.054, η^2=.067), revealing the same pattern of effects as in Study 2a.

Specifically, independent-samples t-tests on dependent measures showed that condition predicted whistleblowing for *family member* (M=4.04, SD=1.43 vs. M=3.45, SD=1.45; p=.012), *friend* (M=4.34, SD=1.27 vs. M=3.84, SD=1.28; p=.017), marginally for *acquaintance* (M=5.03, SD=1.07 vs. M=4.74, SD=1.06; p=.099), and non-significantly for *stranger* (M=5.29, SD=1.07 vs. M=5.17, SD=1.04; p>.47). Again, these findings reveal that priming fairness versus loyalty norms corresponds to subsequent whistleblowing decisions.

Study 3: Recalling past whistleblowing opportunities

Studies 1 and 2 measured whistleblowing behavior in hypothetical scenarios. Study 3 extended this work by asking participants about whether fairness and loyalty have contributed to real-life whistleblowing opportunities.

Method

One thirty-five MTurk workers (M_{age}=30.15, SD_{age}=9.82; 34% female) participated. Participants were randomly assigned to either the whistleblowing (WB+) condition or to the no whistleblowing (WB-) condition. In the WB+ condition, participants were asked to recall "an incident when you became aware that someone you knew—a friend, family member, co-worker, or acquaintance—did something that was morally wrong, and you had the option to report that person, and chose to report the person." In the WB-, participants were asked to recall a similar incident in which they had the option to report that person and chose *not* to do anything.

Participants were asked to write a short paragraph about what was going through their mind when they did or did not decide to report that person, their reasons for their decision and whether their decision was driven by any particular moral or ethical values. On a separate screen, participants indicated, "How difficult was your decision to (not) report the person?" (1=*Very*

easy; 7=*Very difficult*), and the extent to which their decision was driven by (1) the value of "fairness/justice" and (2) the value of "loyalty" (1=*Not at all*; 10=*Very much*).

Results and Discussion

In exploring participants' free responses, we created categories for any violation that appeared more than once; violations that appeared only once were categorized as "other". Participants were also given the opportunity *not* to identify the violation (to prevent distress over revealing sensitive information and to prevent data loss from people dropping out). Frequencies for this categorization scheme for the WB+ vs. WB- conditions were: stealing/theft (25 vs. 20), cheating (6 vs. 8), infidelity/sex (5 vs. 9), substance abuse (1 vs. 6), property damage (4 vs. 2), lying (3 vs. 3), physical abuse (3 vs. 3), sexual abuse/harassment (2 vs. 2), missing work (0 vs. 3), other (3 vs. 7), and those who chose not to identify the violation (9 vs. 11). Binomial tests confirmed that these counts did not differ by condition (all ps>.12). In addition, the difficulty of the decision did not differ by condition, $t(133)=0.530$, $p>.50$.

We analyzed participants' written responses using Linguistic Inquiry and Word Count (LIWC; Pennebaker Booth, & Francis, 2007), a text analysis software program. To investigate the values driving whistleblowing behavior, we used the Moral Foundations Dictionary (MFD, developed by Graham, Haidt, & Nosek, 2009), which includes lexical categories for fairness and loyalty. For each participant, we calculated the proportion of words in their response that fell into each MFD category. This analysis, described in supplementary materials, revealed that, as predicted, WB- participants used a larger proportion of loyalty versus fairness words. However, given that a low percentage of participants used words captured by the MFD and that this analysis appeared to be driven by outliers, we pursued a more sensitive analysis strategy next, analyzing the open-ended data by employing human coders.

Coders were blind to our hypotheses; however, because condition (WB+, WB-) was easily determined from participants' responses, we assigned coders to only one condition. Four independent coders (paid participants; two per condition) characterized the reason each participant gave for reporting/not reporting. Specifically, each response was coded using four binary-measures, i.e., whether or not (1) fairness, (2) loyalty, (3) a disregard for fairness, or (4) a disregard for loyalty played a part in the participant's decision. We emphasized that the codes should be based on the participant's final decision whether to report or not, and that any response could be coded as any combination of the four measures. Some participants were coded as having mentioned none of the principles, and some were coded as having mentioned more than one principle. This coding scheme allowed us to account for the richness of participants' responses. For example, some participants appeared to be contemplating the merits of both fairness and loyalty, whereas others showed both regard and disregard for loyalty (or fairness) simultaneously. Coders were given an additional section to comment if they felt any other reason played a role in participants' decisions; this approach revealed no notable differences. After coding the stories independently, coders within each pair discussed discrepancies between their codes and came to an agreement.

We analyzed final codes by conducting binomial tests on each variable—fairness, loyalty, disregard for fairness, disregard for loyalty—for each condition (see Table 3 for frequencies below; see supplementary materials for an alternative analysis). For the WB+ condition, significant differences emerged in the predicted direction for fairness (72.1% were identified as using this principle, p=.001), loyalty (23%, p<.0001), and disregard for fairness (78.7%, p<.0001) (disregard for loyalty, p=.31). For the WB- condition, significant differences emerged in the predicted direction for fairness (5.4%, p<.0001), disregard for fairness (62.2%,

p=.047), and disregard for loyalty (94.6%, p<.0001) (loyalty, p=.91). Chi-square tests that compare each variable by condition were significant as well, χ^2s>9.46, ps<.01, Φs>.26.

We also analyzed the Likert scale questions of "to what extent the decision was driven by fairness or loyalty." A 2(condition: WB+ vs. WB-) x 2(value: fairness vs. loyalty) mixed-effects ANOVA revealed the predicted condition x value interaction, $F(1,133)=112.191$, p<.001, η^2=.46. WB+ Participants expressed higher levels of fairness (M=7.98, SD=2.11) and lower levels of loyalty (M=4.72, SD=3.03), compared to WB- participants; WB- participants expressed higher levels of loyalty (M=7.55, SD=2.42) and lower levels of fairness (M=4.51, SD=2.24). All simple effects were significant (independent-samples t-tests, ps<.001) (Figure 1). No main effects emerged. Overall, these findings suggest that fairness and loyalty constitute core values considered for whistleblowing decisions.

Study 4: Modulating whistleblowing behavior

Study 4 combined methods from Studies 2 and 3 to test whether priming fairness versus loyalty could modulate actual whistleblowing behavior in an online environment.

Method

One hundred fifty-seven MTurk workers (M_{age}=35.09, SD_{age}=13.61; 48% female) were randomly assigned to the fairness or loyalty condition, in which they wrote an essay about the importance of promoting a sense of fairness or loyalty amongst Amazon.com Mechanical Turk's workplace. For example, in the fairness condition, participants were asked to "write about why it is important to be just and fair amongst your fellow workers", whereas in the loyalty condition, participants were asked to "write about why it is important to be loyal amongst your fellow workers" (see supplementary materials for complete essay prompts).

Following the essay task, participants were told they would be instructed to complete a typing task for piloting purposes. In this task, they were asked to write out, in English, all numbers, 1-30. Participants read that they would then see the work of the previous participant. All participants were presented with an example of substandard work, which participants later had the chance to report (Figure 2).

Participants proceeded with the task and then afterward wrote about their impressions of the task and the previous participant's work. In these open-ended questions, six participants expressed suspicion about the existence of this previous participant (e.g., "It was not a real person") or the purpose of the study (e.g., "I'm guessing you were trying to prime me for fairness and seeing how I reacted to a person who acted unfairly"); excluding these participants did not alter the results. Finally, participants responded to two key items on a 7-point scale (1=*definitely not*; 7=*definitely yes*). The first item stated, "Please tell us, would you recommend that we block this previous participant from future MTURK studies?" The second item stated, "The present study is a pilot study for future studies. In the future we plan to weed out rule violators. Please report: Did the previous participant violate any rules or codes of MTURK?" These items were averaged together ($r(155)=.53$, $p<.0001$) as a composite measure of whistleblowing.

Results and Discussion

An independent-samples *t*-test revealed that participants in the fairness condition ($M=4.36$, $SD=1.83$) engaged in more whistleblowing behavior than participants in the loyalty condition ($M=3.72$, $SD=1.71$), $t(155)=2.29$, $p=.023$, $d=.37$. Only ten participants (6% of the sample) did not properly complete the typing task, five per condition (thus, completion rates seem unlikely to have affected whistleblowing behavior). We suggest instead that the present effects resulted from experimental priming of fairness versus loyalty prior to participants'

evaluation of a substandard worker. This result provides further support that the fairness-loyalty tradeoff contributes to real-world whistleblowing decisions.

General Discussion

The present research provides a novel characterization of whistleblowing as a tradeoff between acting fairly and acting loyally. Five studies demonstrate that moral behavior appears to be influenced not simply by the moral norms that people hold but by how people *trade off* different moral norms. The current work thus follows a tradition of research that describes decision-making as involving tradeoffs between competing values (Keeney & Raiffa, 1976; von Neumann & Morgenstern, 1947). In this case, we examine the influence of values that people may naturally trade off (Jost, 2009; see also, Graham, 2011). The present goal was to examine the fairness-loyalty tradeoff for whistleblowing decisions, yet it will be important for future work to determine the separate influence of these values, the influence of additional values (e.g., concerns regarding harm, authority, purity, and liberty), as well as these values' influence on decisions related to whistleblowing, such as confronting the offender directly (Czopp, Monteith, & Mark, 2006).

Our research raises a question concerning the precise meaning of loyalty in the context of whistleblowing. Theoretical discussion of organizational whistleblowing sees "whistleblowing as entirely compatible with employee loyalty" (Larmer, 1992, p. 125) and indicates that "whistle-blowers have been argued to be more loyal to the organization than inactive observers" (Vadera et al., 2009, p. 558). Indeed, loyalty may reflect loyalty not only to an offending individual but also to one's organization at large; thus the whistleblower may believe that reporting wrongdoing benefits the organization and its reputation. Loyalty may therefore be targeted at groups of

varying scope – from one's immediate friends and family to society at large. Fairness too may reflect diverse values, from equality to meritocracy (Rai & Fiske, 2011). Notably, the present studies relied on fairly broad definitions of fairness and loyalty, in line with existing literature in moral psychology (e.g., Moral Foundations Theory; Graham, Haidt, & Nosek, 2009; Graham et al., 2011; Haidt & Josephs, 2004). Importantly, our findings suggest that valuation of fairness, broadly construed, over loyalty, broadly construed, increased whistleblowing decisions.

Given the results of Studies 1 and 2, we also qualify that this tradeoff may be most pertinent when the perpetrator of wrongdoing and the whistleblower are socially close (e.g., relatives). Although unpredicted, this interaction effect can be understood in the context of prior work on relationship types and moral norms (Haidt, 2007; Rai & Fiske, 2011), which suggests that the principle of loyalty may apply more to close relationships than distant relationships. Because loyalty represents a key component of the fairness-loyalty tradeoff, this tradeoff can be expected to play a significant role in cases of close relationships. Future work should additionally examine whether this effect is modulated by the nature of the loyalty – loyalty to the individual (e.g., family member, friend) or the relevant group (e.g., family, social circle).

Implications of the Present Research

The tradeoff illustrated here has important implications for research on the influence of moral emotions, such as guilt. On one hand, guilt proneness is associated with unwillingness to behave unethically (Cohen, Panter, & Turan, 2012), suggesting guilt *increases* willingness to blow the whistle on fairness violations (see also, Ellemers, Spears, & Doosje, 2002). On the other hand, when ingroup concerns are made salient, guilt-prone individuals show increased concern for the welfare of their ingroup (at the expense of the outgroup) (Cohen, Montoya, & Insko, 2006), suggesting guilt may also *decrease* whistleblowing in adherence to loyalty norms.

The current results illuminate these seemingly contradictory findings. When moral norms conflict, as in the case of fairness and loyalty, the relationship between moral emotions and behavioral outcomes will vary depending on which moral norm takes precedent.

The current work also informs our understanding of the cross-cultural perceptions of unethical behavior. For example, although most Western cultures are individualistic, endorsing fairness over loyalty, many Asian cultures are collectivistic, exalting the importance of loyalty (Graham et al., 2011; see also, Miller & Bersoff, 1992). Asian cultures might therefore preferentially tolerate unfair but "loyal" behavior. Indeed, existing research suggests that the United States, an individualistic country, views whistleblowing as more ethical than collectivistic countries such as Japan (Brody, Coulter, & Mihalek, 2001), China (Chiu, 2003), and Taiwan (Brody, Coulter, & Lin, 1999; see also Christie, Kwon, Stoeberl, & Baumhart, 2003).

The present research may also be relevant for ongoing normative debates about the very nature of morality. On one account, a key aspect of morality is impartiality (DeScioli & Kurzban, 2009; 2012; Kurzban, DeScioli, & Fein, 2012). This account suggests that being moral consists of being fair, rather than being loyal. Despite this, adults and children are prone to behaving loyally rather than fairly or impartially (Bloom, 2004; Haidt, 2012; Shaw, et al., 2012; Hamlin, et al, 2013). Indeed, our moral senses may have evolved from processes that function to favor the ingroup, such as kin selection (Bloom, 2011; Churchland, 2011).

Finally, the current findings offer recommendations for how to promote fairness and to encourage whistleblowing. One suggestion is to engage brute-force deliberate reasoning to override prepotent partiality-based responses of ingroup favoritism and bias (Feinberg, Willer, Antonenko, & John, 2012; Kahneman, 2011). Another method may be to reframe whistleblowing as demonstrating a "larger loyalty" (Rorty, 1997) – an allegiance to the

superordinate group of society as a whole and the *greater good*. Loyalty can reflect allegiance toward a distinctive ingroup or toward a more universal social circle, and promoting loyalty toward a larger social circle may in fact reflect norms of fairness and promote whistleblowing behavior. Reconciling the conflict between fairness and loyalty in these terms may improve perceptions of whistleblowing and ultimately encourage ethical behavior across cultures.

Acknowledgments

We thank Davis Vo and Ellen Winner for advice and assistance on this work.

References

Baron, J., Ritov, I., & Greene, J. D. (2011). The duty to support nationalistic policies. *Journal of Behavioral Decision Making*.

Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology*, *68*, 1042-1054.

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review*, *99*, 26-43.

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York, NY: Basic Books.

Brody, R.G., Coulter, J.M., & Lin, S. (1999). The Effect of national culture on whistle-blowing perceptions. *Teaching Business Ethics*, *3*, 385-400.

Brody, R.G., Coulter, J.M., & Mihalek, P.H. (1998). Whistle-Blowing: A Cross-Cultural comparison of ethical perceptions of U.S. and Japanese accounting students. *American Business Review*, *16*, 14-21.

Brooks, D. (2013). Katrina's silver lining. *New York Times*, 23.

Brosnan, S.F., Schiff, H.C. & de Waal, F.B.M. (2005). Tolerance for inequity may increase with social closeness in chimpanzees. *Proceedings of the Royal Society: Biological Sciences*, *272*, 253-258.

Chiu, R.K. (2003). Ethical judgment and whistleblowing intention: Examining the moderating role of locus of control. *Journal of Business Ethics*, *43*, 65-74.

Choi, J. -K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science*, *318*, 636-640.

Christie, P.M.J., Kwon, I.G., Stoeberl, P.A., & Baumhart, R. (2003). A Cross-Cultural comparison of ethical attitudes of business managers: India, Korea, and the United States. *Journal of Business Ethics*, *46*, 263-287.

Churchland, P. (2011). *Braintrust: What neuroscience tells us about morality*. Princeton University Press.

Cohen, T. R., Montoya, R. M., & Insko, C. A. (2006). Group morality and intergroup relations: Cross-cultural and experimental evidence. *Personality and Social Psychology Bulletin*, *32*, 1559-1572.

Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. *Current Directions in Psychological Science*, *21*, 355-359.

Czopp, A.M., Monteith, M.J., & Mark, A.Y. (2006). Standing Up for a Change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, 90(5), 784-803.

DeScioli, P. & Kurzban, R. (2012). A Solution to the mysteries of morality. *Psychological Bulletin*, *22*, 1151-1164.

DeScioli, P. & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*, 281-299.

Dozier, J.B. & Miceli, M.P. (1985). Potential predictors of whistleblowing: A prosocial behavior perspective. *Academy of Management Review, 10*, 823-836.

Dyck, A., Adair, M., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *Journal of Finance*, *65*, 2213-2254.

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and Social Identity. *Annual Review of Psychology*, *53*, 161-186.

Feinberg, M., Willer, R., Antonenko, O., & John, O.P. (2012). Liberating reason from the

passions: Overriding intuitionist moral judgments through emotion reappraisal. *Psychological Science*, *23*, 788-795.

Friman, P.C., Woods, D.W., Freeman, K.A., Gilman, R., Short, M., McGrath, A.M. & Handwerk, M.L. (2004). Relationships between tattling, likeability, and social classification: A preliminary investigation of adolescents in residential care. *Behavior Modification*, *28*, 331-348.

Gino, F. & Bazerman, M.H. (2009). When misconduct goes unnoticed: The acceptability of gradual erosion in others' unethical behavior. *Journal of Experimental Social Psychology*, *45*, 708-719.

Graham, J. (2011). *Left Gut, Right Gut: Ideology and Automatic Moral Reactions* (Doctoral dissertation, University of Virginia).

Graham, J., Haidt, J. & Nosek, B. (2009). Liberals and conservatives use different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029-1046.

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.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998-1002.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.

Hamlin, J.K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science*.

Hamlin, J.K., Wynn, K., Bloom, P. & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Science, U.S.A.*, *108*, 19931-19936.

Inbar, Y., Pizarro, D. A., Gilovich, T., & Ariely, D. (2013). Moral masochism: On the connection between guilt and self-punishment. *Emotion*, *13*, 14-18

Ingram, G.P.D. & Bering, J.M. (2010). Children's tattling: The reporting of everyday norm violations in preschool settings. *Child Development*, *81*, 945-957.

Johnson, R.A. (2003). *Whistleblowing: When it works and why*, Boulder, CO: Lynne Riener Publishers.

Jost, J. T. (2009). Group morality and ideology: left and right, right and wrong. Paper presented to the Society for Personality and Social Psychology annual conference, Tampa, FL.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus, & Giroux, New York.

Kanngiesser, P. & Warneken, F. (2012). Young children consider merit when sharing resources with others. *PLoS ONE*, *7*, 1 – 5.

Keeney, R. L., & Raiffa, H. (1976). Decisions with multiple objectives: Preferences and value tradeoffs. New York: Wiley.

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, *33*, 323-333.

Larmer, R. A. (1992). Whistleblowing and employee loyalty. *Journal of Business Ethics*, *11*, 125-128.

Lun, J., Oishi, S., & Tenney, E. R. (2012). Residential mobility moderates preferences for egalitarian versus loyal helpers. *Journal of Experimental Social Psychology*, *48*, 291-297.

Mahajan, N., Martinez, M.A., Gutierrez, N.L., Diesendruck, G., Banaji, M.R. & Santos, L. (2011). The evolution of intergroup bias: Perceptions and attitudes in rhesus macaques. *Journal of Personality and Social Psychology*, *100*, 387-405.

Miceli, M.P. & Near, J.P. (1992). *Blowing the Whistle*, New York, NY: Lexington.

Miller, J.G. & Bersoff, D.M. (1992). Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology*, *62*, 541-554.

Minson, J. A. & Monin, B. (2011). Do-gooder derogation: Putting down morally-motivated others to defuse implicit moral reproach. *Social and Psychological and Personality Science*, *3*, 200-207.

Monin, B., Sawyer, P.J. & Marquez, M.J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95*, 76-93.

Natapoff, A. (2004). Snitching: The institutional and communal consequences. *University of Cincinnati Law Review*, *73*, 645-703.

Near, J.P. & Miceli, M.P. (1985). Organizational dissidence: The case of whistle-blowing. *Journal of Business Ethics*, *4*, 1-16.

Olson, K.R. & Spelke, E.S. (2008). Foundations of cooperation in young children. *Cognition*. *108*, 222-231.

Parks, C. D., & Stone, A. B. (2010). The desire to expel unselfish members from the group. *Journal of Personality and Social Psychology*, *99*, 303-310.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007. *Austin, TX, LIWC. Net*.

Rai, T. S. & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57-75.

Rorty, R. (1997). 'Justice as a larger loyalty' in *Justice and democracy: Cross-cultural perspectives*, Eds. Bonketoe R, Stepaniants M (Hawaii, University of Hawaii Press), 9 –

22.

Shaw, A., DeScioli, P. & Olson, K.R. (2012). Fairness versus favoritism in children. *Evolution and Human Behavior*, *33*, 736 – 745.

Simmons, J., Nelson, L., & Simonsohn, U. (2012). A 21-word Solution. *Dialogue*, *26*, 4-12.

Sloane, S., Baillargeon, R. & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, *23*, 196-204.

Smetana, J.G., Killen, M. & Turiel, E. (1991). Children's reasoning about interpersonal and moral conflicts. *Child Development*, *62*, 629-644.

Vadera, A.K., Vadera, R.V. & Caza, B.B. (2009). Making sense of whistle-blowing's antecedents: Learning from research on identity and ethics programs. *Business Ethics Quarterly*, *19*, 553-586.

Vogel G (2011). Psychologist accused of fraud on an astonishing scale. *Science 334*, 579.

von Neumann, J, & Morgenstern, O. (1947). Theory of games and economic behavior (2nd ed.). Princeton, NJ: Princeton University Press.

Walker, L.J. & Hennig, K.H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology*, *86*, 629-647.

Yong, E. (2012). The data detective. *Nature*, *487*, 18-19.

Table 1. Correlations between fairness-minus-loyalty score and whistleblowing score for each offense (Study 1).

Stealing	Embezzlement	Robbery	Cheating	Graffiti	Marijuana	Stabbing
.21 (p=.055)	.26 (p=.019)	.20 (p=.073)	.26 (p=.019)	.29 (p=.008)	.07 (p=.52)	.13 (p=.26)

Table 2. Whistleblowing scores for each offense, by condition (Study 2)

Study/Condition	Stealing *M (SD)*	Embezzlement *M (SD)*	Robbery *M (SD)*	Cheating *M (SD)*	Graffiti *M (SD)*	Marijuana *M (SD)*	Stabbing *M (SD)*
2a/Fairness	4.34_a (1.63)	5.61_a (1.36)	6.01_a (1.15)	4.27_a (1.94)	4.57_a (1.78)	3.56_a (2.08)	6.48_a (0.90)
2a/Loyalty	3.82_b (1.51)	5.02_b (1.27)	5.88_a (1.13)	3.37_b (1.73)	3.95_b (1.79)	2.91_a (1.96)	6.33_a (1.05)
2b/Fairness	3.78_a (1.80)	5.37_a (1.46)	5.96_a (1.02)	3.76_a (1.77)	4.35_a (1.72)	3.17_a (2.12)	6.33_a (1.04)
2b/Loyalty	3.65_a (1.69)	4.96_a (1.30)	5.35_b (1.25)	3.38_a (1.66)	3.85_a (1.64)	2.91_a (1.95)	5.97_b (1.07)

Means that do not share a subscript across column differ significantly, within each Study. In Study 2a, marijuana differs by condition at p=.060. In Study 2b, embezzlement differs at p=.073 and graffiti varies at p=.069.

Table 3. Frequency of principles that coders identified by condition, *N*=61 in WB+ condition, *N*=74 in WB- condition (Study 3).

Condition	Fairness	Loyalty	Disregard for Fairness	Disregard for Loyalty
WB+	44	14	13	26
WB-	4	36	46	4

Figure captions:

Figure 1. Participants instructed to recall past decisions to blow the whistle (WB+ condition) reported that their decision was driven more by fairness than by loyalty, whereas participants instructed to recall past decisions to *not* blow the whistle (WB- condition) reported that their decision was driven more by loyalty than by fairness (Study 3). Error bars indicate standard error.

Figure 2. The substandard work of the ostensible prior participant that all actual participants saw in Study 4.

Figure 1.

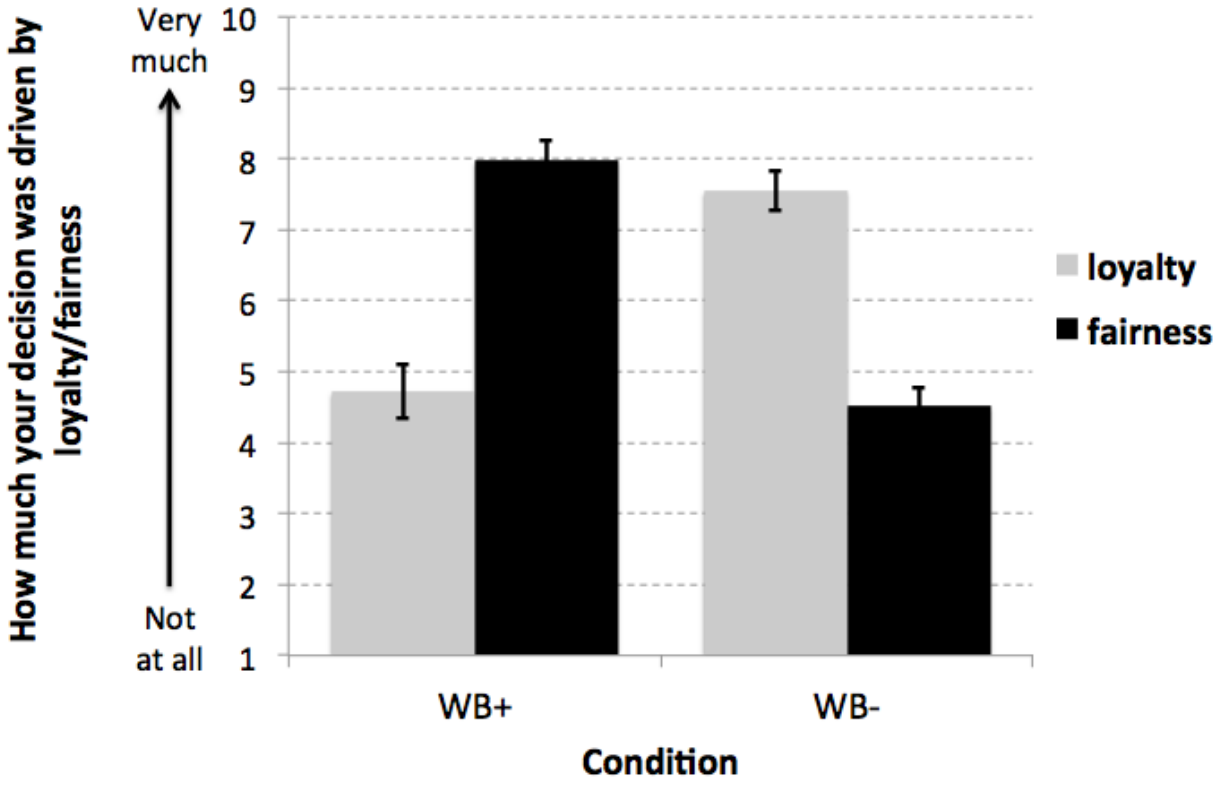
How much your decision was driven by loyalty/fairness
Very much
Not at all
10
9
8
7
6
5
4
3
2
1
WB+
WB-
Condition
loyalty
fairness

Figure 2.