Business Ethics Journal Review
SCHOLARLY COMMENTS ON ACADEMIC BUSINESS ETHICS
businessethicsjournalreview.com

Edited by **Chris MacDonald**

ISSN: 2326-7526

TRUST, FIRMS, AND EXPRESSIVE RESPONSIBILITY: A RESPONSE TO POURYOUSEFI AND TALLANT

Tae Wan Kim and Bryan Routledge*

AN INVITED RESPONSE TO Sareh Pouryousefi and Jonathan Tallant (2025), "Trust, the Participant Stance, and Firms," ***Bus Ethics J Rev*** 12(2): 8–13, https://doi.org/10.12747/j1l02

ABSTRACT

We agree with Pouryousefi and Tallant that a participant-stance account of trust requires a clear bearer of reactive attitudes, and that firms do not literally have such attitudes. We argue that this point is compatible with a trust-based right to explanation.

TO TRUST SOMEONE is to be ready to feel resentment, a sense of betrayal, and related emotions if they let us down. When we ask a friend to drive us home, we trust our friend, not the car. We rely on the car, but we have no reason to feel betrayed by it if something goes wrong. We blame the friend, expect an apology, and might even end the friendship if the breach is serious enough. The car is crucial to what happens, but it is not a fitting target of resentment, nor a possible source of apology. It cannot regret anything.

In their Commentary on our work, Pouryousefi and Tallant worry that many cases of the so-called "trusting firms" in our paper look more like trusting the car than trusting the driver. People say that they trust a social media platform, a bank, or a ride sharing company. But when there is a breach of privacy, an unfair algorithmic decision, or some other morally salient failure, it is not obvious whether the fitting

* Carnegie Mellon University (both).
Email: twkim@cmu.edu, routledge@cmu.edu
Discuss this commentary at https://wp.me/p2x7zx-MX
Cite as: **Bus Ethics J Rev** 13(8) (2026): 44–49, https://doi.org/10.12747/j1m08

object of blame and apology is the firm itself, the CEO, a group of engineers, or someone else. If we cannot say who owes regret and remedy, we do not yet have a genuinely participant-stance form of trust.

Humans, institutions, and expressive responsibility

Firms do not feel resentment or remorse. If the participant stance required the trusted party literally to have such feelings, firms could not be trustees at all. However, that conclusion is too strong. There is a distinction between being the psychological subject who has the attitudes and being the locus, at which those attitudes are expressed and, to which they are publicly attributed. Only humans fit the first description. Institutions can fit the second.

Consider what happens when an airline cancels your flight at the last minute due to a scheduling algorithm. At the gate, the staff makes an announcement in the airline's name. Later, you receive an email that begins, "We are sorry for the disruption to your journey," offers an explanation, and provides compensation. Most passengers take the apology to come from the airline itself. They do not care which manager drafted the text. If the offer is grudging or opaque, they say that the airline has not really apologized. If it is generous and candid, they say that the airline has done the decent thing. Here, the underlying attitudes are human, but the institution is the organized channel through which these attitudes are expressed. It is also the addressee of our reactive attitudes. We resent the airline, not the anonymous copywriter. We demand that the airline explain what happened, apologize properly, and change its systems.

Expressive theories of law capture this idea. In Anderson and Pildes' (2000) account, laws and legal decisions do not only change incentives; they also express social meanings such as condemnation, assurance of concern, or equal standing. A criminal conviction expresses censure. An official apology expresses acknowledgment and regret. The state does not need to feel literally anything for its actions to play this expressive role. Firms can function in a similar way. They can adopt stable practices, such as apology letters, internal investigations, policy changes, and forms of compensation, expressing regret and a commitment to repair relationships when trust is breached. Severance pay after a badly managed restructuring is a

familiar example: by offering more than the bare legal minimum, and by explicitly recognizing its own strategic mistakes, a firm may express something like agent-regret for the harm it has helped to bring about. Once this expressive layer is visible, the gap between a human-centered participant stance and firm-level trust may narrow. Reactive attitudes remain human. But firms can still be the organized sites at which these attitudes are expressed and to which our resentment, expectations, and demands for explanation are directed.

Nothing in Pouryousefi and Tallant's Commentary rules out this expressive route.

Finance, AI, and the practice of explanation

A natural question, especially for work at the intersection of finance and AI, is whether this picture helps us to think concretely about financial institutions that rely on algorithmic systems. We think it does. Evidence from financial economics shows that trust in intermediaries and markets is a central determinant of whether households are willing to participate in the stock market at all (Guiso et al 2008). Trust is not an optional moral extra, but a condition for the proper functioning of financial institutions.

Consider automated credit scoring. A bank uses a machine-learning model to decide whether to grant mortgages or consumer loans. From a finance perspective, such models promise efficiency: better risk pricing, lower underwriting costs, and more consistent decisions. From a trust perspective, they create a familiar vulnerability. Customers must hand over sensitive data, know that opaque models will shape life-changing decisions, and are largely unable to monitor the process. The bank asks to be trusted.

In our view, a trust-based right to explanation can be built into this financial relationship in two complementary ways. First, individual decision-makers within the institution, credit officers, model risk managers, executives on risk committees can see themselves as trustees who are personally responsible for the fairness and accountability of the models they approve. They can feel and appropriately express regret and apologize when the system treats someone unfairly. Second, the bank as an institution can adopt stable practices that express those attitudes: clear explanations of adverse

decisions, channels for challenge and review, and explicit policy changes when biases are found.

Similar points apply in other areas where finance and AI meet. Robo-advisors recommend portfolios to retail investors, trading algorithms execute orders in fragmented markets, and fraud-detection systems flag transactions for review. In each case, there is a familiar financial story about information asymmetry and agency costs. Our contribution is to add a trust-based layer. If the institution asks investors and customers to rely on AI systems that they cannot fully understand, then either the humans designing and overseeing those systems or the firm itself acting expressively must be prepared to explain, apologize and offer remedies, where appropriate. In macro terms, this connects to economic theories in which low generalized trust and heavy regulation can become mutually strengthening equilibria, so that institutional design reflects and shapes social trust (Aghion et al 2010).

Two ways to read the trustee

Seen this way, our trust-based right to explanation is compatible with both a human-centric and a firm-centered reading of the trustee. Consider again a typical algorithmic decision. A bank uses a machine-learning model to decide whether to grant us a mortgage. Our application is rejected. We ask why. Who owes us an explanation?

In strict human-centric reading, only individuals can be trustees. The relevant trustees are then the identifiable humans in the bank: the executives who approved the system, the risk officers who monitor it, and perhaps the data scientists who designed it. They benefit from our trust, know that we are vulnerable to their decisions, and can feel and express regret and apology appropriately. The right to explanation is a right against these humans in their professional roles. If we reject expressive institutionalism, this is the natural way to read our view. In a more institutional expressive reading, the trustee is the bank itself. The human inside the bank still has the attitudes, but the bank is the agent that publicly commits to treating customers fairly, that sends the rejection letter, that can change its systems, and that can act in ways that apologize in its own name. It is true in the relevant normative sense that the bank owes us an explanation and an apology when

things go wrong. The bank, not just a particular manager, is the organized bearer of regret and remedy.

In either case, the basic normative conclusion is the same. Those who stand as trustees in data-driven relationships, whether described as individuals in roles or as firms acting expressively, owe those they affect a right to an explanation when their algorithmic decisions go wrong. Pouryousefi and Tallant's challenge forces us to clarify who the trustee is; it does not show that there is no trustee.

What must be proved

We can now state what remains to be shown and where we disagree with Pouryousefi and Tallant. Those authors argue that, in a participant-stance account of trust, it is not clear that firms can be fitting targets and sources of reactive attitudes such as blame, regret, and apology, and that a fully worked-out account of trust in firms is not yet forthcoming. Their strengthened condition, according to which the very *x* we trust must itself be the appropriate source of regret and remedy, puts real pressure on any attempt to treat firms as trustees in the participant sense. On their side, the safe position is that only humans are uncontroversially proper objects of participant-stance trust. On our side, the burden is clear. We must show two things. First, at a conceptual level, we must defend the claim that the participant-stance can be extended to institutions via expressive practices. Second, at a more applied level, we must show that actual firms, especially data-driven financial institutions, do or can adopt the kinds of practices that meet this condition. We accept that the burden is on us.

Received 5 January 2026 / Posted 2 November 2026

References

Aghion, P., Y. Algan, P. Cahuc, and A. Shleifer. 2010. "Regulation and Distrust." ***Q J Econ*** 125(3): 1015–1049, https://doi.org/10.1162/qjec.2010.125.3.1015

Anderson, E. S. and R. H. Pildes. 2000. "Expressive Theories of Law: A General Restatement." ***U Pa L Rev*** 148(5): 1503–1575, https://doi.org/10.2307/3312748

Guiso, L., P. Sapienza, and L. Zingales. 2008. “Trusting the Stock Market.” ***J Finance*** 63(6): 2557–2600, https://doi.org/10.1111/j.1540-6261.2008.01408.x

Pouryousefi, S. and J. Tallant. 2025. “Trust, the Participant Stance, and Firms. ***Bus Ethics J Rev*** 12(2): 8–13, https://doi.org/10.12747/j1l02