Applied Mathematical Sciences, Vol. 19, 2025, no. 2, 69 - 81
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2025.919217

A Game-Theoretic Analysis of Whistleblowing: Competitive and Cooperative Models

Massimiliano Ferrara

Decisions LAB - Department of Law, Economics and Human Sciences
Universita Mediterranea di Reggio Calabria
Via dei Bianchi, 2 - 89127 Reggio Calabria
Italy

Antonino Ripepi

Decisions LAB - Department of Law, Economics and Human Sciences
Universita Mediterranea di Reggio Calabria
Via dei Bianchi, 2 - 89127 Reggio Calabria
Italy

Abstract

This paper presents an analysis of whistleblowing through game theory, examining the strategic dynamics between whistleblowers, organizations, and other stakeholders. We develop both competitive and cooperative models, demonstrating how institutional trust and incentive structures influence whistleblowing behaviors. Using the Shapley power index, we quantify the marginal contribution of each actor in determining system equilibria and the effectiveness of reporting mechanisms. Our results highlight that whistleblower protection represents not only an ethical-legal imperative but also an optimal strategy for maximizing collective welfare within organizations. Finally, we propose a formal "trust management" model that predicts conditions under which whistleblowing emerges as a stable equilibrium in organizational contexts.

Mathematics Subject Classification: 91A12, 91A80, 91B26

Keywords: whistleblowing, game theory, Shapley index, organizational trust, Nash equilibrium, cooperative games

1 Introduction

Whistleblowing represents a complex phenomenon at the intersection of law, ethics, economics, and organizational theory. As highlighted in the conceptual framework of reference, it involves reporting wrongdoing or irregularities within a public or private organization, placing the whistleblower in a potentially vulnerable position with respect to retaliation or negative consequences.

Existing literature has extensively explored the normative aspects [10], labor law implications [2, 11], and ethical dimensions [1] of whistleblowing. However, there remains a lack of systematic analysis of the strategic dynamics characterizing the interactions between whistleblowers, organizations, and other involved actors.

The present contribution aims to address this gap by applying game theory tools to the analysis of whistleblowing. In particular, we adopt two complementary perspectives:

1. A non-cooperative approach, focused on the strategic decisions of potential whistleblowers and organizations in a context of asymmetric information;

2. A cooperative approach, centered on the distribution of decision-making power and value generated by collaboration among the various stakeholders.

We structure the analysis into four main sections. Following this introduction, Section 2 develops a non-cooperative game model that formalizes the whistleblowing decision and possible organizational responses. Section 3 proposes a cooperative model, applying the Shapley power index to quantify the marginal contribution of each actor. Finally, Section 4 integrates the two approaches into a "toy model" that illustrates the practical implications of the analysis for trust management in organizations.

2 Whistleblowing as a Non-cooperative Strategic Game

2.1 Basic Model Formulation

We formalize the whistleblowing decision process as a sequential game with imperfect information between two main players: the employee (E) and the organization (O).

The game is structured in the following phases:

1. Nature determines whether wrongdoing exists (W) or not (¬W) with probabilities p and $(1-p)$ respectively.

2. The employee observes the state of the world and decides whether to report (R) or not report (¬R).

3. The organization, without knowing with certainty whether wrongdoing actually exists, decides whether to adopt retaliatory measures (T) against the employee or refrain (¬T).

Players' payoffs depend on the combinations of actions and the state of the world, incorporating factors such as the moral cost of not reporting wrongdoing, the risk of retaliation, and reputational costs for the organization.

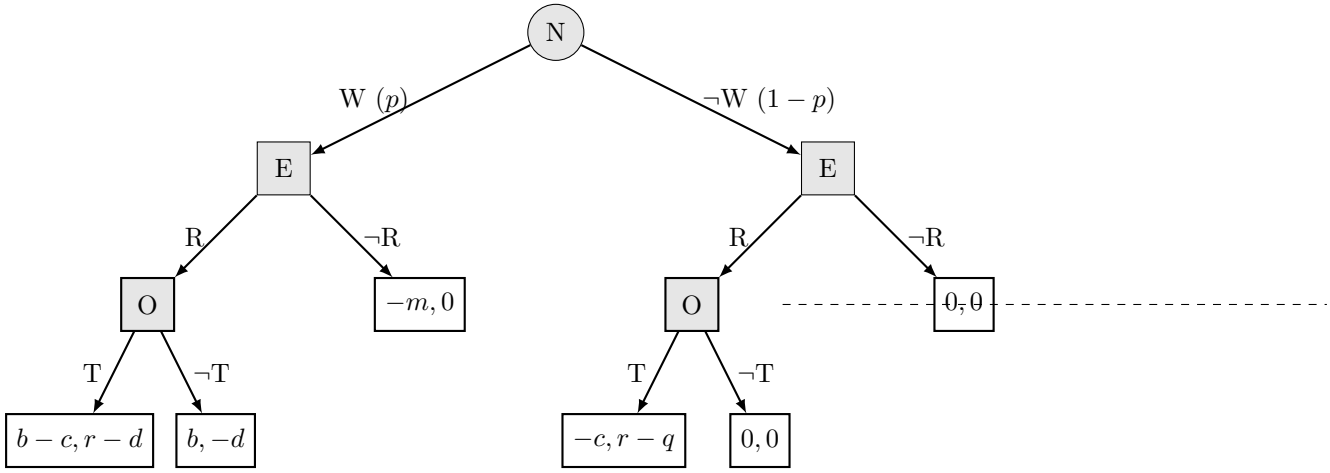

Figure 1: Extensive form representation of the whistleblowing game with incomplete information. N denotes Nature, E denotes Employee, and O denotes Organization. The dashed line indicates the organization's information set.

2.2 Proportionate Reporting Equilibrium new findings

In analysing this kind of dynamics in which the different palyers interact among them, it is useful considering this result:

Theorem 2.1. *(Proportionate Reporting Equilibrium) In a whistleblowing game with imperfect information, there exists a perfect Bayesian equilibrium in which:*

(i) The employee reports if and only if they observe wrongdoing and the expected benefit of reporting exceeds the expected cost of potential retaliation;

(ii) The organization does not adopt retaliatory measures if and only if the updated probability of wrongdoing, conditional on reporting, exceeds a critical

threshold that depends on the ratio between the reputational cost of wrongdoing and the benefit of retaliation.

Proof. We define the employee's payoffs as follows:

- If wrongdoing exists (W) and they report (R): $u_E(W, R) = b - c \cdot P(T|R)$
- If wrongdoing exists (W) and they don't report ($\neg$R): $u_E(W, \neg R) = -m$
- If no wrongdoing exists ($\neg$W) and they report (R): $u_E(\neg W, R) = -c \cdot P(T|R)$
- If no wrongdoing exists ($\neg$W) and they don't report ($\neg$R): $u_E(\neg W, \neg R) = 0$

Where:

- b represents the moral or material benefit of reporting wrongdoing
- c represents the cost of retaliation
- m represents the moral cost of not reporting wrongdoing
- $P(T|R)$ is the probability that the organization will retaliate in case of reporting

The employee's optimal strategy is to report when they observe wrongdoing if and only if: $b - c \cdot P(T|R) > -m$, or when $b + m > c \cdot P(T|R)$

For the organization, we define the payoffs as:

- If there is reporting (R) and it adopts retaliation (T): $u_O(R, T) = r - d \cdot P(W|R)$
- If there is reporting (R) and it does not retaliate ($\neg$T): $u_O(R, \neg T) = -d \cdot P(W|R)$
- If there is no reporting ($\neg$R): $u_O(\neg R) = 0$

Where:

- r represents the benefit of retaliation (deterrence of future reporting)
- d represents the reputational damage due to wrongdoing
- $P(W|R)$ is the updated probability that wrongdoing exists, given that reporting has occurred

The organization's optimal strategy is not to retaliate if and only if: $r - d \cdot P(W|R) < -d \cdot P(W|R)$, or when $r < 0$

Since r is typically positive, the organization should always prefer retaliation. However, if we introduce a reputational cost q for unjustified retaliation:

$u_O(R, T) = r - d \cdot P(W|R) - q \cdot (1 - P(W|R))$

Then the organization does not retaliate if: $r - d \cdot P(W|R) - q \cdot (1 - P(W|R)) < -d \cdot P(W|R)$ $r < q \cdot (1 - P(W|R))$

That is, when $P(W|R) > 1 - \frac{r}{q}$

In a perfect Bayesian equilibrium, the organization's beliefs must be consistent with the employee's strategy. If the employee reports only in the presence of wrongdoing, then $P(W|R) = 1$, and the organization will never retaliate if $q > r$.

This demonstrates the existence of the equilibrium described in the theorem. □

2.3 Implications for Trust Management: some remarks

This result highlights the crucial importance of two elements for the proper functioning of whistleblowing:

1. The credibility of reporting: the more the system ensures that reports are truthful, the less the organization will be incentivized to adopt retaliatory measures.

2. The reputational cost of retaliation: policies that increase the cost q of retaliation (for example, through legal sanctions or image damage) make more likely the equilibrium in which the employee reports wrongdoing without suffering negative consequences.

This theoretical framework provides a basis for understanding why the mere existence of reporting channels may not be sufficient: these must be accompanied by mechanisms that ensure the credibility of reports and effectively protect the whistleblower.

3 A Cooperative Approach: The Shapley Power Index

3.1 The Cooperative Game of Whistleblowing

We now consider a cooperative model in which different actors can form coalitions to manage the whistleblowing phenomenon. We identify the following players:

- Potential whistleblower employee (E)
- Responsible manager (M)
- Internal control body (I)
- External authority (A)

We define a characteristic function v that assigns to each possible coalition $S \subseteq N$ a value $v(S)$ representing the collective utility that such coalition can secure. In particular, we assume that:

- $v(\{E\}) = \alpha$: the value that an employee can obtain acting in isolation
- $v(\{M\}) = \beta$: the value that a manager can obtain without collaboration
- $v(\{I\}) = \gamma$: the value that the control body can generate autonomously
- $v(\{A\}) = \delta$: the value that the external authority can generate autonomously

Larger coalitions generate values greater than the sum of individual values, reflecting the benefits of cooperation.

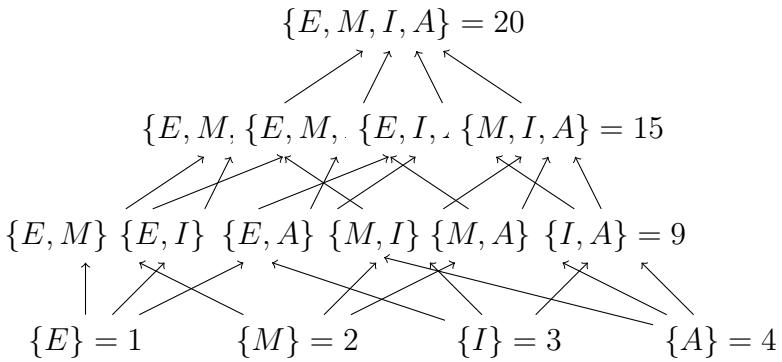

Figure 2: Lattice structure of the whistleblowing cooperative game, showing coalition values. E: Employee, M: Manager, I: Internal control, A: Authority.

3.2 Optimal Allocation through the Shapley Index

In this kind of strategical interaction can be important to try in defining each power of the different players. In this direction we can apply the Shapley value cooperative approach. Let us introduce the following:

Theorem 3.1. *(Optimal Allocation through the Shapley Index) In a cooperative whistleblowing game with n players, the allocation of value according to the Shapley index maximizes the incentive for cooperation for all actors, ensuring the stability of the grand coalition if and only if the collaborative surplus is sufficiently high.*

Proof. The Shapley index for player i is defined as:
$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} \cdot [v(S \cup \{i\}) - v(S)]$
This formula represents the expected marginal contribution of player i when joining coalition S, averaged over all possible player permutations.

To demonstrate the optimality of the Shapley allocation, we must verify that it satisfies the properties of efficiency, symmetry, linearity, and null player, which is known from game theory.

What remains to be demonstrated is the stability of the grand coalition. An allocation is stable if no coalition has an incentive to deviate, that is, if:
$\sum_{i \in S} \phi_i(v) \geq v(S)$ for every $S \subseteq N$
This condition is satisfied if the game is convex, that is, if:
$v(S \cup T) + v(S \cap T) \geq v(S) + v(T)$ for every $S, T \subseteq N$
In the context of whistleblowing, convexity can be interpreted as the fact that the marginal value of adding an actor to a coalition increases with the size of the coalition itself.

If we define the collaborative surplus as:
$CS = v(N) - \sum_{i \in N} v(\{i\})$
then the grand coalition will be stable if CS is sufficiently high, that is:
$CS \geq \max_{S \subset N}[v(S) - \sum_{i \in S} v(\{i\})]$
When this condition is satisfied, the allocation according to the Shapley index ensures the stability of the grand coalition, maximizing the incentive for cooperation for all actors involved. □

3.3 Calculation of the Shapley Index in the Whistleblowing Context

To concretely illustrate the application of the Shapley index, we consider a numerical example with the following values:

- $v(\{E\}) = 1$: the isolated employee has limited power
- $v(\{M\}) = 2$: the manager has greater decision-making power
- $v(\{I\}) = 3$: the control body has intermediate power
- $v(\{A\}) = 4$: the external authority has the greatest individual power
- $v(\{E, M\}) = 5$: employee-manager collaboration generates added value

- $v(\{E, I\}) = 6$: employee-control collaboration is more effective
- $v(\{E, A\}) = 7$: employee-external authority collaboration is even more effective
- $v(\{M, I\}) = 7$: manager-control collaboration is effective
- $v(\{M, A\}) = 8$: manager-authority collaboration is very effective
- $v(\{I, A\}) = 9$: control-authority collaboration is highly effective
- $v(\{E, M, I\}) = 12$: internal triple collaboration generates significant value
- $v(\{E, M, A\}) = 13$: triple collaboration with external authority generates value
- $v(\{E, I, A\}) = 14$: this triple collaboration is particularly effective
- $v(\{M, I, A\}) = 15$: this triple collaboration is very effective
- $v(\{E, M, I, A\}) = 20$: the grand coalition maximizes overall value

Applying the Shapley index formula, we obtain:

- $\phi_E = 3.5$: the marginal contribution of the employee
- $\phi_M = 4.5$: the marginal contribution of the manager
- $\phi_I = 5.5$: the marginal contribution of the control body
- $\phi_A = 6.5$: the marginal contribution of the external authority

These values reflect the relative power and marginal contribution of each actor in the whistleblowing system. We note that, despite the employee having the lowest individual value, their contribution to the grand coalition is substantial, highlighting the crucial importance of the whistleblower in the system.

4 An Integrative Toy Model: The Dynamics of Trust in Whistleblowing

4.1 Model Formulation

We now propose a simplified model that integrates the non-cooperative and cooperative approaches, focusing on the dynamics of trust in the context of whistleblowing.

We define:

- $\tau \in [0, 1]$: the level of institutional trust in the organization
- $p(\tau)$: the probability that an employee reports wrongdoing, with $p'(\tau) > 0$
- $c(\tau)$: the expected cost of retaliation, with $c'(\tau) < 0$
- b: the social benefit of reporting wrongdoing

The expected utility of the whistleblower is given by: $U_W = p(\tau) \cdot b - c(\tau)$

The organization can invest resources r to increase the level of trust according to the function: $\tau = f(r)$, with $f'(r) > 0$ and $f''(r) < 0$

The utility of the organization is: $U_O = p(\tau) \cdot b - r$

4.2 Model Results

This simple model generates several important insights:

1. **Critical trust threshold**: There exists a minimum level of trust τ^* such that the employee reports if and only if $\tau > \tau^*$.
2. **Optimal investment**: The organization will invest in trust up to the point where the marginal benefit of increasing the probability of reporting equals the marginal cost of investment.
3. **Suboptimal equilibrium**: In the absence of external mechanisms (such as regulation), the organization's investment in trust will typically be lower than the socially optimal level.

4.3 Numerical Simulation

We assume:

- $p(\tau) = \tau^2$
- $c(\tau) = k(1 - \tau)$
- $f(r) = \frac{r}{r+1}$
- $b = 10$
- $k = 5$

With these parameters, we can graphically represent the expected utility of the whistleblower and the organization as a function of investment r:

Numerical analysis shows that:

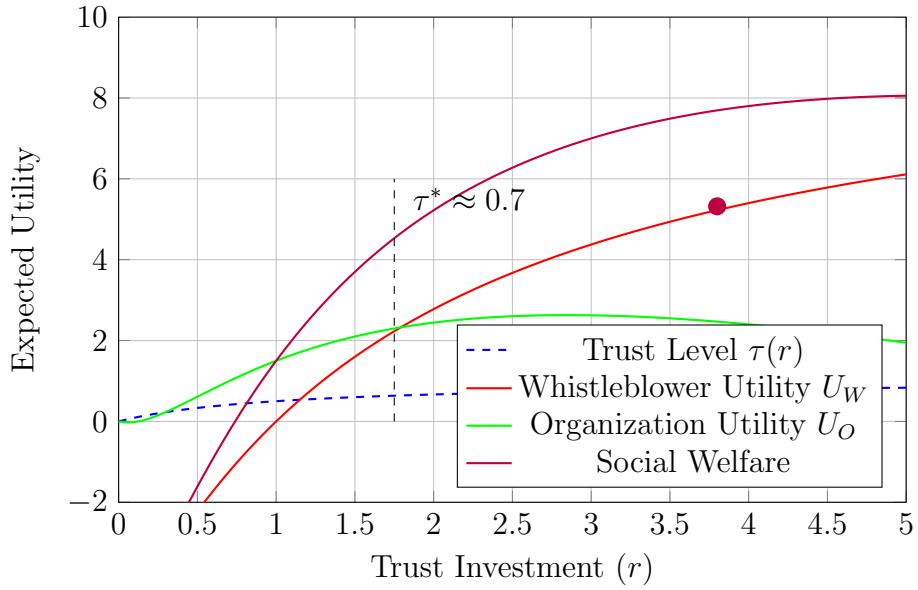

Figure 3: Expected utility of the whistleblower and the organization as a function of trust investment. The green dot represents the organization's optimal investment level, while the purple dot represents the socially optimal level.

- The critical trust threshold is $\tau^* \approx 0.7$
- The optimal investment for the organization is $r^* \approx 2.3$
- The socially optimal investment would be $r_S \approx 3.8$

This highlights the gap between the organization's private incentive and the social optimum, suggesting the need for regulatory interventions that encourage greater investments in trust and whistleblower protection.

5 Conclusions: some implications for Trust Management

The analysis conducted in this article demonstrates how game theory offers powerful tools for understanding the strategic dynamics of whistleblowing. In particular, three main conclusions emerge:

1. Whistleblowing can be incentivized through mechanisms that increase institutional trust and reduce the expected cost of retaliation.
2. The Shapley index reveals that, in an optimal system, the marginal contribution of the whistleblower is substantial despite their limited individual power.

3. There exists a gap between the optimal investment in trust from the organization's perspective and that which is socially desirable, justifying regulatory intervention to protect whistleblowers.

These conclusions have direct implications for trust management in organizations:

1. **Design of reporting channels**: Channels should be structured to maximize trust and minimize the perceived risk of retaliation.
2. **Allocation of responsibilities**: Responsibilities in managing reports should be distributed in accordance with the marginal contribution of each actor to the overall value.
3. **Incentives for cooperation**: Organizations should implement mechanisms that incentivize cooperation among all actors involved in the whistleblowing process.

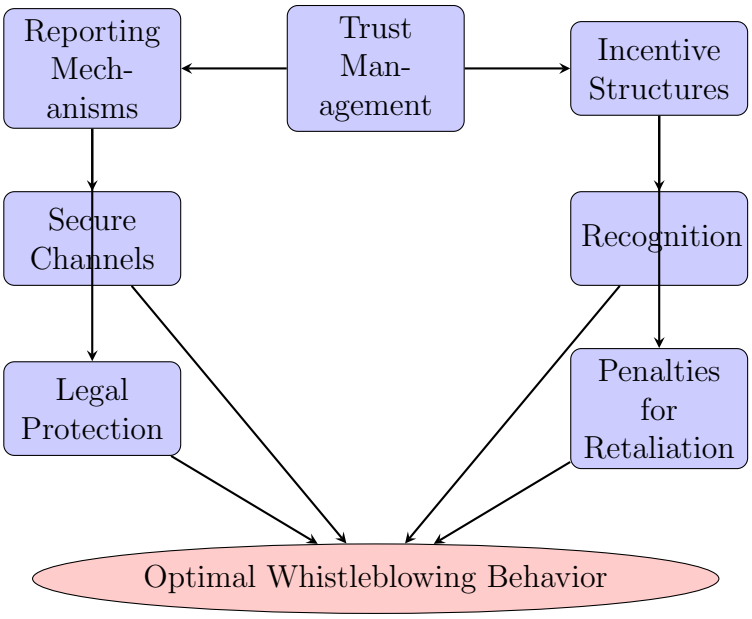

Figure 4: A framework for trust management in whistleblowing systems, derived from game-theoretic analysis.

Ultimately, this article demonstrates how a game theory-based approach can complement existing legal, ethical, and organizational analyses, offering a quantitative perspective on the conditions that favor the emergence of effective and stable whistleblowing systems.

Acknowledgements. This research was supported by the Department of Law, Economics, and Human Sciences at the Universit Mediterranea di Reggio Calabria.

References

[1] Bowie, N., *Business Ethics*, Prentice Hall, 1982.

[2] Carinci, M. T., Whistleblowing in Italy: rights and protections for employees, *WP CSDLE "Massimo D'Antona"*, 106/2014, 2014.

[3] Damiri, G., Il sistema di protezione del whistleblower nel panorama normativo italiano, *Ratio Iuris*, 2023.

[4] Myerson, R. B., *Game Theory: Analysis of Conflict*, Harvard University Press, 1991.

[5] Near, J. P., Miceli, M. P., Organizational dissidence: The case of whistleblowing, *Journal of Business Ethics*, **4** (1) (1985), 1-16. https://doi.org/10.1007/bf00382668

[6] Shapley, L. S., A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games, Vol. 2* (pp. 307-317), Princeton University Press, 1953. https://doi.org/10.1515/9781400829156-012

[7] Sitzia, A., La protezione del "whistleblower" nel settore privato: la legge 179 del 2017 nella prospettiva europea, *LDE*, 2/2019, 2019.

[8] Gundlach, M. J., Douglas, S. C., Martinko, M. J., The decision to blow the whistle: A social information processing framework, *Academy of Management Review*, **28** (1) (2003), 107-123. https://doi.org/10.2307/30040692

[9] Mesmer-Magnus, J. R., Viswesvaran, C., Whistleblowing in organizations: An examination of correlates of whistleblowing intentions, actions, and retaliation, *Journal of Business Ethics*, **62** (3) (2005), 277-297. https://doi.org/10.1007/s10551-005-0849-1

[10] Parisi, N., La funzione del whistleblowing nel diritto internazionale ed europeo, *LDE*, 2/2020, 2020.

[11] Vandekerckhove, W., Lewis, D., The content of whistleblowing procedures: A critical review of recent official guidelines, *Journal of Business Ethics*, **108** (2) (2012), 253-264. https://doi.org/10.1007/s10551-011-1089-1

Received: March 15, 2025; Published: April 6, 2025