The whistleblower's dilemma and the fairness–loyalty tradeoff

Adam Waytz [a],[*], James Dungan [b], Liane Young [b]

[a] Northwestern University, USA
[b] Boston College, USA

HIGHLIGHTS

- The tradeoff between fairness and loyalty corresponds to whistleblowing decisions.
- Experimental and dispositional variation in this tradeoff maps onto whistleblowing.
- Five studies demonstrate this previously undocumented relationship.
- These results shed light on a novel psychological determinant of whistleblowing.

ARTICLE INFO

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.

© 2013 Elsevier Inc. All rights reserved.

Introduction

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 (Dozier & Miceli, 1985; Near & 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 (Kanngiesser & Warneken, 2012; Sloane, Baillargeon, & Premack, 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

* Corresponding author at: Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208, USA.
E-mail address: a-waytz@kellogg.northwestern.edu (A. Waytz).

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 overridden 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, 2013), and relationship type (Rai & Fiske, 2011) modulate people's preference for loyalty versus fairness (see also Shaw et al., 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 the one hand, whistleblowers may act in the service of fairness and justice when exposing corporate wrongdoing (Miceli & Near, 1992; Near & Miceli, 1985), 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., Minson & Monin, 2012; Monin, Sawyer, & Marquez, 2008; Parks & Stone, 2010). 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]

---

[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.

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 et al., 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 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 among 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 ($\alpha = .64$) 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/

---

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