

Referee Report AER-2019-1641

The article studies beliefs about, and preferences toward, carbon taxation using a representative survey sample and an associated survey experiment in France.

The authors find:

1. Individuals largely oppose a carbon tax with lump sum transfers, despite most citizens benefitting from the policy.
2. Opposition is a result of respondents estimating *losses* from the policy, even when respondents are estimated to gain from it.
3. The degree of overestimation of losses is strongly associated with respondent support for the Yellow Vest movement, suggesting either belief-driven political affiliation or politically-motivated belief formation.
4. Experimental assignment to a hypothetical policy producing private benefits leads to increased beliefs that the policy will benefit survey respondents, and also to increased support for the policy.
5. IV estimates based on (4) suggest that correctly-calibrated beliefs would generate overwhelming support for the carbon tax.

This is a fascinating paper, which provides compelling evidence that belief distortions are associated with citizens' failure to support policy outcomes that would make them better off.

However, the article has two clear weaknesses that make it unsuitable for the *AER*:

1. Frontier research implementing survey experiments examines real-stakes subject choices related to the policy question of interest (see work by Leonardo Bursztyn, for example), while responses in the current experiment are all low-stakes.
2. The main empirical result – that recalibration of beliefs would be sufficient to generate “correct” policy choices – is interesting, but does not speak directly to the conceptual analysis, which emphasizes that beliefs may be an outcome of preferences. The authors suggest that belief changes can generate large effects on policy support; but the conceptual question is what would it take to shape *preferences* in a manner that would produce changed *beliefs*? The article examines the *local* effects of experimentally manipulated changes in hypothetical, private incentives. But this is not very informative about the possibility of preference change more generally (e.g., among “never takers” in the IV analysis, and at different parts of the belief distribution), which is crucial to interpreting the elasticities estimated with respect to beliefs.

In sum, I think this is a very nice paper and should be published in a good journal (e.g., *JEEA*, the *EJ*, or a political science journal) but I do not believe it is up the standard expected at the *AER*.