INCENTIVES: A MULTIDISCIPLINARY APPROACH

MARCIN DZIUBIŃSKI

1. Research Project Objectives

When individual rational agents make choices, they follow their preferences over outcomes of these choices. Their incentives are to make choices that lead to their most preferred outcomes. When such choices are made in an environment consisting of many agents, collections of individual choices lead to outcomes that affect all the agents. These outcomes may or may not be desirable, be it from the perspective of the welfare of all the agents or from the perspective of some external entity (as social planner) that would like to ensure that, depending on the preferences of the agents, only outcomes from a particular set are obtained. This requires creating incentives for the agents to make the right choices. When individual preferences of the agents are not known to the social planner, she can follow a mechanism design approach and create a mechanism consisting of sets of choices (messages) for each of the agents as well as the outcome function that connects the individual choices to the outcomes. The goal is to create incentives for the agents to make choices that will map their true preferences to the desired outcomes. Often the agents are embedded in a social network and the shape of this network (its topology) affects how choices of the agents result in different outcomes. This, in turn, affects agents incentives to make choices that lead to more or less desirable outcomes. The right features of network topologies may result in better outcomes. The main focus of the proposed project will be on some of the new frontiers of mechanism design as well as on designing the topology of social networks in order to create incentives for the agents to make choices that lead to more desirable outcomes.

Very recent papers by Kattwinkel (2020), Pereyra and Silva (2020), and Bloch et al. (2021) show, in specific settings where monetary transfers are not available, that if the social planner has access to information from unbiased outside source that is positively correlated with agents' private information then this can be used by the planner to his advantage. More precisely, in a principal-agent model, Kattwinkel (2020) shows that the principal can exploit the correlation in costs and values to design a direct mechanism that induces truth-telling and allows the principal to decide on allocation of an indivisible project, the cost of which he knows, to the agent that improves upon a constant mechanism that ignores the information about the value of the project reported by the agent. Pereyra and Silva (2020), inspired by the problem of assigning students to colleges, consider a setting where colleges receive an imperfect signal correlated with the students' abilities. They show that access to imperfect signals can improve the allocation mechanism. In Bloch et al. (2021), we show how access to an informative signal from less informed external referee can improve selection of a winner from a group of better informed agents reporting on their own quality.

All these papers show that there exist settings of mechanism design without transfers, in which access to an external, imperfect, signal can be used by the social planner to design better mechanisms. How general are these results? What are the factors that affect the extent of this improvement? What are

Date: September 14, 2021.

the mechanisms that allow for the best use of the external information? Answers to these questions are largely unknown. Motivated by them, we plan to study mechanism design without transfers and with external source of information, aiming to generalize existing results and obtain a better understanding of when and how this external information can be used best.

The research mentioned above is related to the older literature on allocation without transfers but with costly verification Ben-Porath et al. (2014), Mylovanov and Zapechelnyuk (2017) and Li (2020), where the social planner can verify types of agents ex post at some cost. Like in the case of access to an external signal ex ante, this allows for improving the quality of allocation. However, the results and methods are different between those to strands of literature. A more comprehensive study of scenarios with access to external information ex ante will help to understand their relation to scenarios with access to external information ex post.

Another topic in mechanism design that gains interest recently are mechanisms in social networks. In this setting agents are embedded in a social network and connections in the network allow them to share some (or all) their private information with their neighbours. A simple example of such a private information that is shared in social networks are the links in the networks. In Dziubiński et al. (2016) we studied the problem of optimal protection (e.g. vaccination) allocation in a network which is unknown to the social planner and where protection is scarce and is demanded by all the agents. Not allowing for monetary transfers but allowing for waste, we proposed a dominant strategy incentive compatible mechanism that allows for optimal protection allocation even though each agent knows his neighbours in the network only and does not know the topology of the whole network. Baumann (2018) studies prize allocation mechanisms without monetary transfers in a setting where agents are embedded in a social network and know their own value as well as values of their neighbours. The social planner wants to allocate a single prize to the most valuable agent. She proposes a mechanisms which, depending on the network topologies, fully implement the planner's objective. Most recently, Bloch and Olckers (2019) studied a mechanism for extracting a ranking of agents based on a single individual characteristic. Agents are embedded in a social network, which allows them to observe a local ranking, that is the ranking including them and all their neighbours. Assuming that each agent cares only about its own rank in the global ranking, Bloch and Olckers (2019) propose an ex post incentive compatible and ex post efficient mechanism that allows the social planner to extract the global ranking in network topologies where every two agents have a common neighbour. The also show that this feature of network topology is necessary for ex post incentive compatibility and efficiency in their setting. In the following paper Bloch and Olckers (2020) they provide an experimental evaluate of their mechanism. Lastly, in a related classic result, Saijo (1988) shows that full Nash implementation is possible in a setting where agents are embedded in a directed cycle where each agent observes her own and her successor's preferences.

The papers described above exploit the fact that agents in a social network share some relevant information with their neighbours. They all show that, depending on the network topology, it is possible to construct mechanisms that implement the desire objectives. However, in all the three cases there are no externalities: the agents care about their own outcome only. We plan to investigate more settings of implementation where agents are embedded in a social network. In particular, we would like to obtain a better understanding the connection between network topologies and implementability of different social choice functions or correspondences. We would also like to investigate the problem of mechanism design in networks in settings where externalities are present.

Very little is known about automated design of mechanisms implementing given social choice rules. Sjöström (1991) presented an algorithm for checking Nash implementability and designing implementing mechanism for given social choice correspondences. Sang-Chul (1996) proposed an algorithm for strong Nash implementation. This problem is closely related to the problem of obtaining small mechanisms implementing given social choice correspondences or functions. In recent years, researchers in mechanism design addressed the problem of simplicity of mechanisms (Li (2017), Pycia and Troyan (2021)) which focuses on the simplicity of decision making of agents participating in the mechanism. A different aspect of simplicity is related to mechanism size or the size of mechanism description. This problem is closely related to automated design of mechanism. We plan to study the question of automated design of small mechanism not only for the settings where finite mechanisms are possible but also for the settings where it is known that infinite mechanisms are necessary (Palfrey and Srivastava (1991); Jackson (1992); Dutta and Sen (1994)). In the infinite cases, it is possible that although a mechanism must be infinite it may admit a finite representation (e.g. in the form of a regular expression). Is such a finite representation possible for the cases where it is known that infinite mechanisms are necessary? Is it possible to automatically obtain such mechanisms with minimal or close to minimal such representation? These questions we would like to address as well.

A different problem of creating incentives arises in scenarios where preferences of agents are known but where agents, embedded in a social network, engage in local interactions that affect welfare of the population globally. Such local, self-oriented interactions lead to global effects that affect the welfare of all the agents in the network. This often lead to inefficiencies, as compared to the centralized variant, where the choices are made by a social planner. The reasons for these inefficiencies may be many: restricted interaction structure may lead to nodes failing to coordinate, payoffs of different nodes may depend positively or negatively on the choices made by other nodes (network externalities), lastly, the strategic choices of different nodes may depend positively or negatively on the choices of other nodes (strategic complementarities). Can the right choice of this topology reduce this cost? How does it depend on the type of local interactions?

For example, consider a setting where agents decide on individual protection (e.g. taking a vaccine or taking security measures in a computer network) that is costly and that provides not only individual protection but also helps protecting others in the network. In such a setting decentralized decisions made by strategic agents are known to be inefficient and these inefficiencies can be significant (Kunreuther and Heal (2003),Varian (2004), Acemoglu et al. (2016)). We addressed this problem in Cerdeiro et al. (2017) and showed that the inefficiencies can be largely mitigated by choosing the right network topology. In particular, this design can create a cascade of incentives to protect, which addresses the problem of underprotection.

We shall study the problem of interaction design. In particular, we would like to systematically examine how different types of dependencies between strategic choices of the players (strategic substitutes, strategic complements) affect the effectiveness of the interaction design. The ultimate goal would be to learn how effective can interaction design be depending on the properties of the local interactions that the individuals engage into.

Our key objectives in the project are as follows:

(1) To study problems of mechanism design with external source of information in order to understand to what extent and how can an external imperfect signal be used by the social planner to design better mechanisms.
(2) To study problems of mechanism design in social networks in order to obtain a better understanding the role of network topologies and network externalities in implementability of social choice functions or correspondences under different concepts of implementation.
(3) To study automated mechanism design and small mechanism in order to propose algorithms that construct mechanisms which are either small or have small representation.
(4) To study the problem of interaction design in order to understand how different types of network effects affect the effectiveness of such a design.

2. Significance of the project

Research on mechanism design and implementation is among the main fields of research in economics for more than 60 years now, since the seminal work of Hurwicz (1959). The field of networks, although much younger in economics, is well established and very active. We aim to contribute to the most recent research efforts in both these fields by studying mechanism design with external sources of information and mechanism design in networks. We also plan to study a largely unexplored area of automated mechanism design. Lastly, in the fourth of our research objectives we plan to develop a novel idea of interaction design which provides a new tool for creating incentives among connected agents.

Our approach is going to be multidisciplinary. We plan to combine methods of theoretical economics, methods and tools from computer science, as well as rigorous methods of economic experiments. In particular, computational tools can help in designing mechanisms in complex setting such as the settings of mechanism design with external sources of information or mechanism design in networks. For example, we successfully used a computer program to help us obtaining the full characterization of optimal mechanisms in the problem of winner selection with external referees in Bloch et al. (2021). The problem of automated mechanism design will also benefit from methods and concepts of theoretical computer science.

Experimental approach to test and evaluate mechanisms and other methods of creating incentives among self-interested agents is still largely underdeveloped (with an exception of auctions, which are widely used and studied, also experimentally). There are examples of the experimental approach in the literature, like Sefton and Yavas (1996), Cabrales et al. (2003) and Chen (2005), but examples are still scarce. An exception is the recent paper by Bloch and Olckers (2020) who employ experiments to test their mechanism from Bloch and Olckers (2019). We plan to follow this approach and use experiments to test the mechanism that we obtain through theoretical research. In particular, we plan to apply this approach to testing small mechanisms obtained by the automated approach.

To summarize, we plan to contribute to the state of the art research in mechanism design and networks and to develop and use methods and tools from computer science as well as to contribute to experimental evaluation of mechanisms and interaction topologies. Because of important applications of mechanisms and other methods for creating incentives (like winner selection, prize allocation, decentralized protection/vaccination), the significance of the objectives exceeds the mere theoretical interest.

3. WORK PLAN

We plan to start with objectives (1) and (2). In the case of objective (1), we plan to start by building on our earlier work Bloch et al. (2021) and extend it to address the problem if winner selection in the cases where qualities of the agents are multivalued, beyond binary *high* and *low*. We will also attempt to study the extension where each agent observes own quality as well as the qualities of two or more different agents and reports all of them. This extension connects objectives (1) and 2. We also plan to research other concrete scenarios of mechanism design with external information. In the later phase of the project we hope to build on the results in these concrete settings as well as on the scenarios studied in the literature in order to attempt general results concerning mechanism design with external information.

In the case of objective (2), we plan to start with a model, possibly extending our earlier work Dziubiński et al. (2016), that would address the problem of optimal prize allocation in networks where the network topology is unknown or only partially known to the social planner. One such problem that we plan to consider is the problem of rewards allocation in context of viral marketing, where the social planner aims to choose and rewards a subset of agents as seeds of new ideas or products. We also plan to propose and study a model of prize allocation in networks where agents care not only about receiving the prize themselves but also benefit from other agents (e.g. their neighbours) receiving the prize. In later stages of the project we plan to propose and study similar such problems with hope of obtaining insights that would lead to general results and characterizations.

When work on objectives (1) and (2) is sufficiently advanced, possibly at the end of the second year of the project, we plan to start working on objectives (3) and (4). In the case of Objective (3), we plan to start with studying automated design of mechanisms for Bayesian implementation as well as for Nash implementation. As shown in Dutta and Sen (1994), Bayesian implementation may require infinite mechanism. We first plan to investigate whether automated design of such mechanisms is possible in general (i.e. whether the problem if Bayesian implementability is decidable) and, if it is possible, whether such mechanisms always allow finite representation. In the case of positive answer to both these questions, we plan to propose algorithms for constructing mechanisms with small representation. In parallel to that study, we plan to investigate automated construction of finite mechanisms for Nash implementation with finite environments. Sjöström (1991) proposed an algorithm for constructing mechanisms for Nash implementation. However, this algorithm is not designed to minimize the size of the constructed mechanisms. We plan to propose algorithms that would construct mechanisms with minimal size. After completing these two tasks, we plan to use insights from them to address automated design of mechanisms for other solution concepts as well as for truthful instead of full implementation.

In the case of objective (4), building on experience with our earlier paper Cerdeiro et al. (2017), we plan for formulate the general problem of interaction design. We expect that this will lead to formulation of theoretical results characterising the connection between different features of bilateral interactions with features of outcomes that can be obtained by choosing the right interaction topology. Work on this point will most likely require studying additional, practically motivated, concrete problems of interaction design, different to the one studied in Cerdeiro et al. (2017). We plan work on such problems as well.

In parallel to theoretical work on objectives (1)–(4), research on experimental evaluation of mechanisms and interaction topologies will be conducted. This research will be conducted by a member of the team with experience in experimental economics. The research will start with designing experiments to evaluate mechanisms in our earlier theoretical works Bloch et al. (2021) and Dziubiński et al. (2016).

With progress in work on theoretical part of the research objectives, we plan research on experimental evaluation of the mechanisms obtained there.

3.1. **Risks analysis.** As for the theoretical research, we did some very preliminary study related to objectives (1) and (2) and we estimate the associated risk as intermediate. We also conducted some preliminary analysis of the first task in objective (3), automated design of mechanisms for Bayesian implementation, and we expect that deciding Bayesian implementability is decidable. The remaining questions are challenging and lack similar research in the literature. Therefore we estimate the associated risk as high. Similarly, objective (4) is a novel idea which may, potentially, turn out to be very challenging. Therefore we estimate the associated risk as high.

As for the experimental research, although the laboratory experiments have their own limitations, this component of the project is of low risk. We start with dealing with the reluctance of the respondents to take part in experiments and the possible homogeneity of the obtained sample. To combat this, the recruitment of participants will take place through the ORSEE run by Laboratory of Experimental Economics at the University of Warsaw. This database currently has more than 2000 diverse users interested in participating in economic experiments. We are aware that the experiments on auctions might be easier to comprehend for participants than the experiments evaluating effectiveness of mechanisms. In the later ones, the task is often artificial and unfamiliar to respondents, which is raised as one of the main risks in the literature Dickhaut et al. (2013). Relatively low stakes for non-trivial tasks may not be an sufficient incentive for participants. To mitigate those risks, we will design the experiments in compliance with the good practices in the experimental economics. Fortunately, the team members have prior experience in conducting experiments; they pilot experiments on auctions or mechanism design in their classroom activities to get feedback. To ensure that participants understand the tasks, before the experiment starts, we provide a few exercise turns. Moreover, the planned remuneration for participants reflects the estimated difficulty of the task. The expected value of the reward was suggested based on the distribution of the rewards for ORSEE participants (their opportunity costs).

4. Research Methodology

The purpose of our project is deepening and extending knowledge and understanding of methods of mechanism design and network design for creating incentives among rational agents. The main objectives are designing mechanisms, network topologies, and algorithms for automated design of mechanisms as well as developing methods of experimental evaluation of such mechanisms. Therefore the outcomes of our research will be a combination of formal results and experiments.

For formal results, we shall use a variety of techniques from theoretical economics (mainly from implementation theory, mechanism design and game theory), computer science (mainly algorithms and theory of computation), and mathematics (linear algebra, optimization, and probability theory).

Given that it could be a hard taks, for an outside observer, to validate the effectiveness of mechanisms on the basis of observations from, e.g., enterprises or any other real-life environments, to assess the effectiveness of the mechanisms, we will conduct laboratory experiments. We expect that our laboratory experiments will be individual choice tasks. We will follow a mixed-method approach. From the statistical point of view, the experiments will supply us with grouped results. That is why we expect that comparisons of distributions (e.g., Kolmogorow-Smirnov tests), relationships in medians (Wilcoxon

tests), correlations (Spearman, Kendall coefficients), linear regressions, discrete choice models or social network analysis will be sufficient to validate the effectiveness of the mechanisms. We suppose the samples may violate assumptions for the parametric techniques, that is why we suggest usage of non-parametric alternatives. However, it it hard to predict in advance the exact statistical models we will use; we will refine them after examining the characteristics of data samples.

We plan to conduct our research in collaboration with research units abroad: University of Warwick, Paris School of Economics, Indian Statistical Institute (Delhi Center), and New York University Abu Dhabi.

REFERENCES

D. Acemoglu, A. Malekian, and A. Ozdaglar. Network security and contagion. *Journal of Economic Theory*, 166:536–585, 2016.

L. Baumann. Self-ratings and peer review, 2018. Working paper.

E. Ben-Porath, E. Dekel, and B. Lipman. Optimal allocation with costly verification. *American Economic Review*, 104(12):3779–3813, 2014.

F. Bloch and M. Olckers. Friend-based ranking, 2019. forthcoming in American Economic Journal: Microeconomics.

F. Bloch and M. Olckers. Friend-based ranking in practice, 2020. forthcoming, American Economic Review.

F. Bloch, B. Dutta, and M. Dziubiński. Selecting a winner with external referees, 2021. Working paper.

A. Cabrales, G. Charness, and L. Corchón. An experiment on Nash implementation. *Journal of Economic Behavior and Organization*, 51:161–193, 2003.

D. Cerdeiro, M. Dziubiński, and S. Goyal. Individual security, contagion, and network design. *Journal of Economic Theory*, 170:182–226, 2017.

Y. Chen. Dynamic stability of Nash-efficient public goods mechanisms: Reconciling theory and experiments. In A. Rapoport and R. Zwick, editors, *Experimental Business Research*, pages 185–200, Boston, MA, 2005. Springer US.

J. Dickhaut, D. Houser, J. Aimone, D. Tila, and C. Johnson. High stakes behavior with low payoffs: Inducing preferences with Holt–Laury gambles. *Journal of Economic Behavior and Organization*, 94 (C):183–189, 2013.

B. Dutta and A. Sen. Bayesian implementation: The necessity of infinite mechanisms. *Journal of Economic Theory*, 64:130–141, 1994.

M. Dziubiński, P. Sankowski, and Q. Zhang. Network elicitation in adversarial environment. In Q. Zhu, T. Alpcan, E. Panaousis, M. Tambe, and W. Casey, editors, *Decision and Game Theory for Security*, pages 397–414, Cham, 2016. Springer International Publishing.

L. Hurwicz. Optimality and informational efficiency in resource allocation processes. In K. Arrow, editor, *Mathematical Methods in the Social Sciences*, pages 27–46. Stanford University Press, Stanford, 1959.

M. Jackson. Implementation in undominated strategies: A look at bounded mechanisms. *The Review of Economic Studies*, 59:757–775, 1992.

D. Kattwinkel. Allocation with correlated information: Too good to be true. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 109–110, 2020.

H. Kunreuther and G. Heal. Interdependent security. *Journal of Risk and Uncertainty*, 26:231–249, 2003.

S. Li. Obviously strategy-proof mechanisms. *American Economic Review*, 107:3257–3287, 2017.

Y. Li. Mechanism design with costly verification and limited punishments. *Journal of Economic Theory*, 186:105000, 2020.

T. Mylovanov and A. Zapechelnyuk. Optimal allocation with ex post verification and limited penalties. *American Economic Review*, 107(9):2666–94, 2017.

T. Palfrey and S. Srivastava. Nash implementation using undominated strategies. *Econometrica*, 59: 479–502, 1991.

J. Pereyra and F. Silva. Optimal assignment mechanisms with imperfect verification. *Documento de Trabajo/FCS-Decon; 04/20*, 2020.

M. Pycia and P. Troyan. A theory of simplicity in games and mechanism design, 2021. Working paper.

T. Saijo. Strategy space reduction in maskin's theorem: Sufficient conditions for nash implementation. *Econometrica*, 56:693–700, 1988.

S. Sang-Chul. An algorithm for checking strong nash implementability. *Journal of Mathematical Economics*, 25:109–122, 1996.

M. Sefton and A. Yavas. Abreu-Matsushima mechanisms: Experimental evidence. *Games and Economic Behavior*, 16(2):280–302, 1996.

T. Sjöström. On the necessary and sufficient conditions for nash implementation. *Social Choice and Welfare*, 8:333–340, 1991.

H. Varian. System reliability and free riding. In J. Camp and S. Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, pages 1–15, Boston, MA, 2004. Springer US.