

Referee Report on: Guilt Aversion and Redistributive Politics: A Moral Intuitionist Approach

Gilles Le Garrec

31-12-2013

General Comment

This work is aimed at modeling the voting behavior over redistribution as the interaction between an automatic cognitive process which quickly generates intuitions on the fair level of redistribution and a rational self-oriented reasoning which controls the feeling of guilt associated with fair institution.

An interesting feature of the paper is its attempt in formalizing psychological concepts so as to use them to improve the modeling of redistribution preferences. The inclusion of psychological aspects of individuals' behavior into the mainstream economic framework is indeed helpful in providing a mathematical formalization to belief and context-dependent motivations, other than self-interest, that are experimentally shown to be of great relevance in shaping individuals' behavior.

Further, cultural contexts and historical background are presented as relevant in modeling individuals' fairness and guilt aversion.

The contribution of the paper is then to set up a comprehensive model, that entails both psychological (and neuroscience) and economic theories.

Although the issue proposed is much interesting, the syntactical and linguistic outlook of the paper is not satisfactory and makes it hard to be followed. Therefore, I think that the current version of the paper does not deserve to be published in *Economics-E-Journal*. The author should into a major revision. The paper should be rewritten in order to be understandable by the large and miscellaneous audience that a publication in an open journal brings. In the next sections of this report, I try to highlight the main drawbacks of the paper, at the same time suggesting lines of improvement of the current version.

Specific Comments

1. I found the rich list of references contained in the introduction of the paper quite confusing in the way it is presented. The author goes back and forth from traditional economic papers to social-preference studies and psychological/neuroscience studies. At some point of my reading (end of the introduction), I totally lost his line of reasoning. The introduction looks more like a series of thoughts sometimes not linked among them. I think that the introduction should be rewritten by more clearly stating, at the beginning of each paragraph/subsection, what the author wants to show, e.g. "this is what traditional economic theories conclude", "this is

what empirical studies have shown", "these are the issues where traditional theories and empirical data do not match", "these are the new insights brought by psychological and neuroscience studies into these issues", "this is how these insights could help explain puzzling subjects' behavior in the field", and so on and so forth.

2. In the abstract, the author declares that the main goal of the paper is modelling "the voting behavior over redistribution as the interaction between an automatic cognitive process [...] and a rational self-oriented reasoning [...]". However, the literature on voting is almost completely missed in the introduction (apart from Campante 2010 in footnote 1) and in the rest of the paper. Furthermore, the word "voting" itself appears the first time in the text at page 6, when describing Figure 1, where "voting" looks like a rather external feature "magically" inserted in the dual-process decision-making approach. Then, one should "wait up to" page 18 to read again the word "voting", that throughout the paper appears no more than 5 times. Therefore, if the author wants to keep "voting over redistribution" as a main feature of the paper, he should stress this point better. Unfortunately, the idea that the current version of the paper gave to me is that the author disregarded the economic (game-theoretical) literature on voting. And the fact that the only reference to this literature (Campante 2010 forthcoming) is not updated (the paper was published in 2011, more than 2 years ago) reinforces this idea.
3. As follow-up on the previous comment, a more general comment about the motivation of the paper should be made. When reading the beginning of the introduction, my guess was that redistribution (and voting over it) was the central issue of the paper. However, at the end of the paper, the feeling I had is that voting over redistribution looks more as an application (rather, an analytical exercise) of Haidt's (2007) new synthesis in moral psychology. Indeed, all theoretical results in the paper (Propositions 1, 2, 3) are "preparatory" to voting behavior, in the sense that they concern preferences over redistribution, rather voting behavior over redistribution. Actually, neither a voting game nor a voting rule is ever introduced in the paper. To summarize, I think that the author should be clearer about under which light the paper has to be seen. Obviously, the contribution of the paper would be greater if the author would be more convincing about the relevance of his model in the economic literature, and, in particular, about its usefulness for voting behavior over redistribution.
4. The fact that guilt aversion is context-dependent (see page 5 and page 15) is not only "the cultural trend in psychology". Indeed, it has also been experimentally proved (see Attanasi, Battigalli and Nagel 2013) and theoretically modeled (see Attanasi, Battigalli and Manzoni 2013) in a trust game. These studies show that guilt aversion might be role-dependent, i.e. it can emerge according to the specific role a subject assume in the strategic interaction. Moreover, the fact that guilt aversion plays almost no effect in specific games (e.g., prisoner's dilemma) and a significant effect in others (e.g., dictator game and trust game as trustee) – a rather acknowledged result in experimental economics – should lead to conclude that guilt aversion is "context"-dependent also when with the term "context" one identifies a specific interactive strategic situation. I think that – for the sake of completeness – the author should discuss also these aspects in section 3.1 of the paper.
5. The author often cites concepts and other researchers' theories without

giving a proper explanation to them (not every reader has a background in this field). As an example, see page 5 lines 4 and following, what does *System1* and *System2* in Kahneman (2003) represent? What is an *automatic and domain-specific cognitive process*?

6. The author often makes use of neuroscience/psychology technical terms, without properly explaining their meaning to a general audience of economists. As an example, see page 8 lines 7-8: What is the "*anterior insula*"? What does "*skin conductance*" refer to? It is likely that economists have not enough knowledge in these fields.
7. At page 9, after Figure 2, I did not get the relation between the reference to the literature about native vs immigrant behavior and the point – cultural side of human behavior – that the author wants to make in the light of his model. I see this reference as too specific and not related to the modelling assumptions the author makes in the same paragraph. Further, the author recalls this point ("immigrants from high-preference redistribution countries") when commenting equation (13) at page 16. But also there I do not see the link.
8. This comment is the most important in terms of the revision I would require for the paper. I do not like the way in which the model is presented. The different features of the model are introduced one after the other in subsequent paragraphs, each time placing before each feature a short – sometimes unrelated – discussion of the literature behind the motivation of a specific assumption. This made quite difficult for me to follow the author's line of reasoning and to understand whether – apart from Section 4 – the feature introduced in a specific paragraph is a completion or an extension of the model presented in the previous paragraph (e.g.: is guilt aversion an extension of the basic model introduced in Section 2 or is it already assumed – though not made explicit – in the previous section?). Hence, I strongly encourage the author to implement a major revision of Sections 2, 3 and 4, by: first, indicating all the (theoretical and behavioral) assumptions behind the model; second, discussing the motivation of each of these assumptions, by making reference to the relevant literature; third, solving the model and introducing the main propositions; fourth, discussing the economic implications of the theoretical results.
9. At the beginning of Section 3.2, the author compares the way guilt aversion is integrated in (12) and guilt aversion *à la* Charness and Dufwenberg (2006). He says that "such a specification is also consistent with Charness and Dufwenberg (2006) who have experimentally shown in the trust game that guilt aversion was shaping significantly individual behavior". However, in Charness and Dufwenberg (2006) guilt aversion is a belief-dependent motivation (see Dufwenberg 2002 for a deeper discussion of this feature). Indeed, they find a significant positive correlation between the trustee's frequency of cooperation and his/her second-order beliefs about cooperation. Hence, on one side – from a theoretical point of view – I do not see how guilt aversion in (12) can be interpreted as a belief-dependent motivation. On the other side – from an experimental point of view – I do not see how the experimental results in Charness and Dufwenberg (2006) sustain the guilt aversion assumption in the specific game the author analyzes. To tell the truth, I have the impression that there is a confound between what the author calls "guilt" and "fairness" (inequity aversion) *à la* Fehr and Schmidt (1999).

10. At the end of page 16, the author introduces the sufficient condition $\max_i\{a_i\} \leq 2\bar{a}$. I would like the author to provide an economic meaning to this condition, if any. Same for the assumption that the distribution of a is skewed to the right ($a_{med} \leq \bar{a}$), few lines later.

Typos

The text contains several typos. They are too many to be reported here in a way that could help the author implement my suggested changes. Therefore, I have decided to indicate them **directly in the pdf of the paper**, so as for the author to more easily identify the suggested change in correspondence of each typo.

References

- Attanasi, G., Battigalli, P., and E. Manzoni (2013). Incomplete Information Models of Guilt Aversion in the Trust Game. IGIER Working Paper, Bocconi University.
- Attanasi, G., Battigalli, P., and R. Nagel (2013). Disclosure of Belief-Dependent Preferences in the Trust Game. IGIER Working Paper, Bocconi University.
- Dufwenberg, M. (2002). Marital Investment, Time Consistency and Emotions. *Journal of Economic Behavior and Organization*, 48, 57-69.