**Reviewer1 – Reviewer2 – Reviewer3**

I cannot easily think of a half-realistic setting where one would know that the scores need to be

convex, but at the same time one would not know which ones exactly to use.

The other problem is that I can hardly think of a scenario where the chair can answer the questions about the scoring rule, but somehow it is difficult for him to produce the full rule in a single step. So what do we gain?

I can hardly imagine a person who faced with the profile P' from Claim 4 could meaningfully answer if a is better than b or the other way round. A piece of software could do it, though---but probably by precomputing the scoring rule to use (which removed the need for eliciation).

Reviewer 1 addresses the significance of the profile P’ built in Claim 4. We want to clarify that our goal here is not to elicit the complete voting voting rule, but rather elicit the answer to a single specific question. Consider the toy example given in the proof of Claim 4 where we take p = q = 1, m = 4 and r = 2. Our question to the committee would investigate whether w\_2 – w\_3 ≥ w\_3 – w\_4. We can rewrite this by asking whether w\_2 + w\_4 ≥ 2 w\_3 , or, in other words, whether an alternative “a” ranked once in second position and once in fourth position should be preferred to an alternative “b” ranked twice in third position. Therefore we construct a preference profile that represents this situation and we ask the chair to pick a winner for this scenario.

**q p**

- -

a -

b b

- a

If we add two more alternatives to complete the profile we may occur into the situation where the chair will pick one of these alternatives as a winner. We have then to exclude this scenario, and we do that by adding 2 delta = 2 (p + q) = 4 more agents.

**q p d\_1 d\_2 d\_3 d\_4**

c c a a b b

a d b b a a

b b c c d d

d a d d c c

Now asking to the chair whether w\_2 – w\_3 ≥ w\_3 – w\_4 corresponds to ask whether who should be picked as a winner in the following profile:

1° 2° 3° 4°

a 2 3 0 1

b 2 2 2 0

c 2 0 2 2

d 0 1 2 3

Note that c and d will not be preferred over a and b, and the chair is left to choose whether she prefers an alternative never ranked in last position (in this case we can conclude that the response to our question is false) or an alternative ranked more times in second position (in this case the response is true).

This also replies to another point raised by the same reviewer regarding the ability of the chair to respond to our questions but not to formalize the voting rule. Consider the case where the committee would like that the weight associated to the second position is half the weight associated to the first position but then she cannot really associate a precise score to the third position. This situation is not so unrealistic, especially for non expert users. In this case the value of w\_3 would range between 0.25 and 0 to ensure convexity. By asking, for example, who it would be considered the winner in a profile where the alternative “a” is ranked twice in second position and five times in fourth position, and the alternative “b” is ranked seven times in the third position, we are able to narrow the range of w\_3 to [0.25, 0.2] or [0.2, 0].

I am not really convinced that the assumption that a scoring rule is not known makes sense. On the one hand, there is the problem of choosing the scoring rule manipulatively (but since the authors mention that the elections at hand are built into recommendation systems etc., this is not really an issue).

Q3: Can you provide a realistic example where elicitation of the scoring rule is useful?

Another point that is not clear is when, in practice a system like this would be deployed.

Example of reviewing process. People submit their preferences even though the aggregation process is not fixed beforehand. Regarding manipulability, the elicitation processes are independent, the committee has no information regarding the preferences expressed by the voters.

----------------

6. Proof of Claim 4: I do not understand notation (3,m) or (4,m-1) etc.

a) Why do the authors use IC only? Why not other statistical cultures? Why not some real-life data? After all, the experimental part is the main contribution of the paper and it is done in a fairly minimalistic way.  
  
b) Why do the authors look at very small elections only? 15 candidates and 30 voters seems completely inadequate for the motivations from the introduction.

c) The conclusion that the number of questions required to reach low regret cannot be made based on Table 3. There is simply far too little data to have any sort of confidence in claims like this.

Q1: Why did you only look at IC elections?  
Q2: How to use your approach in large elections? Hundreds of candidates, thousands of voters? After all, showing a strategy that can deal with large elections would meaningfully extend the work of Lu and Boutilier?

The lack of theoretical guarantees for the proposed elicitation strategy, as well as missing reasoning behind several of the experimental observations, makes this a weak paper in my opinion.

I believe the paper will be much stronger if the authors could provide theoretical convergence guarantees for the proposed heuristics.

-"For generating the weights we first draw m − 1 numbers uniformly at random"-->What is the range from which you make the draws?

-"We see that the number of questions required to reach a low regret level grows approximately linearly with the number of agents"-->Can you provide a formal regression analysis to support this claim?

1. On Page 1, you write that "The quality of the recommendation increases faster than linearly with the number of questions, after an initial phase with almost no increase in quality, before slowing down and converging to an optimal recommendation." What is the reasoning behind this trend?  
  
2. How did you decide the range of the parameter \lambda to be [1,n]?

1) Could testing mixed strategies yield better results than the pure ones?