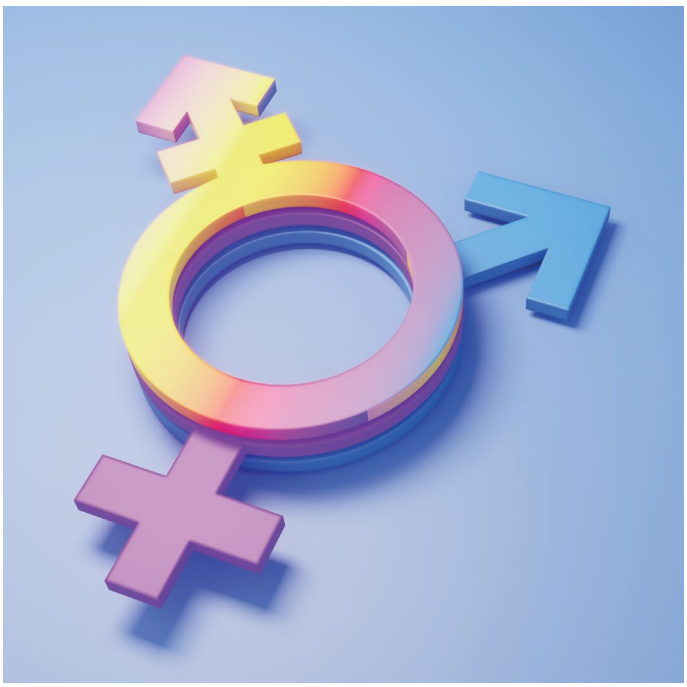
https://doi.org/10.1038/s41562-023-01657-y

Trans-inclusive gender categories are cognitively natural

Andrew Perfors, Steven T. Piantadosi & Celeste Kidd

On the basis of decades of cognitive science research into the nature of lexical concepts, we argue that gender categories that reflect the reality of the experiences of transgender people are more useful and cognitively natural than sex-based category definitions.



The nature of human concepts has recently become central in public discourse. Some prominent thinkers have argued that transgender people deny both the biological reality and the real meaning of words such as 'man' and 'woman' when they categorize themselves on the basis of gender identity rather than their assigned sex at birth (for example, refs. 1,2). We argue that the question of what words such as 'woman' and 'man' mean falls squarely into the domain of cognitive science, which has spent more than half a century investigating the nature of lexical concepts[3]. In this Comment, we review relevant parts of this scientific literature with the hope that it can inform these public debates.

Many people make a distinction between gender and sex: sex-based categories are defined on the basis of biological features, whereas gender-based categories reflect social roles or internal identity. The gender-versus-sex distinction gives rise to the question of whether sex-based categories are inherently better or more objective than gender-based ones. We argue that research shows that the answer to that question is 'no': trans-inclusive words are natural in that they match the form of most of our concepts and effectively communicate meanings that are central to people's lives.

Problems with sex-based categories

A sex-based definition of 'man' and 'woman' is neither as objective nor as simple as intuition might suggest. Attempts to provide objective binary classifications of 'man' and 'woman' in terms of physical features immediately run into the issue that humans are biologically more complex than is permitted by a binary classification. For example, there are a variety of intersex conditions that may account for as much as 1–2% of the population[4]. Moreover, the physical features that many people would intuitively think are necessary (for example, genitalia, chromosomes or childbearing roles) are difficult to use as strict binary definitions, owing to human variation in these properties: people can lose their physical features through surgery or accidents, have unusual karyotypes, be infertile or never have children, and so on.

The difficulties with strict binary definitions are not surprising. Human lexical concepts are almost always vague, fuzzy and graded[3,5], as well as impossible to specify precisely. Even mathematical or scientific concepts, which may be created to have strict definitions, are fuzzy in how we think about them[6]. Similarly, it is easy to think of compelling counterexamples for almost any strict definition — unmarried men who do not seem like bachelors (for example, the pope) or fruits that do not seem like fruits (for instance, tomatoes). These cases also illustrate how categorization invariably rests on convention: there is no truth of the matter about whether the pope really is a 'bachelor', only a social agreement about what we use the term to mean.

This conventionality can be seen in sex and gender concepts as well. Biologists, for example, decided on a certain criterion for defining sex across species: females are the ones with larger gametes. Notably, there is no inherent biological truth about what is the 'right' criterion to use in this division, or even if it always makes sense. Thinkers who reject the concept of gender adopt this definition for humans (for example, ref. 2). Others consider it among several other definitions, all of which they claim "divide humans unambiguously into one of two categories" on the basis of essential properties that are present at birth[1]. These properties do not include genitalia or hormones, which can be acquired by transgender people through medical intervention[1].

This sort of definition is also unusual because few, if any, human categories correspond cleanly to 'objective' or 'unambiguous' partitions of the world. As one example of many, colour and pitch are determined by unidimensional physical quantities — frequencies of light and sound — but our conceptual system does not code them that way. Violet and red look similar despite being opposite ends of the spectrum, and the same notes at different octaves sound more

similar than different notes that are closer in frequency. Conversely, many distinctions that are real in the world are not real in our head: the word 'beetle' refers to thousands of phylogenetically distinct species and a 'koala' is thought of as a kind of bear even though it is genetically more similar to a wombat.

More broadly, the immense cross-linguistic variation that exists in human lexical systems is impossible to reconcile with the assumption that lexical concepts are only sensible or useful to the extent that they directly map onto an objective feature of the physical world. For example, languages vary considerably in how they categorize phenomena such as colour, spatial relationships, family relationships and physical properties such as containment. Even seemingly platonic ideas such as number systems exhibit notable cross-linguistic diversity. The history of words also documents their conventionality, as meanings evolve in response to changing communicative needs. 'Barbers' used to do the work of surgeons and dentists, and the Old English form of 'wife' referred to any woman, not only a married woman.

Socially defined gender categories are cognitively natural

We are not saying that there is no objective reality. Rather, our point is that social role is an important aspect of what people need to communicate. This is reflected by the fact that category terms used for people often reflect their social roles ('friend', 'spouse', 'accountant' or 'cryptobro'). Gender concepts, too, reflect social organization, resulting in languages and cultures that recognize more than just 'woman' and 'man.' Examples of these include *kathoey* in Thai, *māhū* in Hawaiian, *fa'afafine* in Samoan, *femminielli* in Neapolitan, all of which refer to a gender category that does not fit into a binary classification and has a long history of use in each language. Within English, variations in gender terms (for example, 'miss', 'widow', 'husband' and 'fiancé') often highlight social information such as marriage status. Even words such as 'mother' and 'father' prioritize social role over biology, as they apply equally to adoptive parents.

It is no accident that gender concepts are conventionally rooted in social roles: this is precisely the thing that makes them useful. Social relationships are central to human life in every culture. Hidden properties — such as gamete status or chromosomes — are impossible to perceive as well as irrelevant to anybody except a doctor; they are therefore not an adequate basis for assigning useful labels in day-to-day life. A transgender person is no more making a claim about the size of their gametes when they state their gender than a cisgender person is when they state theirs. Instead, both are communicating a social role.

If gender concepts, with their fuzzy boundaries and grounding in social roles, are cognitively natural, why do many people have the strong intuition otherwise? One important factor is likely to be a well-documented cognitive bias known as essentialism[7], which is the belief that concepts have an unobserved core that is responsible for making them what they are. Humans are essentialists about gender starting from a young age[8].

Unfortunately, essentialist logic is often illusory: we feel as though there is an objective core to concepts such as gender — or race or morality — even when there is not one or we have no clear idea what the core could be. Because word meanings do not, at face value, seem conventional, we are overconfident in the objectivity and correctness of our own individual definitions[9]. Thus, although essentialist reasoning may make sense in some situations, it is ethically dangerous to apply to other humans where the conceptual distinctions we make truly matter. Essentialist reasoning has often been used to support discrimination through claims of inherent (often biological or physical) differences[10]. We see this in some works[1,2], which argue that transgender women do not deserve the same protections as cisgender women because they are 'essentially' men — even if they have lived as female since childhood, even if hormones and surgery have greatly altered their body, and even if their social experience has been almost indistinguishable from that of cisgender women.

Implications for policy and society

How should our understanding of categories and concepts inform societal decisions? This is much less clear as it is as much about morality as it is about science. We simply offer a few observations.

First, any justification for using sex-based categories cannot be based on the idea that there is something incoherent or unscientific about gender-based categories. Across languages and cultures, all lexical concepts are conventions that are heavily shaped by communicative need, and people clearly need to communicate social roles and identities.

Second, if lexical concepts are primarily conventions, this means that we should choose conventions that are useful. The usefulness of sex-based categories in domains such as medicine or issues such as participation in sports is often raised as an argument for their value. However, although biological factors certainly matter for some situations, use of these sex-based categories in broad public policy frequently runs into trouble. Gametes or chromosomes are irrelevant in most situations where policy matters — not only do laws prevent sex-based discrimination, but also we almost never know a person's genetics or gametes because they are not easily observed. Some physical features correlate with these things (for example, muscle mass or hormone levels) but exceptions abound, men and women overlap considerably on any given characteristic and many people do not fall into a simple binary classification. As a result, sports bodies sometimes organize along nongender-based dimensions that are related to performance, such as weight classes in boxing or ages in children's sports. Appropriate medical decision-making requires not only knowledge of natal sex but also information about gender and direct measurement of the relevant physiological variables[11]. Rather than basing policy on blunt proxies such as natal sex, decisions should be left to the people who have expertise and knowledge of the directly relevant factors — such as individual patients and their doctors.

Third, the choice of how we use words such as 'man' and 'woman' has real consequences. The debate is not actually about metaphysics or maintaining some imagined objectivity in our lexicon; it is about what rights transgender people should be granted. How we use these words affects the ability of transgender people to be legally protected from losing their job on the basis of their identity, to access the medical and social services they need, to use accurate documents for identification, to have personal privacy, to use public facilities safely and to be treated as equal citizens. Fighting against trans-inclusive language is not fighting for science; it is fighting to deny people civil rights.

Ironically, there is one aspect of biological reality that we feel has been overlooked: the brain. If there is an essence to humanity at all, most people would place it there. Our ideas, our thoughts, our identities, our hopes and fears and loves — those are the things that define us. Trans-inclusive gender categories reflect these aspects of what make us all human, all of which are, of course, grounded in the physical biology of the brain.

Arguments that sex-based categories are more correct rely on the deeply unscientific presumption that our categories are precise and objectively aligned to the world, even though decades of empirical

work shows that this is false. Human lexical concepts are conventions that we choose, and they change as society changes. Changes towards trans-inclusive categories yield linguistic systems that are both natural and useful.

Andrew Perfors [1] ✉, **Steven T. Piantadosi** [2,3] **& Celeste Kidd** [3]
[1]School of Psychological Sciences, University of Melbourne, Parkville, Victoria, Australia. [2]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. [3]Department of Psychology, University of California, Berkeley, Berkeley, CA, USA.
✉e-mail: andrew.perfors@unimelb.edu.au

Published online: 20 October 2023

References

1. Stock, K. *Material Girls: Why Reality Matters For Feminism* (Hachette, 2021).
2. Lawford-Smith, H. *Gender-Critical Feminism* (Oxford Univ. Press, 2022).
3. Laurence, S. & Margolis, E. in *Concepts: Core Readings* (eds. Laurence, S. & Margolis, E.) 3–81 (MIT Press, 1999).
4. Blackless, M. et al. *Am. J. Hum. Biol.* **12**, 151–166 (2000).
5. Smith, E. E. & Medin, D. L. *Categories and Concepts* (Harvard Univ. Press, 2013).
6. Armstrong, S. L., Gleitman, L. R. & Gleitman, H. *Cognition* **13**, 263–308 (1983).
7. Gelman, S. A. *Trends Cogn. Sci.* **8**, 404–409 (2004).
8. Gülgöz, S., Alonso, D. J., Olson, K. R. & Gelman, S. A. *Dev. Sci.* **24**, e13115 (2021).
9. Martí, L., Wu, S., Piantadosi, S. T. & Kidd, C. *Open Mind* **7**, 79–92 (2023).
10. Mahalingam, R. *Hum. Development* **50**, 300–319 (2007).
11. Bale, T. L. & Epperson, C. N. *Neuropsychopharmacology* **42**, 386–396 (2017).

Competing interests

The authors declare no competing interests.