



# My objectivity is better than yours: contextualising debates about gender inequality

Rosie Worsdale<sup>1</sup> · Jack Wright<sup>1</sup>

Received: 1 November 2019 / Accepted: 10 August 2020 / Published online: 12 September 2020  
© The Author(s) 2020

## Abstract

In this paper, we contribute to a growing literature in the philosophy of social science cautioning social scientists against context-independent claims to objectivity, by analyzing the recent proposal of a new Basic Index of Gender Inequality (BIGI) by Gijsbert Stoet and David Geary. Despite the many internal problems with BIGI, Stoet and Geary have had some success in positioning the index as an important corrective to the way in which gender inequality is measured in mainstream metrics like the Global Gender Gap Index (GGGI). We argue that this success is facilitated at least in part by the failure of GGGI's proponents to adequately justify the methodological choices underpinning the index in relation to the context in which the index's findings are intended to be used. In so doing, the authors of GGGI oversell the objectivity of the metric's assessment of the state of global gender inequality—and it is this overselling that allows Stoet and Geary to present BIGI as a metric that corrects what they claim are systematic biases within GGGI. The case of BIGI and GGGI, we argue, suggests that the kind of epistemic modesty exhibited by recent operational approaches to objectivity is particularly important for social research on highly politically contested topics.

**Keywords** Objectivity · Gender inequality · Social indices · Context · Bias

Addressing the systematic inequalities faced by women around the world has become an important aim for policymakers and NGOs in the last few decades. Yet, according to a recent paper by Stoet and Geary (2019), this aim obscures a number of systematic issues that affect men. These issues are so significant, they claim, that existing gender inequality metrics, all of which indicate that women are still doing significantly worse

---

✉ Rosie Worsdale  
rjw229@cam.ac.uk  
Jack Wright  
jw675@cam.ac.uk

<sup>1</sup> Centre for Research in the Arts, Social Sciences and Humanities, University of Cambridge, Alison Richard Building, 7 West Road, Cambridge CB39DP, UK

than men on average, are getting the picture fundamentally wrong. Stoet and Geary offer a new gender inequality metric: the Basic Index of Gender Inequality (henceforth, BIGI), according which men in fact fare worse than women in more countries than the reverse. This surprising finding has not gone unnoticed. Stoet and Geary's results have been reported in a range of news outlets and their paper was viewed and downloaded more than 60,000 times in the 2 weeks following its publication.<sup>1</sup>

As we will show in this paper, BIGI fails to provide compelling evidence to warrant its controversial conclusions. Despite this, we submit, it would be a mistake to simply dismiss BIGI as a poor piece of social science unworthy of serious attention. This is because the traction it has gained in the short time since its publication is indicative of an important problem in the way that mainstream gender inequality metrics tend to be framed. Proponents of metrics like the World Economic Forum's (WEF) Global Gender Gap Index (GGGI)—presented by Stoet and Geary as foil for BIGI—do not adequately justify their methodological choices in relation to the context for which the index's findings are intended to reliably contribute. Instead, GGGI is presented as offering an objective ground for examining gender inequality, irrespective of the context in which the metrics findings are to be used. Claims to objectivity of this sort are not uncommon in social research. But, there is a growing literature in the philosophy of science that cautions against such claims, and makes the case that evaluations of objectivity in scientific research must be made relative to the context in which this research is intended to be useful.<sup>2</sup> Our intention in this paper is to contribute to this literature by using the case of BIGI and GGGI to articulate another reason to worry about context-independent claims to objectivity. The lack of epistemic modesty and adequate contextualism in GGGI, we submit, plays a crucial role in opening up space for people like Stoet and Geary to present their research as offering a mere corrective to the partial perspectives offered by mainstream metrics, rather than having to acknowledge that their research is grounded in a substantive, and highly contestable, perspective on the phenomenon of gender inequality. This has the effect of shielding Stoet and Geary from the need to acknowledge or justify this perspective—shielding them, in effect, from pragmatic contestation.

The paper proceeds as follows. In Sect. 1 we outline how Stoet and Geary position BIGI against existing gender inequality metrics, typified by GGGI. Then, in Sect. 2, we argue that BIGI does not succeed in correcting the kinds of biases that that Stoet and Geary claim to diagnose in GGGI. In Sect. 3, we argue that Stoet and Geary's claim that BIGI is superior to GGGI and other mainstream metrics can be seen as a claim that BIGI is more objective, where the increased degree of objectivity applies irrespective of the context in which the metric is to be used. In Sect. 4, we argue that Stoet and Geary's arguments for BIGI gain traction, despite flaws in BIGI, because GGGI is presented as drawing authority from a similarly context independent understanding of objectivity.

---

<sup>1</sup> Stoet and Geary's paper was given airtime on *BBC Radio* and in newspapers around the world, including in: the *Daily Mail* (Pinkstone 2019 [UK]), *Metro* (Hamill 2019 [UK]), *Toronto Star* (Hunter 2019 [Canada]), *Russia Today* ('Men face MORE' 2019 [International]), and *NRC Handelsblad* (Huygen 2019 [Netherlands]). See PLoS ONE website for updated download figures: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205349>.

<sup>2</sup> See, for examples, Koskinen (2018) and Wright (2018) and Sect. 4 for more.

## 1 Building the basic index of gender inequality (BIGI)

In 1995, the United Nations Development Program (UNDP) launched their first development indices that utilised gender-differentiated development data. In part because they grew out of and were reported yearly alongside of the well-known Human Development Index (HDI) and Human Development Reports, the UNDP's Gender-related Development Index (GDI) and Gender Empowerment Measure (GEM) quickly became key metrics in policy analysis (Hawken and Munck 2013, p. 802). As is often the case with aggregating indices, disagreements about what should be included in GDI and GEM have led to the creation of alternatives. Two of the most widely-used indices that developed after GDI and GEM were the WEF's GGGI and the Organisation for Economic Co-operation and Development's (OECD) Social Institutions and Gender Index (SIGI).

Although GDI, GEM, GGGI, and SIGI all differ in their stated goals, the data they draw upon, the specific issues they highlight, and their methodologies, all four indices are regularly used to highlight the different ways that women can be seen as disadvantaged in comparison to men. Stoet and Geary take this common usage as evidence that GDI, GEM, GGGI, and SIGI are designed to answer the question 'are men or women doing better?' They then argue that the answers GDI, GEM, SIGI, and GGGI give to this question are biased because they fail to adequately consider the aspects of life for which men can be seen to fall behind women. Stoet and Geary offer BIGI as an alternative index that seeks to answer the same question in what they see as a less biased way. They argue that once their less biased perspective has been adopted the commonly held belief that men in general fare better than women is thrown into doubt. Instead of finding that women fare worse than men in most countries of the world, BIGI finds the reverse. It finds that although women in general do worse in less developed countries, it is men that fall behind in more developed ones. Overall it finds that men are disadvantaged in 91 out of 131 countries surveyed (68 percent), with the median country having a 1.7 percent disadvantage against men.

What are we to make of these striking findings? To evaluate BIGI and understand how it can reach the conclusions it does we need to drill down a little deeper into the details.

Given that Stoet and Geary present GGGI as foil for BIGI, the easiest way to understand how BIGI reaches its controversial conclusions is by understanding how and why it departs from GGGI. GGGI, which mostly utilises the HDI data and indicators that GDI and GEM are also based on, is divided into four subindices: economic participation and opportunity, educational attainment, health and survival, and political empowerment. Each of the four subindices contain a collection of relevant indicators.<sup>3</sup> Each indicator is given as a ratio of the female value over the male value, but with

<sup>3</sup> Economic participation is comprised of ratios of: female over male labour force participation; wages for comparable work of women over men; estimated female over male earned income; the number of female legislators, senior officials, and managers over the number of males in such roles; and, the number of female professional and technical workers over the number of males in such roles. Educational attainment is comprised of ratios of: female over male literacy rate; female over male primary-level education enrolment; female over male secondary-level education enrolment; and, female over male tertiary-level education enrolment. Health and survival is comprised of ratios of: female over male healthy life expectancy; and, sex ratio at birth (female over male). Political empowerment is comprised of ratios of: the number of women

the maximum possible value truncated at 1.<sup>4</sup> This means that all indicators appear as scores between 0 and 1, with the distance from 1 being the gender ‘gap’. The authors of GGGI claim this one-sided scale is “more appropriate” (Hausmann et al. 2006, p. 7) than a two-sided, non-truncated scale. Although they do not say explicitly what they mean by this, we assume it has something to do with loose pragmatic reasoning they give elsewhere, where they point to the goal of measuring whether countries are making progress towards closing the existing gender gap “rather than whether women are ‘winning’ the battle of the sexes” (p. 5). After indicator scores are calculated and normalised, a score for each subindex is calculated as the arithmetic mean of the indicators within it, with weights for each indicator inversely proportional to their standard deviation.<sup>5</sup> The final GGGI number is then the arithmetic mean of the four subindex scores, each weighted equally (0.25). See Fig. 1 for a summary of all the subindices, indicators, weights, and sources of GGGI.

Stoet and Geary’s main criticism of GGGI (as well as GDI, GEM, and SIGI) is that the subindices and indicators that it includes bias the answers it gives to the question ‘are men or women doing better?’ They make two general and two more specific points about what we might call *inclusion bias*. Their first general criticism is that the subindices and indicators included in GGGI focus on known ways that women seem to fall behind men (e.g., political representation) but have no way of picking up the many ways that men fall behind women (incarceration rates, drug and alcohol abuse, etc.). Their second general criticism about what is included in GGGI is that the subindices on economic participation and opportunity and political empowerment (as well as similar subindices in GDI, GEM, and SIGI) do not reflect “true gender inequality” (Stoet and Geary 2019, p. 3) but rather individual or cultural choices. They argue that it is plausible that women choose not to enter politics, for example, for cultural or personal reasons and not because of any lack of opportunity.

Stoet and Geary use these criticisms to argue that any indicators included in gender related indices must be at a sufficiently general level to pick out disadvantages for both women and men and to not be skewed by cultural and subjective preferences. They argue that the only way of doing this is to construct a metric based on a picture of all-things-considered well-being for both genders. This is what they take to be the key to the construct ‘gender equality’. Equality between genders is equivalent to men

Footnote 3 continued

over men in parliament; the number of women over men at ministerial level; and, the number of years with a female head of state (in the last fifty years) over the number of years with a male head of state.

<sup>4</sup> The exceptions are life expectancy and gender ratio at birth, which are truncated at 0.944 and 1.06 respectively. These numbers are taken to represent equality for these factors (rather than 1). This is based on what is considered a ‘normal’ sex ratio at birth (0.944) (Klasen and Wink 2003) and the UN HDI’s targets for life expectancy (87.5 for women and 82.5 for men, a ratio of 1.06). Since the rationale for truncation is the same for all indicators, we will simply refer to ‘truncation at 1’ for the remainder of the paper.

<sup>5</sup> This means that the indicators with larger standard deviations are weighted less than those with smaller standard deviations within each subindex. The rationale for this weighting is that it prevents indicators with large deviations from dominating the variation between countries on each subindex. Note, however, that all the weights have been kept stable since the first report in 2006, meaning that the standard deviations in 2006 have defined the weights ever since. For example, within the political empowerment subindex the ratio of women to men in parliament had a standard deviation of 0.166 in 2006, the ratio of women to men at ministerial level had a standard deviation of 0.208, and ratio of years of female heads of state to male had a standard deviation of 0.116. The weight for number of women in parliament is therefore calculated as  $(0.01/0.166)/((0.01/0.166) + (0.01/0.208) + (0.01/0.116)) = 0.310$ .

Subindex	Indicator	Source (for 2018 index)
<b>Economic participation and opportunity</b> [weight in GGGI score: 0.25]	Female labour force participation over male [weight in subindex score: 0.199]	International Labour Organization, <i>ILOSTAT</i> database, 2017 or latest available data
	Number of female legislators, senior officials, and managers over male [0.149]	
	Number of female professional and technical workers over male [0.121]	
	Wage for comparable work of women over men [0.210]	WEF <i>Executive Opinion Survey (EOS)</i> , 2017–2018
	Estimated female earned income over male [0.221]	WEF calculations based on the UNDP methodology (refer to <i>Human Development Report 2007/2008</i> )
<b>Educational attainment</b> [0.25]	Female literacy rate over male [0.191]	United Nations Educational, Scientific and Cultural Organization Institute for Statistics, <i>Education indicators</i> , database, 2017 or latest available data
	Female primary-level education enrolment over male [0.459]	
	Female secondary enrolment over male [0.230]	
	Female tertiary enrolment over male [0.121]	
<b>Health and survival</b> [0.25]	Female healthy life expectancy over male [0.307]	United Nations Population Division, <i>World Population Prospects</i> , 2017 or latest available data
	Sex ratio at birth (female over male) [0.693]	World Health Organization, <i>Global Health Observatory</i> database, 2016 or latest available data
<b>Political empowerment</b> [0.25]	Number of women in parliament over men [0.310]	Inter-Parliamentary Union, <i>Women in National Parliaments</i> , reflecting elections/appointments up to October 2018
	Number of women at ministerial level over men [0.247]	Inter-Parliamentary Union, <i>Women in Politics: 2017</i> , reflecting appointments up to January 2017
	Number of years of female head of state in the last fifty years over number of years of male head of state [0.443]	WEF calculations, reflecting situation as of 30 June 2018

Fig. 1 GGGI subindices, indicators, weights, and sources. Source: Zahidi et al. (2018, pp. 5–6)

and women doing equally well with respect to some “core aspects of life,” which Stoet and Geary define as the “opportunity to live a long and healthy satisfied life that is grounded on educational opportunities in childhood” (2019, p. 3). They break this into three independently necessary components—healthy long life, educational opportunities in childhood, and life satisfaction—and calculate a subindex for each.<sup>6</sup>

<sup>6</sup> Note that this is an odd notion of welfare, one that does not seem to be based on any serious philosophical analysis of what should count as welfare. We criticise it in more detail in Sect. 2.

Within GGGI's health and education subindices, Stoet and Geary make two more specific accusations of inclusion bias. First, GGGI's health subindex is made up of life expectancy and the ratio at which boys and girls are born, both weighted inversely to their standard deviations. Stoet and Geary take issue with this. They argue that the fact that healthy life expectancy is weighted less (at 0.307) than sex ratio at birth (0.693) in GGGI's health and survival subindex "undervalues the health and survival of actually living persons" (2019 p. 3). Moreover, they argue that although gender ratio at birth may indicate negative attitudes towards women,<sup>7</sup> it is an indirect way of measuring such attitudes. Stoet and Geary, therefore, jettison the gender ratio at birth indicator entirely from their metric, leaving the health subindex of BIGI as simply a ratio of healthy life expectancy for men and women.

Second, GGGI's education subindex is comprised of ratios of male and female primary, secondary, and tertiary education enrolment and a ratio of male and female literacy rates. Stoet and Geary argue that, like the economic and political indicators that GGGI includes, tertiary education enrolment rates "may result more from choice than from a disadvantage" (2019, p. 2). Given this they decide to also remove tertiary education enrolment from their education subindex.

Overall BIGI scores are then the arithmetic mean of subindices for healthy long life, educational opportunities in childhood, and life satisfaction (each weighted 0.333). The first two of these subindices use the indicators that remain from GGGI's health and education subindices once sex ratio at birth and tertiary enrolment ratios are excluded. BIGI's 'healthy long life' subindex is a simple ratio of the life expectancy figures that GGGI and GDI also use. To circumvent any worries about weighting choice, the 'educational opportunity in childhood' subindex is taken as the greatest from parity (rather than a weighted average) of the three remaining education ratios: primary enrolment, secondary enrolment, and literacy rates. The additional subindex on 'life satisfaction' is a simple ratio of the average male and female answers to Gallup World Poll's question 'life today'.<sup>8</sup>

In addition to arguing that GGGI is biased in what it includes, Stoet and Geary argue that GGGI exhibits what we might call *calculation bias*. Stoet and Geary take particular issue with GGGI's method of truncating indicator scores at 1. They argue that this automatically obscures any male disadvantage:

GGGI truncates all values such that no country can, by definition, be more favorable for women than for men (for details see below). As a result, existing measures do not fully capture patterns of wellbeing and disadvantage at a national level. This is an important oversight, as there are issues that disproportionately

---

<sup>7</sup> GGGI's authors justify the inclusion of sex ratio at birth as capturing such attitudes. In their own words: "This variable aims specifically to capture the phenomenon of 'missing women' prevalent in many countries with strong son preference." (Hausmann et al. 2006, p. 6).

<sup>8</sup> "Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?" (See Helliwell et al. 2018, note 16).

affect boys and men. ... [T]here is no defensible rationale for truncating scores on an ‘equality’ measure when they disadvantage boys or men (2019, p. 2).<sup>9</sup>

In order to include the possibility that men might be disadvantaged, BIGI uses a two-sided scale between -1 and 1 for each indicator. All indicators (ratios of life expectancy, life satisfaction, primary enrolment, secondary enrolment, and literacy) are given as the distance from 1 of the smallest number over the largest number (e.g.,  $1 - \text{male life expectancy}/\text{female life expectancy}$ ). A negative sign is then added if the smaller number corresponds to the male figure and a positive sign added if the smaller number corresponds to the female figure.<sup>10</sup> In addition to adopting a two-sided scale, Stoet and Geary also choose to shift from yearly data to taking five-year averages—the idea being that this makes BIGI less susceptible any yearly quirks.

Altogether, then, BIGI exhibits eight key differences to GGGI (1–7 are summarised in Fig. 2):

1. The removal of the economic participation and opportunity subindex and related indicators.
2. The removal of the political empowerment subindex and related indicators.
3. The removal of tertiary enrolment ratio within the education subindex.
4. The removal of sex ratio at birth within the health index.
5. The addition of a life satisfaction subindex.
6. The education subindex score is a maximum rather than an average (this means that all subindices are a single number and none contain internal weightings).
7. The shift from yearly reporting to taking five-year averages.
8. Each indicator (and consequently subindex) is calculated as a two-sided positive/negative score rather than a single sided scale truncated at 1.

<sup>9</sup> This seems to misunderstand GGGI’s aim of measuring gender gaps rather than women’s’ empowerment. Although, as we will argue in Sect. 4, whether this is oversight or deliberate misreading on the part of Stoet and Geary, it may be partly explained by an insufficient focus in GGGI reports on the particular pragmatic context for which the specific methodological choices of the index are deemed to provide “the most objective basis” for discussing the different factors related to gender inequality (Zahidi et al. 2018, p. 4).

<sup>10</sup> The sign is added later and the smaller number is always divided by the larger number because  $1 - x/y \neq y/x - 1$ .

Subindex	Indicator	Source (for 2018 index)
<b>Economic participation and opportunity</b>	Female labour force participation over male Number of female legislators, senior officials, and managers over male Number of female professional and technical workers over male Wage for comparable work of women over men Estimated female earned income over male	
<b>Educational opportunities in childhood</b> [ <u>taken as indicator that differs most from parity</u> ; weight in BIGI score: 0.333]	Ratio of female and male literacy rates Ratio of female and male primary-level education enrolment rates Ratio of female and male secondary enrolment rates Female tertiary enrolment over male	United Nations Educational, Scientific and Cultural Organization Institute for Statistics, <i>Education indicators</i> , database, <u>average 2012-16</u>
<b>Healthy life expectancy</b> [0.333]	Ratio of female healthy life expectancy to male Sex ratio at birth (female over male)	United Nations Population Division, <i>World Population Prospects</i> , <u>average 2012-16</u>
<b>Political empowerment</b>	Number of women in parliament over men Number of women at ministerial level over men Number of years of female head of state in the last fifty years over number of years of male head of state	
<b>Overall life satisfaction</b> [0.333]	Ratio of female and male life satisfaction	Gallup World Poll, answers to 'life today' question, <u>average 2012-16</u>

**Fig. 2** BIGI subindices, indicators, weights, and sources, in contrast to GGGI. Omissions from GGGI are greyed out, changes and additions are underlined. Sources: Stoet and Geary (2019) and Zahidi et al. (2018)

## 2 Problems with BIGI

BIGI is presented by Stoet and Geary as a necessary corrective to the inclusion and calculation biases of GGGI. The extent of the supposed skewing of common understandings of gender inequality caused by these biases is brought into sharp relief by one of BIGI's most counterintuitive results. BIGI finds that, contrary to what many people believe, women are actually less disadvantaged than men in Saudi Arabia and that Saudi Arabia is in fact the third most equal country for women and men in the world (after only Italy and Israel). The unexpectedness of this finding warrants scrutiny.

## 2.1 Inclusion bias

The fact that BIGI finds women and men to be almost equal in Saudi Arabia invites skepticism about what Stoet and Geary include in their conceptualization of gender parity. Recall that for Stoet and Geary, parity is the balancing out of all-things-considered well-being, where all things considered well-being is taken to be a combination of life satisfaction, educational opportunities in childhood, and life expectancy. Saudi Arabia performs well in terms of overall parity because although men and women have substantially different scores for these three components—women have a much lower rate of basic education attainment, whereas men have a lower life expectancy and lower life satisfaction scores—when these disparities are aggregated they more or less cancel each other out, leaving Saudi Arabia with a small overall advantage for women.

The reason Stoet and Geary give for focussing on life satisfaction, educational opportunities and life expectancy at the expense of political and economic factors is that they capture the “core aspects of life” while also, they claim, not being affected by different cultural influences or individual choices (2019, p. 3). But it is far from clear that the picture of all-things-considered well-being generated by measuring life satisfaction, education, and life expectancy is not impacted by cultural or individual choice.

Consider BIGI’s educational opportunity measure for Lesotho. According to BIGI., Lesotho has “one of the largest gender gaps in education in favor of girls” (Stoet and Geary 2019, p. 13).<sup>11</sup> But, as Stoet and Geary themselves note, greater educational enrolment of girls in Lesotho is in part due to the greater role men and boys play in the labour market. This may be of benefit to girls on certain understandings of well-being, but a comparative lack of ability to be involved in the labour market also creates an economic dependence on men later in life that may also cut against their well-being.<sup>12</sup> Stoet and Geary might respond that female scores for life satisfaction in Lesotho would reflect this dynamic—i.e., if the disadvantage of exclusion from the labour market outweighs the benefit of their comparatively high rates of education, then female life satisfaction scores would reflect this<sup>13</sup>—but they offer no evidence as to why this would be the case. Such a response, moreover, would undermine Stoet and Geary’s claim that measuring all things considered well-being requires measuring more than life satisfaction alone. If life satisfaction scores pick up on the relevant

<sup>11</sup> This is a score of  $-0.36738$ . Neither Stoet and Geary’s paper nor the companion website (Stoet 2019) provide the full data, just the overall education score, so we do not know whether this is data from primary educational enrolment, secondary enrolment, or literacy (although the fact that Stoet and Geary go on to talk about this score in terms of enrolment means that it is likely not the literacy rate).

<sup>12</sup> Ansell (2002, pp. 96–97) notes, for example, that curricula in rural schools in Lesotho are very often segregated according gender, with vocational subjects geared towards employment, such as woodwork and metalwork, being taught only to boys; furthermore, calls for a greater focus on vocational education for girls tend to be couched in terms of better preparing them for being mothers, wives and homemakers rather than contributors to the waged-labour economy.

<sup>13</sup> Indeed, they seem to make precisely this kind of move when they argue that “while it is very difficult to determine the degree to which men and women are disadvantaged in any particular aspect of life, an overall assessment of life satisfaction likely reflects the combination of advantages and disadvantages they have experienced, whatever they might be” (2019, p. 3, emphasis added).

well-being consequences of differential educational opportunities, then it is not clear why we need to include educational opportunities in our measure of well-being at all. Little is said as to why educational opportunities in childhood are crucial beyond that leaving out any one of the three components:

[M]isses an important aspect of what defines a good life. For example, a person may have a satisfied and long life, but without educational opportunities, such a person might not have had a chance to develop his or her talents (2019, p. 3).

If Stoet and Geary accept a conception of well-being that recognises the importance of capacities and/or opportunities as well as life satisfaction then it is not clear why economic empowerment or freedom should not also be considered as key ingredients of a good or fulfilled life. Stoet and Geary argue that the specific indicators used by GGGI to factor in economic and political empowerment are susceptible to “personal motives or ... cultural, political or religious strictures” (2019, p. 4). Yet, as the example of Lesotho shows, so are the educational opportunity indicators in BIGI. Moreover, Stoet and Geary make no effort to come up with alternative political and economic indicators. If educational opportunities are to be considered part of well-being and political and economic opportunities not, then an argument for that conception of well-being needs to be given. If it is simply an issue with biased indicators for political and economic opportunities, then educational opportunities are not immune and there is no reason not to come up with alternative political and economic opportunity indicators.

There is, or course, a vast literature on the subject of what constitutes human well-being and how that well-being should be measured. This is a literature that Stoet and Geary completely ignore in their (scant) explanation for why they choose to focus on the three factors that they do.<sup>14</sup> A cynical observer might conclude that this move is simply aimed at drawing attention to well known men’s rights issues.

Even if we ignore this lack of sufficient justification for including educational and not political or economic factors in an all-things-considered picture of well-being, there are worries to be had about the heavy use of life satisfaction survey results in BIGI. The extensive literature on adaptive preferences that has flourished in recent decades has drawn critical attention to the ways in which systematic deprivation or disadvantage can substantially impact a person’s sense of life satisfaction—with gendered disadvantages being one of the key examples of this phenomenon.<sup>15</sup> Stoet and Geary make no mention of the challenge this literature presents to the idea that well-being can be effectively measured via life satisfaction. Indeed, they go as far as to suggest that overall life satisfaction scores, in particular, can serve as a kind of proxy measure for the cumulative disadvantages and advantages men and women experience, making it a unique selling point for BIGI that it includes these scores where other metrics do not.

---

<sup>14</sup> For a convenient summary of various theories of well-being and the disparate measures adopted for each theory, see Alexandrova (2017, p. xxxv).

<sup>15</sup> Sen’s pivotal example—of the extent to which Bengali widows, in the aftermath of the famine of 1943, had much lower rates of self-reported ill-health than did widowers, despite the fact that widows in general were considerably more deprived (Sen 1984, p. 309)—is instructive here. See also chapter 5 of Nussbaum (1999), and Khader (2011).

Our point here is not to deny that there may be contexts in which measuring childhood education opportunities, life expectancy rates, or overall life satisfaction scores—or even, potentially, some combination of the three—might be useful for understanding certain aspects of gender inequality. Rather, what we want to emphasise is that BIGI is no less vulnerable to the accusation of inclusion bias in general than GGGI. Like GGGI, BIGI relies on a set of assumptions that may be justifiable in some context but that can be seen as problematic biases in others. Stoet and Geary, recall, accuse GGGI of producing a skewed picture of gender inequality, through the inclusion of indicators which focus exclusively on known women’s issues, and others that are likely to be impacted by individual choice or cultural norms and beliefs. Given well-documented worries about the impact of systematic disadvantages on subjective welfare assessments like life satisfaction, and the demonstrable role that cultural factors can play in a country’s childhood educational opportunities score (just to highlight two examples), there are more than adequate grounds for accusing BIGI of being equally biased as to what it includes and excludes.

## 2.2 Calculation bias

The second set of methodological choices that are called into question by BIGI’s Saudi Arabia score contain its trade-off style, double-sided scale and the weights it uses. Stoet and Geary summarise their observations on Saudi Arabia by saying that: “because men’s and women’s disadvantages average one another out, it [Saudi Arabia] reaches a high level of overall parity” (p. 8). And the BIGI website states:

We believe that the Global Gender Gap Report [GGGI] is biased by ignoring societal problems that affect more men than women, which results in an unrealistically negative outlook for women. The reality is that both men and women are affected by societal issues, and that issues men and women face differently *cancel one another out* (Stoet 2019, ‘Common questions about BIGI’; emphasis added).

Of course, the claim that some issues that disproportionately affect men (prison populations or drug abuse, for example) do not get due discussion in policy circles is perfectly reasonable. But the idea that societal issues which affect different groups of the population cancel one another out seems a decidedly odd way to think about how one might want to improve people’s lives—be they men or women. A metric in which societal issues are tallied up and cancelled out essentially hides those issues.

The logic at play here seems to be based on a misunderstanding about feminist arguments for equality. Feminists typically do not argue that all men do better than all women out of present social relations.<sup>16</sup> They instead argue that in general men have more power, agency, and money in present social relations and that these factors cause patriarchal norms and ways of thinking to dominate society. The fact that patriarchal norms and patterns of thought might hurt both men and women, and might even sometimes hurt men more than women, does not count against such an argument—indeed,

<sup>16</sup> Especially not as the rise of intersectional analysis has focussed attention on the interplay between gender and other axes of identity.

it is a corollary of the feminist critique of the gender binary that restrictive norms are enforced on men, as well as women. This makes it likely that there will be certain harms or disadvantages which disproportionately accrue to men and boys.<sup>17</sup> In opposing patriarchy, feminists oppose the social system characterised by men in general having more power, agency, and money than women—which includes opposition to the harms that this patriarchal system inflicts on men. By equating gender equality with a situation in which the various harms differently affecting men and women cancel each other out—a situation, in other words, where men and women are harmed equally—BIGI totally distorts the feminist analysis of patriarchy which underpins the interest in measuring gender inequality. Especially interesting on this point is the fact that Stoet and Geary do acknowledge that “the lack of gender inequality [represented by BIGI scores closest to 0] does not imply that women or men have abundant opportunities in life [...] and neither does it mean that a country is free of sexist attitudes” (Stoet and Geary 2019, p. 15). To address this fact, they also rank countries in BIGI not, in fact, by their BIGI scores, but by their Average Absolute Deviation from Parity (AADP) scores. They even say of the AADP ranking that it “better reflects the amount of work to be done (in a society) to resolve all relevant gender disparities” (ibid). Given this acknowledgement of the limitations of the trade-off style BIGI scores, it is significant that the authors nonetheless present the BIGI scores themselves, and not the AADP scores, as the flagship innovation of their new index.<sup>18</sup>

Stoet and Geary also argue that because each subindex in BIGI is given as a single number and is not a weighted average of different indicators, they are able to avoid the potential weightings biases that they see as skewing the GGI calculations. But their tripartite system—education, satisfaction, life expectancy—is itself the product of an unacknowledged weighting decision, about which there should be serious reservations. Given that women have a longer life expectancy in nearly all countries of the world and that female mammals live longer in general, there is reason to believe that exact life expectancy parity in any society may indicate disadvantages for women rather than equality. Including healthy life span as one third of the BIGI’s overall score makes it much more difficult to pick up on the possibility that women face a disadvantaged but longer life in some countries. If life span is excluded from BIGI or weighted less than one third then many of BIGI’s headline findings (e.g., on Saudi Arabia, or that women do better in most developed countries) are reversed or become less pronounced.

As with the issue of inclusion bias, our point here is not that nothing useful could ever come from employing BIGI’s trade-off style double-sided scale, or weighting life expectancy heavily, in measuring gender inequality. There may be uses and contexts that could justify such methodological choices. Our point is that BIGI is no more immune to accusations of calculation bias in general than GGI. Stoet and Geary present their method as a response to what they see as the skewed calculations of GGI, positioning BIGI as an unbiased measure of gender equality by contrast (Stoet 2019, ‘Background’). Yet there are clearly grounds to think that BIGI makes a number

<sup>17</sup> Even mainstream media reporting on the kinds of men’s issues that Stoet and Geary highlight acknowledges the role of gender norms in facilitating harms to men. See for example Schumacher (2019), Coles (2019) and Ananthaswamy and Douglas (2018).

<sup>18</sup> Note that, while they avoid the issues involved in giving a trade-off based overall score, the AADP scores can still be accused of containing all of the other biases BIGI can be accused of.

of assumptions in relation to its weightings and calculations that may be considered biased when viewed in certain contexts. The heavy weighting of life expectancy, and the decision to take exact life expectancy parity as the benchmark for gender equality, certainly warrants the suspicion that BIGI scores will exhibit a bias towards minimising or obfuscating female disadvantage; as does the decision to trade-off comparative male and female disadvantage across the three subindices in the calculation of the overall BIGI score.

### 3 Bias, objectivity, and context

BIGI is, thus, at least as vulnerable to the charge of bias as GGGI. Yet, Stoet and Geary appear to have been quite successful so far in positioning BIGI as an antidote to the issues they claim to identify with GGGI. What is going on here? How has BIGI gained critical traction against GGGI and other mainstream approaches to measuring gender inequality (as evidenced by the extent of the attention and media coverage it has received), despite the fact that it fails to successfully overcome the issues it identifies with GGGI?

At least part of the answer to this question, we believe, lies in the way that Stoet and Geary lean on the authority of objectivity in their critique of GGGI.<sup>19</sup> More specifically, it is the way that they associate the measurement of “true gender equality” (Stoet and Geary 2019, p. 3) with the mitigation or elimination of a specific set of (feminist) perspectives, with the resulting objectivity of such a measurement judged independently of context. To see what we mean by this, it is helpful to briefly examine in greater detail the reasoning and language used by Stoet and Geary in presenting BIGI as an antidote to the apparent problems of inclusion and calculation bias in GGGI.

As noted above, the authors argue that GGGI is biased because it is grounded in a conception of gender inequality that contains cultural and political assumptions. Given that specific “disadvantages cannot always be defined *objectively*,” Stoet and Geary choose to base BIGI on life satisfaction scores, educational opportunity metrics and life expectancy—all measures that they argue are “culturally *independent*” (Stoet and Geary 2019, p. 15; emphasis added in both). At the centre of Stoet and Geary’s argument, then is the claim that GGGI is biased by judgements and assumptions that are the product of a set of particular (cultural or political) perspectives, which skew or distort the picture of gender inequality produced by the metric and compromise its objectivity—and that BIGI, by contrast, avoids the infiltration of bias through its culturally independent approach.

The authors argue, for example, that jumping to the conclusion that there must be something wrong with BIGI on account of its finding that Saudi Arabia has a relatively high level of gender parity would simply recapitulate the kind of cultural biases in GGGI. Those of us with a Western perspective will be biased towards a negative assessment of many of the more heavily gendered facets of life in a country like Saudi

---

<sup>19</sup> Another piece of the puzzle is clearly the political context into which the research has been presented – we will return to this at the end of the paper.

Arabia—one that might not be shared by citizens within the country. To decide that the BIGI metric is a flawed or inappropriate measure of gender inequality because its conclusion vis-à-vis life in Saudi Arabia does not accord with our own, culturally influenced assessment of it, is to compromise what, as far as possible, should be a process—measuring gender inequality—*independent of cultural, political, or other perspectives*. A lack of such perspectival independence in the construction of a metric makes for a compromised picture of reality.

Similar reasoning and language (‘skew’, ‘bias’, ‘overestimation’) is present throughout Stoet and Geary’s paper.<sup>20</sup> And, the framing of BIGI in opposition to GGGI is presented even more forcefully on the website accompanying BIGI, where BIGI’s closer approximation to a “true measure of gender inequality” is explicitly positioned against GGGI’s biased and therefore untrue measure:

[GGGI] relies heavily on issues that are often highlighted in women’s rights movements. [...] While the GGGI can be useful as a measure of women’s advancement in the areas of politics and employment, it is too biased to one specific gender to consider it a **true measure** of gender *equality* [...] the BIGI aims to provide a simplified and *unbiased* measure by focusing on key indicators that are relevant to all men and women in any society (Stoet 2019, ‘Background’; their italics, our bold).

What is evident in these passages is that Stoet and Geary’s twofold argument, against GGGI and in favour of BIGI by comparison, relies on the idea that truth in social scientific research is compromised by the presence of considerations that stem from certain social, political, or otherwise pragmatic, perspectives. GGGI, on their view, does not give an accurate picture of the phenomenon of gender inequality because the various decisions about what to include in the index and how to calculate country scores are informed by a cultural or political agenda (as they see it) and this biases the picture of gender inequality generated by the metric. BIGI, by contrast, is comparably independent of such considerations, which makes the picture it generates less biased and hence more accurate.

This equation of the presence of any perspective with bias and less truth appears to draw on a way of thinking about the nature of objectivity that is common in scientific research, in which objectivity is associated with the idea of a ‘view-point from nowhere’ (Nagel 1989; Rorty 1979). According to this view, objectivity (whether it is actually attainable in practice or can only be aimed towards) consists in achieving a view-point from no perspective in particular and that such a view-point represents the true state of affairs in the world. Since perspectives are partial, anything that introduces a particular perspective into a research process skews that research and compromises its objectivity. In the case of BIGI and GGGI, Stoet and Geary’s argument is that GGGI’s findings are biased because the metric is designed, at least in part, from the perspective

<sup>20</sup> It is in the justifications for their indicator choices and the two-sided scale, and in their accusation that GGGI exhibits calculation bias. For example: “[C]ertain subindices may result more from choice than from a disadvantage. ... [D]ifferences in occupational preferences and strategic divisions of labor in family life may skew quantifications of *true gender inequality*.” (pp. 2–3; emphasis added).

of a particular political (i.e. feminist) agenda. Since BIGI—so the authors claim—is designed independently of such a perspective, its findings are more objective.<sup>21</sup>

Central to the objectivity-as-view-from-nowhere ideal that Stoet and Geary are drawing on is the idea that the evaluation of objectivity is context-independent. That is to say, how objective we deem something (a research process, a measurement tool, a picture of the world) to be does not depend on the context for which that thing is designed. The point of the view-from-nowhere ideal is to strive to shed all and every partial perspectival on the thing in question, with the aim of achieving the most objective view of it. How objective the end product of this process is thus has nothing to do with the context in which we want to make use of the thing (research process, measurement tool, etc.) in question; all that matters is that we get as close as possible to the total elimination of partial perspective, since all such perspectives, regardless of context, undermine objectivity.

The view-from-nowhere ideal of objectivity has been heavily criticised by philosophers of science.<sup>22</sup> It is beyond the scope of this paper to give an overview or assessment of all of these various critiques; it suffices for our purposes only to note that contemporary philosophy of science has in general moved away from the more ontological concerns about access to the “really real” (Lloyd 1995) that underpin the view-from-nowhere ideal. Instead, increased attention has been paid to developing what Heather Douglas (2004) calls “operational” notions of objectivity—ways of articulating what objectivity might require that are more closely tied to the methods and processes of regular scientific research.<sup>23</sup> Among these discussions three factors are particularly relevant for our purposes. First, there are scenarios in which including more perspectives (of the right sort<sup>24</sup>) might be said to offer more objectivity (Harding 2015; Wylie 2003). Second, that the elimination of any and all perspectives is too often implausible (Douglas 2009; Longino 1990, 2001), if not impossible (Nagel 1989), so that in practice trade-offs are likely to be made about which perspectives compromise (or aid) objectivity. Third, in part because of these two points, what is deemed to threaten or boost the objectivity of something, as well as what is deemed to be the most appropriate strategies for mitigating those threats or attaining those benefits, will typically be determined by specific features of the context in which that thing is intended to

---

<sup>21</sup> For the most part, we are reading this notion of objectivity from Stoet and Geary’s equation of the elimination of biasing perspectives with the attainment of truth; but it should be noted that they do use the term ‘objectivity’ in such a way that implies this view-from-nowhere ideal (they caution that what counts as a life disadvantage “cannot be defined objectively” (p. 15), and for this reason their use of life satisfaction scores is preferable because these scores are culturally and politically independent).

<sup>22</sup> See, for examples, Megill (1994), Nagel (1989), and Reiss and Sprenger (2014).

<sup>23</sup> See Burch and Furman (2019), Koskinen (2020), and Zahle (2020) for examples of philosophers putting this call for operational versions of objectivity into practice for specific areas of social research.

<sup>24</sup> Those trained by experience to impartially (i.e., without any incentives to skew their judgement) spot key phenomena, like an expert witness in a trial, for example.

be used.<sup>25</sup> Wright illustrates this contextualised way of thinking about evaluations of objectivity with the following example:

If energy usage projections are to be used to determine future government investment in energy infrastructure, that goal determines what the projections should step back from. In the context of that goal, the interests of energy companies might be relevant to step back from. Accordingly, one way in which the objectivity of energy usage projections can be judged is how much they step back from those interests. This stepping back then occurs in the production of the projections—maybe they can be produced by an independent body with no connections to the energy industry. (2019, p. 398)

In contextualised understandings of objectivity, then, the key idea is that attempts to increase the objectivity of something should be dictated in part by factors relating to the context in which that thing—a measurement tool, a research process, a picture of the world—is intended to be reliably used. What does this entail for Stoet and Geary's accusation of bias against GGGI? Recall that the basis of their accusation is that GGGI's design is informed by certain perspectives (political and cultural), and the influence of these perspectives serves to skew the index's rankings and make them less objective. They juxtapose this with what they see as the independence of BIGI from such perspectives. Crucially, however, Stoet and Geary fail to adequately contextualise their justification for these claims. In order to substantiate the claim that the various perspectival considerations that they see as informing the construction of GGGI serve to bias the index, Stoet and Geary would need to show that these considerations undermine the index's findings relative to the context in which they are intended to be useful. Similarly, in order to justify the claim that BIGI is more objective, they would need to explain why the mitigation of the particular perspectival considerations that they take themselves to have addressed in comparison to GGGI serves to remove a particular threat to the objectivity of the index, relative to its intended use and goals.

Absent such a contextual justification, the claim to BIGI's greater objectivity must implicitly rely on something like the bigger view-from-nowhere ideal, where context of intended use plays no role in evaluations of objectivity. This, we submit, is part of how Stoet and Geary are able to level the charges of calculation and inclusion bias against GGGI, despite the fact that BIGI is no less vulnerable to these charges once it is subject to methodological scrutiny. By appealing to the idea of objectivity as the maximal elimination of any and all perspective, they perform a kind of sleight

---

<sup>25</sup> Different authors interpret the context relevant for judging objectivity differently (see, for example, Zahle 2020). Although we follow Koskinen (2018) and Wright (2018) in focusing on the contexts in which things (e.g. metrics) are used, our argument does not rest on this. Philosophers also disagree whether objectivity should be thought of as incorporating positive epistemic elements (Wylie 2003; Zahle 2020) or only mitigating negative epistemic elements (Hacking 2015; Koskinen 2020). Our discussion mainly focusses on objectivity understood in terms of mitigating negative elements, simply because this is the approach to objectivity implied by Stoet and Geary's focus on the mitigation of bias in relation to truth and accuracy. However, we remain agnostic on the broader question as to whether positive or negative accounts of objectivity are preferable in general.

Also see Daston and Galison (2007), Fine (1998), Kincaid et al. (2007), Kuhn (1977), Montuschi (2004, 2014, 2017), and Padovani et al. (2015), for a series of discussions on objectivity and value-freedom that foreshadow the contextualised accounts of objectivity we refer to here.

of hand: drawing attention to the perspectival considerations informing GGGI and framing them as biasing the index, and presenting their supposed mitigation of these same considerations in BIGI as a consequent achievement of a greater degree of objectivity—all whilst ignoring the different perspectives informing their construction of BIGI.

Stoet and Geary, thus, fail to offer a proper, contextualised assessment of the comparative objectivity BIGI and GGGI. Such an assessment would require attending to the context in which each of the indices is intended to offer reliable information, and evaluating whether the mitigation or elimination of certain perspectives in the construction of each index successfully diffuses the most significant threats to the reliability of that index's findings, relative to that intended context of use.<sup>26</sup> It is precisely this detailed, contextual work that Stoet and Geary fail to do in their reductively simple presentation of BIGI as providing a more objective, truer picture of the state of global gender inequality than GGGI.

Stoet and Geary's chicanery in their articulation of the charge of bias against GGGI, however, is only one part of the explanation as to why the charge has proved capable of generating considerable traction. The other part, we argue, concerns GGGI itself—specifically, the way in which GGGI's authors themselves fail to adequately justify their various methodological choices relative to the pragmatic context to which the index is intended to contribute. This failure, we believe, leaves the door open to Stoet and Geary's sleight of hand vis-à-vis their claim to superior objectivity.

#### 4 Leaving the door ajar: inadequate contextualism in GGGI

Each of the yearly WEF Global Gender Gap Reports includes, as part of its methodology section, a brief elaboration of three basic concepts which underpin GGGI's approach to measuring gender inequality. These concepts are: (1) measuring gender-based gaps, rather than absolute levels of achievement; (2) measuring gender-based outcomes rather than inputs; and, (3) measuring proximity to gender equality rather than the empowerment of women. Each concept is accompanied by some justificatory remarks.

The focus on measuring gaps rather than absolute levels (1) for each indicator is justified on the basis that to do otherwise would be to penalise countries which are less developed generally, but might nonetheless have made great progress in equalising the quality of life of men and women in many respects. This means that countries with smaller gaps between men and women (in tertiary education enrolment, for example) will be ranked higher than those with larger gaps but higher overall achievement (more women enrolling in tertiary education overall). The decision to measure proximity to gender equality rather than women's empowerment (3) is justified on the basis that the goal of the index is to measure the extent to which countries make, or fail to make, progress in closing the various gaps between men and women, rather than to assess which gender might be coming out on top in various respects. This is why GGGI is

<sup>26</sup> The intended context of use for GGGI and BIGI may or may not be the same, and this will obviously impact the ease with which an evaluation of their comparative objectivity could be conducted—we will return to this question at the end of the paper.

reported on a scale from 0 to 1 (with 1 being equality or a higher level for women) rather than on a two-sided scale.

For conceptual components (1) and (3), the justification offered thus amounts to a clarification of what the index is and is not intended to be useful for—namely, it is intended, narrowly, to provide comparable information on the extent to which women’s attainment is lagging behind men in a number of indicators. The methodology sections of the reports do not contain a further explanation as to why tracking women’s performance relative to men’s in the chosen indicators is important (and for what purpose); but some explanation to that effect can be deduced, in the case of the 2018 report, from remarks made in the Preface. Here, the goal of achieving gender equality is framed in terms of the need to ensure that men and women can make an “equal contribution [... to the] process of deep economic and societal transformation” and the need for societies not to “lose out on the skills, ideas and perspectives of half of humanity to realize the promise of a more prosperous [...] future” (Zahidi et al. 2018, p. v).<sup>27</sup> These remarks reveal that the overall perspective from which the index is conceived and constructed is a developmental one, where the phenomenon of gender inequality is problematised primarily as a barrier to economic development and prosperity.

It is the developmental goals of the index, and the specific policy context to which it is contributing, then, that dictates that the index measures gaps in attainment rather than absolute levels, and proximity to gender parity rather than whether women are doing better than men. The justification for measuring gender-based outcomes rather than inputs or means (2), by contrast, is quite different. The focus on outcomes is justified on the basis that this approach provides a “snapshot of where men and women stand with regard to some fundamental outcome indicators related to basic rights such as health, education, economic participation and political empowerment”—and this snapshot provides “the most objective basis” for discussing other more contextual factors related to gender inequality (Zahidi et al. 2018, p. 4). It is this that is used to justify the specific indicators used in the metric—“outcomes” like number of female heads of state rather than “inputs” like the length of paid maternity leave.

The difference between how (1) and (3) and (2) are justified is subtle but important. Justifications for methodological choices take the form of explanations as to why particular research approaches should be preferred over others (Crasnow 2015, p. 638). In the case of conceptual components (1) and (3), the answer offered to the question of why GGGI’s approach to measuring gender inequality is preferable makes reference (albeit in a rather scattered and superficial way) to the policy context in which the issue of gender inequality is problematised and the index’s findings are intended to be useful. In the case of component (2), the answer goes further: namely, that the kind of evidence that is gathered by measuring gendered outcomes, using the specific data and methods of the index, allows us to obtain the most objective, most accurate ‘snapshot’ of the state of global gender inequality, irrespective of how the snapshot should be used.

The idea that the most objective picture of the phenomenon of gender inequality can be attained through a focus on measuring outcomes comes from Ricardo Hausmann (a

<sup>27</sup> A similar focus can be identified in the 2006 report Preface, which emphasises the need for “tapping into the extensive reservoir of women’s skills” in order to meet the challenges facing businesses (Hausmann et al. 2006, p. v).

contributor to the WEF reports), who proposes that progress can be made in establishing how certain policy outcomes can be achieved—for example, sustainable growth or gender equality—by adopting a strategy of ‘learning without theory’ (Hausmann 2016). Too much research, Hausmann contends, gets caught up in trying to produce or verify causal theories about complex social phenomena; theories we are either not in a position to construct, or whose accuracy we cannot determine. When we construct metrics on the basis of such theories, which measure purported or hypothetical causes for the outcomes we seek, we risk instigating misguided and unhelpful policy agendas, by incentivising actors to improve their performance with respect to indicators which may or may not lead to more successful outcomes. In such cases, in other words, we “attempt to be more theory-driven than our knowledge allows” (Hausmann 2016). According to Hausmann, GGGI is an example of a good index which limits itself only to measuring what it can know for sure: whether and to what extent countries have improved their performance with respect to the various outcome indicators. On the basis of this benchmarking, less well-performing countries can improve by imitating the practices of higher ranked countries.

It is Hausmann’s worries about the potential pitfalls of theory-driven research that underpin the claim that GGGI’s methodological strategy gives “the most objective basis” for analysing gender inequality. Gender inequality is a complex and multifaceted phenomenon; if we attempt to track countries’ progress towards overcoming it by measuring anything other than change in the relevant outcome variables, according to Hausmann, we risk quantifying things that we don’t truly know—namely, the salience of various possible causal factors for the outcomes we seek. By only measuring outcomes, the authors of GGGI want to avoid endorsing any causal claims or assumptions about how gender inequality functions or what might be done about it. They do so to avoid GGGI being biased by preconceived ideas about what the right solutions to the problem of gender inequality might be. There are a number of ways we can make sense of this. We might worry that our theories about what factors causally contribute to gender inequality are wrong for all kinds of reasons: idiosyncratic cognitive biases or errors of reasoning on the part of individual researchers; the contingent fact that certain causal theories have received more attention than others, and might thus seem more compelling because of this disproportionate focus; or particular social, cultural, and political outlooks from which researchers might consciously or subconsciously approach a topic. We might also just get things wrong because the task is a difficult one and there are limits to what we can know about how complex social phenomena work. The point is that, if we allow judgements or assumptions about the causes of gender inequality to impact what we measure and how we measure it, each of these ways in which those judgements and assumptions might be mistaken could serve to distort or skew the picture of gender inequality that our metric would generate—meaning that the resulting picture of the state of gender inequality will be less objective than it would be without these considerations.

The problem with this line of argument, we suggest, is that it obscures the necessarily pragmatic and contextual considerations that go into choosing what to include in an index like GGGI. The development-policy context informs which outcomes are taken to be salient for capturing the phenomenon of gender inequality. Only against a set of background theoretical hypotheses, about the particular reasons why gendered

disparities should be of concern to policy makers and politicians, can the choice of indicators in GGGI be justified. Without these theoretical foundations, there is no justification for why subjective well-being, for example, is not taken to be one of the “fundamental outcome variables” (Hausmann et al. 2006, p. 5) according to which gender parity should be measured.

Similarly, the developmental context of GGGI must also inform the distinction between ‘outcome’ and ‘input’ variables. In a different pragmatic context—one in which gender inequality is problematised principally as matter of injustice, for example—access to government-funded childcare in different countries might be identified as a fundamental outcome variable, on the basis that the provision of universally accessible childcare is a necessary component of achieving justice for women.<sup>28</sup> In GGGI, by contrast, access to government-funded childcare is included in country profiles only as one of a list of additional indicators whose data might help explain a country’s gender inequality score (Hausmann et al. 2006 p. 29), since parity is defined from a development perspective whereby what is desirable to achieve is not primarily justice, but the “equal contribution of men and women” to a process of “deep economic and societal transformation” (Zahidi et al. 2018, p. viii). Access to childcare thus does not factor into the gender equality league table that GGGI generates—and this impacts on the ‘snapshot’ of reality generated by the index.

Of course, relative to the broader context of a global policy and funding emphasis on development, into which the findings of GGGI are intended to contribute, it may well represent a sensible pragmatic decision to adopt a perspective on gender inequality through which it is problematised fundamentally as a barrier to the equal economic contribution of men and women. From this perspective, it might make sense to worry that measuring factors such as access to childcare risks measuring things that we don’t know for sure are causally connected to expanding women’s participation in different economic sectors. We do not wish to deny this. The point, though, is that it is *only* within this pragmatic context that the strategy of measuring outcomes can properly be understood to constitute *a step towards increased objectivity in the index*. From a different contextual perspective—a concern with a more holistic assessment of the various harms associated with the differential treatment of men and women, for example—this strategy might be considered to make the findings of the index less objective, since it eliminates the possibility of measuring factors which should be considered essential to understanding the nature and extent of gender inequality. Evaluated from such a context, GGGI’s rankings might look unacceptably skewed or biased.

Thus, from a contextualist perspective, the justifications for methodological choices involved in GGGI’s design leave a lot to be desired. Although, as we have seen, some of the index’s methodological features are justified in relation to the pragmatic context to which its scores and rankings are intended to contribute, this contextualisation is scant. It also underplays (i) the extent to which the specific development-policy context is just one way of framing the phenomenon of gender inequality as a problem to be tackle; and

---

<sup>28</sup> Universal access to childcare might be thought of as essential for helping to mitigate the inequality in various life opportunities that women experience as a result of being disproportionately burdened with reproductive and childcare labour—making rectifying this inequality a matter of achieving a just social settlement for women.

(ii) the fact that the developmental policy context directly informs several of the key methodological commitments of the index, including the merely-measuring-outcomes approach, which the reports do not acknowledge.

This, then, is the way in which GGGI leaves the door open to Stoet and Geary's accusation of bias. Because the GGGI reports contain only minimal reference to the policy context into which the index is designed to contribute<sup>29</sup>—and totally fail to justify the supposedly objectivity-generating methodological choice of merely measuring outcomes in relation to this context—the effect is to oversell the objectivity of the index's findings, which paves the way for sceptics of the metric to make accusations of bias.

From this perspective, indeed, there is a narrow way in which Stoet and Geary's accusations against GGGI might be seen to have some kind of validity. The inadequate contextual justification of the methodological choices underpinning GGGI creates space for Stoet and Geary to point out—rightly—the ways in which the picture of gender inequality generated by the index is only partial (for example, that there are many heavily gendered phenomena that the index doesn't measure). Since GGGI doesn't do enough to justify these methodological decisions in relation to its contextual goals, Stoet and Geary can thus claim that GGGI is unacceptably biased and present BIGI, by contrast, as a less biased alternative.

Of course, as we have already shown, BIGI itself is replete with assumptions that undercut its claims to be unbiased *in general*. Insofar as Stoet and Geary also fail to justify their various methodological choices in terms of them being appropriate pragmatic choices relative to a certain goal, their claim that BIGI produces a more truthful picture of gender inequality than GGGI is no less dubious than those of GGGI's proponents. Furthermore, given the issues outlined in Sect. 2, it is much harder to envisage what an adequate pragmatic justification would look like in the case of BIGI, such that, relative to a certain pragmatic use, BIGI's measurement of gender inequality could be deemed sufficiently objective. BIGI certainly doesn't seem like a particularly useful guide to policy. Given the way that its two-sided scale trades-off different factors against one-another, obscuring inequality for both genders in the process, it is hard to see how BIGI could illuminate important factors to change in any given country. Given the heavy weighting of life expectancy within BIGI, and the generally longer life expectancies of women (which may well be due to biological differences), the policies recommended by aiming for BIGI parity seem likely to ensure overall lower female educational opportunities and/or life satisfaction. And, given Stoet and Geary's lack of justifications for educational indicators being included as BIGI's only opportunity measurements, it is hard to see how BIGI might be used in any serious debate as to how opportunities should be more equally distributed between genders. Thus, BIGI seems likely to face more of a challenge than GGGI to produce an adequate defence that its findings are to some significant extent objective relative to a specific use.

Our point, however, is that BIGI is, in a sense, saved from having to do this substantive work by the fact that GGGI largely fails to do it. Since the methodological decisions underpinning GGGI are insufficiently contextualised, and its claims to objec-

<sup>29</sup> The water is further muddied by fact that, beyond what little contextual information is provided in the prefaces, elsewhere the reports occasionally slip into selling the index as a means for contributing towards the more widely ranging goals of 'justice' and 'social equality'.

tivity insufficiently qualified, Stoet and Geary are able to position BIGI as offering a corrective to the seeming biases of GGGI—a corrective which, once administered, succeeds in attaining something closer to the kind of view-from-nowhere picture of gender inequality that GGGI, as they frame it, seeks but fails to generate. The authors can thus piggyback on GGGI’s insufficiently contextualised claim to objectivity, gaining leverage for their promotion of BIGI as a favourable alternative by pointing out GGGI’s apparent failings. If, on the other hand, more was done to emphasise that the indicators chosen in GGGI represent “the most objective” way of measuring gender inequality *relative to the specific context in which its finding are intended to be used*, BIGI would have no access to such easy leverage. This would mean that Stoet and Geary would have to make a substantive case for why their approach to conceptualising and measuring gender inequality is best for achieving a particular pragmatic goal—which would entail putting their own perspectival cards much more firmly on the table and opening up them up to pragmatic and political contestation.

## 5 Conclusion

In this paper we have sought to achieve two things. The first was to expose the unwarranted nature of the assessment of the state of global gender inequality offered by BIGI, which we achieved by a critical examination of Stoet and Geary’s methodological choices in constructing the index. We demonstrated that BIGI is no less vulnerable to the charges of inclusion and calculation bias that Stoet and Geary level at GGGI, and is thus no better placed to claim that it produces a truer picture of gender inequality. Our second aim was reveal that Stoet and Geary’s critique of mainstream gender inequality metrics, and their presentation of BIGI as a favourable alternative, was facilitated in part by GGGI itself. We argued for this by showing how GGGI’s claims to provide the most objective picture of gender inequality are insufficiently qualified with respect to the particular pragmatic, policy context in which the metric’s rankings are intended to be used. It is this context independent claim to objectivity, we argue, that Stoet and Geary are able to make use of in positioning BIGI as a corrective to the perceived bias in GGGI.

As discussed above, there is an emerging group of contemporary philosophers of social science that caution against the use of context independent notions of objectivity in social scientific practice, and advocate instead for the adoption of context-sensitive and partial notions of objectivity. Without wishing to nail our colours firmly to any one mast within this group—that is, without advocating any one particular contextualist account of objectivity—we broadly agree that social scientists would do well to embrace a kind of epistemic modesty with respect to their claims to objectivity. This means, among other things, being clear both about the context in which a given piece of research is intended to be used, and the steps taken to increase the objectivity of the research for that context. The purpose of this paper has been to use the recent case study of BIGI’s publication, and its take-up in popular media and discussion, to articulate an additional worry which we think should be added to this growing voice of caution against context-independent claims to objectivity in social research. The case of BIGI and GGGI is instructive, we believe, because it demonstrates the dangers

posed when mainstream social science that engages highly politically sensitive topics claims to offer an objective view of these topics without adequately contextualising these claims. This positioning obscures the fact that methodological justifications in social scientific research must typically be context-dependent, and in doing so leaves this social science open to accusations of bias and creates space for reactionary social science to position itself as a necessary corrective.

Stoet and Geary clearly have the perspective that attempts to measure global gender inequality are operating in an environment where feminism may have gone too far in focussing the attention of politicians and policy makers on women's issues at the exclusion of those that affect men. This perspective taps into a broader context of backlash and pushback against the progress of feminism and other progressive movements into the political and policy mainstream—a zeitgeist that, clearly, provides another crucial facet of the opportunity structure that has enabled Stoet and Geary's intervention. Our point is that, for those of us who think that this perspective is misguided, it would be a mistake to think that we can avoid this disagreement simply by reasserting the methodological superiority and objectivity of GGGI (and other more established metrics). Stoet and Geary correctly highlight some important blind spots in GGGI that may undermine the objectivity of the picture of gender inequality the metric generates, given certain policy goals. The task of devising an objective metric for gender inequality is always going to depend, in part, on how the phenomenon itself is conceived and problematised, and this framing will very likely be susceptible to political contestation—and we should not avoid this contestation by claiming for our own preferred tools and measures a kind of objectivity that isn't context sensitive. This episode suggests that it is especially important for social scientists working on highly politically contested topics such as gender inequality, immigration, or the impact of climate change, where the goals of research are subject to intense disagreement, to be clear on the contexts within which their claims to objectivity should be evaluated.

**Acknowledgements** We are grateful to Christopher Clarke, Inkeri Koskinen, Eleonora Montouschi, Julie Zahle, Meghan Tinsley and the participants at the *Objectivity and Social Research* workshop in Bergen for their feedback and suggestions; and to our two anonymous reviewers for their helpful comments.

**Funding** This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 715530).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alexandrova, A. (2017). *A philosophy for the science of well-being*. New York: Oxford University Press.
- Ananthaswamy, A., & Douglas, K. (2018). Why the patriarchy isn't good for men and how to fix it.' *New Scientist*, 18th April. Available at: <https://institutions.newscientist.com/article/mg23831740-900-why-the-patriarchy-isnt-good-for-men-and-how-to-fix-it/>. Accessed 4 July 2019
- Ansell, N. (2002). Secondary education reform in Lesotho and Zimbabwe and the needs of rural girls: Pronouncements, policy and practice. *Comparative Education*, 38, 91–112.
- Burch, M., & Furman, K. (2019). Objectivity in science and law: A shared rescue strategy. *International Journal of Law and Psychiatry*, 64, 60–70.
- Coles, T. (2019) 'How the patriarchy harms men and boys, too.' *Huffington Post*, 11th October. Available at: [https://www.huffingtonpost.ca/2017/11/10/patriarchy-men-boys\\_a\\_23273251/](https://www.huffingtonpost.ca/2017/11/10/patriarchy-men-boys_a_23273251/). Accessed 4 July 2019
- Crasnow, S. (2015) Feminism, causation and mixed methods research. In Hesse-Biber, S., & Johnson, R. (eds) *The Oxford handbook of multimethod and mixed methods research inquiry*. Oxford: Oxford University Press. Retrieved 7 Aug. 2019, from: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199933624.001.0001/oxfordhb-9780199933624-e-47>.
- Daston, L., & Galison, P. (2007). *Objectivity*. New York: Zone Books.
- Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, 138(3), 453–473.
- Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Pre.
- Fine, A. (1998). The viewpoint of no-one in particular. *Proceedings and Addresses of the APA*, 72(2), 9–20.
- Hacking, I. (2015). Let's not talk about objectivity. In F. Padovani, A. Richardson, & J. Y. Tsou (Eds.), *Objectivity in science* (pp. 19–33). New York: Springer.
- Hamill, J. (2019) Men are more disadvantaged than women in the UK, US and most of Europe, scientists claim. *Metro*, 4th January. Available at: <https://metro.co.uk/2019/01/04/men-disadvantaged-women-uk-us-europe-scientists-claim-8309361/>. Accessed 3 July 2019
- Harding, S. (2015). *Objectivity and diversity: Another logic of scientific research*. Chicago: University of Chicago Press.
- Hausmann, R. (2016) Learning without theory. *Project Syndicate*, March 30 2016. Available: <https://www.project-syndicate.org/commentary/learning-without-theory-by-ricardo-hausmann-2016-03?barrier=accesspaylog>. Retrieved May 9 2019.
- Hausmann, R., Tyson, L., & Zahidi, S. (2006). *The global gender gap report 2006*. Switzerland: World Economic Forum.
- Hawken, A., & Munck, G. L. (2013). Cross-national indices with gender-differentiated data: what do they measure? How valid are they? *Social Indicators Research*, 111, 801–838.
- Helliwell, J., Layard, R., & Sachs, J. (2018). *World happiness report 2018*. New York: Sustainable Development Solutions Network.
- Hunter, B. (2019) SHOCK REPORT: Women have it BETTER than men in most nations. *Toronto Sun*, 7th January. Available at: <https://torontosun.com/news/world/shock-report-women-have-it-better-than-men-in-most-nations>. Accessed 31 Oct 2019
- Huygen, M. (2019) Wie heeft het beter, man of vrouw? *NRC Handelsblad*, 4th January. Available at: <https://www.nrc.nl/nieuws/2019/01/04/wie-heeft-het-beter-man-of-vrouw-a3127902>. Accessed 3 July 2019
- Khader, S. (2011). *Adaptive preferences and women's empowerment*. New York: Oxford University Press.
- Kincaid, H., Dupré, J., & Wylie, A. (2007). *Value-free science: Ideals and illusions*. Oxford: Oxford University Press.
- Klasen, S., & Wink, C. (2003). Missing women: Revisiting the debate. *Feminist Economics*, 9(2–3), 263–299.
- Koskinen, I. (2018). Defending a risk account of scientific objectivity. *The British Journal for the Philosophy of Science*. <https://doi.org/10.1093/bjps/axy053>.
- Koskinen, I. (2020). Objectivity in contexts: Withholding epistemic judgement as a strategy for mitigating collective bias. *Synthese*. <https://doi.org/10.1007/s11229-020-02645-9>.
- Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension* (pp. 320–339). Chicago: University of Chicago Press.
- Lloyd, E. (1995). Objectivity and the double standard for feminist epistemologies. *Synthese*, 104, 351–381.
- Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.
- Longino, H. E. (2001). *The fate of knowledge*. Princeton, NJ: Princeton University Press.
- Megill, A. (1994). *Rethinking objectivity*. NC: Duke University Press Durham.

- Men face MORE discrimination than women according to new research. (2019). *Russia today*, 6th January. Available at: <https://www.rt.com/news/448208-men-more-disadvantaged-than-women/>. Accessed 3 July 2019
- Montuschi, E. (2004). Rethinking objectivity in social science. *Social Epistemology*, 18(2–3), 109–122.
- Montuschi, E. (2014). Scientific objectivity. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: A new introduction* (pp. 123–144).
- Montuschi, E. (2017). Using science, making policy: What should we worry about. *European Journal for Philosophy of Science*, 7(1), 57–78.
- Nagel, T. (1989). *The view from nowhere*. Oxford: Oxford University Press.
- Nussbaum, M. (1999). *Sex and social justice*. New York: Oxford University Press.
- Padovani, F., Richardson, A., & Tsou, J. Y. (2015). *Objectivity in Science: New perspectives from science and technology studies (310)*. Berlin: Springer.
- Pinkstine, J. (2019). ‘Men ‘face MORE discrimination than women’: Global study claims males receive the raw end of the deal with harsher punishments for the same crime, compulsory military service and more deaths at work.’ *MailOnline*, 7th January. Available at: <https://www.dailymail.co.uk/sciencetech/article-6564767/Men-face-discrimination-women.html>. Accessed 3 July 2019
- Reiss, J., & Sprenger, J. (2014). Scientific objectivity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University.
- Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton: Princeton University Press.
- Schumacher, H. (2019) Why more men than women die by suicide. BBC, 18th March. Available at: <http://www.bbc.com/future/story/20190313-why-more-men-kill-themselves-than-women>. Accessed 3 July 2019
- Sen, A. (1984). *Resources, values and development*. Oxford: Clarendon Press.
- Stoet, G. (2019) Basic Indicator of Gender Inequality (BIGI). *Gender Equality Info*. Available at: <https://bigi.genderequality.info/>. Accessed 4 July 2019
- Stoet, G., & Geary, D. (2019). A simplified approach to measuring national gender inequality. *PLoS ONE*, 14(1), e0205349.
- Wright, J. (2018). Rescuing objectivity: A contextualist proposal. *Philosophy of the Social Sciences*, 48(4), 385–406.
- Wylie, A. (2003). Why standpoint matters. In R. Figueroa & S. Harding (Eds.), *Science and other cultures: Issues in philosophies of science and technology* (pp. 26–48). New York, NY: Routledge.
- Zahidi, S., Geiger, T., & Crotti, R. (2018). *The global gender gap report 2018*. Switzerland: World Economic Forum.
- Zahle, J. (2020). Objective data sets in qualitative research. *Synthese*. <https://doi.org/10.1007/s11229-020-02630-2>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.