A call for epistemic analysis of cultural theories for AI methods

Masoumeh Mansouri[1]

Robots are not any more only in factories. They are guides in museums or airports, waiters in restaurants, and will soon become carers: robots begin to operate in environments that have hitherto been only human territory. Many argue that for these robots to be trusted and accepted by humans, they should be culturally aware/-sensitive/-competent, i.e., robots should have "culture". In such arguments, we read that Arabs prefer a robot standing closer to them than Germans do, as this interpersonal distance is following the social norms of their respective cultures. Chinese are more likely to accept recommendations from a robot that communicates implicitly as opposed to Americans, who mostly heed explicit advice from a robot, or Americans also feel more comfortable towards anthropomorphic robots than Japanese who rather prefer more conventional types of robots.

Social robotics sees an increasing number of studies suggesting people prefer robots that comply with the social norms of their own culture. The vast majority of these studies introduce cultural factors into robotics by relying on what is perceived to be a national culture. Although it seems only fair to roboticists to use nations as heuristics to define systems of knowledge, beliefs, behaviours and norms used by large groups of people, using this heuristic has ramifications. It ignores cultures within a nation-state that are sometimes more distinct from the dominant culture than cultures of other geographically distant nation-states. In addition to ignoring marginal cultures within a nation-state, the nationality-based stance of culture fails to recognise those that fall outside of this definition, e.g., stateless persons. Also, it promotes the need for assimilation and abandonment of marginal culture in favour of a generalised culture. Moreover, the oversimplified confounding of culture and nationality may result in implicit support for conservative social policies, and the reproduction of cultural stereotypes.

✉ Masoumeh Mansouri
   m.mansouri@bham.ac.uk

1   School of Computer Science, University of Birmingham, Birmingham, UK

So if nations are not sufficient to define culture, then these important questions arise: *what is culture?* and *how should roboticists endow robots with culture?*

Fortunately, we have no scarcity of definitions that take account of more than national differences. The book "Redefining Culture" written by Baldwin et al. in 2006 lists 313 definitions from disciplines including psychology, linguistics, anthropology and political science to philosophy, to name a few. Although these definitions overlap in content, each implicates a different design or set of techniques for use within AI/robotics, and may not even be reifiable in robots. The latter is because there are many knowledge representations, automated reasoning, and machine learning techniques, each of which has a different expressivity and varying computational capacities. Not all of them can be harnessed for representing a given interpretation of culture. In this article, I advocate for an epistemic analysis of cultural theories through the lens of AI/robotics methods, as an essential step towards endowing robots with culture. I argue that such an analysis not only allows us to identify which fundamental theories of culture are candidates to be programmed into, or learned by robots, but also provides transparency in how an intelligent robot behaves in different cultural contexts.

For an epistemic analysis of such a pool of definitions, we need to delineate the major themes. As can be imagined, defining such themes is highly debatable depending on which academic areas you are from, so I will stick to those introduced by Baldwin et al. The *structure/pattern* theme looks at culture in terms of a system of ideas, behaviour, symbols, or any combination of these. For instance, in this theme, culture can be seen as a cognitive structure inside the minds of the individuals in a community of people, or "whole way of life" such as stereotyped patterns which are handed from one generation to the next through the means of language and imitation. Many anthropologists hold this view of culture. As a case in point, Goodenough on cultural evolution in 1961 referred to culture as "standards for deciding what is … what can be … what one feels about it … what to do about it, and … how to go about doing it".

The *function-based* definitions see culture as a tool for achieving some end like culture as a function to provide people with a shared sense of identity/belonging, or of difference from other groups. Culture as *process* focuses on the ongoing social construction of culture. In this theme, culture is framed as the process of sense-making, producing group meaning, or that of relating to others. Culture is defined as an artefact for the *product* theme, for instance, art, architecture or books. The *refinement* theme frames culture as a sense of individual or group cultivation of higher intellect or morality. In this theme, culture can be interpreted as any human efforts to distinguish humans from other species, which can encapsulate some other previous definitions as well. Definitions based on *power or ideology* move the focus from what culture is or how it arises to questions of whom it serves. For this theme, culture is not the artefacts of a group, rather is concerned with the dominant politics and ideology of the group. Finally, the *group-membership* theme speaks of culture concerning a place or group of people, or belonging to such a place or group, e.g., country or identity.

All these themes are within the scope of human culture. The current aim of the field on defining culture for human–robot interactions is for them to be culturally consistent with human–human interactions, and yet, human culture is reduced to a single dimension, nationality, for most of this robotics research. Although these themes overlap in what they consider culture to be, with certain caveats they can be a useful guide for our epistemic analysis of what culture could mean for robots in interactions with humans. Now, let's analyse a possible implication of some of these themes for AI and robotics research.

Within the culture as structure/pattern theme, Hofstede's theory has received significant attention from roboticists. For him, "culture is a code we learn" or "a program for behaviour". The 'national culture' is also a formulation of his theory. However, implementing a "cultural" code for robots is not straightforward. Let's assume the cultural code that a robot wants to learn is 'eating habits of people who live in England'. To learn the national habit, based on this specific theme of culture, we should have sufficient data collected in many contexts that feature different types of behaviours. The domain of these contexts can be huge: spatial context of eating such as the various rooms of a house, indoor or outdoor, which region or city, etc.; temporal context like the time of day, day of the week, occasion, etc.; or the emotional context such as celebration or grieving. The list can go on. These contexts within which cultural patterns are shaped go well beyond, for instance, our shopping behaviours as manifested in social media, based on which companies like Google or Facebook learn a pattern. One could argue that we can employ machine learning methods only if we isolate the context and behavioural traits for part of the history we have data for, e.g., online food or restaurant shopping, and correlate them to non-shopping data e.g., location data via Google Map. In this way, stereotypically speaking, the robot may learn that English people eat curry on Friday nights at their homes. However, learning this cultural habit is not at all straightforward. The first problem with such an approach is the issues related to the legality and ethics of collecting such a huge amount of personal data, especially the non-commercial ones. The second problem concerns the technical adequacy of the machine learning methods to reason about the existing correlations among the data and deduce meaningful causal relations. If the current methods turn out to be technically inadequate, we cannot effectively learn the eating habits of England under this definition of culture. In summary, the analysis of this example serves to illustrate how a seemingly computer-friendly definition of culture, e.g., Hofstede's, must be butchered to find use in robotics and AI.

As an alternative to learning methods and their problem of lacking ample comprehensive data, logic-based AI methods can be used to encode cultural norms, behaviours or beliefs within the limit of their expressivity and computational capacities. For example, we can encode the breakfast preferences of a British pensioner using a non-monotonic logic, as the preference might change over time. Then, the problem lies with who gets to code the meanings and what counts as culture. If we commit to a power/ideology theme of definitions, one implication might be to ascertain if our implementation does not privilege some groups over others.

Similarly, seeing culture as a product or a process has a great impact on how we develop cultural robotics. Within the process theme, robots must be active participants in creating culture, indeed culture is *the process* of interactions among robots and humans. For instance, based on this definition, if we want to set up a cultured robot in a factory to help human co-workers in assembly tasks, we do not program culture into that robot. Rather, we equip the robot with the capability to interact and adapt over time to the habits of human co-workers. In this context, the nationality or any identity of human co-workers is irrelevant to the process of learning culture. Whereas culture as a product does not necessarily implicate the co-participatory role of robots in the making of culture. In the example before, robots should already be equipped with some sort of interaction culture, such as a specific way of task coordination and movements, prior to setting up in the factory under the *product* definition.

To conclude, I do not provide these examples to pick one out as the most suitable definition of culture for robotics. I rather aim to emphasise the multiplicity of interpretations and the necessity of analysing them epistemically in relation to AI/robotics methods. Through this analysis, we can assess, for instance, whether our robotics development can lead to the marginalisation of minority groups or contribute to the further propagation of stereotypes. Leaving the moral

responsibility aside, I believe that the interdisciplinary epistemic analysis of culture and robotics will stimulate fresh thinking on existing challenges, and pose whole new ones.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.