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

Alayo Tripp[1]

1 University of Minnesota

Potential competing interests: No potential competing interests to declare.

The article attempts to demonstrate that although linguistic informants may demonstrate bias in reporting acceptability judgements, bias does not inevitably negatively impact data collection. The author argues that the optimal strategy for collecting intuitive judgements is “universal acceptance,” i.e. the inclusion of all available judgements. However the discussion of “bias,” its formal definition, and instantiation in the presented models leave much to be desired. My opinion is that the evidence presented does not substantiate the conclusions. The use of the word “bias” borders on misleading, and the manuscript title should be revised to better reflect what seems to be the intended narrow scope of evaluating the importance of *metalinguistic naïveté* in linguistic consultants. My misgivings lead me to focus my comments on the theoretical grounding of the work.

The author writes that “while universal acceptance does not safeguard researchers against data distorted by bias, it yields legitimate data in a clear majority of cases under almost all reasonable assumptions about the distribution of bias in the population of consultants.”

My key objections are threefold:

1. The manuscript does not make clear what linguistic problems can be fruitfully investigated with the kind of data under discussion (intuitive judgments), and therefore neither effectively defines “legitimate” or “corrupted” data, nor effectively communicates a justification for applying game and decision theory.
2. The manuscript neglects to make clear theoretical commitments to the relationship between idiolectic data and data on named languages. Assumptions about this relationship underlie the presented models and ought to be discussed.
3. The presented results rely upon *unreasonable* assumptions about the significance and distribution of bias in linguistic informants.

In what scope are intuitive judgements relevant?

The competence/performance distinction is invoked, however, the cited literature is fairly old, and does not reflect contemporary disagreement about the usefulness of this theoretical distinction, or bias inherent in its conception. Thus, although intuitive linguistic judgements are discussed as a source of data with questionable reliability, the manuscript relies upon an assumption it does not justify: that the concept of an “unbiased consultant” is theoretically relevant to answer research questions in linguistics. The advantages of leveraging this abstraction should be explicitly stated.

- Hymes, D. (1992). The concept of communicative competence revisited. *Thirty years of linguistic evolution*, 31-57.
- Saville-Troike, M. (2008). *The ethnography of communication: An introduction*. John Wiley & Sons.


Ultimately, intuitive judgements are operationalized and leveraged differently in different subfields of linguistics, and it is crucial to carefully identify which fields of linguistics are meant to be represented in this manuscript, noting at least in a cursory way that the relevant research questions and methodologies necessarily diverge from those employed by linguists with different goals. Keeping the scope of the manuscript narrow, it remains important to establish a clear purpose for collecting intuitive judgements as data, such that discussion of legitimate, biased or "corrupted" judgements and the consequences of including them is explicitly grounded with respect to specific linguistic research goals. As is, the manuscript does not clearly explain the import of such data.

The presented work relies upon theorizing knowledge of underlying linguistic structure as distinctive from factors associated with noisy and variable performance. The *purpose* of this abstraction should be explicitly justified, rather than evoked as dogmatic.

We know that contexts manipulated to evoke social contrasts can impact linguistic perception, producing categorically different responses to identical stimuli. The following literature presents such evidence suggesting that the conception of linguistic structure as structurally dissociable from "non-linguistic" social knowledge is itself a biased perspective. Whether and when linguistic knowledge can be fruitfully examined independent of other socio-communicative competencies is relevant to a discussion of how intuitive judgements about language may exhibit inevitable bias.


- Casasanto, L. S. (2008). Does social information influence sentence processing?. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 30, No. 30).
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, *137*(5), 2823-2833.
- Rosa, J., & Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in society*, *46*(5), 621-647.
- Flores, N., & Rosa, J. (2022). Undoing competence: Coloniality, homogeneity, and the overrepresentation of whiteness in applied linguistics. *Language Learning*.
- Tripp, A., & Munson, B. (2022). Perceiving gender while perceiving language: Integrating psycholinguistics and gender theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(2), e1583.


One reason to elicit intuitive judgments comes from evidence that performance differences in response to varying task demands can reveal structural linguistic knowledge. However, it is also well known that non-linguistic context can impact linguistic representations, and may do so differently at different levels of analysis. To begin addressing this tension, it is crucial to carefully specify the intended scope of the present paper. The expected usefulness of introspective judgements must be contextualized within *specific contexts*, i.e. informant populations defined by some criterion of homogeneity, or specific methodologies for presenting stimuli for judgment (i.e. self-elicitation). This kind of clarification is needed throughout the text, as general terms like "intuitive judgement" and "elicitation" ambiguously evoke a wide range of

methodologies. However, even if the intent were to focus exclusively on self-elicitation, examining this methodology for bias demands engagement with the process by which particular informants are(n't) credibly selected as authoritative sources for describing a named language such as English, while other consultants who nonetheless also speak English, would only be recruited to provide data on more narrowly construed lects, i.e. ideolects and sociolects. Confronting this reality requires making explicit the kinds of theoretical commitments and research questions which shape the manuscript's present definition of “bias."


- Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2011). You had me at “Hello”: Rapid extraction of dialect information from spoken words. *Neuroimage*, *56*(4), 2329-2338.
- Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2014). Processing effects in linguistic judgment data:(Super-) additivity and reading span scores. *Language and Cognition*, *6*(1), 111-145.
- Reese, H., & Reinisch, E. (2022). Cognitive load does not increase reliance on speaker information in phonetic categorization. *JASA Express Letters*, *2*(5), 055203.


Inclusion Criteria - Whose Acceptability Judgements are Acceptable?

Although it remains unclear how representations of speaker-specific and group-specific linguistic variation should be conceived of with respect to more generalized linguistic competence, it is well known that sociolinguistic variation in presented stimuli can impact the results of language experiments. The assumption that undue variation can interfere with discovering generalizable linguistic principles is reflected in the traditional design of empirical studies *and the preference of some linguistics to rely on the kind of intuitive judgement data which is discussed in this manuscript.*

The author notes that [some] linguists encounter tension between the “desire to achieve valuable generalizations about language and unavoidable idiolectic differences across speakers.” The manuscript would be improved by acknowledging that intuitive judgements offered of an informant’s *own idiolect* offer the advantage of minimizing irrelevant sociolinguistic variation during data collection, since each idiolect can have only one consultant. The use of self-elicitation could be therefore be discussed as potentially desirable for controlling variation in the informant population, and potentially problematic in its potential to invisibilize relevant and impactful differences in social perception.

However, “intuitive judgments” as it is used throughout the manuscript, does not explicitly pertain to idiolects, but rather to a more broadly construed construct of “language,” which presumably indexes some linguistic knowledge shared with some category of languagers with the ability to use that same variety. Again, from the introduction: “In the following part of the article It is assumed that speakers are capable of forming intuitive judgments simply by virtue of being speakers of a **given language**.” (emphasis mine.)

I believe the manuscript would be improved by explicitly noting that when ideolectic data is systematically screened for inclusion in aggregated data on a “given language,” this process is itself also subject to effects of bias.

The presented work implicitly relies upon **[named] languages** as a foundational theoretical construct, however this approach ought to be justified and contextualized with respect to alternative views. There is no scientific consensus

regarding the nature and proper usage of such constructions. As Otheguy, García and Reid (2015) put it, “whereas the idiolect of a particular individual is a linguistic object defined in terms of lexical and structural features, the named language of a nation or social group is not; its boundaries and membership *cannot* be established on the basis of lexical and structural features.”


- Otheguy, R., García, O., & Reid, W. (2015). Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review*, *6*(3), 281-307.
- Erker, D. (2017). The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language and Linguistics Compass*, *11*(1), e12232.
- Saraceni, M., & Jacob, C. (2019). Revisiting borders: Named languages and de-colonization. *Language Sciences*, *76*, 101170.


Clarifying the scope of “bias” in intuitive judgements of a named language, e.g. "English"

In its current state, the manuscript entirely ignores the complex social construction of named languages and their borders. Again from the introduction: “A speaker of English may entertain and report the impression that the sentence *The cat is on the mat* is a well-formed sentence in English without knowing what makes the sentence well-formed and without explicit declarative knowledge of the mechanisms of their mental grammar.”

However, in African-American varieties of English “The cat on the mat” would be an equally well-formed sentence with identical meaning, despite the copula drop. Although we may expect that not all “English” speakers will indicate this is an acceptable utterance to index that meaning, patterns in rejection of AAE syntax as unacceptable are not dissociable from ideological and political stances regarding the competence of AAE languagers as “speakers of the English language.” Reports of impressions regarding the well-formedness of sentences therefore necessarily entail an impact of social environment, pre-existing sociolinguistic ideologies and attitudes towards persons *other than the linguist conducting the study.* Even in a self-elicitation procedure!

In the presented models, it is only the linguistic informant’s attitude *towards the linguist* and towards the *linguist’s goal* of collecting accurate data which are instantiated as potential sources of “consultant bias.” Notions of consultants' legitimacy are confined to contrasts between consultants who were already chosen by the linguist as representatives of a given language community, who is then described as further refereeing the relative authenticity of their reports. The entire inquiry rests upon the notion that informants can be implicitly identified as users “of a given language,” when this process, preceding the elicitation, is itself laden with bias. However, any discussion of potential bias in the recruitment of consultants has been omitted entirely.

We could also imagine that informants may have biases which are deployed wholly independently of the linguist’s intentions, or in perfect cooperation with linguists having matched biases. Supposing that the linguist could be biased in their initial selection of consultants would necessitate further analyses. Potential sources of bias in the design and execution of elicitation studies are well documented,  and traditionally mitigated in linguistic data collection. However, in placing these phenomena beyond its scope, the manuscript is left with an extremely rarified and somewhat useless

interpretation of the word “bias."

I would agree with the conclusion that the linguist cannot effectively adjudicate between potentially biased consultants, when those consultants are drawn from a pool specifically constructed according to that linguist's pre-existing biases. However, that is not the conclusion that is drawn here. Questioning how the linguist's work may be differently affected by the inclusion of data from various consultants *begs the question* of how the linguist had the ability in the first place to construct a consultant pool (of whatever size) putatively only containing consultants competent in the relevant language.

Summary recommendations

The manuscript would be improved were the introduction to present and justify an explicit definition of “bias,” making clear how and why it will later be operationalized. It would also be improved by explicitly identifying the relevance of these intuitive judgments to investigations of any specific linguistic phenomena. There should be a discussion of what role bias plays in the way idiolectic contributions are constructed into corpora taken to evince the character of named languages. These changes would make clear the basis for ascribing usefulness to the kind of data and data evaluation under discussion. As is, the pertinence of the work is not effectively communicated. The title and abstract would likewise benefit from increased specificity.