

The Mechanism is Truthful, Why aren't You?

Avinatan Hassidim, Déborah Marciano, Assaf Rommm,
and Ran I. Shorrer*

January 16, 2017

Traditionally, *strategy-proofness* is considered an extreme form of strategic simplicity. When a strategy-proof mechanism is in place, honesty is the best policy: no matter what actions others are taking, the best course of action is to report one's true preferences. For this reason, such mechanisms are often referred to as *truthful*, and the preferences that are reported to them are interpreted at face value.

Truthfulness is thought to reduce the costs of strategizing, promote equity (by not giving advantage to more sophisticated players), provide robustness, eliminate the costs associated with the collection of information on others, and simplify the interpretation of reported preferences.¹ These desirable features led many centralized markets, especially in education and entry-level labor markets, to adopt truthful mechanisms. More specifically, the mechanisms in use are often based on the applicant-proposing version of the Deferred Acceptance (DA) algorithm, which is strategy-proof for applicants.

But do truthful mechanisms actually induce truthful reporting? Recent evidence that emerged from the Israeli Psychology Master's Match (IPMM) strongly indicates that the assumption of truthful reporting is false. Additional evidence from the field and from the lab suggests that this phenomenon of preference misrepresentation is pervasive. A recurring finding is that misinterpretation rates are higher in weaker segments of markets. This motivates us to further investigate who are the individuals that misrepresent their preferences under DA and what drives this behavior.

Our hope is that our findings will inform market designers and policy makers about the prevalence of misrepresentation and its systematic nature. A better understanding of when and why individuals misrepresent their preferences can guide these practitioners in designing specific and targeted interventions to promote truthful reporting. For example, if members of particular groups are more likely to (err and) misrepresent their preferences, then this behavior may have

*Hassidim: Bar-Ilan Univ., Ramat Gan, Israel, avinatanh@gmail.com. Marciano: Hebrew Univ., Jerusalem, Israel, davorah.marciano@gmail.com. Romm: Hebrew Univ., Jerusalem, Israel, assaf.romm@mail.huji.ac.il. Shorrer: Penn State Univ., State College, PA 16801, USA, rshorrer@gmail.com. Hassidim is supported by ISF grant 1394/16. Marciano is supported by the Azrieli Foundation. Romm is supported by a Falk Institute grant and ISF grant 1780/16.

¹See Roth (2008) and Pathak and Sönmez (2008).

equity implications, and targeted interventions may be advisable. If misrepresentation increases when applicants expect to face stronger competition (as we indeed find), then releasing information about affirmative action interventions may amplify their effect.

Understanding the causes of misrepresentation can also inform the interpretation of choice data, which, in turn, influences policy. For example, if students tend to rank schools higher where they have priority, and students have priority in schools in their neighborhood (which is a typical case), then a naïve interpretation of students’ reported preferences would overstate the importance that students attach to studying in close proximity to where they live.

In what follows, we review existing evidence of preference misrepresentation under DA and its correlates, and evaluate possible explanations as to why individuals do not report truthfully.

1 Preference Misrepresentation

Preference misrepresentation under DA has been studied both in the field and in the lab. In the field, stakes are typically high, information is abundant, and applicants give the matching process their full attention. However, evaluating the truthfulness of reported preferences in the field is an inherently difficult task. Alternatives are often not objectively ranked, and any preference report could be reasonable under some circumstances. That is, even if most participants prefer alternative X to alternative Y , a participant ranking Y as his first choice need not be untruthful — he might simply have different preferences.

In unique settings, however, alternatives *are* objectively ranked, thus allowing one to detect evident preference misrepresentations. Hassidim, Romm and Shorrer (2016) study one such setting. They use data from the IPMM, a matching-with-contracts environment (Hatfield and Milgrom, 2005) in which applicants express their preferences over combinations of study track and contractual terms such as funding and degree type (Master’s or Ph.D.).² As a result of this feature, in some cases applicants were asked to report their preferences over naturally ranked contracts, i.e., contracts that offered admission to the exact same study track, but in one case students would receive a prestigious no-strings-attached scholarship, and in the other case they would not.³ Hassidim, Romm and Shorrer examine applicants’ reported preferences, given as Rank Ordered Lists (ROLs), and dub ROLs that contradict the natural ranking *obvious misrepresentations*. In such ROLs, the funded contract may not appear on the ROL (*obvious dropping*) or may be ranked lower than the unfunded contract (*obvious flipping*). They report that out of 704 relevant ROLs, 19.5% obviously

²This feature is necessary to assure that the mechanism has good properties such as stability and strategy-proofness, as contractual terms could potentially affect applicants’ preferences over programs (see, e.g., Sönmez, 2013). For details on the IPMM see Hassidim, Romm and Shorrer (2017).

³By contrast, if there is stigma attributed to funding, as in Aygün and Turhan (2016), contracts are not naturally ranked.

misrepresented the applicant’s true preferences, with obvious dropping and obvious flipping each accounting for about one half of these observations. Obvious misrepresentation was negatively correlated with various measures of cognitive ability and desirability to programs.

Importantly, the rate of obvious misrepresentation is only a lower bound for the prevalence of preference misrepresentation, as applicants may also misrepresent their preferences over programs. In a post-match survey, 20.2% of the respondents reported misrepresenting their preferences over programs. The vast majority of these applicants had not submitted an obvious misrepresentation.

Additional evidence comes from the National Resident Matching Program (NRMP), the clearinghouse through which physicians find their first job in the U.S.⁴ Rees-Jones (2015) conducted a survey among medical school seniors who applied for a position through the NRMP. Seventeen percent of the respondents reported that their ROL did not represent their true preferences, with about one third reporting that the reason for their behavior was “strategic.” Rees-Jones also documents a negative correlation between misrepresentation and academic achievement, and a positive correlation between misrepresentation and the competitiveness of the residencies being applied to.

In the 2015 NRMP applicant survey, which covers the entire population of applicants, 5% of U.S. medical seniors and 17% of independent applicants reported that they “ranked the programs based on the likelihood of matching (*most likely first, etc.*)” Similarly, only 75% of independent applicants reported that they had ranked the programs in the order of their preferences, while the number among U.S. seniors is 92%. The independent applicant category comprises the weaker segments of the market.⁵

Other evidence comes from school choice settings. In Mexico City, a centralized process is used to assign more than 200,000 students to public high schools (11th and 12th grade) every summer. Seats are allocated through a variant of Serial Dictatorship,⁶ using student ROLs, which are reported in March, and with priorities based on performance in a centralized screening exam held in June. After students are assigned, the minimal score that was sufficient to gain admission to each program is made public.

Chen and Pereyra (2016) compare reported preferences with responses to a voluntary survey held in conjunction with the submission of ROLs. They find that at least 20% of the students in their relevant sample submitted an ROL that appeared contradictory to their survey responses: they declared that they wished to attend a certain university, but did not rank first any of the elite high schools that guarantee admission to this institution. Chen and Pereyra interpret this behavior as preference misrepresentation, and report that it is more common among academically weaker applicants and among applicants of

⁴The NRMP famously uses a variant of the doctor-proposing DA (Roth and Peranson, 1999).

⁵The independent applicant category includes prior graduates (cf. seniors) of allopathic medical schools, students and graduates of international medical schools, and students and graduates of schools of osteopathy.

⁶Serial Dictatorship is a special case of DA where all schools share the same ROL.

lower socioeconomic status.

Another indication of preference misrepresentation in the Mexico City high school match comes from the field experiment of Bobba and Frisncho (2015). The authors show that giving students information about how well they are expected to perform on the centralized screening exam that determines priority changes their submitted ROL. Prior to the submission of ROLs, about three months before the screening exam was to be held, the authors held a mock exam in numerous classrooms in Mexico City. A subset of the students who took the exam were informed about their performance. The authors show that informing students about their performance in the mock exam made ROLs more aligned with performance on the exam; i.e., the match became more assortative. They further show that the effect of their intervention is concentrated in the group of students who had initially been pessimistic about their performance on the exam.

Truthfulness under DA has also been evaluated in the lab, an approach pioneered by Chen and Sönmez (2006). In the lab, participants are endowed with (known) preferences. Thus, while the stakes are typically significantly lower than in the field, detecting misrepresentation is simple, and parameters of the environment can be easily manipulated.

Chen and Sönmez find that a significant fraction of subjects (on average 36%) were not truthful under DA. This finding was corroborated in numerous follow-up studies, which also connected misrepresentation with having more information about other players' payoffs (Pais and Pintér, 2008), having lower cognitive ability (Basteck and Mantovani, 2016), and having lower priority (Hassidim, Romm and Shorrer, 2016; Li, 2016).

2 Possible Explanations

We now turn to the question of why individuals do not report their true preferences under DA. We suggest four potential explanations to this phenomenon.⁷

Failing to identify the dominant strategy.— Participants may be unaware that truthful reporting is the dominant strategy. In some cases, such as the Mexico City high school match, the formal instructions do not spell out the dominant strategy. In the common case in which instructions are explicit about the dominant strategy, participants may be unable to understand the meaning of a dominant strategy.

This explanation is supported by the numerous findings relating low cognitive ability with higher rates of misrepresentation (Basteck and Mantovani, 2016; Hassidim, Romm and Shorrer, 2016). In the 2014 IPMM post-match survey, 18% of the 310 respondents thought they could possibly gain from “behaving

⁷We do not think that inputting mistakes account for much of the misrepresentation observed in the field. This is supported not only by the fact that user interfaces are typically simple, but also by the significant percentage of individuals who report intentionally misrepresenting their preferences in different markets and countries.

strategically,” which indicates that they did not understand that truth-telling was dominant. An additional 21% of the respondents reported that while they thought that truth-telling was dominant, they did not understand why.

Mistrust.— Participants may doubt the veracity of the advice provided by the market maker. They may also be skeptical about the market maker’s commitment to use the advertised mechanism. In the 2015 IPMM post-match survey, out of the 51 respondents who provided a verbal explanation of misrepresentation, 3 (5.8%) mentioned that they mistrusted the system. According to Gross, DeArmond and Denice (2015), in Denver and New Orleans, where strategy-proof mechanisms are being used to allocate schools seats, parents report that “*they were told to provide the full number of choices in their true order of preference.*” However, these authors add that “*few parents actually trusted this message. Instead, they commonly pursued strategies that matched their own inaccurate explanations of how the match worked.*”

When individuals do not fully understand the mechanism, or do not trust the information they were given, they may resort to naïve theories of matching mechanisms. A natural idea in such theories is that the mechanism rewards higher ranking with increased probability of matching (provided that higher-ranked alternatives are not available). The behavior that such theories suggest is optimal is consistent with the optimal behavior in many decentralized markets, as well as with the optimal behavior in centralized markets that use mechanisms such as Immediate Acceptance (Boston Mechanism).

Self-selection.— If participants assign a zero probability to being matched to a desired alternative, then they believe that there is no loss in ranking this alternative lower, or dropping it from their ROL altogether (Chen and Pereyra, 2016). Moreover, participants may decide to drop some alternatives which they think they are very unlikely to get. In the 2015 IPMM post-match survey, 18 out of the 51 respondents who provided a verbal explanation of preference misrepresentation mentioned that they did not rank some program because they thought that chances of admission to the program were low. While self-selection considerations could explain why some applicants in the IPMM did not include the funded option in their ROL, it is less likely that such considerations account for the flipping observed in 10.4% of the relevant ROLs.

Non-classical utility.— Participants may derive utility from things other than their assignment. We suggest two such utilities: *ego utility* and *social preferences*.

Ego utility (Köszegi, 2006): Participants may distort their choices to avoid receiving information about their desirability or priority, as this may hurt their self-image. In particular, they may lower the ranking of some alternatives or even not rank them at all. It is unlikely that concerns about self-image preservation account for misrepresentation in environments such as the Mexico City high school match, where students know their screening exam scores and cut-off scores for admission are made public shortly after the match is finalized.

Social preferences: Participants may derive utility from altruistic behavior.

In the 2015 IPMM post-match survey, three respondents claimed that they did not rank a funded position since they preferred to see the money allocated to someone who needed it more. We believe that such considerations are not likely to arise when money is not involved.

3 Conclusion

Misrepresentation under DA is prevalent across markets and countries. We presented possible explanations as to why individuals do not report truthfully. It is worth noting that while some of the explanations are related to personal traits (e.g., trust, altruism, or cognitive ability), some, such as self-selection, are related to the subjective perception of desirability to the other side of the market. The explanations are not mutually exclusive — they may co-exist and interact. We believe that further efforts should be devoted to understanding this phenomenon.

The past decade has witnessed significant advances in the study of matching mechanisms, and we now have a much better grasp of the incentives induced by various designs. This stands in contrast with our limited understanding of the effects of human psychology on behavior in matching markets, and the resulting economic implications. In order to fully capitalize on the advances in the theoretical study of matching markets and to be able to apply them in real life, a better understanding of the interaction between users and matching mechanisms is indispensable.

References

- Aygün, Orhan, and Bertan Turhan.** 2016. “Conflict Resolution for India’s IIT Admission Problem.” Mimeo.
- Basteck, Christian, and Marco Mantovani.** 2016. “Cognitive Ability and Games of School Choice.” Mimeo.
- Bobba, Matteo, and Veronica Frisancho.** 2015. “Learning about Oneself: The Effects of Signaling Academic Ability on School Choice.” Working paper.
- Chen, Li, and Juan Sebastian Pereyra.** 2016. “Self-Selection in School Choice.” Available at SSRN 2880182.
- Chen, Yan, and Tayfun Sönmez.** 2006. “School choice: An experimental study.” *Journal of Economic Theory*, 127(1): 202–231.
- Committee, NRMP Data Release and Research.** 2015. “Results of the 2015 NRMP Applicant Survey by Preferred Specialty and Applicant Type.” National Resident Matching Program.
- Gross, Betheny, Michael DeArmond, and Patrick Denice.** 2015. “Common Enrollment, Parents, and School Choice: Early Evidence from Denver

- and New Orleans.” Center on Reinventing Public Education Making School Choice Work Series.
- Hassidim, Avinatan, Assaf Romm, and Ran I. Shorrer.** 2016. “Strategic’ Behavior in a Strategy-proof Environment.” Working paper.
- Hassidim, Avinatan, Assaf Romm, and Ran I. Shorrer.** 2017. “Redesigning the Israeli Psychology Master’s Match.” *American Economic Review Papers and Proceedings*, 107(5): ???-???
- Hatfield, John W., and Paul R. Milgrom.** 2005. “Matching with contracts.” *American Economic Review*, 95(4): 913–935.
- Köszegi, Botond.** 2006. “Ego utility, overconfidence, and task choice.” *Journal of the European Economic Association*, 4(4): 673–707.
- Li, Shengwu.** 2016. “Obviously strategy-proof mechanisms.” Working paper.
- Pais, Joana, and Ágnes Pintér.** 2008. “School choice and information: An experimental study on matching mechanisms.” *Games and Economic Behavior*, 64(1): 303–328.
- Pathak, Parag A., and Tayfun Sönmez.** 2008. “Leveling the playing field: Sincere and sophisticated players in the Boston mechanism.” *American Economic Review*, 98(4): 1636–1652.
- Rees-Jones, Alex.** 2015. “Suboptimal Behavior in Strategy-proof Mechanisms: Evidence from the Residency Match.” Available at SSRN 2662670.
- Roth, Alvin E.** 2008. “What have we learned from market design?” *The Economic Journal*, 118(527): 285–310.
- Roth, Alvin E., and Elliott Peranson.** 1999. “The redesign of the matching market for American physicians: Some engineering aspects of economic design.” *American Economic Review*, 89(4): 748–782.
- Sönmez, Tayfun.** 2013. “Bidding for army career specialties: Improving the ROTC branching mechanism.” *Journal of Political Economy*, 121(1): 186–219.