Information Design in the Principal-Agent Problem

YAKOV BABICHENKO, Technion – Israel Institute of Technology, Israel
INBAL TALGAM-COHEN*, Tel Aviv University, Israel
HAIFENG XU, University of Chicago, USA
KONSTANTIN ZABARNYI, Technion – Israel Institute of Technology, Israel

We study a variant of the principal-agent problem in which the principal does not directly observe the agent's effort outcome; rather, she gets a signal about the agent's action according to a *variable* information structure designed by a regulator. We consider both the case of a risk-neutral and of a risk-averse agent, focusing mainly on a setting with a limited liability assumption. We ask the following question – which actions and utility profiles can be implemented by *some* information structure? Surprisingly, even though the principal-agent problem with unobserved outcomes has appeared in previous work, ours is the first work to study the implementability of utility profiles and expected transfers.

We provide a clean characterization for implementability, which turns out to be simple thresholds on the utilities. Specifically, every possible utility profile that exceeds some natural thresholds for the principal and agent can be implemented by some information structure. Furthermore, every action the welfare of which exceeds the welfare of the least costly action can be implemented. These conditions are necessary and sufficient. We further show that this characterization holds for both a risk-neutral agent and a risk-averse agent, but the set of implementable utility pairs under risk aversion becomes significantly richer. The richness of the sets of implementable actions and utility profiles prescribes the power of the regulator with the ability to design the information structure. All our results are constructive, and thus naturally admit polynomial-time algorithms.

Moreover, we consider a natural constrained version of the problem motivated by common limitations on information structures. We identify the set of signals directly as the set of actions $\{1, \ldots, n\}$. For any fixed *noise level* described by an integer $d \geq 0$, if the agent takes action a, the induced signal is assumed to be supported on signals from the set $[a - d, a + d]$ (with obvious truncation if $a - d$ or $a + d$ exceeds the boundary). When $d = 0$, this captures one extreme case in which the principal observes the agent's action, and the regulator choosing the information structure has no flexibility at all. When $d = n - 1$, this captures the other extreme case in which all information structures are available for the regulator. We characterize the set of implementable actions for all intermediate values of d. Specifically, an action is implementable if and only if its welfare is weakly above the minimal welfare of every $2d + 1$ consequent actions. This result nicely demonstrates the efficiency loss that can be caused by the inaccuracy of the principal's information. We also prove that a slightly enriched class of constraints on the information structure – namely, upper-bounding the allowed probability of mapping an action to some signal – would make it NP-complete to decide whether a certain action is implementable.

A full version of this paper can be found at https://arxiv.org/abs/2209.13688.

---

*Also with Technion – Israel Institute of Technology.

Authors' Contact Information: Yakov Babichenko, yakovbab@technion.ac.il, Technion – Israel Institute of Technology, Haifa, Israel; Inbal Talgam-Cohen, inbaltalgam@gmail.com, Tel Aviv University, Tel Aviv, Israel; Haifeng Xu, haifengxu@uchicago.edu, University of Chicago, Chicago, Illinois, USA; Konstantin Zabarnyi, konstzab@gmail.com, Technion – Israel Institute of Technology, Haifa, Israel.

Yakov Babichenko, Inbal Talgam-Cohen, Haifeng Xu, and Konstantin Zabarnyi

CCS Concepts: • **Theory of computation → Algorithmic game theory and mechanism design**.

Additional Key Words and Phrases: Contract design, Information design

Acknowledgments