Measuring gender in surveys

Social psychological perspectives

Anna Lindqvist
Department of Psychology
Lund University

Emma A. Bäck
Department of Psychology
University of Gothenburg

Hanna Bäck
Department of Political Science
Lund University

Marie Gustafsson Sendén
Department of Psychology
Stockholm University

Abstract

A lot of research in the social sciences is more concerned with effects of adherence to gender stereotypes rather than actual sex differences. For instance, women and men vary greatly in gender traits and expressions, which in turn also varies with context. Moreover, there are also other gender identities that cut across or go beyond the traditional dichotomy of woman/man, which is therefore an unsatisfactory measure. The problem of measuring gender becomes increasingly complex as gender identity is less straightforward to measure than sex. In our own research, gender identity, defined as social identification with the own gender, is a better predictor of attitudes to feminist activities, such as, the use of gender-fair language, than categorical gender, even if the latter includes more than two options (Gustafsson Sendén, Bäck & Lindqvist, 2015). Another venue is to assess individual-level sexism and to use this as a predictor of political behaviour. Again, our research show that sexism, regardless of gender, affected voting intentions in the 2016 U.S presidential election (Bäck, Carroll, Hansen & Bäck, 2017). This indicates that other measures, such as gender identity and sexism should be more regularly included in social science research. The challenge is to find appropriately short, validated versions of such measurements.

Introduction

In the social sciences, many research findings are related to participants' gender, but what is gender and what do we measure when we ask about gender in surveys? As we will argue, gender may reflect a number of things, and how it is measured should depend on the present research question. The gender variable includes many facets and dimensions important to reflect upon in order to understand how and why gender is associated with attitudes and behaviours. Nevertheless, gender is a variable that is often included in analyses without much deliberate reflection of why. For instance, gender differences are often reported without much concern for the causes of such differences. Moreover, gender is seldom operationalized in any more detail than categorizing individuals into a binary system of being either a woman or a man.

It is important that every researcher reflects upon why they include 'gender' as a variable, how it is connected to their research question, and what aspect/s of gender that best may serve as a predictor for the attitudes or behaviour the researcher aims to explain. If gender seems to be a relevant factor – how is it relevant? Is it the bodily attributions, the assigned gender at birth or the self-identified gender identity? Or are attitudes related to gender identity more relevant?

This paper consists of two parts. First, we will cover different approaches to how we can categorize and ask about participants' gender in survey research. When browsing the literature on this question one finds many different practices and recommendations that not always correspond to each other. Hence, this is a complex matter. In the second part, we introduce some measures that can be used in addition to categorical measures, or instead of them, which also seem to be better predictors to many outcome variables than participants' gender. In both parts, we will give empirical examples from our own research. Our ambition with this paper is not to provide an exhaustive literature review of measuring gender in social sciences, nor are there any definitive solutions. Instead, we hope that this paper will provide a basis for further discussion and development of how gender is used in social science research.

Part 1: Categorizing gender

Defining and including gender variables

As already mentioned, the definition of gender that we use is highly important. *The Swedish Federation for Lesbian, Gay, Bisexual, Transgender and Queer Rights* (RFSL, 2014) defines 'gender' as constituted of four aspects: physiological/bodily aspects (*sex*); legal gender; social gender and gender expression; gender identity or self-defined gender. Similar ideas can be found in the work by scientific scholars in various fields. Some researchers add a timeline to the concept of gender, to demonstrate how these aspects of gender may change over a life-time, due to external impact from, e.g., society which includes social norms (Reisner et al., 2015). Others focus on how these aspects affect each other (Moerman & van Mens-Verhulst, 2004).

Regardless of how the complexity of the sex/gender feature is described, one common conclusion is that analysing gender differences alone is not enough (Joel, Tarrasch, Berman, Mukamel, & Ziv, 2014; Nowatzki & Grant, 2011). Instead, it is important that researchers carefully consider which dimensions of sex and/or gender that are important to their research question, and how to best operationalize these when designing studies (Reisner et al., 2015). In this text, we will use the term 'gender' as an umbrella term referring to any of these aspects.

In much social science research, the participants' gender may not be relevant *per se*. However, sometimes it is important to include a gender variable, for example when establishing salary differences or other structural disadvantages. Much official statistics use legal gender to assess such gender differences. That is one way to operationalize gender. In this part of the text, we discuss other approaches to the categorization of gender. Our main message is that these are aspects that every researcher needs to consider if they want to include gender as a variable.

The confusion between sex and gender

Both on a theoretical and a linguistic level, the terms 'sex' and 'gender' are confused with each other. In English, the biological/physiological aspect is often referred to as 'sex' while the social category is referred to as 'gender' (Frohard‑Dourlent, Dobson, Clark, Doull, & Saewyc, 2017), including cultural meanings associated with behaviour, personality and expressions labelled as feminine or masculine (Reisner et al., 2015).

The traditional dichotomous response alternatives to 'sex' are 'female/male' and the traditional dichotomous response alternatives to 'gender' might be 'woman/man' or 'masculine/feminine' (Ansara & Hegarty, 2014; Westbrook & Saperstein, 2015). However, most of us have participated in, seen or maybe even constructed surveys where the question asks about 'gender', but has two possible response alternatives consisting of 'male/female'. In fact, this seems to be the most common way of asking about participants' gender in the social sciences (Westbrook & Saperstein, 2015). This way of asking about gender visualizes how researchers, and people in general, make assumptions that links sex to gender. However, it is an incorrect assumption that sex precedes and thus determines gender (Butler, 1990; A Fausto-Sterling, 2012; Westbrook & Saperstein, 2015). Instead, this link is normative and thereby excludes many individuals with other experiences and/or identities.

In most societies, an individual's gender is categorized as 'female' or 'male' at birth, a categorization which is mostly based on a visual inspection of the baby's genitalia (A. Fausto-Sterling, Coll, & Lamarre, 2012). This procedure defines the individual's legal gender, which is salient in for example a passport or a birth certificate. However, how individuals identify or express their gender does not have to correspond to their assigned gender at birth. Thus, a survey question asking about 'gender', which has 'female' and 'male' as response alternatives makes it quite unclear what the researchers aim at. In addition, it might also be unclear what the participants respond to – bodily attributes, legal gender or self-defined gender identity?

Furthermore, the words 'sex' and 'gender' are also confused. In English, both terms are used in surveys and questionnaires, often interchangeable, even though they imply different meanings. In Swedish, the term for 'gender' (*genus*) is used mostly on a theoretical level (e.g., gender science), or in grammar (e.g., feminine/masculine/neuter). When referring to gender identity, gender segregation and/or gender roles, the Swedish term for 'sex' (*kön*) is instead used. In Swedish, the term for 'sex' also literally refers to the bodily attributes (in fact, the Swedish word *kön* also refers to genitalia). When asking for participants' gender in Swedish the word *kön* is used, most often with two response alternatives representing the words for 'woman' and 'man' – which accordingly are not linked to the biological aspect of sex/gender.

The confusion between sex and gender also implies a binary gender system, both implicitly and explicitly. For instance, many countries only allow two legal genders (female/male), although the biological/physiological variation is larger than this (Lundberg, 2017). Thus, using only two boxes implies that existing variation is not accounted for (Westbrook & Saperstein, 2015). There is a variety of intersex conditions in which individuals are born with anatomies that does not fit the typical dichotomy of female/male according to current medical norms (c.f., A Fausto-Sterling, 2000; Lundberg, 2017). In other words, individuals with an intersex or diversity of sexual development (DSD) fall outside this dichotomy (Nowatzki & Grant, 2011; Richards et al., 2016). Nonetheless, they are still most often assigned one of the binary genders. So far, only a handful of countries allow more than two legal genders (e.g., Australia, Canada, Germany, India, New Zealand).

Asking about gender identity in surveys

Gender

○ Woman
○ Man

Many researchers have already moved beyond the response categories of male/female when they ask about participants' gender identity. Instead, they use the two categories 'woman' or 'man'. In contrast to the response categories 'female' and 'male', the terms 'woman' and 'man' are more open and can refer to an individual's self-identified gender identity regardless of their bodily attributes. Therefore, a transwoman might feel more comfortable and confident in her response when asked to choose either 'woman' or 'man' to define her gender, especially if she does not share all the bodily attributes commonly presumed in the definition of 'female'. Because of this, the terms 'female' and 'male' should not be used interchangeably with the terms 'woman' and 'man' (Ansara & Hegarty, 2014).

However, only two possible categories to choose between still implies that gender is a binary category, consisting of the mutually exclusive alternatives 'woman' or 'man' (Richards et al., 2016; Westbrook & Saperstein, 2015). Since there are many other gender identities, this excludes all individuals who identify between or beyond the "traditional" gender dichotomy; who have a fluid gender identity; who do not identify with gender at all, etc (Nowatzki & Grant, 2011). Moreover, in Sweden, providing only two response options is even illegal according to

the Discrimination Act (DA. 2008:567§4) because it is a form of indirect discrimination. Nonetheless, this is still the standard way of asking about gender in surveys. The procedure appears neutral, but is in fact neglecting individuals of a certain gender identity – namely all other *nonbinary* gender identities outside/between the binary categories of woman/man. A nonbinary gender identity can be defined as belonging to the umbrella term "transgender" (Thanem, 2011) – a term which refers to individuals whose assigned gender at birth does not correspond to their self-identified gender identity. This can be compared to the term "cisgender" which refers to individuals whose assigned gender at birth does correspond to their self-defined gender identity (Frohard‐Dourlent et al., 2017).

Several gender categories as multiple choice

There are several ways of asking about participants' gender without treating gender as a binary category, for example, by adding a third option, which some researchers already do. Here, some third alternatives could be 'transgender', 'nonbinary' or 'other'. This strategy acknowledges that gender is not a binary category. However, it is quite difficult for the researcher to decide what alternatives to include. One problem with "response boxes" is that they imply that the categories in the boxes are mutually exclusive and that the researcher defines what the possible categories are. Some particular problems with the most common third categories are listed below.

    Transgender as a third alternative implies a mutually exclusive category in comparison to the binary categories woman/man. Some suggest that it should not be used (Ansara & Hegarty, 2014), because it falsely implies that all transgender individuals identify as 'trans*', but 'transgender' is not a gender identity, but rather an umbrella term. Many individuals with transgender experiences may not identify as being 'transgender', but also identify as women or men. This procedure might hence force participants to hierarchize their identities (Frohard‐Dourlent et al., 2017) – is a transwoman, for example, primarily 'transgender' or 'woman'?

    Nonbinary as a third option both acknowledges that gender is not a binary category and that 'transgender' is not a sufficient third option to complement 'woman' and 'man'. However, this alternative still implies that 'gender' has fixed categories (Richards et al., 2016). In fact, there are many other gender identities than 'women', 'men' and 'nonbinary' (e.g., genderfluid, genderqueer, etc). Because of this, the term 'nonbinary' might in fact be seen both as a descriptor or umbrella term, or as a possible gender identity (Frohard‐Dourlent et al., 2017).

    Other has also been used as third option. Using this procedure has the advantage of not defining what 'other' includes, meaning that individuals not identifying with the categories 'woman' or 'man' can fit here. However, this is also a disadvantage – if adding an 'other' option, the researcher does not know what it means (c.f., Ansara & Hegarty, 2014). One solution might be to add a free-text response connected to the 'other' option, making it possible for the participants to specify how they define 'other' (Reisner et al., 2015). However, there is still an implicit notion of gender as a

binary category: There are 'women'. 'men' and 'others' with 'a gender not listed here'; however, these 'other' people are clearly 'the deviant' (for a discussion on certain gender/s as norm, see e.g., Pratto, Korchmaros, & Hegarty, 2007). The deviant aspect is also included in the presentation order such that the norm is most often presented first (Kahneman & Miller, 1986).

Yet another option might be to *maximize the number of response categories*, such as for example 'cisgender', 'transgender', 'genderqueer', etc. (Broussard, Warner, & Pope, 2018). This procedure could be used, with an instruction to 'Please check all that apply' (Harrison, Grant, & Herman, 2012). However, even though this strategy is both ambitious and well-meant, it could be seen as one form of reductionism (Westbrook & Saperstein, 2015). Because of the countless terms individuals use to self-identify, it is probably impossible to create a question including all possible responses (Magliozzi, Saperstein, & Westbrook, 2016).

Gender as free-text response

Instead of defining what categories the participants can choose from when responding to a question about their gender, one possibility is to provide them with a blank text box where they write their self-defined gender (Ansara & Hegarty, 2014). Depending on what the researcher aims for, this question may need instructions. If the researcher is interested in the participants' self-identified gender, they may specify that. But, for many cispersons, it may be unusual to respond to questions about one's 'gender identity' – not all cispersons have reflected upon the relationship between (their) bodily attributes, assigned gender at birth and gender identity. Because of this, one suggestion is to provide a short description such as "if you identify as a woman, you may write 'woman' as your response".

The next step is to categorize the free-text responses. Because this paper addresses how to measure gender quantitatively, some categorization of qualitative responses are needed (Frohard‑Dourlent et al., 2017). If one uses a standard, population-based sample, the majority of participants will most likely be included in the categories 'woman' or 'man'. If the sample is not extremely large, there will probably not be enough statistical power to differentiate the sub-groups within an 'other' category. Such a third 'left-over' other-category is not optimal since it may consist of a diversity of identities. However, if power is not enough to detect statistical differences between more fine-grained categories, such a left-over category may be warranted since it may differ from the binary categories of woman/man. Again, if this is interesting or important, essentially depends on the research question. In this case, a new data collection aiming at a more gender-diverse population may be a solution to better understand the results.

Other times, a much more detailed categorization is warranted. For instance, if the sample is composed of a relatively large number of gender identities other than the common genders woman/man. This could be the case if the research question targets other identities. Categorization based on free-text choices is a bit more time consuming than response boxes with fixed categories. However, our own experience is that categorization is made quite fast by computerized methods that also develop rapidly within the social sciences.

Also, a response that "fits" with a binary gender should not be assumed to represent a cisgender person. In order not to make (binary) transgender individuals (even more) invisible, some researchers recommend to also include a question about assigned gender at birth, or if the participants have trans* experience (Frohard‑Dourlent et al., 2017; Magliozzi et al., 2016; Reisner et al., 2015). Others recommend this practice only when it is relevant to the research question (Ansara & Hegarty, 2014), for example if the aim is to analyse discrimination of transgender individuals, transgender individuals' experiences, etc.

Own data on gender identity as a free-text response

Free-text responses might trigger some participants to provide ridiculing responses, for example because they do not see the need for an open category and/or do not approve of other possible responses than woman/man. One possible solution to the potential problem with ridiculing responses, is that the researcher might prefer to give all participants the opportunity to make a satisfying response to the question about their gender, instead of forcing them to choose between response categories that might not be adequate for them. Using free-text responses will unavoidably lead to some missing data, but this should not be more systematic than missing data due to the response options excluding certain individuals.

Another response to that potential problem is empirical. During the last years when we have asked about participants' gender, we have used a free-text response. In Table 1, we show the free-text responses from four different studies (N = 1 146). In total, 5.4 % of the participants gave another response than 'woman' or 'man'[1]. From this table, it is obvious that only a few responses could be categorized as ridiculing (such as e.g., apache helicopter, pussy cat). This means that the missing data would be greater if participants would have been forced to respond woman or man (and left this blank) as compared to the ridiculing responses. Moreover, if this response option would become standard it should be perceived as less provocative.

---

[1] Even though responses like 'feminine', 'masculine', 'woman toward nonbinary', '*hen* (woman)', '*hen* (man)', 'man, I think…' etc might be interpreted as one of the common genders of woman/man, we did not categorize them as such. To avoid any interpretations of the participants' gender, only responses explicitly consisting of 'woman' or 'man' (or synonyms, such as the Swedish words for girl/boy, *tjej/kille*) were included in these two categories.

Table 1. Categorized free-text responses from 1 146 participants.

Gender identity	Frequency	Percent
woman	722	63.0
man	362	31.6
hen	7	0.6
feminine	2	0.2
masculine	2	0.2
?	1	0.1
not important	1	0.1
born with penis	1	0.1
hen (woman)	1	0.1
hen (man)	1	0.1
nonbinary	1	0.1
I don't know	1	0.1
pussy cat	1	0.1
woman and man	1	0.1
woman toward nonbinary	1	0.1
a little bit woman	1	0.1
man, I think…	1	0.1
most often woman	1	0.1
you shall not care about it	1	0.1
trans	1	0.1
outside	1	0.1
don't want to say	1	0.1
humapx (typo? Swedish: *människpx*)	1	0.1
I identify with an apache helicopter but am a biological man	1	0.1
Total	1 146	100.0

Part 2. Alternatives to gender identity as a predictor

Even when the participants' assigned gender at birth, bodily attributes and their self-identified gender identities correspond, 'sex' in terms of bodily differences is most often a poor proxy for many observed gender differences in social sciences, because this type of variable does not capture the full range of social and political dynamics that may affect the outcome variables (Nowatzki & Grant, 2011). In other words, even though the researcher manages to navigate the difficulties in asking about participants' gender, as discussed above, there are still other considerations left. Despite the wording used, in terms of, for example, avoiding the terms female/male, 'gender' itself implies several layers which can be illustrated in different ways.

Beside the gender categorization, other measures could be more relevant, again depending on the research question. There are a number of measures that could add knowledge about the respondent's gender identity or gender attitudes. Here, we present some of the most commonly used scales for measuring gender identity in social psychology. Following this, we also present some alternatives to using gender as a predictor.

Diverse measures of gender identification

There are some existing alternatives to treating gender as a categorical variable, and instead using continuous scales. The earliest and maybe most famous alternative is the *Bem Sex-Role Inventory*, BSRI (Bem, 1974). One important assumption in BSRI is the notion of gender as *not* constituting of two mutually exclusive categories or opposite poles in terms of femininity/masculinity. Instead, these are seen as two orthogonal dimensions of which an individual can have high values of one, both or none. In BSRI, femininity and masculinity are hence measured and analysed separately, and constituted of personality traits related to femininity and masculinity respectively. In other words, BSRI can be defined as measuring gender on a social level in terms of personality and behaviour, which may be related to gender expression. One critique of BSRI is that the traits impose definitions of femininity and masculinity which may be relying on stereotypes when assigning scale scores (Connell, 2005). Previous research has shown that BSRI and related instruments measuring traits associated with femininity and masculinity (such as Personal Attributes Questionnaire; Spence & Helmreich, 1978) may explain more variance than gender. For example, masculinity has been identified as a better predictor for aggression than gender (Hammock & Richardson, 1992), and instrumental managerial style as a better predictor of leadership aspiration (Marongiu & Ekehammar, 1999).

Another alternative is *Multi-Gender Identity Questionnaire*, Multi-GIQ (Joel et al., 2014). Just as BSRI, this measurement builds on the notion that femininity/masculinity are not two poles of a continuum, but rather independent dimensions. But, instead of measuring personality traits, Multi-GIQ has items related to how the participants self-identify with femininity and masculinity on levels related to gender identity; gender expression; legal gender; bodily aspects. All participants, despite gender identity, are asked to indicate to what extent they feel like a

woman *and* feel like a man; to what extent they wish to be a woman *and* a man, etc. One advantage of Multi-GIQ is that the instrument treats gender as a complex matter which may differ between contexts and levels. This is also a disadvantage – Multi-GIQ is an extensive instrument with 32 items which are not straight-forward to interpret and analyse. A simpler way is to just ask the participants about how feminine *and* masculine they see themselves, and how feminine *and* masculine they believe others see them (Magliozzi et al., 2016). Of course, some of the diversity of 'gender' is being lost by this, but it still treats femininity and masculinity as two independent and self-defined categories.

Own experiences of using continuous measures of gender identity

As discussed above, gender identity as a categorical variable may not be a sufficient predictor of some attitudes or behaviours. We have positive experiences of including continuous measures of gender in our own research. For example, we have included how strongly the participants identify with their gender, using the *gender identity subscale*, along with participants' gender as a free-text response. This is an adapted version of Luthanen and Crocker's (1992) well-established scale for measuring the strength of a social identity (Tajfel & Turner, 1986).

According to *Social identity theory*, identity consists of both individual and social factors. Individuals identify with a range of social groups that all constitute parts of their social identity. These groups can vary in definition and boundary from being very specific (member of the local football team) to highly loose (parent). They may also be of different importance in different situations, for example identifying as a Swedish citizen may be more salient when being abroad. Nonetheless, all social groups that an individual belongs to are seen as constituting parts of their identity – how they define and view themselves (Tajfel & Turner, 1986). The identification with social categories is also associated with attitudes toward both the own and other groups. Most importantly, categorization may lead to ingroup favouritism (Mullen, Brown, & Smith, 1992). It could also, but does not necessarily have to, lead to outgroup negativity (Brewer, 1999). Hence, an individual that strongly identify as a woman or a man (i.e., who strongly identify within a binary gender system) may be negative to violations of this binary system.

Below is the full scale that we have used in our research to measure gender identity. The question and items were phrased as follows (R = reversed item). Responses are made on 7-point Likert scales, and a mean index provides a general level of how important their gender identity is to them, despite what gender identity that may be.

> "Below are some questions about your gender identity. If you for example identify as a woman, you respond in relation to your experience as a woman."
> 1. My gender identity has very little to do with how I feel about myself (R).
> 2. My gender identity is an important part of my self-image.
> 3. My gender identity is an important reflection of who I am.
> 4. My gender identity has no significance for my sense of what kind of person I am (R).

When studying attitudes toward, and use of, the recently implemented Swedish gender-neutral third-person pronoun *hen* (Gustafsson Sendén, Bäck, & Lindqvist, 2015), we found that the gender categories woman/man[2] partly explained attitudes toward *hen* (such that participants self-identified as 'men' were more negative than participants self-identified as 'women'), but this effect disappeared when strength of gender identity was included in the statistical model. Instead, we found that the stronger a person identifies with their gender identity, the more negative are their attitudes toward *hen.* In another study, participants' self-identified gender identity (as binary gender[2]) did not predict attitudes toward *hen* at all, but identification with gender identity was still a significant predictor (Lindqvist, Gustafsson Sendén, & Bäck, 2016).

Additional related measures – sexism

So far, we have discussed how participants' self-identified gender identity account for differences in attitudes or experiences. However, also participants' attitudes and beliefs about gender-related issues may predict certain outcomes better than gender. In this part of the paper we present three different measurements of *sexism* and how they relate to political values and attitudes. The term *sexism* is similar to racism in that it indicates devaluation and prejudice toward a social category based on a certain aspect – here, gender (Fiske & Northm 2015). Thus, sexism can focus either on women, men or individuals with other identities. Nonetheless, the research on sexism has traditionally focused on negative attitudes toward women.

Currently, three scales are the most commonly used; (1) *modern (and old-fashioned) sexism* (2) *neosexism*, and (3) *ambivalent sexism* which consists of two dimensions (*benevolent/hostile*). The first two scales only include negative attitudes toward women, whereas the third also has a separate version about men (Glick & Fiske, 1999). For sexism toward women, participants self-identified as men generally score higher than individuals self-identified as women (these differences are smaller for benevolent sexism, see Fiske & North, 2015). Common for all three scales is that they try to capture sexism as a dynamic construct which evolves and takes new forms over time, for example by including dimensions that either emphasize traditional values or consider the feminist movement as unnecessary (or that equality work is enough). Also, all three scales relate to gender as constituting of the common genders woman/man. In other words, they can be described as testing the notion of gender as the mutually exclusive categories of women and men, and to what extent traditional gender roles are seen as important. Ambivalent sexism is also related to, and tests, heteronormativity. The three scales are described in more detail below.

The *old-fashioned and Modern sexism scales* (Swim, Aikin, Hall, & Hunter, 1995) show that sexism is dynamic and take different forms. Old-fashioned sexism is blatant and the items

---

[2] Because of too few participants having another gender identity than woman or man, we could unfortunately not include them in these analyses, for statistical reasons.

refer to attitudes like endorsement of traditional gender roles, polarizing women and men as well as devaluing women's competence. In comparison, modern sexism is subtler with items referring to attitudes that deny discrimination, devalue the importance of policies designed to help women, and feminist movements. Modern sexism can be especially useful because of its links to political, employment and harassment attitudes. Past studies have shown modern sexism to be predictive of less collective feminist action (Becker & Wagner, 2009), appreciating sexist humour (Eyssel & Bohner, 2007), gender harassment and bias (Hitlan, Pryor, Hesson-McInnis, & Olson, 2009), and of use and detection of sexist language (Parks & Roberton, 2004).

Neosexism (Tougas, Brown, Beaton, & Joly, 1995) is a unidimensional scale similar to modern sexism but with an emphasis on gender-related attitudes in society. Neosexism correlates more strongly with blatant sexist attitudes than benevolent and modern scales (Fiske & North, 2015). Neosexism also predicts acceptance of sexist language (Parks & Roberton, 2004), less feminist-movement support, negative attitudes toward lesbians and gay men, lower humanitarian-egalitarian values (Masser & Abrams, 1999), negative attitudes toward women's rights (Masser & Abrams, 1999), and toward affirmative action (Tougas et al., 1999). Individuals self-defined as men score higher on neosexism when they feel threatened, and view women as intrusive. Neosexism also correlates with pro-male bias in evaluating women's and men's competence, and unwillingness to support women (Beaton, Tougas, & Joly, 1996).

Ambivalent sexism builds specifically on ambivalent racism (Katz & Hass, 1988). The sources of ambivalence stem from the tight relationships between women and men (Peter Glick & Fiske, 1996), such as paternalism, gender differentiation and heterosexuality. Thus, although men are supposed to have higher societal status than women, the notion is that women might earn from being protected by men, being feminine and heterosexual. This dimension of sexism is labelled *benevolent sexism* and includes items assessing behaviours in which women cooperate with traditional gender roles. In contrast, *hostile sexism* is more similar to old-fashioned sexism where women who do not follow gender traditions are being punished. Thus, hostile sexism punishes some women whereas benevolent sexism rewards other women, together forming an ideology of how women should behave. In comparison to the other sexism scales, ambivalent sexism focuses on more intimate, relational aspects of sexism, consistent with its analysis of heterosexual woman-man interdependence.

Benevolent sexism differs from other measures of sexism in that gender differences are often smaller for benevolent sexism than hostile sexism. Hence, many women are positive to the traditional gender relations presented in the benevolent sexism scale. Examples of items are 'Women, compared to men, tend to have a superior moral sensibility', 'Men are incomplete without women', and 'Every man ought to have a woman whom he adores'. In one study, benevolent sexism predicted women's attitudes, such that high ratings on benevolent sexism predicted women's system-maintaining ideology and behaviour, such as preferring a high-resource partner (Sibley & Overall, 2011), whereas hostile sexism in men predicted stronger preferences for physically attractive partners. Furthermore, benevolent sexism predicted

women's acceptance of a partner's protective but sexist and restrictive paternalism (Moya, Glick, Exposito, de Lemus, & Hart, 2007) and acceptance of partners that react negatively to a wife's workplace success (Expósito, Herrera, Moya, & Glick, 2010).

Both hostile and benevolent sexism sees women as unfit for leadership. While hostile sexists openly acknowledge that this is due to women being incompetent at agentic tasks, benevolent sexism provides a comfortable rationalization for restricting women to domestic work (Glick & Fiske 1996). In the latter view, women are seen as weak and sensitive beings that need to be cared and provided for. This reinforces men's dominance. Sexism is also related to other traits, such as *Social dominance orientation, Right-wing authoritarianism,* and some *Big Five facets* (Akrami, Ekehammar & Yang-Wallentin 2011).

Own study on ambivalent sexism and support for presidential candidates

In a recent study, we explored how hostile and benevolent sexism predicted traits ascribed to political candidates, feelings toward the candidates and vote choice (Bäck, Carroll, Hansen, & Bäck, 2017). The study focused on the 2016 US Presidential candidates. In all analyses, ambivalent sexism was a better predictor for the outcome variables than gender (see table A1 in the appendix). Specifically, hostile sexism predicted support for Donald Trump, while benevolent sexism predicted support for Clinton. Sexism exerted significant effects even when controlling for a wide range of other possible predictors.

Figure 1 below shows the effect of sexism on probability to vote for Hillary Clinton, while controlling for age, gender (binary), if the participant was born outside the US, ethnicity, education, religiosity, ideology, political interest, and party identification.
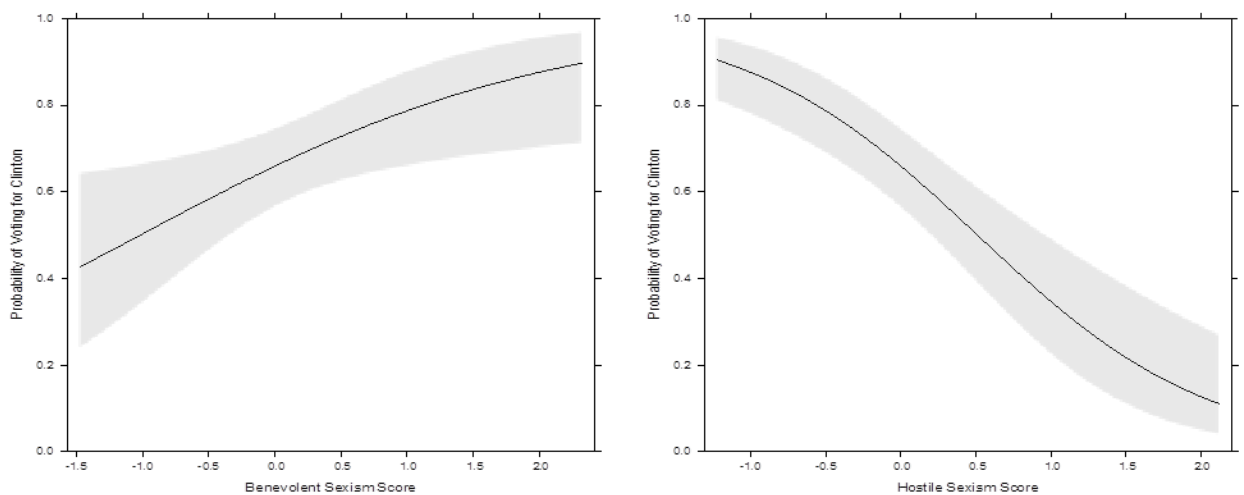

Figure 1. Probability of voting for Clinton as predicted by benevolent (left panel) and hostile sexism (right panel), controlling for other predictors.

Moreover, the effect of hostile sexism on support for Trump was explained by participants attributing a range of positive candidate traits to him, such as knowledge, leadership skills and intelligence. This was not the case for Clinton. The effect of benevolent sexism on support for Clinton, was explained by participants attributing 'communal' traits, such as honesty, morality and compassion to her. Hence, even though benevolent sexists seem to support Clinton, they do not think she possesses 'agentic' leadership traits (Eagly & Karau 2002).

Again, these results indicate that what we may really be after when measuring gender and predicting political behavior, is attitudes related to gender (roles). This particular empirical result also shows that it is important to not only consider the role of hostile, and more 'blatant' sexism when analysing political behaviour – in this case, benevolent sexism in fact had the opposite effect to what we might expect. We suggest that this result could potentially be explained by the particular context of the 2016 US election, which was characterized by the fact that a female candidate was running for president, facing an opponent who is likely to have been perceived as holding hostile sexist views and 'attacking' women as a group. We suggest that in such a situation, benevolent sexists may react positively towards a female presidential candidate, resulting in a higher likelihood of voting for her, since they should feel protective of women as a group (Bäck, Carroll, Hansen & Bäck 2017).

We have also identified modern sexism as a significant predictor for negative attitudes toward *hen* – while (as already mentioned) participants' self-identified gender was not a significant predictor (Lindqvist et al., 2016).

Even though sexism seems to be a relevant variable that explains more than participants' gender identity, some problems are worth noting. Much of the sexism research has focused on the denigration of women, however the term sexism should be more broadly defined. It should relate to denigrating attitudes due to gender of any identity. Moreover, the benevolent sexism scale, taking into account the positive relations between women and men do so with the implicit assumption that this heterosexual constellation is default.

Other additional measures

In our previous research, we have mainly included sexism as a relevant predictor, as discussed above. However, there is a range of other relevant measures as well, which could be included. One example of such a measure is the one-item question about *interest in gender issues*, which can be assessed on a 5-point Likert scale. In our own research, we have shown that also this measure was a better predictor of attitudes toward *hen*, compared to participants' gender identity (Gustafsson Sendén et al., 2015; Lindqvist et al., 2016). Also *Attitudes toward sexist language* (Parks & Roberton, 2000), which is the attitudes toward for example the use of masculine generics, was found to explain attitudes toward *hen* where an acceptance of sexist language predicted negative attitudes toward *hen* (Lindqvist et al., 2016).

Other examples of possible measures to include may be *Gender Lay Theory*, which tests the notion of gender as being socially constructed or biologically determined (Coleman & Hong,

2008) or *Gender Beliefs Scale* (Tee & Hegarty, 2006) which is the belief that there is only two, biologically determined, genders. There are most likely other scales that we have not covered in this overview, as well as scales that still needs to be developed.

Conclusion

In social sciences, much research is somehow related to participants' gender. We believe this is an important factor to take into account, but/and because of this we wish the operationalization of 'gender' to be sufficient. We believe that the decisive factor in deciding how to include gender comes down to what the researcher is interested in. If the goal is to map out pay gap differences, occurrences of discrimination, attitudes, behaviours, etc., the researcher needs to divide gender into a set of sub-categories. In relation to this, we have discussed different options of how this may be done. Researchers in social sciences are probably quite seldom interested in biological/physiological sex (genitalia, chromosomes, bodily attributes), but are more often interested in how individuals identify or express themselves from a social perspective, as well as how they may be seen and interpreted by others. Legal gender might be of interest in comparisons to population data (e.g., *Statistics Sweden*, SCB), whereas self-defined gender identity might be of interest in relation to the participants' experiences regarding social roles, division of labour, etc. However, asking about gender identity could be complex. In this paper, we have addressed this complexity, and how questions about gender should be more specific regarding what dimensions of gender they aim to capture.

These results also imply that other aspects may be more important than gender identity. One example may be identification with one's own gender. Our results imply that identification with the own gender, as measured with social identification, could be more important for attitudes and behaviour than the gender identity *per se*. More specifically, participants who strongly identified with their gender identity – regardless of what gender that was – tended to be more negative toward the gender-neutral pronoun *hen* (which challenges the notion of gender as a binary category) and less willing to use it. Essentially, this means that individuals who find it particularly important to be members of a certain gender category seem to be more negative to gender fair language, compared to those who do not think their gender identity is of particular importance. If we had measured nothing but participants' gender in terms of gender identity, this deeper knowledge about what predicts the outcome variable had gone missing.

Moreover, a number of attitudinal or behaviour variables may also be better predicted by other measures than participants' gender identity. For example, when we included measures of sexism in our research, we found that sexist attitudes were better predictors of support for the presidential candidates in the 2016 US election, as well as toward gender fair language (in terms of the pronoun *hen*). Other examples of this kind include interest in gender issues, which also may be a better predictor of attitudes, compared to participants' gender identity.

To sum up, our recommendation to researchers in social sciences is to relate their outcome variables to relevant measurements. Sometimes, participants' gender is relevant, such as when

assessing pay gap differences. If so, the researcher needs to operationalize 'gender' and formulate the question/s accordingly. Other times, participants' gender is not interesting *per se*. Instead, other related measures may be of higher predictive value. For example, individual levels of sexism may be more relevant than gender to predict some variables. If 'gender' is believed to be relevant, it is important to define what part of 'gender' that might be relevant – and how to operationalize it. Are there even other aspects that may overlap with gender identity, but which are not solely explained by gender identity, that better predicts the outcome variable?

References

Ansara, Y. G., & Hegarty, P. (2014). Methodologies of misgendering: Recommendations for reducing cisgenderism in psychological research. *Feminism & Psychology, 24*(2), 259-270. doi:10.1177/0959353514526217

Beaton, A. M., Tougas, F., & Joly, S. (1996). Neosexism Among Male Managers: Is It a Matter of Numbers?1. *Journal of Applied Social Psychology, 26*(24), 2189-2203. doi:doi:10.1111/j.1559-1816.1996.tb01795.x

Becker, J. C., & Wagner, U. (2009). Doing gender differently—The interplay of strength of gender identification and content of gender identity in predicting women's endorsement of sexist beliefs. *European Journal of Social Psychology, 39*(4), 487-508. doi:doi:10.1002/ejsp.551

Bem, S. L. (1974). The measurement of psychological androgynous. *Journal of Consulting and Clinical Psychology, 42*, 155-162.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55*(3), 429-444. doi:10.1111/0022-4537.00126

Broussard, K. A., Warner, R. H., & Pope, A. R. (2018). Too Many Boxes, or Not Enough? Preferences for How We Ask About Gender in Cisgender, LGB, and Gender-Diverse Samples. *Sex Roles, 78*(9), 606-624.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York: Routledge.

Bäck, H., Carroll, R., Hansen, M., & Bäck, E. A. (2017). Chivalry or Chauvinism? The impact of benevolent and hostile sexism in the 2016 US Presidential election. Presented at the Nordic Political Science Association, Odense, 2017.

Connell, R. W. (2005). *Masculinities, 2nd ed.* Berkeley, CA: University of California Press.

Expósito, F., Herrera, M. C., Moya, M., & Glick, P. (2010). Don't Rock the Boat: Women's Benevolent Sexism Predicts Fears of Marital Violence. *Psychology of Women Quarterly, 34*(1), 36-42. doi:10.1111/j.1471-6402.2009.01539.x

Eyssel, F., & Bohner, G. (2007). The Rating of Sexist Humor Under Time Pressure as an Indicator of Spontaneous Sexist Attitudes. *Sex Roles, 57*(9), 651-660. doi:10.1007/s11199-007-9302-5

Fausto-Sterling, A. (2000). The five sexes, revisited. *The Sciences, 40*, 18-23.

Fausto-Sterling, A. (2012). *Sex/gender: Biology in a social world*. New York: Routledge.

Fausto-Sterling, A., Coll, C. G., & Lamarre, M. (2012). Sexing the baby: Part 1-What do we really know about sex differentiation in the first three years of life? *Social Science & Medicine, 74*(11), 1684-1692. doi:Doi 10.1016/J.Socscimed.2011.05.051

Fiske, S. T., & North, M. S. (2015). Chapter 24 - Measures of Stereotyping and Prejudice: Barometers of Bias. In G. J. B. H. S. Matthews (Ed.), *Measures of Personality and Social Psychological Constructs* (pp. 684-718). San Diego: Academic Press.

Frohard-Dourlent, H., Dobson, S., Clark, B. A., Doull, M., & Saewyc, E. M. (2017). "I would have preferred more options": accounting for non-binary youth in health research. *Nursing inquiry, 24*(1).

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of personality and social psychology, 70*(3), 491.

Glick, P., & Fiske, S. T. (1999). The ambivalence toward men inventory - Differentiating hostile and benevolent beliefs about men. *Psychology of Women Quarterly, 23*(3), 519-536.

Gustafsson Sendén, M., Bäck, E. A., & Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. *Frontiers in psychology, 6*.

Hammock, G. S., & Richardson, D. R. (1992). Predictors of aggressive behavior. *Aggressive Behavior, 18*(3), 219-229. doi:doi:10.1002/1098-2337(1992)18:3<219::AID-AB2480180305>3.0.CO;2-P

Harrison, J., Grant, J., & Herman, J. L. (2012). A gender not listed here: Genderqueers, gender rebels, and otherwise in the National Transgender Discrimination Survey. *LGBTQ Policy Journal, 2*, 13-24.

Joel, D., Tarrasch, R., Berman, Z., Mukamel, M., & Ziv, E. (2014). Queering gender: studying gender identity in 'normative' individuals. *Psychology & Sexuality, 5*(4), 291-321. doi:10.1080/19419899.2013.830640

Kahneman, D., & Miller, D. T. (1986). Norm theory - Comparing reality to its alternatives. *Psychological Review, 93*(2), 136-153. doi:Doi 10.1037/0033-295x.93.2.136

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology, 55*(6), 893-905. doi:10.1037/0022-3514.55.6.893

Lindqvist, A., Gustafsson Sendén, M., & Bäck, E. A. (2016). Vem tycker om hen? *Språk och stil, 26*, 101-129.

Lundberg, T. (2017). *Knowing bodies: Making sense of Intersex/DSD a decade post-consensus*: Doctoral Thesis. Department of Psychology, University of Oslo.

Magliozzi, D., Saperstein, A., & Westbrook, L. (2016). Scaling Up: Representing Gender Diversity in Survey Research. *Socius: Sociological Research for a Dynamic World, 2*, 1-11.

Marongiu, S., & Ekehammar, B. (1999). Internal and external influences on women's and men's entry into management. *Journal of Managerial Psychology, 14*(5), 421-433. doi:10.1108/02683949910277175

Masser, B., & Abrams, D. (1999). Contemporary Sexism: The Relationships Among Hostility, Benevolence, and Neosexism. *Psychology of Women Quarterly, 23*(3), 503-517. doi:10.1111/j.1471-6402.1999.tb00378.x

Moerman, C. J., & van Mens-Verhulst, J. (2004). Gender-sensitive epidemiological research: suggestions for a gender-sensitive approach towards problem definition, data collection and analysis in epidemiological research. *Psychology, Health & Medicine, 9*(1), 41-52.

Moya, M., Glick, P., Exposito, F., de Lemus, S., & Hart, J. (2007). It's for your own good: benevolent sexism and women's reactions to protectively justified restrictions. *Pers Soc Psychol Bull, 33*(10), 1421-1434. doi:10.1177/0146167207304790

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status - an integration. *European Journal of Social Psychology, 22*(2), 103-122.

Nowatzki, N., & Grant, K. R. (2011). Sex is not enough: the need for gender-based analysis in health research. *Health care for women international, 32*(4), 263-277.

Pratto, F., Korchmaros, J. D., & Hegarty, P. (2007). When Race and Gender Go Without Saying. *Social Cognition, 25*(2), 221-247.

Reisner, S. L., Conron, K. J., Baker, K., Herman, J. L., Lombardi, E., Greytak, E. A., . . . Matthews, A. K. (2015). Counting" transgender and gender-nonconforming adults in health research: Recommendations from the Gender Identity in US Surveillance Group. *Transgender Studies Quarterly, 2*(1), 34-57.

RFSL. (2014). Kön. http://www.transformering.se/vad-ar-trans/kon.

Richards, C., Bouman, W. P., Seal, L., Barker, M. J., Nieder, T. O., & T'Sjoen, G. (2016). Non-binary or genderqueer genders. *International Review of Psychiatry, 28*(1), 95-102.

Spence, J. T., & Helmreich, R. (1978). *Masculinity and femininity - Their psychological dimensions, correlates, and antecedents*. Austin Texas: University of Texas.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism - old-fashioned and modern prejudices. *Journal of personality and social psychology, 68*(2), 199-214.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7-24). Chicago: Nelson-Hall.

Thanem, T. (2011). Embodying transgender in studies of gender, work and organization. In J. Knights, D & P. Y. Martin (Eds.), *Handbook of gender, work and organization* (pp. 191–204). Chichester: Wiley.

Tougas, F., Brown, R., Beaton, A. M., & Joly, S. (1995). Neosexism - Plus-Ca-Change, Plus Cest-Pareil. *Personality and Social Psychology Bulletin, 21*(8), 842-849. doi:Doi 10.1177/0146167295218007

Westbrook, L., & Saperstein, A. (2015). New categories are not enough: Rethinking the measurement of sex and gender in social surveys. *Gender & Society, 29*(4), 534-560.

Appendix

Table A1: Sexism, candidate traits and vote choice in the 2016 US election

	Vote Choice (Clinton = 1)	
	Model 1	Model 2
Benevolent Sexism	0.65*	0.45
	(0.26)	(0.38)
Hostile Sexism	-1.29*	-0.26
	(0.26)	(0.43)
Clinton traits		0.59*
		(0.15)
Trump traits		-0.40*
		(0.07)
Age	-0.01	0.03
	(0.02)	(0.03)
Woman	0.12	0.42
	(0.43)	(0.65)
Foreign-born	-0.28	-1.13
	(0.50)	(0.88)
White	-0.70	-1.00
	(0.52)	(0.95)
Education	0.32*	0.09
	(0.16)	(0.26)
Religiosity	-0.22	-0.38
	(0.16)	(0.23)
Political Ideology	-0.36*	-0.38*
	(0.09)	(0.18)
Political Interest	0.08	-0.15
	(0.08)	(0.15)
Independent	-1.60*	0.12
	(0.47)	(0.77)
Republican	-3.27*	-0.30
	(0.62)	(1.08)
Intercept	3.90*	3.32*
	(0.91)	(1.61)
N	322	315
PRE	0.692	0.868
ePRE	0.597	0.841

Coefficents are unstandardized logistic regression coefficients. Standard errors in parentheses. ∗ indicates significance at $p < 0.05$. Table from Bäck, Carroll, Hansen & Bäck (2017).