# Science Communication and the Problematic Impact of Descriptive Norms

Uwe Peters

When scientists or science reporters communicate research results to the public, this often involves ethical and epistemic risks. One such risk arises when scientific claims cause cognitive or behavioural changes in the audience that contribute to the self-fulfilment of these claims. I argue that the ethical and epistemic problems that such self-fulfilment effects may pose are much broader and more common than hitherto appreciated. Moreover, these problems are often due to a specific psychological phenomenon that has been neglected in the research on science communication: many people tend to conform to 'descriptive norms', that is, norms capturing (perceptions of) what others commonly do, think, or feel. Because of this tendency, science communication may frequently produce significant social harm. I contend that scientists have a responsibility to assess the risk of this potential harm and consider adopting strategies to mitigate it. I introduce one such strategy and argue that its implementation is independently well motivated by the fact that it helps improve scientific accuracy.

## 1. Introduction

Science plays an important role in liberal democracies. One of its key functions is to inform public deliberation and decision making (Anderson [2011]; Kitcher [2011]). To fulfil that function, scientists need to produce 'public scientific testimony', that is, oral or written claims about scientific results, theories, or research that are directed at laypeople (Gerken [2020]). In the context of COVID-19, climate change, and other pressing social problems, public scientific testimony was and is critical and highly salient, as the public depends on scientific expertise to understand and tackle these problems (Kitcher [2020]; Posetti and Bontcheva [2020]).

While public scientific testimony clearly plays a vital epistemic role in society, many philosophers of science have argued that it also comes with significant ethical and epistemic risks for the public (Forge [2008]; Kitcher [2011]; Keohane et al.

Electronically published August 11, 2023.

*The British Journal for the Philosophy of Science*, volume 74, number 3, September 2023.
 https://doi.org/10.1086/715001

[2014]; Alexandrova [2018]; Gerken [2018]; Keren [2018]).[1] Focusing on one of them, it has been noted that social scientific classifications and claims can in some cases contribute to bringing about their own truth, resulting in negative social consequences (Hacking [2007]; Kourany [2016]; Peters [2020]).

For instance, it has been argued that scientific claims about cognitive differences between men and women (Baron-Cohen [2003]), including assertions about men's higher scores in mathematics tests, can be harmful by leading the recipients of these claims to expect women to be less capable in mathematics (Fine [2012]; Kourany [2016]). This can cause people to think or act in ways that may contribute to the self-fulfilment of these claims: some women may subsequently be or feel discouraged to study mathematics or perform more poorly in this area, due to the fear of confirming (or disconfirming) the expectations of others ('stereotype threat').[2] I shall refer to cognitive and/or behavioural changes that are caused by or contribute to the confirmation of scientific claims as 'self-fulfilment effects'.

The harm of public scientific testimony tied to these kinds of effects has so far only been discussed in the context of testimony about gender and race differences (Fine [2012]; Kourany [2016]), mental disorders (Hacking [2007]), criminal conduct (Kuorikoski and Pöyhönen [2012]), implicit biases (Saul [2018]), and political polarization (Peters [2020]).[3] It might thus seem that the ethical and epistemic problems connected to these effects of scientific testimony are confined to a relatively small subset of such testimony. Furthermore, questions as to whether there is a particular and robust psychological process driving pernicious self-fulfilment effects, whether scientists are responsible for them, and how to counteract these effects remain largely unexplored in the theorizing on scientific testimony.[4] Here I want to start changing this. I shall argue for the following points:

(1) The ethical and epistemic problems with self-fulfilment effects of public scientific testimony are much broader and more common than hitherto appreciated, arising with respect to a wide range of ordinary public scientific generalizations about negative features of individuals, groups, society, and social structures.

(2) The reason for this is that self-fulfilment effects may often arise because of a common psychological phenomenon that has been neglected in the work on scientific testimony, namely, that many people tend to conform to 'descriptive norms'—norms capturing (perceptions of ) what others commonly

[1] For instance, in aiming to communicate their findings effectively to laypeople so that they can understand and use them, scientists might oversimplify results (Forge [2008]), ignore the value judgements underlying their conclusions (Alexandrova [2018]), or fail to acknowledge uncertainty (Keohane et al. [2014]).

[2] I will return to and assess the empirical data on stereotype threat as well as Fine's and Kourany's arguments in section 4.2.

[3] This article builds on, generalizes, and further explores the normative implications of a line of thought that I developed in (Peters [2020]).

[4] There is, for instance, no mentioning of them by Forge ([2008]), Anderson ([2011]), Kitcher ([2011]), Jamieson et al. ([2017]), or Gerken ([2018]).

do, think, or feel (Prentice [2007]).[5] Due to this tendency, scientific testimony can produce significant social harm.

(3) Scientists have a responsibility to assess the risk of this potential harm and consider adopting strategies to mitigate it. One such strategy involves linguistic restrictions of the generality of scientific claims. The implementation of this strategy is independently well motivated by the fact that it can help improve scientific accuracy.

In arguing for these three points, I shall not defend the claim that problematic self-fulfilment effects will arise in all cases of public scientific testimony about negative features of individuals, groups, society, and social structures. The overall point is more modest. It is that often, in a wider range of cases than so far noted, the effects at issue are likely to emerge.

I begin the discussion by specifying the kind of public scientific testimony relevant here. I will then connect it to descriptive norms, before making the case for these three points.

## 2. The Relevant Type of Scientific Claims

There are different kinds of public scientific testimony (Jamieson et al. [2017]; Alexandrova [2018]; Gerken ([2020]). It is useful to characterize the one relevant here by focusing on three of its dimensions: its source, content, and scope.

### 2.1. The source of the claims

When it comes to the source of public scientific testimony, we can distinguish between 'scientific expert testimony' and 'science reporting' (Gerken [2020]): scientific expert testimony is public scientific testimony whose direct sources are scientific experts themselves (including research institutions) and whose audience is laypeople. In contrast, science reporting is indirect, mediated, and offered by agents who frequently (but not always) lack scientific expertise about the relevant domain. Newspaper journalists, business associations, or governments reporting scientific findings could be examples. Since many science reporters and some government actors do have expertise and training in the scientific fields they report on, the two groups of testifiers referred to here are not always mutually exclusive but might overlap.

Moreover, and importantly for the following discussion, both types of public scientific testimony share a key feature: overall, the public assigns a high degree of trust and authority to them compared to non-scientific claims (Pew Research Center [2015]; Jamieson [2017]; Cacciatore et al. [2018]; Sanz-Menéndez and Cruz-Castro [2019]).[6]

[5] I use 'perception' in this article broadly to include non-perceptual beliefs.

[6] The public should not be reified as a monolithic entity though; it might be highly heterogeneous with respect to the individuals and groups it contains; see (O'Connor and Joffe [2014]).

With these points in mind, in the following, whenever I use the term 'public scientific testimony' (PST), I shall refer to both scientific expert testimony and science reporting.

## 2.2. The content of the claims

The content of PST may differ depending on the science it concerns. The sciences can be divided into four main branches (Trefil and Hazen [2009]): (i) the natural sciences, which study nature in the broadest sense (for example, biology, chemistry, physics), (ii) the social (including medical) sciences, which study individuals, groups, society, or relationships between them (for example, economics, psychology, health sciences), (iii) the formal sciences, which study abstract concepts (for example, logic, mathematics, computer science), and (iv) interdisciplinary science, which combines elements of the previous three branches. Corresponding to these four branches, four different kinds of PST can be distinguished. I will only be concerned with PST from ii and iv, more specifically, with testimony that involves social scientific and/or interdisciplinary claims about individuals, groups, society, or social structures. These claims are henceforth the sole referents of the term 'PST'. The reason for this narrow focus is that the kind of self-fulfilment effects discussed below are best illustrated and corroborated by empirical data with respect to these particular claims.[7]

Among them, we can further distinguish between assertions about properties viewed as positive or conducive to the functioning of individuals, groups, society, and/or social structures, and claims about properties viewed as negative or detrimental to it. While the positive versus negative distinction might not always be easy to draw, it is clear enough that it can be drawn. The following discussion will focus primarily on PST about negative (individually or socially undesirable or maladaptive) features. This is because even though self-fulfilment effects might also occur with respect to positive features (Peters [2022]), the potential harm of PST is (as will become clearer below) more closely linked to claims about negative than positive features.

## 2.3. The scope of the claims

While some social scientific claims might only be about a single individual, the ones relevant here involve, as PST typically does, generalizations about several individuals, that is, social groups, nations, social structures, and so on (Little [1993]). Two kinds of such social generalizations may be distinguished. These are (1) generalizations

[7] Some of the effects relevant here might also arise from natural scientific generalizations (for example, that endocrine disrupters cause infertility). However, the underlying social psychological processes are likely to be distinct; see also (Turnwald et al. [2019]).

expressed with explicit quantifiers (for example, 'many', 'most', 'all'), collective nouns (for example, 'US Americans'), or percentiles to describe the sample or functional relations at issue in the testimony (for example, '90% of people are biased'), and (2) generalizations expressed by social generics, which are broad, explicitly unquantified claims about a kind of people (that is, a category as a whole; for example, men, teachers, African-Americans) as opposed to individual members of the kind (Leslie [2017]). Generics are pervasive in many social scientific publications. For instance, DeJesus et al. ([2019], p. 18370) analysed 1,149 psychology articles published during 2015–16 and found that 'generics were ubiquitously used to convey results' in claims about a 'wide range of categories and constructs: People, women, children, adult people with schizophrenia', and so on. Given their pervasiveness, it becomes interesting to explore the social effects of such broad claims. In what follows, I will thus focus on PST with generalizations that involve generics, large (greater than 50%) percentiles, collective nouns, or majority quantifiers (for example, 'most', 'all', 'always').

In sum, then, the type of PST relevant here has three features. It (a) comes directly or indirectly from a scientific source, (b) captures social scientific and/or interdisciplinary claims pertaining to negative properties of individuals, groups, society, or social structures, and (c) expresses broad generalizations of the types just mentioned. From now on, I shall use the term 'PST' as shorthand for claims displaying these three features.

Notice that even though PST is but a subset of scientific testimony, it is large and highly relevant. After all, communicating to the public research results that pertain to negative or harmful characteristics of individuals, groups, society, or social structures is important for enabling informed policy-making to counteract them. For concrete examples of the type of claims that will be in the centre of the discussion, consider the following ten instances of them (found via a quick Google search):

(1) 'Men resist green behavior as unmanly' (Boroughs and Wilkie [2017]).

(2) 'Americans eat too much processed meat' (American Association for the Advancement of Science [2019]).

(3) 'People [on social media] are quicker to repeat something that's wrong than something that's true' (Fox [2018]).

(4) 'US voter turnout is low' (DeSilver [2022]).

(5) '90% of people are biased against women' (BBC [2020]).

(6) 'Most Republicans and Democrats view each other as more closed-minded than other Americans' (Pew Research Center [2015]).

(7) 'Smart people are more susceptible to [accepting] fake news' (Robson [2019]).

(8) 'Conservatives in the United States are substantially less likely than liberals to accept that human-caused climate change is happening, and less likely to support climate policies' (Goldberg et al. [2019]).

(9) 'Britons are uniquely reluctant to wear [COVID-19] face masks' (Smith [2020]).

(10) 'Unreliable and false information is spreading around the world to such an extent, that some commentators are now referring to the new avalanche of misinformation that has accompanied the COVID-19 pandemic as a "disinfodemic"' (United Nations [2020]).

I shall remain agnostic on the truth, plausibility, or evidential support of these ten claims, and any other claims displaying features a–c. What I am interested in is how people respond to statements of this kind, that is, to PST.

## 3. From Public Scientific Testimony to Descriptive Norms

It is useful to begin by considering psychological research on how people process broad generalizations about individuals or groups, in general, before returning to PST in particular. Psychological research on social norms will be especially relevant in this context.

### 3.1. The psychology of social norms

In investigating social norms, psychologists distinguish between 'descriptive norms' and 'prescriptive norms' (Cialdini et al. [1990]; Cialdini [2003]). Prescriptive norms are taken to capture (perceptions of) what characteristics and/or behaviour members of a group should (or should not) display according to some moral standard (for example, politicians should be honest).

In contrast, descriptive norms are norms capturing (perceptions of) what features members of a group in fact commonly do display, no matter whether they should or should not (Cialdini [2003]). That is, *prima facie*, descriptive norms merely describe what is widespread or typical in a social environment, including 'what most people in a group think, feel, or do' (for example, college students party a lot during spring break, CEOs sleep less than 8 hours a night; Prentice [2007], p. 629). This is the notion of descriptive norms often used by psychologists. There are other ones proposed by other researchers (for example, philosophers; see Muldoon et al. [2014]; Bicchieri [2017]).[8] I shall here adopt the psychologists' notion and terminology because they are now standard in empirical research on social norms.

With these clarifications in place, the key point to note for present purposes is that PST, including the ten claims listed above, is likely to convey descriptive norms. This is because of its broad and in many cases ambiguous scope. For instance, on

[8] Bicchieri ([2017], p. 19) defines them as a 'pattern of behavior such that individuals prefer to conform to it on condition that they believe that most people in their reference network conform to it'.

a natural reading, the ten examples quoted above convey that men commonly resist green behaviour (claim 1), Americans commonly consume unhealthy amounts of processed meat (claim 2), social media users commonly spread falsehoods faster than truths (claim 3), US voters are commonly less inclined to vote (claim 4), and so on.

Claims 1–10 do not have to be read in these ways. But given their unconstrained scope, they are likely to be interpreted thus by the public. That is, they are likely to convey (perceptions of) what features are common among certain groups of people. Since, in doing so, such generalizations convey descriptive norms, members of the public are likely to take PST, such as claims 1–10 to indicate descriptive norms. What, then, do we know about the effects of descriptive norms on people's behaviour and cognition?

### 3.2. Empirical data on the impact of descriptive norms

A wide range of different studies suggests that many people tend to conform to descriptive norms. I shall briefly review a selection of experiments testing effects of descriptive norms on behaviour and cognition.

As for behaviour-related research, in one study, (Californian) households informed that they used more electricity than all the other households in their neighbourhood subsequently reduced their energy consumption, whereas households informed that they used less than all others increased it (Schultz et al. [2007]). In another study, hotel guests informed that most people reuse their hotel towels were subsequently significantly more likely to do so than guests given environmental reasons for that action (Goldstein et al. [2008]).[9] Similarly, US lunch-goers informed that most people in the US consume high levels of meat tended to order more meaty lunch meals than controls (Sparkman and Walton [2017]), and students told that only a minority of their peers engaged in water conservation subsequently increased the amount of water they used compared to controls (Mortensen et al. [2019]). The same type of conformist behavioural response has been found with respect to majority information about various socially highly relevant actions, for instance, paying taxes (Hallsworth et al. [2017]), as well as corrupt behaviour (Köbis et al. [2015]) and even readiness to steal (Cialdini et al. [2006]).

Turning from descriptive-norm effects on behaviour to effects on cognition (for example, motivation, intentions, biases), presenting subjects with statements conveying descriptive norms emphasizing low (expected) voter turnout (in an election) elicited less motivation to vote than messages emphasizing high turnout (Gerber and Rogers [2009]). Also, students told that most of their peers do not eat healthily (but think that people should do so) subsequently reported significantly lower intentions

[9] The number of towels washed decreased by 26%.

to eat healthily themselves compared to controls (Staunton et al. [2014]). Additionally, individuals informed that the vast majority of people harbour stereotypes that bias their social evaluations subsequently showed themselves more stereotyping in their own social evaluations than those told that only very few people are biased by their stereotypes (and controls; Duguid and Thomas-Hunt [2015]).

Intriguingly, descriptive norms have also been found to affect subjects' perception of how others should act. In one study (Roberts et al. [2019]), children (4- to 9-year-olds) and adults were presented with two groups of fictional characters, 'Hibbles' and 'Glerks', described in terms of their positive/negative behaviours (for example, Hibbles make babies smile/cry; Glerks give people flowers/punches). The test participants were then asked to evaluate a dissenting individual (that is, a Hibble/Glerk not conforming to what its group was doing). Across ages, participants tended to judge, for instance, a Hibble that, unlike other Hibbles, made babies cry as worse than a Glerk that, like other Glerks, did the same thing. Similarly, individuals tended to view a Glerk that, unlike other Glerks, made babies smile as worse than a Hibble that, like other Hibbles, did the same thing. That is, many study participants expected individuals to conform to a group's descriptive norms, even though the participants themselves did not belong to that group, and even when these norms were morally problematic.

Notice that many of the experiments just mentioned were field studies, that is, they were conducted in naturalistic contexts ensuring high ecological validity (see Schultz et al. [2007]; Goldstein et al. [2008]; Gerber and Rogers [2009]; Hallsworth et al. [2017]; Sparkman and Walton [2017]). Moreover, there are various other studies on descriptive norms specifically venturing outside the laboratory, reporting the same kind of overall findings from a wide range of different domains (see, for example, Cialdini et al. [1990]; Mollen et al. [2013]). Given this, there is reason to assume that many subjects tend to conform to descriptive norms in many different contexts.

Notice too that while most of the studies just reviewed manipulated descriptive norms by directly providing summary information in the messages (for example, 'most people reuse their towels', 'the vast majority stereotype'), this is not required for the effects at issue to occur. Indeed, or other research found that subjects frequently form perceptions of descriptive norms by inferring these norms (often incorrectly) from simply observing others, their comments, and the media (for example, advertisements suggesting pervasiveness of smoking, or drinking; Borsari and Carey [2003]; Nan and Zhao [2016]; Liu and Shi [2019]).

### 3.3. Strength of the data

Given recent replication failures in the psychological sciences (Bird [2021]), it is worth noting that when it comes to the effects of the kind at issue here, meta-analyses of experiments on descriptive norms confirm the reality of the phenomenon. For instance, in the most recent meta-analysis, Melnyk et al. ([2019]) assessed 297 studies

on the impact of descriptive norms on consumer decision-making processes. They also contrasted the effects of descriptive norms with those of prescriptive norms so as to attain comparative insights into their distinctive causal contributions. They found that, overall, descriptive norms 'directly influence behavior' (not only intention formation) and their effect on behaviour is generally 'stronger than that of prescriptive norms' (pp. 4, 13). This meta-analysis and several individual experimental studies that manipulated the scope of descriptive norms to track causal norm-behaviour links (Kormos et al. [2015]; von Wagner et al. [2019]) suggest that descriptive norms do often exert a significant causal influence on behaviour.[10] Similarly, but related to these norms' effects on cognitions, Rivis and Sheeran ([2003], p. 228) assessed fourteen studies (total sample size $N = 5{,}810$) covering a wide range of behavioural domains, and they found 'strong evidence in support of the predictive validity of descriptive norms' in 'intention formation'.

The efficacy of these norms is perhaps also not difficult to explain. For instance, as Cialdini et al. ([1990], p. 203) note, they seem to 'motivate by providing evidence as to what will likely be an effective and adaptive action', the underlying rationale being, 'If everyone is doing or thinking or believing it, it must be a sensible thing to do or think or believe'. Similarly, Bicchieri ([2006], p. 29) writes that 'we conform because such norms make life easier for us, because we want to "fit in"'.

However, while there is good ground to hold that descriptive norms often prompt conformist behaviour, they do not always do so, and do not always do so equally powerfully across domains. There are contexts and domains where they have not been effective (Richter et al. [2018]; Paryavi et al. [2019]). And some kinds of descriptive norms (for example, dynamic norms, or don't- versus do-norms) have been found to be more powerful than others (Sparkman and Walton [2017]; Bergquist and Nilsson [2019]; Mortensen et al. [2019]). I shall thus assume a qualified and explicitly quantified view here: that in many (not necessarily all) domains descriptive norms are often (not necessarily always) likely to have the effects discussed.

## 4. Revisiting Public Scientific Testimony

The data just mentioned are relevant for the normative theorizing on PST because, as noted, PST frequently communicates descriptive norms. In fact, some of the studies just mentioned can be viewed as involving actual cases of PST doing precisely that. The reason is that the participants were laypeople who were presented, in naturalistic settings, with statements capturing broad, social scientific generalizations (for example, about energy consumption of their neighbourhood, people's voting behaviour, stereotyping) that came directly or indirectly from scientific sources (psychologists, political scientists, science reporters, and so on). The study participants were thus

[10] For instance, test subjects were told that 20%, 50%, or 80% of previous participants acted in a certain way; the broader the scope of the descriptive norm (for example, 50% versus 80%), the stronger the effect on intention to act (Kormos et al. [2015]; von Wagner et al. [2019]).

in much the same situation as people are when they receive PST in everyday life via TV, online, newspapers, and so on.

Importantly, the descriptive-norm effects were in many cases harmful. For instance, when participants learned about higher energy consumption in their neighbourhood, a low voter turnout, or the prevalence of stereotyping, this increased their energy consumption, decreased their voting inclination, and increased their stereotyping. Since, as noted, various different studies found the same phenomenon with different property ascriptions in diverse domains, there is an inductive basis for assuming that the ethical (behaviour-related) and epistemic (truth-, belief-, or knowledge-related) problems that this raises are likely to arise with respect to a wide range of PST.

To make this more concrete, consider again the ten claims mentioned above. *Prima facie*, these claims and instances of PST seem ethically and epistemically innocuous. But they also explicitly or indirectly (for example, claim 10) indicate to an audience that the cognitive and/or behavioural features described in the claims are common among people. That is, they involve broad social scientific generalizations (for example, '90%', 'most Republicans') or generic nouns ('men', 'Americans'), and so indicate descriptive norms. And importantly, these claims come from authoritative sources, that is, sources that the public commonly trusts (for example, news outlets, science associations, scientists, the United Nations). Since that is so, given the data on the directive impact of descriptive norms just reviewed, there is reason to believe that these ten claims are also likely to contribute to self-fulfilment effects among the public receiving these claims and belonging to the demographic groups concerned. These effects include (1) reluctance towards green behaviour among men, (2) increased meat consumption among Americans, a (3) proliferation of falsehoods among social media users, (4) low US voter turnout,[11] (5) bias against women, (6) affective polarization among Republicans and Democrats,[12] (7) acceptance of fake news among 'smart' people, (8) decreased support for climate policies among conservatives, and so on.[13] Some of these consequences are clearly ethically problematic, contributing, for instance, to the destruction of the environment, to bias against some individuals, and so on. Others are clearly epistemically problematic, contributing, for instance, to the spread of misinformation, thus hindering reliable belief formation or knowledge acquisition.

To be sure, no experiment yet exists that has tested whether the kind of self-fulfilment effects encountered in the studies I reviewed here will also occur with

[11] This example could be construed as conveying a negative descriptive norm: 'Americans don't vote'. It is worth noting that there is evidence that don't-descriptive norms are in fact more powerful in eliciting conformity than do-descriptive norms (see Bergquist and Nilsson [2019]).

[12] For a development of this point with respect to political polarization, see (Peters [2020]).

[13] What are the cues and immediate social environments that activate norm-conformist responses? Just as in the studies discussed above, they will differ depending on the content of the descriptive norm; for example, after exposure to a descriptive norm about green behaviour, the cues and immediate environments might include green versus non-green products and choice situations in supermarket.

respect to the specific PST in these ten claims. But independently of these claims, as mentioned, many of the experiments on descriptive norms discussed earlier can in fact be viewed as involving actual instances of PST on their own: they involve public (descriptive-norms conveying) claims by scientific or otherwise authoritative sources about negative features (for example, Schultz et al. [2007]; Goldstein et al. [2008]; Gerber and Rogers [2009]; Staunton et al. [2014]; Duguid and Thomas-Hunt [2015]). So even if one is sceptical about generalizing from the reviewed studies to the ten claims, the preceding discussion already directly supports the claim that in some cases PST leads to negative, descriptive norm related self-fulfilment effects.

And it is not unmotivated to go further to the assumption that such effects are likely to also arise at least in some cases with respect to the ten claims. This is because some of these claims are very closely related to the statements used in the studies reviewed (see claim 2 and Sparkman and Walton [2017]; claim 4 and Gerber and Rogers [2009]; claim 5 and Duguid and Thomas-Hunt [2015]), or are supported by other studies (for example, descriptive norms in the context of promotions of health-risk behaviour; see Cialdini [2003]; Rivis and Sheeran [2003]).[14]

Moreover, we have so far no reason to believe that when it comes to the rest of the ten claims, the descriptive norms mechanism discussed is interrupted or absent. And even if in the case of these ten claims, descriptive norms are less powerful and self-fulfilment effects less frequent or weaker than those found in the studies discussed, statements of the type at issue are very widely broadcast nationally and internationally (for example, on the BBC). This means that even only very modest self-fulfilment effects in individual cases might still accumulate to significant overall social harm.

Finally, suppose there is only a slight chance that the preceding considerations are correct and that claims such as 1–10 result in self-fulfilment effects. Given what is at stake (for example, environmentally friendly behaviour, spread of misinformation, voting), this would arguably still make it reasonable, if not rational, to take the considerations above seriously and reflect on how to avoid the potential effects outlined.

Adding to this point, while in some of the ten claims, the PST at issue concerns only a particular subset of individuals (for example, men, US voters), even when it comes to recipients of PST who are not themselves mentioned in the PST, the claims are still likely to incline the recipients to expect members of groups who are mentioned in the PST to act in ways conforming to the descriptive norm. The support for this assumption comes from the above-mentioned study by Roberts et al. ([2019]). Roberts et al. found that study participants presented with descriptive norms of a group to which the participants themselves did not even belong still subsequently tended to disapprove of norm-violating behaviour among members of that group.

[14] For instance, Cialdini ([2003], p. 105) mentioned the negative effects of descriptive norms in the context of claims that 'alcohol and drug use is intolerably high, that adolescent suicide rates are alarming', and 'that rampant polluters are spoiling the environment'.

It is thus not unreasonable to assume that people's expectations to the effect that, for instance, men resist green behaviour, Americans eat too much meat, 'fake news' spread quickly on social media, and so on might in subtle ways bias their social interactions such that self-fulfilment effects become more likely—not least in the sense of reducing people's surprise when they are actually encountering such behaviour. Additionally, as a result of being informed that certain negative features are common among people, recipients of that information might feel absolved from responsibility if they subsequently display such features themselves (see also 'moral licensing effects'; Blanken et al. [2015]; Saul [2018]). This, too, can further increase the spread of undesirable features.

### 4.1. Qualifications of the argument

What I have said so far is compatible with granting that there are many cases in which PST does not produce self-fulfilment effects. As noted at the outset, the claim here is not that receiving PST will always trigger such effects in an audience. For instance, there might be PST claims that do not capture any specific behaviour that people could conform to. Or there might be PST that the recipient distrusts, or motivated cognition might prevent an uptake of the claims (Gerken [2020]).[15] The people involved might then not interpret the PST as conveying descriptive norms, or might not conform to them. The argument here is only meant to support the view that, given the data discussed, there are also a wide range of cases in which the mentioned effects are likely to arise, that is, when the PST captures specific behaviour and the public trusts the source, which is frequently the case (Sanz-Menéndez and Cruz-Castro [2019]).

The overall argument here is qualified in another way. It grants that claims such as the ten discussed here might have overall more significant ethical and epistemic benefits than the costs linked to self-fulfilment effects. Relatedly, the preceding considerations are not intended to suggest that claims of the type I discuss here should never be made. The point is just that even if the ethical and/or epistemic benefits of the proliferation of PST, such as in these ten claims, outweigh the costs tied to self-fulfilment effects, we still have empirical reasons to believe that these effects are real and often problematic in a wide range of cases of PST. They should thus be taken into account in the theorizing on the ethical and epistemic consequences of PST.

[15] Recent studies suggest, however, that people's tendency to follow descriptive norms might in some cases even be stronger than their motivated cognition in favour of (for example) their political in-group: Pryor et al. ([2019], p. 1) found that 'counter to self-categorization theory's prediction', informing 'participants that an action was unpopular amongst people they did not [politically] identify with led participants' preferences to shift away from that action. These results suggest that a general desire to conform with others may out-power the common in-group vs. out-group mentality'.

### 4.2. A neglected issue

The preceding considerations have been largely overlooked in the literature on descriptive norms and the theorizing on PST. For instance, in the empirical literature on descriptive norms, the focus in discussions of experimental results is typically only on how descriptive norms can be used by, for instance, policy-makers to bring about positive social change (for explicit statements, see Gerber and Rogers [2009]; Duguid and Thomas-Hunt [2015]; Walton and Wilson [2018]; Lede and Meleady [2019]). This is an important question. But it seems equally vital to consider the responsibility of scientists for the negative social change that their testimony might (inadvertently) cause due to the directive impact of descriptive norms.

Some detrimental effects of descriptive norms have been discussed in the context of 'public service announcements' (that is, messages directed at the public, produced by governments or private institutions to raise awareness of, and alter attitudes and behaviour towards social issues; Cialdini [2003]). But the problem at issue has so far gone largely unnoticed in the context of scientific testimony and, more specifically, in the discussion of the responsibility of scientists. None of the articles cited so far relates the work on descriptive norms to the impact of PST on the public.

Of course, this is not a criticism of the scientists who conducted the studies. The link between descriptive norms and PST is not obvious. Moreover, the problem with self-fulfilment effects highlighted here is largely a normative one. As such, it may not be of interest to social scientists, who might, even if they have noticed the problem, view it as falling into a field of inquiry outside their expertise or training such as ethics or philosophy more generally (Nagel [1961]; Wolpe [2006]).

But unfortunately, philosophers have so far not paid much attention to this issue either. To be sure, in the philosophy of science, Hacking ([1995], [2007]) has influentially argued that social scientific classifications of individuals as autistic, depressive, criminal, immigrant, and so on can cause behavioural changes in these individuals such that a 'new scientific classification may bring into being a new kind of person, conceived of and experienced as a way to be a person' ([2007], p. 286). Hacking holds that the behavioural and cognitive changes triggered by the classifications may in turn trigger revisions in the classifications in order to ensure that they reflect these changes, a phenomenon he calls 'looping effects'. While Hacking's ideas are clearly connected and congenial to the points made here about self-fulfilment effects, he has not yet considered the role of descriptive norms in how social scientific classifications might 'make people up', and how this relates to the responsibilities that scientists have in their science communication.

But some other philosophers of science have touched on similar issues. For instance, Fine ([2012]) argues that neuroscientific claims might have pernicious self-fulfilment effects. Yet, she focuses 'primarily only on claims of the type that the

female brain is hardwired for empathizing while the male brain is hardwired for systemizing' (p. 286). Similarly, in her argument that certain scientific research and claims 'should be banned' (*inter alia*) because of their harmful self-fulfilment effects, Kourany ([2016], p. 779) concentrates primarily on research and testimony related to cognitive differences between gender and race groups. Neither Fine nor Kourany consider whether the problem of a self-fulfilment of scientific claims might in fact be highly general, potentially arising even for *prima facie* unproblematic claims (for example, about meat eating, voting, social media use).

Moreover, neither Fine nor Kourany consider empirical data specifically on descriptive norms but rely on results from studies pertaining to, for instance, 'stereotype threat' (Fine [2012], p. 288; Kourany [2016], p. 781). Stereotype threat is the fear that one's behaviour may confirm, or be interpreted in terms of, a negative stereotype linked to one's social group. This fear is thought to decrease one's performance on tasks associated with a stereotype-relevant domain by drawing cognitive resources away from task performance towards self-regulatory, fear-suppression processes (Steele [1997]).

No doubt, stereotype threat may be linked to self-fulfilment effects: by leading subjects to under-perform in stereotype-related tasks, it may contribute to a confirmation of stereotypes (for example, when the activation of the stereotype that women are bad at math reduces women's performance in math tests; Schmader et al. [2008]; Guyll et al. [2010]). As Fine and Kourany rightly emphasize, some PST might trigger stereotype threat and so cause pernicious self-fulfilment effects.

However, in these cases, the underlying psychological mechanism is distinct from the one that I highlighted here. For instance, descriptive norms need not capture stereotypes, and acting in line with them need not involve any under-performance anxiety but might just be based on self-interest (Bicchieri [2006], p. 26). Indeed, most of the PST and descriptive norms mentioned above (for example, 'men resist green behavior as unmanly', 'Americans eat too much processed meat') are unlikely to elicit the specific type of self-related anxiety found in stereotype threat, for they do not invoke any kind of under-performance to begin with. This means that the detrimental self-fulfilment effects of the PST at issue here are not covered or easily explicable by stereotype threat, and so the problem with PST self-fulfilment effects that I highlighted here is much more general than that of the stereotype-threat related effects that Fine and Kourany mention.

Moreover, the empirical basis of the argument I introduce above is more robust than that of Fine's and Kourany's arguments. Because even though the initial studies on stereotype threat found statistically significant effects (Nguyen and Ryan [2008]), subsequent re-analyses (Zigerell [2017]), and other studies failed to replicate them (Finnigan and Corker [2016]; Flore et al. [2019]). In contrast, meta-analyses of studies on descriptive norms repeatedly confirmed the reality of the type of self-fulfilment effects discussed above (for example, Melnyk et al. [2019]). The argument developed here thus does not only invoke a different psychological mechanism

and is broader in scope. It also enjoys stronger evidential support than Fine's, and Kourany's points.

Suppose, then, the argument is on the right track. To what extent are social scientists responsible for the pernicious self-fulfilment effects that might arise from their PST?

### 4.3. Scientists' responsibility for self-fulfilment effects

It is widely accepted that scientists have a moral duty to avoid causing harms to society and prevent wrongful outcomes and omissions that they are in a position to foresee (Douglas [2009]; Resnik and Elliott [2016]). That this is a moral obligation on social scientists, in particular, rather than a mere suggestion is supported by the fact that social scientists typically need to obtain approval for their experiments from an ethics committee. This requires them to reflect on, and commit to avoiding or mitigating, risks of harm to study participants and the public (Wassenaar and Mamotte [2012], p. 268).

However, scientists are not under an obligation to consider all possible harms of their professional actions. It is commonly accepted that the obligation only extends to harm that they can reasonably foresee (Douglas [2009]). Would the harm linked to self-fulfilment effects qualify as reasonably foreseeable?

Many social scientists or their peers might not be aware of the connection between their PST and its potential self-fulfilment effects. But consider, for instance, the members of a hiring committee. Suppose they lack an awareness of implicit bias. Given (i) their social role and (ii) the now available wealth of empirical evidence on implicit biases, it seems clear that the committee members should be aware of implicit bias even if they are not (see also Washington and Kelly [2016]).

Similarly, the data on descriptive norms and their impact on behaviour and cognition, too, have been known for about 30 years now (Cialdini et al. [1990]). Given this point and the fact that institutional boards (that is, ethics committees) explicitly require scientists to assess and limit potential risks of social harm related to their research, it seems equally clear that social scientists should be aware of potential PST-related self-fulfilment effects too. After all, it is hard to see why the social-role specific obligation for scientists to assess and limit risks of social harm that ethics committees confer on them prior to conducting an experiment should cease to hold afterward and with respect to other professional actions including the communication of the results to the public. It is plausible, then, to hold that social scientists have the following responsibility[16]:

[16] It seems plausible to hold that science reporters (not only social scientists) providing PST have this kind of responsibility and related obligation too. However, the basis for this normative claim is less clear than in the case of scientists. To keep the discussion focused, I shall set science reporters aside here.

**Communicative Risk Anticipation:** Scientists should assess whether their PST is likely to have detrimental self-fulfilment effects and consider adopting strategies to mitigate the risks related to these effects.

The principle of communicative risk anticipation is modest. While it requires scientists to analyse the risks of their PST and consider implementing ways of reducing them, it does not yet ask them to in fact adopt strategies to avoid problematic self-fulfilment effects related to descriptive norms. This matters because, as acknowledged above, there could be cases where the negative consequences of PST-based self-fulfilment effects are outweighed by the PST's benefits. Consider, for instance, the earlier PST example: 'Britons are uniquely reluctant to wear COVID-19 face masks'. This claim might disincline Britons in the audience from wearing masks. But it could also be precisely what policymakers need to know in order to implement a mask mandate. If so, the public communication of that information need not necessarily reduce mask wearing among the public. Given these (and other) complexities, a blanket requirement on scientists to adopt ways of avoiding descriptive norms related negative self-fulfilment effects would be too strong.[17]

Still, the preceding sections suggest that these effects are likely to arise in a wide range of cases. Scientists should thus factor them in. The principle of communicative risk anticipation captures this point. Suppose, then, that in line with this principle, scientists have assessed the self-fulfilment effects of their PST and wish to adopt strategies to mitigate the related risks. What strategies are there?

## 5. Counteracting the Problem: A Proposal

One possible way of reducing the risk of problematic self-fulfilment effects of PST might be to change the linguistic structure of the PST. There are different ways of doing so depending on the content of the claims at issue and the type of descriptive norm that they convey. Elsewhere I focus on PST about political polarization and propose three different linguistic modifications of it (Peters [2020]). Here I shall elaborate, refine, and further defend one of them, the one that strikes me as the most attractive and easily adoptable strategy. It involves explicit restrictions of the generality of PST.

[17] Adding further complications, the effects of PST depend partly on who the PST audience is (for example, different political orientation might lead to different trust assignments and information uptake; Nisbet et al. [2015]). And while some PST might necessarily be directed at a particular audience with particular needs, it may be picked up and disseminated beyond those for whom it is intended (for example, by other individuals who believe they understand the study but lack the expertise to understand the caveats and limitations). Thus, in some cases, the potential effects of PST are not easy to determine. Still, in line with the principle of communicative risk anticipation, scientists ought to assess these effects when thinking about their PST.

### 5.1. Mind the generics and broad generalizations

The guiding thought is that the easier it is for subjects to generalize social scientific claims to most members of groups, the higher the likelihood that they will construe these claims as conveying descriptive norms. This is because these norms refer to features or behaviour common (rather than infrequent) among people (Cialdini et al. [1990]). There is evidence that social scientific results expressed with social generics (explicitly unquantified generalizations about kinds of people as such, not individual members or subsets of them) tend to be viewed as more generalizable than findings expressed with non-generic language (DeJesus et al. [2019]). To reduce the risk of harmful self-fulfilment effects of their PST, then, scientists and science communicators might decrease their use of generics and replace them with quantified statements referring to specific samples, narrow target populations, or subsets of them.

Granted, if the quantified expressions to replace the generics and other broad generalizations in the PST could only be statements with 'all' or 'most' so as to preserve accuracy, then even with this rephrasing, PST would remain problematic for the same reason as before. However, Henrich et al. ([2010, p. 61]) found that 'behavioral scientists routinely publish broad claims about human psychology and behavior in the world's top journals based on samples drawn entirely from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies. Researchers—often implicitly—assume that either there is little variation across human populations, or that these 'standard subjects' are as representative of the species as any other population'. Henrich et al. went on to show that these assumptions are not justified. By reviewing the comparative database from across the behavioural sciences, they found that there is 'substantial variability in experimental results across populations and that WEIRD subjects are particularly unusual compared with the rest of the species—frequent outliers' (p. 61).

While ten years have passed since Henrich et al.'s article, there is reason to believe that key aspects of the problem that the article highlighted still persist today. For instance, Simons et al. ([2017], p. 1123) note: 'Psychological scientists draw inferences about populations based on samples—of people, situations, and stimuli—from those populations. Yet, few articles identify their target populations, and even fewer justify how or why the tested samples are representative of broader populations'.

Indeed, in a study already mentioned above, DeJesus et al. ([2019], p. 18375) analysed more than 1,000 psychology articles and found not only that 'generics were ubiquitously used to convey results', but also that there was 'no evidence that [the use of generics] was warranted by stronger evidence, as it was uncorrelated with sample size. Instead, authors showed an overwhelming tendency to treat limited samples as supporting general conclusions, by means of universalizing statements'. In fact, scientists often used generics despite omitting information on sample features (for example socioeconomic status) or having only small or unrepresentative samples (for

example, Western, white, middle-class) as a basis. Strikingly, authors failing to mention, for instance, socioeconomic status tended to use more generics than those who did mention it (p. 18373).

Taken together, Henrich et al.'s, Simons et al.'s, and DeJesus et al.'s observations suggest that many generics and other broad generalizations currently found in PST in scientific articles that are often freely accessible online can be replaced with claims involving quantifiers other than 'all' or 'most' in the result reports. Since doing so helps reduce the risk that the public is exposed to socially harmful self-fulfilment effects related to descriptive norms, there is (given the principle of communicative risk anticipation) a reason for scientists to adopt this strategy.

## 5.2. An independent epistemic rationale

Such a replacement would, in fact, not only help reduce harmful self-fulfilment effects of PST for the public, but also contribute to epistemically better scientific conduct. After all, if the sample of a study is only small or the results only pertain to people with, say, a certain socioeconomic status, then clearly scientists ought to communicate these facts to their audience, be it other scientists or the public, and not gloss over exceptions and variability by using, say, generics. Basic epistemic norms of scientific accuracy seem to require this already (Resnik [2005]; Forge [2008]). Relatedly, focusing specifically on intra-scientific communication, Simons et al. ([2017], p. 1123) propose that publications should include 'constraints on generality' (COG) statements that identify and justify target populations for the reported findings because 'Explicitly defining the target populations will help other researchers to sample from the same populations when conducting a direct replication, and it could encourage follow-up studies that test the boundary conditions of the original finding. Universal adoption of COG statements would change publishing incentives to favor a more cumulative science'. Notice that broad, generic generalizations formed on the basis of only small samples are not necessarily inaccurate. This is because generics do not imply that all members of the group described have a certain property. They allow for exceptions and counterevidence ('Mosquitos carry malaria' is a true generic, yet it only applies to less than 10% of all mosquitos; Leslie [2017]). It might thus seem that when scientists use generics even though their samples are only small (or outliers), they do not yet violate norms of scientific accuracy.

However, this would overlook the following. Cimpian et al. ([2010]) found that while generic claims of the form '*P*s (for example, men) are *f* (for example, resist green behaviour)' are generally accepted on relatively weak evidence, when they are unfamiliar with *P*, study participants construed such claims strongly, as conveying that almost all *P*s are *f*. Since they may be interpreted as referring to 'some', 'many', 'most', or 'all' *P*s, generics in PST involve an ambiguity that may lead different people (for example, scientific experts versus non-experts) to different interpretations depending on their expertise concerning *P*. This ambiguity or openness to

different interpretations itself is in tension with scientific accuracy. Because if a particular empirical claim has been experimentally supported with, for example, a WEIRD sample only, then even though using a generic might not necessarily involve making a false claim, it would still be inaccurate in that this leaves the social reference class of the claim more indeterminate than the experimental data warrant. It inaccurately suggests generalizability or 'projectability' (Munton [2019]) where the data may not support it. Replacing broad generic generalizations with explicitly quantified claims can help avoid this and may thus improve scientific accuracy while also facilitating result verification by reducing scope ambiguity.

It is worth noting too that a number of philosophers have argued that hearing and using social generics can also have ethically problematic consequences by making people more likely to 'essentialize' social kinds, for instance, racial or gender groups (Langton et al. [2012]; Leslie [2017]): generics are thought to lead people to believe that there is some hidden and stable property or underlying nature shared by members of the kind at issue that causally grounds their common properties and dispositions (for example, 'Latinos are temperamental', 'women are nurturing'). Langton et al. ([2012], p. 765) hold that generics in claims about gender or race groups should thus be rejected because they are 'false, and politically problematic', presenting 'social artifacts as racial [gender, and so on] essences'. To counteract such essentialization, Langton et al. (p. 765) propose that we should convey the content of generics by an 'explicitly quantified statement' involving 'some', 'most', or 'all'. To the extent that this essentialization-related difference between generics and quantified statements is empirically robust, there is thus an additional ethical reason for scientists to try to avoid generic generalizations in science communication.

### 5.3. Combining ethical and epistemic considerations

Given the preceding points, I thus propose the following strategy for reducing the risk of negative self-fulfilment effects of PST. It extends Simons et al.'s suggestion that scientists should include COG statements in their publications from intra-scientific to PST:

**PST Generality Constraint:** In the absence of overriding ethical, epistemic, or feasibility considerations, scientists should ensure that the generalizations in their PST (a) contain an explicitly quantified relativization to the population(s) to which the claims pertain and (b) are not more ambiguous or broader than warranted by the evidence.

Following this PST generality constraint principle has significant benefits. Given the points from the previous section, it can help (1) decrease the risk of negative self-fulfilment effects tied to PST, (2) reduce scientific inaccuracy related to over-generalizations, and (3) counteract social essentialization and stereotyping (assuming that generics are more likely than quantified claims to feed into these two processes).

Moreover, the PST generality constraint principle is attractively modest. It does not say that scientists should always avoid broad generalizations in their PST. Due to the qualifier 'in the absence of overriding ethical, epistemic, or feasibility considerations', it allows for cases in which using, for instance, generics would be justified, say, to ensure people's well-being, to counteract social injustice (Ritchie [2019]), or for feasibility reasons (for example, when scientists or science reporters have little time/space to give an opinion online, on TV, and so on).

Granted, settling whether certain considerations would override the communicative practice proposed might not always be easy. But it seems clear that this is often unproblematic. With respect to several of the ten claims discussed earlier, there are unlikely to be ethical, epistemic, or feasibility concerns overriding the benefits of rephrasing these claims with explicit quantifications and relativization (including past tense reference) to specific samples or narrow target populations. For example, 'Men resist green behaviour as unmanly', 'Americans eat too much processed meat', and 'People on social media are quick to repeat something that's wrong' could, if the evidence supports it, be easily rephrased thus: 'Many 20–40-year old men in Western societies resist green behaviour as unmanly', '65% of White, middle-class Americans may eat too much processed meat', and '1/3 of Facebook users in Britain were found to be quick to repeat something that's wrong'. In fact, even if there are overriding conditions against such rephrasing, following the PST generality constraint principle would still help improve the current provision of PST because scientists would then first need to consider the potential self-fulfilment effects of their PST and check them against other potential harms and benefits before testifying.[18]

Having said that, following the PST generality constraint does not eliminate all possible detrimental self-fulfilment effects of PST. For even when it involves explicitly quantified claims, PST might often pertain to a majority of individuals. In these cases, the risk of problematic effects related to descriptive norms will persist. But even so, since currently much PST involves over-generalizations (Simons et al. [2017]; DeJesus et al. [2019]) following the PST generality constraint principle may help significantly ethically and epistemically improve the current practice of providing PST.

## 6. Conclusion

Communicating scientific data to the public is important to inform public deliberation and democratic decision making. Yet, it also involves serious risks. Here, I focused on one intriguing such risk: social scientific claims about negative features of individuals, groups, society, or social structures can cause cognitive and behavioural changes that contribute to the self-fulfilment of these claims. I provided reasons to

[18] PST generality constraint also coheres well with, and can be added to, other recently defended guidelines for scientists and science reporters to avoid causing ethical and/or epistemic harm via PST (see, for example, Gerken [2018], [2020]).

believe that this is likely to happen frequently and can result in a wide range of harmful upshots (for example, environmentally unfriendly behaviour, low voter turnout, proliferation of 'fake news', bias against women). I argued that such problematic self-fulfilment effects are tied to many people's tendency to conform to descriptive norms. Social scientists should factor this tendency in when they are about to provide PST. This is because they have a responsibility to consider the potential harm resulting from their professional actions, including the communication of research results to the public. I introduced one possible strategy with which the risk of negative self-fulfilment effects of PST can be decreased, namely, to ensure that the generalizations in the PST contain an explicitly quantified relativization to the population(s) to which they pertain, and are not more ambiguous or wider in scope than warranted by the evidence. This strategy can help counteract the harmful effects related to people's conformist responds to descriptive norms and is independently well motivated by the fact that it increases scientific accuracy.

## Acknowledgements

Many thanks for comments on the article to two reviewers of the journal, Dan Kelly, Mattias Skipper, Olivier Lemeire, Ken Boyd, and Esben Nedenskov Petersen. The research for this article was partly funded by the Danmarks Frie Forskningsfond (grant no. 8018-00053B).

*Leverhulme Centre for the Future of Intelligence*
*University of Cambridge*
*Cambridge, UK*
and
*Center for Science and Thought*
*University of Bonn*
*Bonn, Germany*
*up228@cam.ac.uk*

## References

American Association for the Advancement of Science [2019]: 'Americans Still Eat Too Much Processed Meat and Too Little Fish', EurekAlert, 21 June, available at <https://www.eurekalert.org/pub_releases/2019-06/e-ase061719.php>.

Alexandrova, A. [2018]: 'Can the Science of Well-Being Be Objective?', *British Journal for the Philosophy of Science*, **69**, pp. 421–45.

Anderson, E. [2011]: 'Democracy, Public Policy, and Lay Assessments of Public Scientific Testimony', *Episteme*, **8**, pp. 144–64.

Baron-Cohen, S. [2003]: *The Essential Difference: Men, Women, and the Extreme Male Brain*, London: Lane.

BBC [2020]: 'Gender Study Finds 90% of People Are Biased against Women', BBC News, 5 March, available at <https://www.bbc.com/news/world-51751915>.

Bergquist, M. and Nilsson, A. [2019]: 'The Dos and Don'ts in Social Norms: A Descriptive Don't-Norm Increases Conformity', *Journal of Theoretical Social Psychology*, **3**, pp. 158–66.

Bicchieri, C. [2006]: *The Grammar of Society*, Cambridge: Cambridge University Press.

Bicchieri, C. [2017]: *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*, New York: Oxford University Press.

Bird, A. [2021]: 'Understanding the Replication Crisis as a Base Rate Fallacy', *British Journal for the Philosophy of Science*, **72**, pp. 965–93.

Blanken, I., van de Ven, N. and Zeelenberg, M. [2015]: 'A Meta-analytic Review of Moral Licensing', *Personality and Social Psychology Bulletin*, **41**, pp. 540–58.

Boroughs, A. R. and Wilkie, J. E. B. [2017]: 'Men Resist Green Behavior as Unmanly', *Scientific American*, 26 December, available at <https://www.scientificamerican.com/article/men-resist-green-behavior-as-unmanly/>.

Borsari, B. and Carey, K. B. [2003]: 'Descriptive and Injunctive Norms in College Drinking: A Meta-analytic Integration', *Journal of Studies on Alcohol*, **64**, pp. 331–41.

Cacciatore, M., Browning, N., Scheufele, D., Brossard, D., Xenos, M. and Corley, E. [2018]: 'Opposing Ends of the Spectrum: Exploring Trust in Scientific and Religious Authorities', *Public Understanding of Science*, **27**, pp. 11–28.

Cialdini, R. B. [2003]: 'Crafting Normative Messages to Protect the Environment', *Current Directions in Psychological Science*, **12**, pp. 105–9.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K. and Winter, P. L. [2006]: 'Managing Social Norms for Persuasive Impact', *Social Influence*, **1**, pp. 3–15.

Cialdini, R. B., Kallgren, C. A. and Reno, R. R. [1990]: 'A Focus Theory of Normative Conduct: Theoretical Refinement and Re-evaluation of the Role of Norms in Human Behavior', *Advances in Experimental Social Psychology*, **24**, pp. 201–33.

Cimpian, A., Brandone, A. and Gelman, S. [2010]: 'Generic Statements Require Little Evidence for Acceptance but Have Powerful Implications', *Cognitive Science*, **34**, pp. 1452–82.

DeJesus, J. M., Callanan, M. A., Solis, G. and Gelman, S. A. [2019]: 'Generic Language in Scientific Communication', *Proceedings of the National Academy of Sciences of the USA*, **116**, pp. 18370–77.

DeSilver, D. [2022]: 'Turnout in U.S. Has Soared in Recent Elections but by Some Measures Still Trails That of Many Other Countries', Pew Research Center, 1 November, available at <https://www.pewresearch.org/fact-tank/2018/05/21/u-s-voter-turnout-trails-most-developed-countries/>.

Douglas, H. [2009]: *Science, Policy, and the Value-Free Ideal*, Pittsburgh, PA: University of Pittsburgh Press.

Duguid, M. and Thomas-Hunt, M. [2015]: 'Condoning Stereotyping? How Awareness of Stereotyping Prevalence Impacts Expression of Stereotypes', *Journal of Applied Psychology*, **100**, pp. 343–59.

Fine, C. [2012]: 'Explaining, or Sustaining, the Status Quo? The Potentially Self-Fulfilling Effects of "Hardwired" Accounts of Sex Differences', *Neuroethics*, **5**, pp. 285–94.

Finnigan, K. M. and Corker, K. S. [2016]: 'Do Performance Avoidance Goals Moderate the Effect of Different Types of Stereotype Threat on Women's Math Performance?', *Journal of Research in Personality*, **63**, pp. 36–43.

Flore, P. C., Mulder, J. and Wicherts, J. M. [2019]: 'The Influence of Gender Stereotype Threat on Mathematics Test Scores of Dutch High School Students: A Registered Report', *Comprehensive Results in Social Psychology*, **3**, pp. 140–74.

Forge, J. [2008]: *The Responsible Scientist*, Pittsburgh, PA: Pittsburgh University Press.

Fox, M. [2018]: 'Fake News: Lies Spread Faster on Social Media than Truth Does', NBC News, 8 March, available at <https://www.nbcnews.com/health/health-news/fake-news-lies-spread-faster-social-media-truth-does-n854896>.

Gerber, A. and Rogers, T. [2009]: 'Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You', *Journal of Politics*, **71**, pp. 178–91.

Gerken, M. [2018]: 'Expert Trespassing Testimony and the Ethics of Science Communication', *Journal of General Philosophy of Science*, **49**, pp. 299–318.

Gerken, M. [2020]: 'Public Scientific Testimony in the Scientific Image', *Studies in History and Philosophy of Science* A, **80**, pp. 90–101.

Goldberg, M. H., van der Linden, S., Leiserowitz, A. and Maibach, E. [2019]: 'Reducing Ideological Bias on Climate Change', Yale Program on Climate Change Communication, 1 June, available at <https://climatecommunication.yale.edu/publications/reducing-ideological-bias-on-climate-change/>.

Goldstein, N., Cialdini, R. and Griskevicius, V. [2008]: 'A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels', *Journal of Consumer Research*, **35**, pp. 472–82.

Guyll, M., Madon, S., Prieto, L. and Scherr, K. C. [2010]: 'The Potential Roles of Self-Fulfilling Prophecies, Stigma Consciousness, and Stereotype Threat in Linking Latino/a Ethnicity and Educational Outcomes', *Journal of Social Issues*, **66**, pp. 113–30.

Hacking, I. [1995]: 'The Looping Effects of Human Kinds', in D. Sperber, D. Premack and A. Premack (*eds*), *Causal Cognition*, Oxford: Clarendon, pp. 351–83.

Hacking, I. [2007]: 'Kinds of People: Moving Targets', *Proceedings of the British Academy*, **151**, pp. 285–318.

Hallsworth, M., List, J. A., Metcalfe, R. D. and Vlaev, I. [2017]: 'The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance', *Journal of Public Economics*, **148**, pp. 14–31.

Henrich, J., Heine, S. J. and Norenzayan, A. [2010]: 'The Weirdest People in the World?', *Behavioral and Brain Sciences*, **33**, pp. 61–83.

Jamieson, K., Kahan, D. and Scheufele, D. [2017]: *The Oxford Handbook of the Science of Science Communication*, Oxford: Oxford University Press.

Jamieson, K. H. [2017]: 'The Need for a Science of Science Communication: Communicating Science's Values and Norms', in K. Jamieson, D. Kahan and D. Scheufele (*eds*), *The Oxford Handbook of the Science of Science Communication*, Oxford: Oxford University Press, pp. 15–25.

Keohane, R. O., Lane, M. and Oppenheimer, M. [2014]: 'The Ethics of Scientific Communication under Uncertainty', *Politics, Philosophy, and Economics*, **13**, pp. 343–68.

Keren, A. [2018]: 'The Public Understanding of What? Laypersons' Epistemic Needs, the Division of Cognitive Labor, and the Demarcation of Science', *Philosophy of Science*, **85**, pp. 781–92.

Kitcher, P. [2011]: *Science in a Democratic Society*, Amherst, NY: Prometheus.

Kitcher, P. [2020]: 'Can We Sustain Democracy and the Planet Too?', in J. Kourany and M. Carrier (*eds*), *Science and the Production of Ignorance*, Cambridge, MA: MIT Press, pp. 89–120.

Köbis, N. C., van Prooijen, J. W., Righetti, F. and Van Lange, P. A. [2015]: '"Who Doesn't?": The Impact of Descriptive Norms on Corruption', *PLOS One*, **10**, available at <doi.org/10.1371/journal.pone.0131830>.

Kormos, C., Gifford, R. and Brown, E. [2015]: 'The Influence of Descriptive Social Norm Information on Sustainable Transportation Behavior: A Field Experiment', *Environment and Behavior*, **47**, pp. 479–501.

Kourany, J. [2016]: 'Should Some Knowledge Be Forbidden? The Case of Cognitive Differences Research', *Philosophy of Science*, **83**, pp. 779–90.

Kuorikoski, J. and Pöyhönen, S. [2012]: 'Looping Kinds and Social Mechanisms', *Sociological Theory*, **30**, pp. 187–205.

Langton, R., Haslanger, S. and Anderson, L. [2012]: 'Language and Race', in G. Russell and D. Graff Fara (*eds*), *The Routledge Companion to Philosophy of Language*, New York: Routledge, pp. 753–67.

Lede, E. and Meleady, R. [2019]: 'Applying Social Influence Insights to Encourage Climate Resilient Domestic Water Behavior: Bridging the Theory-Practice Gap', *Wiley Interdisciplinary Review of Climate Change*, **10**, available at <doi.org/10.1002/wcc.562>.

Leslie, S. J. [2017]: 'The Original Sin of Cognition: Fear, Prejudice, and Generalization', *Journal of Philosophy*, **114**, pp. 393–421.

Little, D. [1993]: 'On the Scope and Limits of Generalizations in the Social Sciences', *Synthese*, **97**, pp. 183–207.

Liu, J. and Shi, R. [2019]: 'How Do Online Comments Affect Perceived Descriptive Norms of E-Cigarette Use? The Role of Quasi-Statistical Sense, Valence Perceptions, and Exposure Dosage', *Journal of Computer-Mediated Communication*, **24**, pp. 1–20.

Melnyk, V., Van Herpen, E., Jak, S. and Van Trijp, H. [2019]: 'The Mechanisms of Social Norms' Influence on Consumer Decision Making: A Meta-analysis', *Journal of Psychology*, **227**, pp. 4–17.

Mollen, S., Rimal, R. N., Ruiter, R. A. and Kok, G. [2013]: 'Healthy and Unhealthy Social Norms and Food Selection: Findings from a Field-Experiment', *Appetite*, **65**, pp. 83–89.

Mortensen, C. R., Neel, R., Cialdini, R. B., Jaeger, C. M., Jacobson, R. P. and Ringel, M. [2019]: 'Trending Norms: A Lever for Encouraging Behaviors Performed by the Minority', *Social Psychological and Personality Science*, **10**, pp. 201–10.

Muldoon, R., Lisciandra, C., Bicchieri, C., Hartmann, S. and Sprenger, J. [2014]: 'On the Emergence of Descriptive Norms', *Politics, Philosophy, and Economics*, **13**, pp. 3–22.

Munton, J. [2019]: 'Beyond Accuracy: Epistemic Flaws with Statistical Generalizations', *Philosophical Issues*, **29**, pp. 228–40.

Nagel, E. [1961]: *The Structure of Science*, New York: Harcourt, Brace, and World.

Nan, X. and Zhao, X. [2016]: 'The Mediating Role of Perceived Descriptive and Injunctive Norms in the Effects of Media Messages on Youth Smoking', *Journal of Health Communication*, **21**, pp. 56–66.

Nguyen, H. H. D. and Ryan, A. M. [2008]: 'Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-analysis of Experimental Evidence', *Journal of Applied Psychology*, **93**, pp. 1314–34.

Nisbet, E. C., Cooper, K. E. and Garrett, R. K. [2015]: 'The Partisan Brain: How Dissonant Science Messages Lead Conservatives and Liberals to (Dis)Trust Science', *Annals of the American Academy of Political and Social Science*, **658**, pp. 36–66.

O'Connor, C. and Joffe, H. [2014]: 'Social Representations of Brain Research: Exploring Public (Dis)engagement with Contemporary Neuroscience', *Science Communication*, **36**, pp. 617–45.

Paryavi, M., Bohnet, I. and van Geen, A. [2019]: 'Descriptive Norms and Gender Diversity: Reactance from Men', *Journal of Behavioral Public Administration*, **2**, pp. 1–16.

Peters, U. [2020]: 'How (Many) Descriptive Claims about Political Polarization Exacerbate Polarization', *Journal of Social and Political Psychology*, **9**, pp. 42–36.

Peters, U. [2022]: 'What Is the Function of Confirmation Bias?', *Erkenntnis*, **87**, pp. 1351–76.

Pew Research Center [2015]: 'Public and Scientists' Views on Science and Society', Pew Research Center, 29 January, available at <https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2015/01/PI_ScienceandSociety_Report_012915.pdf>.

Posetti, J. and Bontcheva, K. [2020]: *Disinfodemic: Deciphering COVID-19 Disinformation*, Paris: UNESCO.

Prentice, D. A. [2007]: 'Prescriptive vs. Descriptive Norms', in R. Baumeister and K. Vohs (*eds*), *Encyclopedia of Social Psychology*, Newbury Park, CA: Sage, pp. 629–30.

Pryor, C., Perfors, A. and Howe, P. [2019]: 'Conformity to the Descriptive Norms of People with Opposing Political or Social Beliefs', *PLOS One*, **14**, available at <doi.org/10.1371/journal.pone.0219464>.

Resnik, D. B. [2005]: *The Ethics of Science: An Introduction*, London: Routledge.

Resnik, D. B. and Elliott, K. C. [2016]: 'The Ethical Challenges of Socially Responsible Science', *Accountability Research*, **23**, pp. 31–46.

Richter, I., Thøgersen, J. and Klöckner, C. A. [2018]: 'A Social Norms Intervention Going Wrong: Boomerang Effects from Descriptive Norms Information', *Sustainability*, **10**, available at <doi.org/10.3390/su10082848>.

Ritchie, K. [2019]: 'Should We Use Racial and Gender Generics?', *Thought*, **8**, pp. 33–41.

Rivis, A. and Sheeran, P. [2003]: 'Descriptive Norms as an Additional Predictor in the Theory of Planned Behaviour: A Meta-analysis', *Current Psychology*, **22**, pp. 218–33.

Roberts, S. O., Ho, A. K. and Gelman, S. A. [2019]: 'The Role of Group Norms in Evaluating Uncommon and Negative Behaviors', *Journal of Experimental Psychology: General*, **148**, pp. 374–87.

Robson, D. [2019]: 'Why Smart People Are More Likely to Believe Fake News', *Guardian*, 1 April, available at <https://www.theguardian.com/books/2019/apr/01/why-smart-people-are-more-likely-to-believe-fake-news>.

Sanz-Menéndez, L. and Cruz-Castro, L. [2019]: 'The Credibility of Scientific Communication Sources Regarding Climate Change: A Population-Based Survey Experiment', *Public Understanding of Science*, **28**, pp. 534–53.

Saul, J. [2018]: '(How) Should We Tell Implicit Bias Stories?', *Disputatio*, **10**, pp. 217–44.

Schmader, T., Johns, M. and Forbes, C. [2008]: 'An Integrated Process Model of Stereotype Threat Effects on Performance', *Psychological Review*, **115**, pp. 336–56.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J. and Griskevicius, V. [2007]: 'The Constructive, Destructive, and Reconstructive Power of Social Norms', *Psychological Science*, **18**, pp. 429–34.

Simons, D. J., Shoda, Y. and Lindsay, D. S. [2017]: 'Constraints on Generality (COG): A Proposed Addition to All Empirical Papers', *Perspectives on Psychological Science*, **12**, pp. 1123–28.

Smith, M. [2020]: 'International COVID-19 Tracker Update', YouGov, 2 May, available at <https://yougov.co.uk/topics/international/articles-reports/2020/05/02/international-covid-19-tracker-update-2-may>.

Sparkman, G. and Walton, G. [2017]: 'Dynamic Norms Promote Sustainable Behavior, Even If It Is Counternormative', *Psychological Science*, **28**, pp. 1663–74.

Staunton, M., Louis, W. R., Smith, J. R., Terry, D. J. and McDonald, R. I. [2014]: 'How Negative Descriptive Norms for Healthy Eating Undermine the Effects of Positive Injunctive Norms', *Journal of Applied Social Psychology*, **44**, pp. 319–30.

Steele, C. M. [1997]: 'A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance', *American Psychologist*, **52**, pp. 613–29.

Trefil, J. and Hazen, R. [2009]: *The Sciences: An Integrated Approach*, Hoboken, NJ: Wiley.

Turnwald, B. P., Goyer, J. P., Boles, D. Z., Silder, A., Delp, S. L. and Crum, A. J. [2019]: 'Learning One's Genetic Risk Changes Physiology Independent of Actual Genetic Risk', *Nature Human Behaviour*, **3**, pp. 48–56.

United Nations [2020]: 'During This Coronavirus Pandemic, "Fake News" Is Putting Lives at Risk: UNESCO', United Nations, 13 April, available at <https://news.un.org/en/story/2020/04/1061592>.

von Wagner, C., Hirst, Y., Waller, J., Ghanouni, A., McGregor, L. M., Kerrison, R. S., Verstraete, W., Vlaev, I., Sieverding, M. and Stoffel, S. T. [2019]: 'The Impact of Descriptive Norms on Motivation to Participate in Cancer Screening: Evidence from Online Experiments', *Patient Education and Counseling*, **102**, pp. 1621–28.

Walton, G. M. and Wilson, T. D. [2018]: 'Wise Interventions: Psychological Remedies for Social and Personal Problems', *Psychological Review*, **125**, pp. 617–55.

Washington, N. and Kelly, D. [2016]: 'Who's Responsible for This? Moral Responsibility, Externalism, and Knowledge about Implicit Bias', in M. Brownstein and J. Saul (*eds*), *Implicit Bias and Philosophy: Moral Responsibility, Structural Injustice, and Ethics*, Vol. 2, Oxford: Oxford University Press, pp. 11–37.

Wassenaar, D. R. and Mamotte, N. [2012]: 'Ethical Issues and Ethics Reviews in Social Science Research', in M. Leach, M. Stevens, G. Lindsay, A. Ferrero and Y. Korkut (*eds*), *The Oxford Handbook of International Psychological Ethics*, New York: Oxford University Press, pp. 268–82.

Wolpe, P. [2006]: 'Reasons Scientists Avoid Thinking about Ethics', *Cell*, **125**, pp. 1023–25.

Zigerell, L. J. [2017]: 'Potential Publication Bias in the Stereotype Threat Literature: Comment on Nguyen and Ryan (2008)', *Journal of Applied Psychology*, **102**, pp. 1159–68.