Why blame?

Mehment Y. Gurdal, Joshua Miller, Aldo Rustichini

Extended abstract

We provide evidence that subjects in an experimental setup blame others for outcomes for which the others cannot be considered responsible. We then provide an explanation of this apparently irrational behavior on the basis of a principal agent model.

In the experiment a subject acting as agent chooses between a lottery and a safe asset, paying a fixed amount. The payment of the chosen asset goes to a player, who observes the choice of the agent and the outcome, and can then decide how much of a fixed amount to allocate between the agent and an involved third party, who is one of the subjects in the experiment participating as an agent in another pairing. Since no fraction of the amount could be retained by the principal, he has no other monetary incentive to favor one or the other; hence a shift in the allocation of rewards is a reliable measure of the valence of the affective response of the principal to the agent’s choice and the outcome. Since the agent is clearly not responsible for the outcome of the lottery, an allocation norm that obeys a merit principle ("rewards and penalties depend on responsibility") prescribes that payments depend on the choice of the safe or risky asset, but not on the outcome.

We instead observe widespread blame. The money allocated to agents depends on the outcome of the lottery, in a way that seems to suggest that principals “punish” agents for the outcome: when the agent chooses the lottery, payments to him are lower when the outcome is low than when the outcome is high. A similar pattern is observed when the agent chooses the safe asset: payments are lower when the outcome of the lottery is high than when the lottery outcome is low. The affective response of the principal seems to follow a counterfactual evaluation made by the principal, qualitatively similar in nature to regret and envy. He compares the outcome received with the outcome he would have received had the choice of the agent been different. His affective response and payments are monotonic in this difference. Is this behavior irrational?

We claim it is not once it is put in the appropriate prospective. Consider the optimal contract in a delegated expertise principal agent problem. The agent has to put effort to investigate the profitability of assets, and choose for the principal, who collects the outcome of the asset, and pays the agent taking into account the information he has available, namely the choice of the agent and the outcome of the assets. The payments in the optimal contract have typically precisely the pattern followed by the payments of our subjects playing as principals. This is true even if the principal knows, in equilibrium, that the agent is not responsible. More than that: in equilibrium the principal knows that the agent chose the effort, so his payments treat differently and in full awareness two agents who made the same choice. Thus, equal treatment of equals, or fairness, is an irrational “philosophical” position, and the behavior of our subjects is rational. We interpret the results in the light of the theory of Salient Perturbations (Myerson): individuals in a non familiar environment behave in a way which is similar to the best equilibrium behavior in an environment which is more familiar (the Salient Perturbation) than
the one actually in front of them (there is no measure of the environment that is most salient, the best match).