Review of: "Trust is the best policy. Game theoretical analysis of bias in elicitation procedures in linguistics"

Maria Caamaño-Alegre[1]

1 Universidad de Valladolid

Potential competing interests: No potential competing interests to declare.

This is an interesting, stimulating paper, where original ideas are presented in a clear way. Kowalewski applies the frameworks of game theory and decision theory to demonstrate that the common view on the problem of bias in elicitation procedures in linguistics is wrong. In particular, he argues against the view that, in such elicitation procedures, “naïve" consultants provide more reliable data than consultants who are linguists. While the author endorses the common assumption that self-elicitation is affected by the fact that researchers are inherently biased, he questions that this fact entails a higher risk of unreliable data in comparison to cases where data are collected from naïve consultants. Moreover, Kowalewski convincingly shows that both cases are similar in that in both the best strategy is to accept all the data.

According to the author, the above two situations are similar because in both cases:

a) a Laplace principle can be applied to assign equal probabilities to three possible situations: unbiased, positively biased, negatively biased,

b) the researcher has to choose between accepting the judgment as reliable or reject it as unreliable,

c) the consultant has to choose between delivering a veridical judgment or a falsidical judgment,

d) there is reciprocal incomplete information (neither the researcher nor the consultant know each other’s strategies),

e) there is a 50% chance that biased consultants report falsidical judgments, and a 50% chance that they report veridical judgments,

f) in the biased scenario, there is a 50% chance that researchers accept stabilized judgments (i.e. veridical biased judgments) and a 50% chance that they accept altered judgments (i.e. falsidical biased judgments), hence researchers’ payoffs cancel each other in either situation,

g) the expected overall payoff from consistent acceptance of unbiased judgment is 0.33, and the expected overall payoff from consistent acceptance of biased judgment is 0.33, so overall the acceptance is the best strategy in approximately 67% of cases.

A key point in Kowalewski’s argument is that being biased is independent from reporting falsical or veridical judgments. This is the idea behind assumption e) above. Regardless of whether the biased consultant is naïve or not, and of whether the bias is positive or negative, the probability of reporting a falsical judgement is not higher than the probability of

reporting a veridical one. In principle, the judgment delivered may be stabilized or altered regardless of any strategic considerations on the part of the consultant. The naïve biased consultant may have the goal of satisfying the expectations of the researcher or, rather, that of disappointing them, but, either strategy is in principle compatible with the consultants reporting veridical or falsical judgments. Biased consultants who are linguists are in a similar position, even if the cause of their bias is different. They are not trying to satisfy or disappoint the researcher's expectations, instead they are unconsciously reporting what agrees (positive bias) or disagrees (negative bias) with the theory they themselves assume. But either way, because of the unconscious bias, they do not know whether their judgments are stabilized or not, thus there is a 50% chance of either outcome, just like in the cases of "naïve" consultants. Therefore, in all cases it would adequate to assume that there is a 50% chance that judgments are stabilized and an equal chance that they are altered.

Now, assumption a) could be questioned on the grounds that both biased and unbiased liars should be included there at least jointly as a fourth possibility. After all, elicitation procedures are "artificial", non-ordinary situations where both spontaneous unbiased liars and interested biased liars may feel specially motivated to lie. In the first case, even if they don't have a special interest in the subject itself about which they are asked about, it seems that there could be a conscious interest in providing wrong data as a way to disappoint the researcher's expectations about the very reliability of the data. In the second case, there could be a conscious motivation for negative bias, specially when consultants are linguists. Moreover, since the consultant does not know what the researcher's theoretically-shaped expectations are, the best strategy would be to give up the idea of disappointing those expectations, and lie to be sure that the researcher gets unreliable data in the form of altered judgments. In all these cases of dishonest consultants, there would be a desire to defeat the researcher, not by disappointing their theoretically-shaped expectations, but by fighting their expectation that they will get reliable data.

Possible cases of dishonest consultants might be more important than suggested in the paper, and, if so, Kowalewski's restriction regarding the application of his analysis only to cases of honest consultants (see footnote 4) comes across as too narrow. This is specially clear in the case of dishonest negatively biased consultants who are linguists, which seems a more likely scenario than that of honest negatively biased consultants, despite the latter (not the former) being the one included in Kowalewski's discussion.

It must be noted that, if a forth case corresponding to dishonest consultants were included in a), then, accepting all data would be the best strategy only in 50% of the cases, in particular, in those where unbiased consultants don't lie and in half of the honest biased cases. Accepting all cases is not thus an optimal strategy.

Certainly, some enlightening considerations made by the author in section 5 suggest that if some idealizations were replaced by "realistic" assumptions, and a) were assumed without modification, acceptance would be the best strategy in approximately 70% of cases. But, again, this is so only if we endorse a) as stated by the author. By contrast, if the Laplace principle is applied to the four cases described above, instead of just the three contemplated by the author, getting rid of idealizations would not make a significant difference.

There is a second point that could be questionable as well, namely, the one related to the author's view that payoffs from

consistent acceptance are greater than payoffs from consistent rejection. Although Kowalewski does not apply this view to his game-theoretical analysis of bias, which is made under the idealization that the payoff from accepting a veridical judgment is equal to the payoff from rejecting a falsidical judgment, he regards such idealization as potentially eliminable in favor of what he considers as a more realistic assumption. He argues that when linguists accept as much as 50% of unreliable data they can still make progress in their research and correct the effect of unreliable data in the future, on the other hand, if a linguist rejects all judgments, which included 50% of reliable judgments, “their research will not even get off the ground due to lack of data”. It is difficult to see how fruitful a research started with 50% of unreliable data could be. The author is right that some corrections could be made in the future, but maybe it would take a long way to realize about the problem and, by then, many wrong inferences and unsuccessful attempts at providing explanations for the confusing data could have already been made. In comparison, being more careful and demanding on the reliability of data from the start seems like a better option. Overall, the payoff of rejecting unreliable judgments seems underestimated by the author.

Notwithstanding the above criticisms, I found Kowalewski’s paper very valuable and thought-provoking, not only because it challenges some common unwarranted assumptions about bias in elicitation procedures in linguistics, but also because his analysis is extremely clear and coherent. It has been a very interesting, helpful reading.