# The Role of Common and Private Signals in Two-Sided Matching With Interviews

Sanmay Das and Zhuoshu Li

Washington University in St. Louis, St. Louis, USA
sanmay@seas.wustl.edu,zhuoshuli@wustl.edu

**Abstract.** We study two-sided matching markets where the matching is preceded by a costly interviewing stage in which firms acquire information about the qualities of candidates. Our focus is on the impact of the signals of quality available prior to the interviewing stage. Using a mixture of simulation, numerical, and empirical game theoretic analysis, we show that more commonality in the quality signals can be harmful, yielding fewer matches as some firms make the same mistakes in choosing whom to interview. Relatively high and medium quality candidates are most likely to suffer lower match probabilities. The effect can be mitigated when firms use "more rational" interviewing strategies, or through the availability of private signals of candidate quality to the firms.

**Keywords:** matching, information acquisition, empirical game analysis

## 1 Introduction

The matching literature typically assumes that agents know their own preferences before the mechanism is run. Recently, there have been some papers that try to relax this stringent assumption [1, 2, 7], or to look at cases where the mechanism does not wish to elicit complete preference information [4]. In one-shot settings, agents come into the matching setting with unknown (or partially known) true preferences, but can learn more through a costly information acquisition (*interviewing*) stage before the actual matching happens (for example, academic job markets) [7, 13]; in repeated settings, the "match" is not final, but conveys information to participants on quality [5, 2].

Lee and Schwartz proposed what may be the first model of matching with an interviewing stage, where employers first simultaneously choose a subset of workers to interview, and then, in a second stage, submit preferences to a (Gale-Shapley) matching algorithm that then forms the matching [7]. The basic question that Lee and Schwartz ask is about the employer's decision of whom to interview, given that interviews are costly and all employers and workers on either side of the market are ex ante identical. The main complexity is then that the marginal benefit of interviewing a worker goes down as her number of other interviews goes up. The major result is that in symmetric equilibria (where each employer and worker has the same number of interviews), the number of agents matched goes up in the overlap, a measure characterizing the number of common interview partners among agents.

Our focus is on labor markets with interviewing, and in particular, the role of information in the interviewing and matching process. Consider the matching process that academic departments go through when interviewing and hiring faculty candidates. Typically, departments have a budget, say they can interview three or four candidates for a position. They start off the process by receiving a noisy signal about their preferences over candidates – CVs, letters of recommendation, and word-of-mouth can yield much information about candidates, but not nearly as much as an in-person interview. Once they have received these noisy signals, each department chooses which candidates to interview. Following Lee and Schwartz [7], after all the interviews have taken place, we can model the matching process as Gale-Shapley matching with departments submitting ranked lists of the candidates they interviewed. While this ignores some frictions (like exploding offers [3, 8]) that can be important, those are likely to be a second-order effect compared with the choice of candidates to interview. In contrast with Lee and Schwartz, who consider ex ante identical firms and workers, we are interested in situations where firms and workers are of different qualities, and some quality signals are available prior to the interviewing stage.

We look at a stylized model where there is a universally shared, common knowledge ranking of all firms, and there is a "true" universally shared ranking of all candidates, but this true ranking is not known – instead, firms receive different signals of candidates' rankings or qualities. If the true ranking were known to everyone, there would be only one stable matching, the assortative one, and any rational interviewing process would lead to the stable outcome in the matching stage. When signals of quality or ranking are noisy, firms must reason both about the true quality of candidates and about strategic issues in deciding whom to interview. This can lead to inefficiencies, where some candidates and firms do not end up getting matched whereas they would have with better information; these inefficiencies may fall disproportionately on some portion of the population of candidates and firms.

We are particularly interested in the roles of *common* and *private* information on aggregate and distributional outcomes in such matching markets. Common signals are shared across firms – for example, the quality of a CV, number of publications, LinkedIn endorsements, or public contributions to open source projects, can all be thought of as common signals of varying precision. Private signals can be generated through phone screens, preliminary interviews, etc. We assume that common and private signals are conditionally independent given the true ranking or value of the candidate. The central question of this paper is the effect of the relative precision of common signals and private signals on market outcomes. While a perfect common signal would reduce the problem to one with known rankings of both firms and candidates (and lead to the assortative matching and no inefficiencies under any reasonable model), our main finding is that the presence of a strong, but imperfect, common signal in addition to existing private signals can actually have significant negative effects, with fewer matchings occurring than with a private signal alone. The burden of this is typically borne by the candidates who are ranked relatively high (but not in the highest echelon). The mechanism is interesting – when these candidates end up with a common signal that is "too high", they interview at firms that are ranked too high for their actual quality. The firms that are closer to their true range choose not to interview them, but when these candidates' true qualities are revealed, they often don't get offers from the places that did interview them.

These findings are robust to several different choices of how signals of rankings and values are generated, and to different strategic choices by firms of whom to interview. The latter question is independently interesting – we demonstrate the intuition for our main result with a simple, but realistic, interviewing strategy where firms interview candidates "around" their true ranking. We then turn to a form of empirical game theoretic analysis [14] to explore a richer space of interviewing strategies, yielding "more rational" strategic decisions. This can alleviate the problem, with more agents being matched when there is only a common signal, but does not provide much benefit in terms of the number of agents matched when firms have access to both common and private signals.

## 2 Model and Inference

There are $n$ workers and $n$ firms, represented by the sets $W = \{w_1, ..., w_n\}$ and $F = \{f_1, ..., f_n\}$. The matching market operates in two stages, following the model of Lee and Schwarz [7]. In the first stage, each firm selects $k$ workers (or candidates) to interview; this decision is made on the basis of information present in the signals received by firms (described below). During the interview process, the true ranking of the set of candidates that is interviewed is revealed to each firm. The second stage can then be thought of as a Gale-Shapley matching where each firm submits a ranked list of the candidates it interviewed (others are unacceptable), and each candidate submits a ranked list of firms.

All workers know their preference rankings over employers with certainty. We assume that the workers all have exactly the same preferences over potential employers. Further, there exists a universal "true" ranking of all the workers as well, but this ranking is unobserved. Employers receive a private signal of their preferences as well as a common signal. In this paper we consider two possibilities. In **random-utility models** $w_i$ has a true value $v_i$ (which is drawn from a normal distribution). $f_j$'s private signal $\mathbf{s_j} = (s_1, s_2, \ldots s_n)$. Each $s_i, 1 \leq i \leq n$ is a noisy realization of the true value of $v_i$, corrupted by zero-mean Gaussian or uniform noise. The common signal, received by all employers, is a single (noisy) vector $\mathbf{z_C} = (z_1, z_2, \ldots z_n)$.

In the **Mallows model** [11], signals are directly over the ranking space. Following Lu and Boutilier's [9] description of its form, we say that each employer's private signal is a *ranking* $\Gamma_j$ sampled from the distribution which assigns $\Pr(\Gamma_j|\Gamma, \phi_p) = \frac{1}{Z}{\phi_p}^{d(\Gamma_j,\Gamma)}$, where $\Gamma$ is the modal (true) ranking, $\phi_p \in (0, 1]$ is a dispersion parameter such that the smaller $\phi_p$ is, the more the distribution will be concentrated around the modal ranking, $d$ is a distance function between rankings (we use the Kendall tau distance), and $Z$ is a normalizing factor. The common signal, $\Gamma_C$ is sampled from a Mallows model with the same modal ranking $\Gamma$ and a possibly different dispersion parameter $\phi_C$. In both cases, we assume common knowledge of all the relevant parameters of the distributions; the only unknowns are the true values or rankings.

In both the random utility and Mallows models, it is computationally difficult to perform full Bayesian reasoning over the whole space of possible posterior rankings, so we assume that firms compute the single most likely posterior ranking from the common and private signals, which we denote as $\widetilde{\Gamma_j}$, and use this single ranking for interviewing decisions. We defer details of the inference procedures to a longer version of this paper, but note that, in the random utility models, the procedures follow from those developed by MacQueen [10], while in the Mallows model, we use an algorithm based on the one devised by Qin *et al* [12] in the coset-permutation distance based stagewise (CPS) model (which is equivalent to the Mallows model using the Kendall tau distance).

## 3 Results

We first examine outcomes in a market where firms all use the same simple and intuitive interviewing strategy. They each compute their posterior ranking based on the available signals, and then interview the $k$ candidates who are ranked "around" the firms own ranking (e.g. with $k = 5$, the firm ranked 11 interviews candidates 9 through 13), with the firms at the top and bottom of the rankings adjusting their interview sets downwards and upwards as needed. We run 50000 simulations for each of the random utility and Mallows models; each simulation is of a market with 30 firms and 30 workers, each with interview budget 5. In each run, we hold the private signal parameters fixed, which are $\sigma_p, b_p$ in the random utility models and $\phi_p$ in the Mallows model, and vary the common signal parameters, which are $\sigma_C, b_C$ in the random utility models and $\phi_C$ in the Mallows model.

Based on the observation that the only stable matching if true preferences were known is the assortative matching, and that adding a common signal gives everyone more information about the true ranking, one would assume that adding the common signal always leads to more agents being matched. At the extreme, this is obvious – suppose the common signal had no noise and contained perfect information. Then the rational inference is just to use that signal. In this case, the assortative match would occur for sure.

But it turns out that, as the signal becomes less precise, the number of unmatched agents goes up sharply, and quickly exceeds the expected number of unmatched agents when no common signal is present! Surprisingly, on the candidates' side, the candidates who are less likely to get matched are actually the higher ranked ones (except for the very top ranked ones) (see Figures 1 and 2). What is the mechanism at play? In a more coordinated environment, as created by a common signal, the correlation between employers' estimates of a workers desirability is higher. Thus, it is more likely that several employers *all*

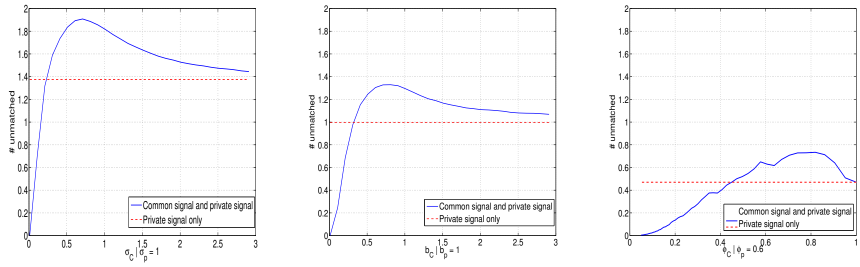


**Fig. 1.** Average number of agents left unmatched (Y axis) versus (decreasing) “precision” of the common signal ($\sigma_C$ for Gaussian noise (**left**), $b_C$ for uniform noise (**middle**), and $\phi_C$ for the Mallows model (**right**)), holding the precision of private signals fixed. The dashed line shows the number that are left unmatched when there is no common signal.

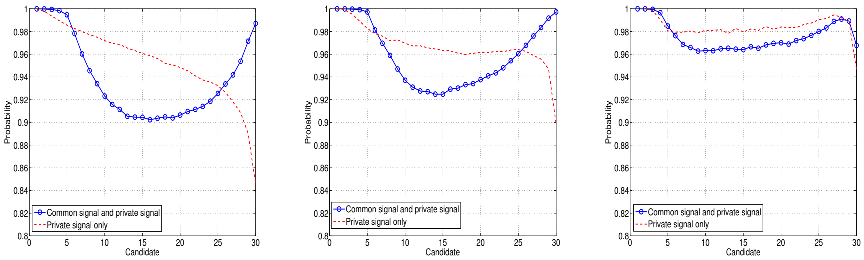


**Fig. 2.** The probability that the candidate of a particular rank is matched when firms have access to both a common signal and a private signal. Left: Gaussian noise ($\sigma_C = 0.6, \sigma_p = 0.5$), Center: Uniform noise ($b_C = 0.6, b_p = 0.5$), Right: Mallows model ($\phi_C = 0.7, \phi_p = 0.6$).

make the mistake of thinking a particular worker is too good or too bad for them. For candidates, the truth-revealing nature of the interview phase means that it can be disadvantageous to “place too high” in the first (interview selection) stage.[1] When opinions are more independent, as is the case when the private signal is stronger, it is less likely that someone will fall through the cracks in this manner. Therefore, more homogeneity of opinion, with even a little bit of noise, can create worse outcomes!

*Alternative Interviewing Strategies* Our results thus far apply to a simple interviewing strategy. What if employers used more sophisticated strategies? We analyze this using the basic idea of empirical game theoretic analysis [14, 6]. The fundamental strategic decision faced by a firm is to choose a set of $k$ candidates to interview. Game-theoretically, an (ex-ante) Bayes-Nash equilibrium would be

[1] For example, suppose a middle-ranked candidate gets early “buzz” on the job market, he may not get interviews from departments actually ranked in his vicinity because they think he is out of reach, but may not get offers once he is interviewed by higher ranked places and they realize he isn’t quite at their level.

one where each firm would not change the set of candidates it chose to interview, given the strategies of other firms, and the information available to them prior to the interview stage. Unfortunately, this game is very complex to analyze. Therefore, we restrict our attention to a manageable set of strategies: each firm can decide on any set of $k$ contiguously ranked candidates (in its posterior private ranking). We can determine (approximate) equilibrium firm strategies using an iterative empirical method (since firm $i$'s best strategy depends only on the choices of the firms ranked above $i$, and also has no effect on the utilities of those firms).

Can these "more rational" interviewing strategies resolve some of the inefficiency in terms of the number of participants left unmatched? Our initial experiments (available in a longer version of this paper) indicate that with only common signals, when the penalty for being unmatched is high enough, the better strategies, do, in fact, reduce the number left unmatched. However, with both common and private signals, the more complex strategies do not provide much, if any, societal benefit.

## References


1. H. Chade. Matching with noise and the acceptance curse. *Journal of Economic Theory*, 129(1):81–113, 2006.
2. S. Das and E. Kamenica. Two-sided bandits and the dating market. In *Proceedings of IJCAI*, pages 947–952, 2005.
3. S. Das and J. N. Tsitsiklis. When is it important to know you've been rejected? A search problem with probabilistic appearance of offers. *Journal of Economic Behavior and Organization*, 74:104–122, 2010.
4. J. Drummond and C. Boutilier. Elicitation and approximately stable matching with partial preferences. In *Proceedings of IJCAI*, 2013.
5. C.-J. Ho, A. Slivkins, and J. W. Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. In *Proceedings of ACM EC*, pages 359–376, 2014.
6. P. R. Jordan, Y. Vorobeychik, and M. P. Wellman. Searching for approximate equilibria in empirical games. In *Proceedings of AAMAS*, pages 1063–1070, 2008.
7. R. S. Lee and M. Schwarz. Interviewing in two-sided matching markets. Technical report, National Bureau of Economic Research, 2009.
8. S. A. Lippman and J. W. Mamer. Exploding offers. *Decision Analysis*, 9(1):6–21, 2012.
9. T. Lu and C. Boutilier. Learning Mallows models with pairwise preferences. In *Proceedings of ICML*, pages 145–152, 2011.
10. J. B. MacQueen. Optimal policies for a class of search and evaluation problems. *Management Science*, 10(4):746–759, 1964.
11. C. L. Mallows. Non-null ranking models. I. *Biometrika*, 44(1/2):114–130, 1957.
12. T. Qin, X. Geng, and T.-Y. Liu. A new probabilistic model for rank aggregation. In *NIPS*, pages 1948–1956, 2010.
13. B. Rastegari, A. Condon, N. Immorlica, and K. Leyton-Brown. Two-sided matching with partial information. In *Proceedings of ACM EC*, pages 733–750, 2013.
14. M. P. Wellman. Methods for empirical game-theoretic analysis. In *Proceedings of AAAI*, volume 21, pages 1552–1555, 2006.