The social construction of non-transitive preferences

Duncan Law

March 18, 2024 - March 21, 2024

Abstract

Most rational choice theory assumes agents with transitive preferences. This strong assumption has sometimes been used as grounds for a critique of the rational choice paradigm, on the basis that real-world individuals frequently do not exhibit preference transitivity. However, individuals with non-transitive preferences can readily be constructed within a rational choice framework by two mechanisms. First, the individual can be treated as itself an aggregation of different sub-individual agents. Second, the individual can form its preferences via a process of aggregation of other individuals' attitudes. Both mechanisms can be modelled within a rational choice framework, provided at least one of two additional assumptions are adopted: a) the individual is not presupposed as the level of analysis at which the attribute of transitivity is applied, or b) expressed preferences are treated as derivative of more basic social motives.

Note

This document emphatically falls within the 'thinking things through as I read' category, and I am fully confident that these issues have been voluminously addressed in literature I have not yet looked at. However, I think it is worth writing thoughts up in a semi-formal way as I go, hence this document.

1   Transitivity of preferences

Mas-Colell, Whinston, and Green (1995) define transitivity of preferences as follows:

> For all $x, y, z \in X$, if $x \succsim y$ and $y \succsim z$, then $x \succsim z$

Here the symbol $\succsim$ represents the weak preference relation "at least as good as", where $\succ$ denotes "preferred to" and $\sim$ denotes "indifferent to".

It is a defining feature of a rational actor's preferences that they are transitive in this sense. If an actor's preferences are non-transitive - if for example $x \succ y$, $y \succ z$, and $z \succ x$ - then the actor is not rational.

This attribute of transitivity does not, however, transmit to preferences aggregated across actors: rational individual preferences do not result in rational social preferences. The Condorcet paradox illustrates this point by considering three social actors, all of whom have transitive preferences, where the aggregation of those preferences by majority voting results in non-transitive collective preferences. Thus:

For voter A, $x \succ y$ and $y \succ z$
For voter B, $y \succ z$ and $z \succ x$
For voter C, $z \succ x$ and $x \succ y$

For the community as a whole, majority voting leads to the non-transitive preference structure:

$x \succ y$, $y \succ z$, and $z \succ x$

Arrow showed that, given a set of plausible restrictions, this result generalises: there is no social choice aggregation mechanism which can translate individual-level transitive preferences to community-wide transitive preferences.

What are the implications of these ideas for the analysis of individual preferences? One approach - the dominant approach in the rational choice tradition - is to fiat by methodological assumption that the individual is the level at which the attribute of transitivity of preferences is applied. But this assumption is not obligatory. We can treat the individual as itself a result of preference aggregation.

2   The divided self

Assume that the individual psyche is composed of multiple subcomponents, each of which exhibits transitive preferences, and that the individual is an 'aggregation' of these subcomponents. The unity of the individual could be called, using Kantian language, a 'synthetic unity', which has been forged in the activity of the thinking and acting subject.

Reproduce the Condorcet scenario with this assumption. Each actor is now a subcomponent of the psyche - one could call them, for convenience, the Id, Ego, and Superego (Freud, 1923). (Of course, the subcomponents in this toy model do not exhibit the features of those elements of the Freudian account.) Assume that the aggregation mechanism is again simple majority rule.

This aggregation mechanism can, if we like, itself be understood as a preference function: imagine that the individual as a whole gains utility by following the preference of a subcomponent of their psyche. The subject wishes, that is, for their actions and decisions to 'integrate' or 'synthesise' as much of their psychic economy as possible. The synthesising individual's goal when confronted with any choice pair is therefore to maximise their utility by maximising the number of psychic subcomponents whose preferences are followed.

Now we have produced an individual with non-transitive preferences on

the basis of rational choice theory, simply by abandoning methodological individualism.

3   The socially constructed self

We can achieve the same result by treating the individual's preferences as aggregations of the preferences of others in their social environment. There is sociological reason to treat preferences in something like this way - Bourdieu (1987), for example, argues that taste should be understood as formed relationally, albeit via the implicit structure of the *habitus* rather than via rational choice decision principles. We are social creatures, and taking our bearings from our social milieu is a central dimension of self-formation.

Consider, then, an agent whose preferences are formed by the aggregation of the preferences of those in their immediate milieu. Again, reframe the Condorcet paradox in these terms. We now have four agents, three of whom exhibit the preference structures attributed to the voters in section 1 above. The fourth agent behaves according to a simple 'majority rule' opinion dynamics decision principle (Galam, 2002). That is to say, in any pairwise choice our target agent simply adopts the majority opinion of the other three agents. Again, this results in our agent exhibiting non-transitive preferences.

Like the 'divided self' model, this 'socially constructed' model can be construed in rational choice terms. Imagine that our agent is keen to enjoy the reputational utility associated with the approval of their peers. Our agent values the opinions of all three peers equally - agreeing with each peer brings our agent equal utility. In any pairwise decision scenario our agent's transitive, rational preference function results in them adopting whichever opinion is held by a majority of their peers.

In this scenario we are treating all our agents as rational, in the sense of exhibiting transitive preference structures. At the level of *approval* our target agent's preferences are fully rational. But this approval-seeking behaviour results in the practical *adoption* of non-transitive attitudes.

This result is of course an artefact of the agent's expressed preferences not being 'pure' or entirely 'individual' preferences. The agent's 'real', transitive preferences relate to social approval; this desire for social approval then leads to the 'downstream' adoption of 'aggregated' preferences, informed by the individual's social milieu. A critic of my approach might object that this is not what we mean - or ought to mean - by preferences.

It seems to me, however, that this dynamic captures a real social and psychological phenomenon. Not only is it psychologically familiar that our preferences may be guided as much by peer approval and disapproval as by some pre-social private preference structure, there are also good philosophical reasons for believing that our attitudes are *in general* inescapably shaped by our social milieu (Brandom, 2019). Separating out 'real' preferences from the preferences formed by our sympathetic inhabiting of others' perspectives is an impossible task.

4 Combining the models

In the previous two sections I have outlined two simple ways of constructing non-transitive individual preferences out of a rational choice framework. Schematically, these approaches involve:

- Treating the agent's (non-transitive) preferences as an aggregation of the (transitive) preferences of subcomponents of the agent's psyche.

- Treating the agent's (non-transitive) preferences as an aggregation of the (transitive) preferences of other social actors.

These approaches can of course be combined. Thus, for example, the subcomponents of the agent's psyche can themselves be understood as formed by the aggregation of the preferences of other agents. Indeed, on the psychoanalytic account, this is how the divided self is formed: introjection of the attitudes of others (paradigmatically but not necessarily the parents) is formative of the (divided) self.

Moreover, if we treat every agent in the model 'symmetrically' - that is, as an 'aggregation' of different preference attitudes - then it may well be that the preferences that are formative of subcomponents of the psyche are already themselves non-transitive, prior to their 'introjection'. In this scenario, we can quickly find ourselves in a situation where non-transitive preferences are 'rationally' aggregated via transitive preference-aggregation functions, resulting in a dizzying layering of transitive and non-transitive preferences.

5 Conclusion

In this document I have briefly outlined a specific approach to analysing individual preferences. I have retained the rational choice assumption of transitivity of preferences but abandoned methodological individualism, applying rational preference structures to sub-individual levels of the psyche.

This approach may prompt the question: why bother? What is gained by reformulating debased versions of Freudian or Bourdieuian ideas in the idiom of rational choice theory? The answer to that question may well be "not much" - but the toolkit of rational choice theory is sufficiently powerful, extensive, and widely used that this domain seems worth exploring, if only for the connections doing so may open up between otherwise quite distant areas of the social-scientific enterprise.

References

Bourdieu, Pierre (1987). *Distinction: A social critique of the judgement of taste*. Harvard University Press.

Brandom, Robert B. (2019). *A spirit of trust: A reading of Hegel's phenomenology*. Belknap Press.

Freud, Sigmund (1923). "The Ego And The Id (1923/1989)". en. In: *TACD Journal* 17.1, pp. 5–22. ISSN: 1046-171X. DOI: 10.1080/1046171X.1989.12034344.

Galam, S. (Feb. 2002). "Minority opinion spreading in random geometry". en. In: *The European Physical Journal B* 25.4, pp. 403–406. ISSN: 1434-6028. DOI: 10.1140/epjb/e20020045.

Mas-Colell, Andreu, Michael Dennis Whinston, and Jerry R. Green (1995). *Microeconomic theory*. Oxford University Press.