Games and Economic Behavior ••• (••••) •••–•••

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match

Alex Rees-Jones [1]

The Wharton School, University of Pennsylvania, United States

ARTICLE INFO

Article history:
Received 10 January 2016
Available online xxxx

JEL classification:
C78
D03

Keywords:
Matching
Deferred acceptance algorithm
Suboptimal behavior

ABSTRACT

Strategy-proof mechanisms eliminate the possibility for gain from strategic misrepresentation of preferences. If market participants respond optimally, these mechanisms permit the observation of true preferences and avoid the implicit punishment of market participants who do not try to "game the system." Using new data from a flagship application of the matching literature—the medical residency match—I study if these potential benefits are fully realized. I present evidence that some students pursue futile attempts at strategic misrepresentation, and I examine the causes and correlates of this behavior. These results inform the assessment of the costs and benefits of strategy-proof mechanisms and demonstrate broad challenges in mechanism design.

A substantial literature in economics has explored mechanism design in two-sided matching markets. The defining characteristic of these markets is the need to accommodate the preferences of the two groups being matched—for example, when matching students to schools. Compared to the one-sided markets more commonly studied, these settings pose unique challenges to reaching desirable outcomes. Difficulty in coordinating on the timing of decisions often leads to "market unraveling" (Roth and Xing, 1994). Furthermore, decentralized approaches often result in unstable matches,[2] which have been empirically shown to be detrimental to the success of these markets (Roth, 1990, 1991). These problems can be avoided by employing a stable matching mechanism to assign a binding match based on preferences reported to a neutral intermediary at an agreed-upon time. However, the use of these mechanisms introduces the new challenge of managing the strategic incentives involved with preference reporting. If market participants can benefit from misrepresenting their preferences, we expect them to do so.

The student-proposing deferred acceptance algorithm (DAA) of Gale and Shapley (1962) provides a partial solution to the issue of strategic misreporting. For students, this mechanism is *strategy-proof*: truthful preference reporting is a weakly dominant strategy (Dubins and Freedman, 1981; Roth, 1982). Furthermore, truth-telling is approximately optimal for all market participants in sufficiently large markets (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Azevedo and Budish, 2013). Strategy-proof mechanisms therefore provide a comparatively simple optimal strategy, which has been viewed as

E-mail address: alre@wharton.upenn.edu.

[1] I thank Aaron Bodoh-Creed, Dan Benjamin, Ori Heffetz, Judd Kessler, Miles Kimball, Ulrike Malmendier, Sean Nicholson, Ted O'Donoghue, Ran Shorrer, and seminar participants at the AEA annual meetings, UC Berkeley, the CHIBE/Roybal Retreat, Cornell University, George Mason University, The Stanford Institute for Theoretical Economics, and the University of Pennsylvania for helpful comments and advice. I thank Allison Ettinger and Matt Hoffman for their help in recruiting schools to the survey, Andrew Sung for excellent research assistance, and the National Institute on Aging (grant T32-AG00186) for generous research support.

[2] That is, matches in which a pair of agents both prefer to be assigned to each other instead of their realized pairing, or where a matched individual prefers to be unmatched.

http://dx.doi.org/10.1016/j.geb.2017.04.011
0899-8256/

especially useful in the student-to-school matching setting. If optimal play is pursued, students may entirely avoid devoting time or effort into figuring out how they should misrepresent their preferences. Students with a poor grasp of game theory are not punished for their failure to optimally "game the system," resulting in a level playing field between strategically sophisticated and strategically unsophisticated market participants (Pathak and Sonmez, 2008). These features, along with other desirable theoretical properties of the student-proposing DAA, have led a number of prominent market designers to assist in deploying this mechanism to the field (Roth and Peranson, 1999; Abdulkadiroğlu et al., 2005a, 2005b).

This paper explores empirically whether the benefits of strategy-proof mechanisms are fully realized. The typically expressed logic suggests that incentivizing the simple truthful reporting strategy will lead to truthful reports. However, even though the optimal strategy in the student-proposing DAA is simple, the strategic environment remains quite complex. In order to deduce the optimal strategy in this environment, students must draw upon a significant degree of game-theoretic sophistication. Among unsophisticated students, failures of optimal behavior might arise.[3] Just as an otherwise-able student might misunderstand the strategic incentives faced in a non-strategy-proof mechanism and fail to optimally "game the system," so too might a student do so in a strategy-proof mechanism. In this environment, the result would be misrepresentation of preferences despite the lack of scope for successful manipulation.

In this paper I document the existence and nature of this suboptimal behavior in a classic setting from the matching literature: the process matching medical students to medical residencies. Analyzing a survey I administered to graduating medical students at 23 medical schools, I find that 17% of students self-assess their preference reporting strategy to be nontruthful, with 5% directly attributing this nontruthful behavior to strategic considerations. To validate these self-reports, I demonstrate that proxies for welfare are less predictive of the submitted preferences of students reporting nontruthful behavior, consistent with a disruption of utility maximization. All else equal, pursuit of strategic misrepresentation is more prevalent among men, among those with lower academic performance, and among those in more competitive specialties.

A growing literature in experimental economics has examined individual behavior in the DAA, and commonly finds a fraction of respondents with nontruthful reporting behavior (see, e.g., Chen and Sönmez, 2006; Pais and Pintér, 2008; Calsamiglia et al., 2010; Klijn et al. 2013; Ding and Schotter, 2015; Featherstone and Niederle, 2016). However, extending the study of this behavior outside of a controlled laboratory environment is challenging. While true preferences may be controlled or assigned—and thus observed—in the lab, the inability to observe true preferences is a defining characteristic of the field settings in which these matching mechanisms are deployed.[4] The validated self-classification approach presented in this paper offers a rare demonstration that failures of truthful reporting persist outside of the lab. These results complement the concurrent work of Hassidim et al. (2016), who study the 2014 roll-out of a DAA matching mechanism in the Israeli psychology match. The authors find that submitted preferences commonly rank an unfunded position higher than the exact same position with funding. Under the reasonable assumption that students prefer more money to less, this finding implies a high lower bound on the rate of suboptimal preference reporting in this nascent matching market. Taking our results together, Hassidim et al. (2016) demonstrate that substantial misunderstanding of optimal play exists when these mechanisms are first deployed, whereas the results presented here demonstrate that this misunderstanding persists even after decades of institutional history and refinement of training interventions. In summary, failures of optimal play persist in perhaps the most well-studied and carefully designed two-sided matching market currently in existence.

Beyond their implications specific to two-sided matching, these results permit a broader assessment of the limits of incentive compatibility. Economists commonly assume that optimal play can be expected when market participants are sufficiently intelligent, when sufficient understanding of the decision-making environment is developed, and when stakes are sufficiently high and outcomes are sufficiently scrutinized.[5] The population considered in this paper is far more educated than most, is acting in a setting with advice readily available and long institutional history with this mechanism, and is extremely invested in the outcome that this algorithm determines. On one hand, the low rate of nontruthful reporting found may be interpreted as a success: most participants appear to respond to incentives as they should. However, the persistence of suboptimal behavior in this setting, even at low rates, suggests the requisite levels of intelligence, information, and incentivization needed to ensure full compliance may never be achieved in practice. Some strategic misunderstanding may be unavoidable in these settings, necessitating attention to the comparative performance of mechanisms in the presence of suboptimal behavior, and to the design of mechanisms that can minimize misunderstanding.

This paper proceeds as follows. In section 1, I provide institutional details about the residency match and discuss the survey data collected for this paper. Sections 2.1 and 2.2 present main results, and 2.3 addresses several robustness concerns. Section 3 concludes by discussing the implications of these results for the practical deployment of matching mechanisms.

[3] However, a lack of sophistication does not necessarily result in suboptimal reporting. Even absent an understanding of the mechanism, truth-telling could arise from, e.g., moral considerations, reliance on correct advice, or the utilization of truth-telling as a default strategy when the optimal strategy is not understood.

[4] Indeed, if true preferences were observed, designing a matching mechanism to incentivize truthful reporting would be unnecessary.

[5] For discussions of this line of logic (as it applies to interpreting and contrasting lab and field experiments) see Levitt and List (2006, 2007, 2008).

Please cite this article in press as: Rees-Jones, A. Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. Games Econ. Behav. (2017), http://dx.doi.org/10.1016/j.geb.2017.04.011