Gender identity as data: Explicating a critical tension within the philosophy of data

By: Lo Min Choong Julian | 18 April 2023 | Edited: 9 December 2023

Written for the partial completion of NTU's HY4110 Philosophy of Data

---

In this essay, I argue that gender identity as data presents us with a critical tension: there exist some classification systems that are simultaneously unsatisfying and inescapable. This dilemma arises when one takes the relational view of data and applies it to certain data, such as gender identity. This may be a central problem of the philosophy of data.

To begin, I shall present a data journey of gender identity. Then, I shall examine gender identity as data. I shall end by explicating a tension that arises and discussing its implications.

A data journey of gender identity

Before I discuss gender identity as data, I shall explicate a data journey of gender identity (Lo & Tan & Tan, 2023). Singapore practices assigned gender at birth [AGAB]. When a baby is born, medical practitioners assign said baby a gender (i.e. male or female, in Singapore) based on the information at hand, and on their medical knowledge. Ms Eliss Chen was assigned male, when in fact, she is female (Chua, 2023). All of her government records reflect her as "male", which is not her actual gender identity. When she grew up, she had to face up to her assigned gender in many aspects of her life—on social media accounts (because they ask if you're male or female), and on legal documents (driver's licence registration, proof of identity to institutions such as banks and universities).

This highlights the many types of interactions that are based on gender identity. For example, gender is taken as an identifier, even when other types of identification are available. This story also highlights the power imbalances social constructs have on people.

Ms Chen could not dictate her gender identity at birth. If she could, she would have had the ability to shape her life significantly.

The process of changing her legal gender to her actual gender is also a significant barrier. The journey is time-consuming, expensive, and mentally taxing. Furthermore, institutions can simply refuse to accept one's actual gender identity, and instead retain one's recorded legal gender identity. Thus, gender as a social construct is imposed on all of us, with or without our consent, because it is institutionalised.

By implementing the data journey format on Ms Chen's story, we are able to identify and enumerate the various impacts of gender identity.

The relational view of data

We are used to thinking of data as concrete objects, i.e. facts or information. However, this is not always the case. Consider the following questions: are all data of the same type? Do they all share some universal properties that make data data? In *Data and Society: A Critical Introduction*, Beaulieu and Lionelli argue that these questions frame data in a problematic way (Beaulieu & Leonelli, 2022, pp. 54-7). They argue such questions quietly hold the assumption that data are facts, things that describe the characteristics and properties of the world. They argue this is not the case. They argue this way of thinking is called the "representational view of data", as data represents characteristics about the world.

Beaulieu and Lionelli argue that data should be seen from the "relational view". The relational view of data argues that data are what we represent objects in order to generate knowledge. In reality, what we consider as data are human-made objects, rather than "pure" representations of the objects and phenomena of the world. This distinction is significant because if we treat data as pure, accurate representations of the natural world, the way we conduct data work is significantly different. For example, when a product of data work is being produced, data workers will take datasets as facts, never to be questioned critically. The

sources of datasets may be questions for reliability, but the data itself is never thought to be examined. This means that traditional data work fails to account for a critical dimension of data. Thus, we should question data itself.

However, before we are able to question data itself, we need to recognise the "gap" between the representational and relational view of data. This gap may be unrecognisable when dealing with certain data. For example, we are used to thinking about velocity in numbers, and taking said numbers as fact upon measurement. When we consider numbers that describe the physical world, numbers that have physical meaning, seeing such data via the representational view or relational view makes no difference, because there is no difference. This shows that for certain types of data, e.g. data that refers to physical phenomenons, the two views agree with each other.

Gender identity as data, on the other hand, provides us with a clear example of how data can be seen differently under the representational view versus the relational view. If we recognise gender identity as a social construct, and then use the relational view to examine such data, a stark contrast can be seen. If we took gender identity as references to concrete, biological phenomena, then the representational view of data holds. However, this is not the case. By arguing the "concrete, biological phenomena" is not as concrete as it seems makes gender identity data not as representational, but relational, as it refers to a nebulus, evolving concept. By viewing data under the relational view, we are reminded that we should not treat data as natural facts about the world.

Thus, a central problem in the philosophy of data is in regard to the relational view of data. Is there a gap between the representational and relational view of data when applied to different types of data? How can we recognise it? What changes when we recognise this gap? How will it impact different areas of data work?

Gender identity as a social construct

In *Data Feminism*, D'Ignazio and Klein argue that gender identity is a social construct (p. 100). That is, the classification of gender is human-made, and not scientifically based. Translating this into the relational view of data, the movement from the objects themselves and how the objects are represented as data is arbitrary. An attempt to "fix" gender is to link it back to sex, a biological phenomenon. However, the gender binary breaks down because of the nature of sex itself (D'Ignazio & Klein, 2020, pp. 113-7; Montañez, 2017). The science of sex determination is incredibly complex. Any attempt to reduce it to a binary will fail to capture the full dimensions of what sex is. Similarly, gender identity "fails" in some sense, because it is necessarily a classification system that was based on the gender binary. Gender identity as a classification is lacking in certain areas. One area is in the way it classifies intersex individuals. They may consider themselves non-binary, but non-binary refers to many distinct identities, apart from intersex individuals. Even the term, "intersex" itself fails in some sense because it glosses over significant details. This is the gap as seen via the relational view of data, versus the representational view. The representational view hides the nuances of biology and the need to question our biological models.

Scientists could construct a classification that perfectly maps all facts about one's sex and gender. But, do we need to capture all distinctions for our human-made, non-medical and desired purposes?

The problem is we require data to talk about concepts. Data is our means of communication. Without making gender identity into data, doctors are unable to provide accurate care. Thus, a defence of gender identity is of its medical purpose. Medical practitioners require knowledge of one's gender in order to accurately diagnose certain conditions (and thus conduct life-saving treatment). Thus, gender identity as data should remain as it is. However, should it really remain?

A critique of this defence is that gender identity as data is misappropriated. It goes beyond medical uses, such as legal identification and social media account creation. It seems that our frequent interactions with gender identity are primarily non-medical, even though our most significant ones are still medical. For example, on our NRICs (Singapore's National Registration Identity Cards, i.e. our national identification cards), we have our NRIC number, full name, gender, race, birthdate, and address, all of which are personal identifiers. Why exactly is gender being used as an identifier? This is a key question Data Feminism is posing.

Furthermore, medical practitioners have their own classification systems when it comes to gender and sex, as seen from the science of sex determinism. Arguably, gender identity as a classification is employed in medical settings as a short-hand, or when it suffices because most medical situations do not require rigorous and precise gender classifications. The rigorous and precise gender descriptions are employed in medical settings when it is appropriate, e.g. for gender dysphoria diagnosis (American Psychiatric Association, n.d.). Thus, the perpetuation of gender identity as an authoritative classification system is due to a non-medical need. What exactly is this need?

Gender identity as data

A motivation of perpetuating gender identity as a classification system is that we desire the ability to talk about gender identity. Data allows for this communication.

Data has two key aspects: "(1) it is treated as potential evidence for one or more claims about phenomena, and (2) it is possible to circulate it among individuals" (Leonelli, 2016; Leonelli, June 2016).

The denial of (1) is unconventional only from a traditional view of data. It becomes possible when taking the relational view of data, especially with data like gender identity. Looking at the history of gender classification, one can see that it did not emerge from science (D'Ignazio & Klein, 2020, pp. 100-4). Looking at studies of sex, gender identity and

its various terms are nowhere to be found (Montañez, 2017). Therefore, the authority of the gender identity spectrum is only derived from itself, i.e. its usage in everyday life. It is only because people take gender identity to be evidence about the phenomena of gender, that the classification system itself is credible.

While people may recognise the problem with (1), (2) is why we cannot dismantle gender as a social construct. Data is integral to the ability to communicate. Data necessitates the employment of classification systems. Therefore, gender identity remains.

Is this really the case? Does data really necessitate the use and retention of classification systems? Is it really not possible to retain the ability to communicate about x, without classifying x in a system?

Explicating the central problem proposed in this essay

I argue that this is a central problem in the philosophy of data. We desire and recognise the shortcomings of certain classification systems, but are unable to deconstruct them because we need them. This need is derived from the need to communicate about x. The way we communicate x is through data. Thus, taking Leonelli's definition of data (data having two key aspects), D'Ignazio and Klein argue that data necessitates classification systems (p. 103-4). Therefore, classifications necessarily arise, and are inescapable and unsatisfying.

The data journey explicates this for a specific token of classifications: gender identity. However, this tension is not exclusive to gender identity.

A possible "solution" is to remove gender identity from all legal documents and social media accounts. Both systems require it not because it is fundamental to one's identity, but because people desire additional identifiers. Gender identity is used as a short-hand of matching one's person to one's NRIC. Gender identity is also used as data in advertising, under the argument that the more data advertisers have on their target audience, the more

effective advertising messages will be (D'Ignazio & Klein, 2020, p. 100). Thus, since such data is inessential, it can be removed.

However, this solution is unsatisfying as people will still "see" gender, even if it is not explicitly stated in one's NRIC. Social media platforms can say they have removed all gender identity data from their algorithms and systems, but in reality have merely shifted gender identity data collection into the background, away from public view.

The need for enforcement is not the reason why such solutions are unsatisfying. It is because attempts to remove gender identity as a classification system from our lives will necessarily fail, to certain degrees. Attempts to remove gender identity do not address the root cause of why data and classifications exist in the first place. The need for data and classification systems supersedes the desire to remove gender identity from the parts of our lives that are not necessary.

Conclusion

This is exactly the tension I have set out to explicate. Data is essential to our contemporary lives because of our desire for convenience and communication. This necessitates data work and its infrastructures. This, in turn, necessitates classification systems. All of this makes it difficult to remove unsatisfying classification systems, such as gender identity, from certain parts of our lives. "What can we do?" is the appropriate response to this tension. It is a question that demands serious consideration, and further exploration into the philosophy of data.

References

American Psychiatric Association. (n.d.). Gender Dysphoria Diagnosis. https://www.psychiatry.org/psychiatrists/diversity/education/transgender-and-gender-nonconforming-patients/gender-dysphoria-diagnosis.

Beaulieu, A., Leonelli, S. (2022). Data and Society: A Critical Introduction. London: SAGE Publications. ISBN: 9781529732535. https://uk.sagepub.com/en-gb/eur/data-and-society/book269709.

Central Japan Railway Company (n.d.). SCMAGLEV [Website]. https://scmaglev.jr-central-global.com/.

Chua M. H. (31 March 2023). For this transgender woman, sharing her story is an act of faith in Singapore. The Straits Times. https://str.sg/iZKS.

D'Ignazio, C., Klein, L. F. (2020). Data Feminism. Cambridge, Massachusetts: The MIT Press. ISBM: 9780262547185. https://mitpress.mit.edu/9780262547185/data-feminism/.

International Bureau of Weights and Measures. (20 May 2019). The International System of Units (SI). https://www.bipm.org/documents/20126/41483022/SI-Brochure-9.pdf/fcf090b2-04e6-88cc-1149-c3e029ad8232.

Leonelli, S. (2016). Data-Centric Biology: A Philosophical Study. Chicago, IL: Chicago University Press. https://press.uchicago.edu/ucp/books/book/chicago/D/bo24957334.html.

Leonelli, S (23 June 2016). The philosophy of data. In Floridi, L. (Ed.), The Routledge Handbook of Philosophy of Information (pp. 191-202). https://www.routledgehandbooks.com/doi/10.4324/9781315757544.ch17.

Lo, M. C. J., Tan, K. H, Tan, P. M. R. (11 April 2023). Data Feminism [Presentation].

Montañez, A. (1 September 2017). Beyond XX and XY: The Extraordinary Complexity of Sex Determination. Scientific American. https://www.scientificamerican.com/article/beyond-xx-and-xy-the-extraordinary-complexity-of-sex-determination/.

Tom Scott (3 April 2023). I rode the world's fastest train [Video]. YouTube. https://youtu.be/4ZX9T0kWb4Y.