ϕ		Avg Queries per Person		# Rounds
		Proposers	Acceptors	
0.0	GS-elicit	86.79 (1.32)	33.61 (1.32)	324.9 (39.3)
	PPGS-F	8.57 (0.16)	8.57 (0.21)	232.4 (21.2)
0.1	GS-elicit	[illegible] (1.68)	34.46 (1.55)	[illegible]
	PPGS-F	8.44 (0.26)	8.46 (0.29)	227.0 (22.1)
0.2	GS-elicit	[illegible] (1.46)	33.98 (1.38)	312.6 (29.4)
	PPGS-F	8.41 (0.13)	8.41 (0.13)	237.1 (22.9)
1.0	GS-elicit	5.90 (0.99)	4.73 (0.97)	408.4 (172.7)
	PPGS-F	7.41 (0.26)	6.42 (0.21)	338.1 (47.6)

Table 3: Average number of queries (std. dev.) until max regret 0, Gaussian riffle model: $n = 250$, 20 trials.

"types" corresponding to some (for ease of exposition) observable characteristic. For each type, we assume a distinct Mallows model over the men of each type. A woman's preference for men is generated by first drawing a ranking from each of these models, then interleaving them using a riffle process, with a parameter p reflecting the woman's bias toward each type: the full ranking is generated by iteratively placing the top "remaining" item from the Type 1 ranking into the next spot in her full ranking with probability p, and the top item from the Type 2 ranking with probability $1 - p$, until all men have been inserted into the ordering. The bias p towards one type of the other is drawn from an equal-weight mixture of two Gaussians (truncated over $[0, 1]$) with variance $\sigma = 0.1$ and means of 0.25 and 0.75 (so women are unlikely to be "indifferent" between the two types of men). Men's preferences are generated in the same way.

We compare PPGS-F to GS in terms of number of queries to reach a stable matching in this model, using three different dispersion parameters for the underlying Mallows models. (We don't use PPGS-ML since maximum likelihood estimation is less straightforward in this mixture process.) Results in Table 3 show that PPGS-F vastly outperforms GS, when $\phi = 0, 0.1, 0.2$; and it remains competitive with GS when $\phi = 1$. It always requires fewer rounds as well.

MovieLens Models. We next consider matching problems in which preferences are generated from the MovieLens collaborative filtering data set.[3] The MovieLens data set consists of 100,000 ordinal ratings (1–5 scale) of 1682 movies by 943 users. We convert this into preference rankings of users for each other by generating an *affinity score* between pairs of users based on the similarity of their movie ratings.

Let $M(a)$ denoted the set of movies rated by user a, and r_a her rating vector. Given two users a and b, we define their affinity score to be $s(a, b) = \sum_{m \in M(a) \cap M(b)} 5 - |r_a(m) - r_b(m)|$, where 5 reflects the maximum rating. This scores creates very correlated preferences: individuals who rate a larger number of movies will tend to be viewed as more desirable across the population. We also create somewhat less correlated affinities by normalizing scores by the number of movies rated in common: $s^N(a, b) = s(a, b)/|M(a) \cap M(b)|$. With these affinities, we create random matching problems by drawing 250 "men" and 250 "women" uniformly at random, and generating preference rankings for the appropriate side of the market using these real-valued affinities to order potential partners (breaking ties arbitrarily). As expected, the Unnormalized Movie Matching (U-MM) problems using score s exhibit more correlation than their unnormalized counterparts (N-MM) generated using s^N.[4]

[3] See http://www.grouplens.org/node/73, the 100K data set.

Data Set		Avg Queries per Person		# Rounds
		Proposers	Acceptors	
U-MM	GS-elicit	43.52 (2.07)	26.62 (1.25)	324.9 (39.3)
	PPGS-F	6.36 (0.12)	6.37 (0.14)	249.7 (22.9)
N-MM	GS-elicit	13.23 (1.09)	10.07 (0.99)	255.5 (47.8)
	PPGS-F	8.11 (0.13)	8.09 (0.15)	349.6 (22.3)

Table 4: Average number of queries (std. dev.) until max regret 0, MovieLens Matching Models: $n = 250$, 20 trials.

Table 4 shows the performance of GS and PPGS-F for both U-MM and N-MM, averaged over 20 random matching instances. PPGS-F outperforms GS with respect to number of queries in both settings, but requires more rounds in N-MM. Not surprisingly the performance gap is greater in when preferences are more correlated. PPGS-F similarly outperforms GS with respect to cognitive cost (setting $\gamma = 0.5, \tau = 5$ as above). In U-MM, average cognitive cost per person for GS is 788.52 for proposers, and 3.56 for acceptors. Using PPGS-F, average costs are 55.31 and 55.40 for proposers and acceptors, respectively. Again, note that the cost of fully sorting all alternatives is 310.34 per person. PPGS-F also outperform GS on N-MM problems: GS has an average cognitive cost of 250.04 and 1.35 for proposers and acceptors, respectively, while PPGS-F has costs of 58.84 and 58.59, respectively.

5 Conclusions and Future Work

We have proposed the use of minimax regret as a robustness criterion for stable matching with incomplete information, and developed several heuristic elicitation schemes designed to quickly reduce regret. These schemes compare favorably to the use of Gale-Shapley for interactive elicitation, especially when preferences exhibit some correlation: they reach stable matching with fewer queries, fewer rounds of elicitation, and with lower cognitive cost. Critically for domains where approximate stability is desirable (e.g., when elicitation/interviewing costs or switching/defection costs are high), our elicitation schemes demonstrate vastly superior anytime performance to GS, allowing approximately stable matchings to be found with very little preference information.

Many interesting research directions remain, including: improved procedures for exact MMR computation; new algorithms tuned to specific forms of partial preferences; analysis of additional probabilistic preference models and the use of priors to further improve elicitation performance; the assessment of our methods for different measures of approximate stability; and the extension to other stable matching problems.

Acknowledgments. Drummond was supported by OGS and a Microsoft Research Graduate Women's Scholarship. We acknowledge the support of NSERC, and thank Tyler Lu for generously sharing his code and the reviewers for their helpful suggestions.

[4] Using Kendall's τ-statistic, the average correlation statistics for U-MM are 0.4539 (proposers) and 0.4653 (acceptors); while for N-MM, they are 0.1462 (proposer) and 0.1184 (acceptors).