Can welfare economics avoid paternalism after the behavioural turn?

In economics, choices, policies, and institutions are evaluated by how well they serve individual preferences. If everybody in the office prefers full fat milk, the welfare economist would recommend we get full fat rather than skim milk to stock the office fridge. This approach is given an anti-paternalist rationale. However, orthodox welfare economics faces a challenge from behavioural economics. My milk choices depend a lot on the default setting in the coffee machine, or on whether I have just been reminded by my surroundings of the need to live more healthily. ***Johanna Thoma*** *asks, if we can't derive a consistent preference from people's choice behaviours, can we still live up to the anti-paternalist ideal?*

Traditionally, welfare economics has proceeded roughly as follows. We start with the idea that, on the basis of people's observed choice behaviour, we can assign preferences to them. If choice behaviours abide by various consistency conditions, we can represent these choices with a preference relation that has all the features standardly assumed in economic theory: It is stable, context-independent, and abides by the axioms of rational choice (Sugden, 2018, p. 7). For instance, if I never pick skim milk to put in my coffee when full fat milk is available, and always pick full fat milk when only skim and full fat are available, I can be ascribed a stable and context-independent preference of full fat over skim milk. Having thus ascribed stable, context-independent and consistent preferences to people on the basis of their choice behaviours, orthodox welfare economics proceeds to use these preference relations as a welfare standard: Choices, policies and institutions are evaluated by how well they serve these preferences. For instance, if everybody in the office shares my milk preferences, the welfare economist would recommend we get full fat rather than skim milk to stock the office fridge.

This approach is commonly given an anti-paternalist rationale: The preference relation is assumed to capture what an agent herself takes to serve her interests best. This seems like a plausible assumption when people's choice behaviours are consistent, and we have no reason to believe that they have false beliefs about crucial aspects of the options available to them. By choosing policies that aim to serve people's preferences, we are thus deferring to their own views of what is in their interests, rather than imposing some external, objective standard of what is good for them. This is just what the anti-paternalist wants.

Orthodox welfare economics faces a challenge from behavioural economics, due to the well-documented and by now widely known examples of choice behaviours that systematically defy representation with stable, context-independent and consistent preference relations. Whether I choose skim or full fat milk may, for instance, depend a lot on the default setting in the coffee machine, or on whether I have just been reminded by my surroundings of the need to live more healthily. If we can't derive a consistent preference relation from people's choice behaviours, can we still live up to the anti-paternalist ideal?

Many behavioural welfare economists are optimistic that we can still do so, at least in the ways that matter most: We can still identify people's subjective interests and override their choices only where these are shown not to serve those interests well. In their optimism, many behavioural welfare economists have presumed that people's true subjective interests must still be representable with a stable, context-independent and consistent preference relation. (*Attempts to explicitly reconstruct latent preferences include Bleichrodt et al. (2001), Bershears et al. (2008), Köszegi and Rabin (2007), Manzini and Mariotti (2012) and Salant and Rubinstein (2008). That reconstruction is possible is presumed by much of the wider literature. Famously, Thaler and Sunstein's (2008) libertarian paternalism claims to intervene only so as to help agents achieve what is best for them, 'as judged by themselves', p. 5.)*

What Bob Sugden, in his 2018 book *The Community of Advantage,* calls the 'New Consensus' is committed to the idea that we can assign latent, true preferences to agents even if they choose inconsistently, and that agents who display outward choice behaviours inconsistent with classic rational choice theory can still, in that sense, be said to have an 'Inner Rational Agent'.

Date originally posted: 2022-06-06

Permalink: https://blogs.lse.ac.uk/businessreview/2022/06/06/can-welfare-economics-avoid-paternalism-after-the-behavioural-turn/

Blog homepage: https://blogs.lse.ac.uk/businessreview/

Take again the case of me inconsistently choosing different kinds of milk in my coffee on different occasions. According to the New Consensus, even if I sometimes choose skim milk and sometimes choose full fat milk, there is still a fact of the matter as to which I truly prefer, and we have practical ways of ascertaining it. But this example already serves to illustrate the basis on which the New Consensus has been criticised. (Next to Sugden 2018, also see Rizzo and Whitman 2020.)

It is often far from clear which of a set of inconsistent choices is more authentically reflective of an agent's true interests. And simply assuming that she must have, for instance, truly preferred the healthier of two options amounts to a more problematic kind of paternalism than most economists want to commit to. In my recent paper, I moreover point to a dilemma for those working with the idea of true latent preferences.

Either these represent people's actual better judgements… *(But then all cases of context-dependent deviation from true preference are cases of weakness of will, of acting against one's own better judgement. This is simply psychologically unrealistic in many cases of context-dependence.)*

Or they represent hypothetical preferences, preferences I would have if I was ideally rational. (*But it is unclear whether such hypothetical preferences are subjective in the way the anti-paternalist would want them to be, that is, if they stand in the right kind of relationship to what I, inconsistent self, actually want.)*

Sugden himself takes the failure of the New Consensus to motivate a more radical rethinking of normative economics: We should not even try to rank options in terms of people's subjective interests. Rather, we should only try to increase their opportunities for choice. But I do not think we have to go that far. Consider these two intuitive ideas:

1. Preferences are not some primitive mental state that forms the starting point of deliberation. Rather, preferences are the result of weighing off various considerations that speak in favour of or against a certain choice. For instance, in my choice of milk, I might be weighing off considerations of taste and health.
2. Many of the basic desires, on the basis of which we form preferences, e.g., desires for health or tastiness, are vague. And there may be no uniquely correct way of aggregating them. As a consequence, there may not be a preference relation that uniquely captures what is in your subjective interests – there may be a number of permissible ways of trading off the things that matter to you.

The practical implication of these two ideas for welfare economics is that inconsistency in an agent's choice behaviour may not be the result of any mistake. It may simply be the result of factors in our environment causing us to trade off the things that matter to us in slightly different, but equally permissible ways on different occasions. And so, unless we have specific grounds for thinking some of an agent's inconsistent set of choices involve a mistake (e.g., are based on false beliefs), we should presume each of a set of inconsistent choices expresses a permissible way for an agent to serve her interests.

Where does that leave the welfare economist, whose aim is to arrive at some measure of subjective interest to use as a standard to evaluate policies by? She may need to accept that there can be indeterminacy in her measure of welfare. But that does not need to mean that she has nothing to go by: While we may need to treat it as indeterminate whether skim or full fat milk serves my interests best, I may quite consistently choose either over cream. Our best attempt at an anti-paternalist welfare economics involves deferring to the coherent aspects of people's choice behaviours when evaluating policies but accepting indeterminacy where context-dependence does not stem from a clear and demonstrable mistake. My paper argues that a choice-theoretic framework developed by Douglas Bernheim and Antonio Rangel (Bernheim and Rangel 2007 and 2009, under the revised interpretation recently offered by Bernheim 2016) can in fact be interpreted in just this way. (Douglas Bernheim also defends his framework along similar lines against Sugden's critique.) Adopting this framework will mean that in practice, compared to the New Consensus, there will be fewer occasions where welfare economists will support overriding people's choices, even where these violate orthodox rational choice theory. But this is as it should be if we take economics' traditional anti-paternalist commitments seriously.

Notes:

- *This blog post is based on On the Possibility of an Anti-Paternalist*

Date originally posted: 2022-06-06

Permalink: https://blogs.lse.ac.uk/businessreview/2022/06/06/can-welfare-economics-avoid-paternalism-after-the-behavioural-turn/

Blog homepage: https://blogs.lse.ac.uk/businessreview/

Behavioural Welfare Economics, Wellbeing and Human Development Project, LSE's Centre for Philosophy of Natural and Social Science (CNPSS) and Journal of Economic Methodology 28 (4), pp. 350-363.

- *The post represents the views of its author(s), not the position of LSE Business Review or the London School of Economics.*
- *Featured image by Trent Erwin on Unsplash*
- *When you leave a comment, you're agreeing to our Comment Policy*

Date originally posted: 2022-06-06

Permalink: https://blogs.lse.ac.uk/businessreview/2022/06/06/can-welfare-economics-avoid-paternalism-after-the-behavioural-turn/

Blog homepage: https://blogs.lse.ac.uk/businessreview/