Science Communication and the Problematic Impact of Descriptive Norms

[Penultimate draft of a paper forthcoming in the *British Journal for Philosophy of Science*]

Uwe Peters
(1) Center for Science and Thought, University of Bonn, Germany
(2) Centre for the Future of Intelligence, University of Cambridge, UK
(3) Department of Psychology, King's College London, UK
Email: u.peters@kcl.ac.uk

Abstract

When scientists or science reporters communicate research results to the public, this often involves ethical and epistemic risks. One such a risk arises when scientific claims cause cognitive or behavioral changes in the audience that contribute to the self-fulfillment of these claims. Focusing on such effects, I argue that the ethical and epistemic problem that they pose is likely to be much broader than hitherto appreciated. Moreover, it is often due to a psychological phenomenon that has been neglected in the research on science communication, namely that 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 can 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 a strategy and argue that its implementation is independently well motivated by the fact that it helps improve scientific accuracy.

1 Introduction

'It ain't what they call you, it's what you answer to.' (W.C. Fields)

Science plays an important role in liberal democracies. One of its key functions is to inform public deliberation and decision-making (Kitcher [2011]; Anderson [2011]). To fulfill 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 lay-people (Gerken [2019], p. 1). In the context of COVID-19, climate change, and other pressing social problems, now more than ever, public scientific testimony is critical and highly salient, as the public depends on scientific expertise to understand and tackle these problems (Posetti and Bontcheva [2020]; Kitcher [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.* [2014]; Alexandrova [2018]; Keren [2018]; Gerken [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 (e.g., Hacking [2007]; Kourany [2016]; Peters [2020a], [2020b]).

For instance, it has been argued that scientific claims about cognitive differences between men and women, 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 bring about a self-fulfillment of these claims: women may subsequently be or feel discouraged to study mathematics or perform more poorly out of fear of confirming (or disconfirming) others' expectations ('stereotype threat').[2] This can contribute to a reproduction of the cognitive difference at issue (Schwartz [1997], p. 24; Fine [2012]). I shall refer to cognitive and/or behavioral changes that are caused by and contribute to the confirmation of scientific claims as *self-fulfilling effects*.

The harm of public scientific testimony tied to self-fulfilling 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 [2020a]).[3] It might thus seem that the ethical and epistemic problems connected to these effects of scientific testimony are confined to a relatively small sub-set of such testimony. Furthermore, questions as to

---

[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 judgments 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 paper builds on, generalizes, and explores the normative implications of the line of thought developed in Peters ([2020a]).

whether there is a particular and robust psychological process driving pernicious self-fulfilling 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 problem with self-fulfilling effects of public scientific testimony is likely to be much broader than hitherto appreciated, arising with respect to a wide range of common public scientific generalizations about negative features of individuals, groups, society, and social structures.

(2) It often does so because of a 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[5] of) what others commonly do, think, or feel (Cialdini *et al.* [1990]; Cialdini [2003]; Prentice [2007]; Peters [2020a]). 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 a strategy involves linguistic restrictions of the generality of scientific claims, and the implementation of this strategy is independently well motivated by the fact that it helps improve scientific accuracy.

In arguing for (1)-(3), I shall not defend the claim that in all cases of public scientific testimony about negative features of individuals, groups, society, and social structures, problematic self-fulfilling effects will arise. 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 (1)-(3).

2 The Relevant Type of Scientific Claims

There are different kinds of public scientific testimony (Alexandrova [2018]; Jamieson *et al.* [2017]; Gerken [2019]). 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 ([2019], p. 2): scientific expert testimony is public scientific testimony whose direct sources are scientific experts themselves (including research institutions) and whose audience is lay-people. 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

---

[4] There is, for instance, no mentioning of them in Forge ([2008]), Anderson ([2011]), Kitcher ([2011]), Jamieson *et al.* ([2017]), or Gerken ([2018]; [2020]).

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

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[6] assigns a high degree of trust and authority to them compared to non-scientific claims (Funk and Rainie [2015]; Jamieson [2017]; Cacciatore *et al.* [2018]; Sanz-Menéndez and Cruz-Castro [2019]). With these points in mind, in the following, whenever I use the term 'public scientific testimony' (henceforth '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 (e.g., biology, chemistry, physics), (ii) the social sciences, which study individuals, groups, society, or relationships between them (e.g., economics, psychology, political science), (iii) the formal sciences, which study abstract concepts (e.g., logic, mathematics, computer science), and (iv) interdisciplinary science, which combines elements of (i)-(iii). Corresponding to (i)-(iv), 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-fulfilling 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 (e.g., maladaptive behaviors). While the positive vs. 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* features. This is because even though self-fulfilling effects might also occur with respect to positive features [e.g., Duguid and Thomas-Hunt [2015]; Peters [2020b]), the potential *harm* of PST is (as will become clearer below) more closely linked to claims about negative than positive features.

---

[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]).

[7] Some of the effects relevant here might also arise from natural scientific generalizations (e.g., that endocrine disruptors cause infertility). However, the underlying social psychological processes are likely to be distinct; see also Turnwald *et al.* ([2018]). Thanks to an anonymous reviewer for raising the issue.

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 generalizations can be distinguished. These are (1) generalizations expressed with explicit quantifiers (e.g., 'many', 'most', 'all', etc.), collective nouns (e.g., 'U.S.A.'), or percentiles to describe the sample or functional relations at issue in the testimony (e.g., '90% of people are biased'), and (2) generalizations expressed by *generics*, which are statements without explicit quantifiers.

Generics capture broad claims about a kind (i.e., a category as a whole; e.g., men, teachers, African-Americans, etc.) as opposed to individuals (Leslie [2017]). Generics are pervasive in many social scientific publications. For instance, DeJesus *et al.* ([2019], pp. 18370–18375) analyzed 1,149 psychology articles published 2015 to 2016 and found that 'generics were ubiquitously used to convey results' in claims about a 'wide range of categories and constructs: People, women, children, adults, people with schizophrenia', and so on. Given their pervasiveness, it becomes interesting to explore the social effects of such claims. In what follows, I will thus focus on PST with board generalizations that involve generics, large percentiles, collective nouns, and/or wide-scope quantifiers (e.g., 'most', 'all', 'always', etc.).

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 (a)-(c).

Notice that even though PST is but a sub-set of scientific testimony, it is large (see, e.g., DeJesus *et al.* [2019]) 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. To illustrate these points and give concrete examples of the type of claims that will be in the center of the discussion, consider the following ten instances of them (found via a quick Google search):

(1) 'Men resist green behavior as unmanly.' (*Scientific American* [2017])[8]
(2) 'Americans eat too much processed meat.' (American Association for the Advancement of Science [2019])[9]
(3) 'People [on social media] are quicker to repeat something that's wrong than something that's true.' (NBC [2018])[10]
(4) 'U.S. voter turnout is low.' (Pew Research Center [2018])[11]

---

[8] https://www.scientificamerican.com/article/men-resist-green-behavior-as-unmanly/
[9] https://www.eurekalert.org/pub_releases/2019-06/e-ase061719.php
[10] https://www.nbcnews.com/health/health-news/fake-news-lies-spread-faster-social-media-truth-does-n854896
[11] https://www.pewresearch.org/fact-tank/2018/05/21/u-s-voter-turnout-trails-most-developed-countries/

(5) '90% of people are biased against women.' (BBC [2020])[12]

(6) 'Most Republicans and Democrats view each others as more closed-minded than other Americans.' (Pew Research Center [2019])

(7) '[S]mart people are more susceptible to [accepting] fake news.' (*The Guardian* [2019])[13]

(8) '[C]onservatives 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.' (Yale Climate Change Communication [2019])[14]

(9) 'Britons are uniquely reluctant to wear [COVID-19] face masks.' (YouGov [2020])[15]

(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])[16]

I shall remain agnostic on the truth, plausibility, or evidential support of (1)-(10) and 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 behavior members of a group *should* (or should not) display according to some moral standard (e.g., politicians should be honest).

In contrast, descriptive norms are taken to be 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' (e.g., college students party a lot during spring break, CEOs sleep less than 8 hours a night, etc., Prentice [2007], p. 629). Notice that this is the notion of descriptive norms often used by psychologists (e.g., Cialdini *et al.*). There are other ones proposed by other researchers (e.g.,

---

[12] https://www.bbc.com/news/world-51751915

[13] https://www.theguardian.com/books/2019/apr/01/why-smart-people-are-more-likely-to-believe-fake-news

[14] https://climatecommunication.yale.edu/publications/reducing-ideological-bias-on-climate-change/

[15] See YouGov graphics at:
https://yougov.co.uk/topics/international/articles-reports/2020/05/02/international-covid-19-tracker-update-2-may?utm_source=twitter&utm_medium=website_article&utm_campaign=covid_intl_tracker_2_May

[16] https://news.un.org/en/story/2020/04/1061592

philosophers, see Muldoon *et al.* [2014]; Bicchieri [2017]). [17] I shall here adopt the psychologists' notion and terminology because they are now standard in psychological research (see studies reviewed below).

With these clarifications in place, the key point to note for present purposes is that PST, including examples such as claims (1)-(10), is likely to convey descriptive norms. This is because of its broad and ambiguous scope. For instance, on a natural reading, claim (1) (see above) conveys that men commonly resist green behavior, claim (2) that Americans commonly consume unhealthy amounts of processed meat, (3) that social-media users spread falsehoods faster than truths, (4) that US voters are commonly less inclined to vote, (5) that people are commonly biased against women, (6) that Republicans and Democrats commonly view each others as closed-minded, [18] (7) that smart people are commonly more susceptible to fake news, (8) that conservatives are commonly less likely than liberals to support climate policies, (9) that Britons are commonly reluctant to wear COVID-19 masks, and (10) that unreliable and false information about COVID-19 is common among people and spreading globally.

Claims (1)-(10) do not have to be read 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 people. Since, in doing so, such generalizations convey descriptive norms, members of the public are likely to take PST such as (1)-(10) to indicate descriptive norms. What, then, do we know about the effects of descriptive norms on people's behavior and/or 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 behavior and cognition.

As for behavior-related research, in one study, (Californian) households informed that they used more electricity than all the others households in their neighborhood 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[19] than guests given environmental reasons for that action (Goldstein *et al.* [2008]). Similarly, U.S. lunch-goers informed that most people in the U.S. 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 behavioral responding has been found with respect to majority information about various

---

[17] 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'.

[18] For further discussion on this particular type of PST and its negative self-fulfilling effects, see Peters [2020a])

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

socially highly relevant actions, for instance, paying taxes (Hallsworth *et al.* [2017]), as well as corrupt behavior (Kȯbis *et al.* [2015]) and readiness to steal (Cialdini *et al.* [2006]).

Turning from descriptive-norm effects on behavior to effects on cognition (e.g., motivation, intentions, biases, etc.), 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 people should do so) subsequently reported significantly lower intentions 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 behaviors (e.g., Hibbles make babies smile/cry; Glerks give people flowers/punches). The test participants were then asked to evaluate a dissenting individual (i.e., 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 (Roberts *et al.* [2019], p. 382). That is, many test 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 clearly 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]; Sparkman and Walton [2017]; Hallsworth *et al.* [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, e.g., Cialdini *et al.* [1990]; Mollen *et al.* [2013]; Peters [2020a]). 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 (e.g., 'most people reuse their towels', 'the vast majority stereotype', etc.), this is not required for the effects at issue to occur. Indeed, 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 (e.g., advertisements suggesting pervasiveness of smoking, drinking, etc., Bosari and Carey [2003]; Nan and Zhoa [2016]; Lui and Shi [2019]).

3.3 Strength of the data

Given recent replication failures in the psychological sciences (Bird [2018]), 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. Melnyk *et al.* ([2019], pp. 4, 13) found that, overall, descriptive norms 'directly influence behavior' (not only intention formation) and their effect on behavior is generally 'stronger than that of prescriptive norms'. This meta-analysis and several individual experimental studies that manipulated the scope[20] of descriptive norms to track causal norm-behavior links (Kormos *et al.* [2014]; van Wagner *et al.* [2019]) suggest that descriptive norms do often exert a significant causal influence on behavior. Similarly, but related to these norms' effects on cognitions, Rivis and Sheeran ([2003]) assessed 14 studies (total sample size N=5810) covering a wide range of behavioral domains, and they found 'strong evidence in support of the predictive validity of descriptive norms' in 'intention formation' (p. 228).

The efficacy of these norms is perhaps also not difficult to explain. For instance, as Cialdini *et al.* ([1990]) notes, 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' (p. 203). Similarly, Bicchieri ([2006]) writes that '[w]e conform because such norms make life easier for us, because we want to "fit in"' (p. 29).

Notice, however, that even though there is good ground to hold that descriptive norms often prompt conformist behavior, 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 (e.g., dynamic norms, or don't- vs. do-norms) have been found to be more powerful than others (Sparkman and Walton [2017]; Mortensen *et al.* [2019]; Bergquist and Nilsson [2019]; Peters [2020a]). I shall thus assume a qualified and explicitly quantified view here: that in many (vs. all) domains descriptive norms are often (vs. always) likely to have the effects discussed.

4 Revisiting Public Scientific Testimony

The data discussed in the preceding section are relevant for the normative theorizing on public scientific testimony. This is 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 in them were laypeople in naturalistic settings presented with statements capturing broad, social scientific generalizations

---

[20] For instance, test subjects were told that 20%, 50%, or 80% of previous participants acted in a certain way; the boarder the scope of the descriptive norm (e.g., 50% vs. 80%), the stronger the effect on intention to act (Kormos *et al.* [2014]; Van Wagner *et al*. [2019]).

(e.g., about energy consumption of their neighborhood, people's voting behavior, stereotyping, etc.) that came directly or indirectly from scientific sources (i.e., psychologists, political scientists, science reporters, etc.). The subjects were thus in much the same situation as they are when receiving PST via TV, online, newspapers etc. in everyday life.

Importantly, the descriptive-norm effects were in many cases clearly harmful. For instance, when participants learned about higher energy consumption in their neighborhood, 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 and epistemic problems that this raises is likely to arise with respect to a wide range of PST.

To make this more concrete, consider again the abovementioned examples (1)-(10). *Prima facie*, these claims and instances of PST seem ethically and epistemically innocuous. But they also explicitly or indirectly (e.g., (10)) indicate to an audience that the cognitive and/or behavioral features described in the claims are common among people. That is, they involve broad social scientific generalizations (e.g., '90%', 'most Republicans', etc.) or generic nouns ('men', 'Americans', etc.), and so indicate descriptive norms. And importantly, these claims come from authoritative sources, that is, sources that the public commonly trusts (e.g., news outlets, science associations, scientists, UN, etc.). Since that is so, given the data on the directive impact of descriptive norms just reviewed, there is reason to believe that claims (1)-(10) are also likely to contribute to self-fulfilling effects among the public receiving these claims and belonging to the demographic groups concerned. These effects include (1) reluctance towards green behavior among men, (2) increased meat consumption among Americans, a (3) proliferation of falsehoods among social media users, (4) low U.S. voter turnout,[21] (5) bias against women, (6) affective polarization among Republicans and Democrats,[22] (7) acceptance of fake news among 'smart' people, (8) decreased support for climate policies among conservatives, and so on.[23] All of these consequences are evidently ethically and/or epistemically problematic.

To be sure, no experiment yet exists that tests whether the kind of self-fulfilling effects encountered in the studies reviewed earlier will also occur with respect to the specific PST in claims (1)-(10). But independently of claims (1)-(10), 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 (see, e.g., Schultz *et al.* [2007];

---

[21] 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]).

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

[23] What are the cues and immediate social environments that activate norm-conformist responding? Just as in the studies discussed above, they will differ depending on the content of the descriptive norm; e.g., after exposure to a descriptive norm about green behavior, the cues and immediate environments might include green vs. non-green products and choice situations in supermarket.

Goldstein *et al.* [2008]; Gerber and Rogers [2009]; Staunton *et al.* [2014]; Duguid and Thomas-Hunt [2015]). So even if one is skeptical about generalizing from the reviewed studies to (1)-(10), the preceding discussion already directly supports the claim that in some cases PST leads to negative, descriptive norm related self-fulfilling 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 (1)-(10). This is because some of (1)-(10) 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 ([2008]), claim (5) and Duguid and Thomas-Hunt ([2015])) or are supported by other studies (e.g., descriptive norms in the context of promotions of health-risk behavior; see Rivis and Sheeran [2003]; Cialdini [2003]).[24]

Moreover, we have so far no reason to believe that when it comes to the other claims belonging to (1)-(10), the descriptive norms mechanism discussed is interrupted or absent. And even if in the cases of (1)-(10), descriptive norms are less powerful and self-fulfilling effects less frequent or weaker than those found in the studies discussed, claims of the type at issue are *very* widely broadcast nationally and internationally (e.g., on the BBC). This means that even only very modest self-fulfilling 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-fulfilling effects. Given what is at stake (e.g., environmentally friendly behavior, spread of misinformation, voting, etc.), this would arguably still make it reasonable, if not rational, to take the considerations above seriously and reflect on ways of how the potential pernicious effects outlined could be avoided.

Adding to this point, while in some of the claims (1)-(10), the PST at issue concerns only a particular sub-set of individuals (e.g., men, U.S. voters, etc.), 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 abovementioned 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 behavior among members of that group. It is thus not unreasonable to assume that people's expectations to the effect that, for instance, men resist green behavior, Americans eat too much meat, 'fake news' spread quickly on social media, etc. might in subtle ways bias their social interactions such that self-fulfillment effects[25] become more likely – not least in the sense of reducing people's surprise when they actually encountering such behavior (Peters [2020a]). Additionally, as a result of

---

[24] 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'.

[25] Such social expectations can also contribute to pernicious instances of 'mindshaping' by eliciting behavioral confirmation and triggering 'self-fulfilling prophecies'; see Peters ([2020b]).

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 distribution of them.

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-fulfilling 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 behavior that subjects could conform to. Or there might be PST that the recipient distrusts, or motivated cognition might prevent an uptake of the claims (Gerken [2019]).[26] The subjects 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 behavior and the public trusts it, 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 (1)-(10) might have overall more significant ethical and epistemic *benefits* than the costs linked to self-fulfilling effects. Relatedly, the preceding considerations are not intended to suggest that claims of the type captured in (1)-(10) should never be made. The point is just that even if the ethical and/or epistemic benefits of the proliferation of PST such as (1)-(10) outweigh the costs tied to self-fulfilling 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.

I have left open a number of interesting questions. For instance, how far into the domain of PST exactly can the point just made be generalized? How significant is the self-fulfillment problem in actual science communication contexts? These questions call for a detailed analysis. I shall set them aside for now. The goal here is to motivate asking them, draw attention to the problem to which they pertain, and provide reasons to believe that this problem is likely to be widespread in the context of PST. In the next section, I will argue that it is in fact more widespread than so far appreciated in work on science communication.

4.2 A neglected issue

The preceding considerations have been largely overlooked in both the literature on descriptive norms and the theorizing on PST. For instance, in the empirical literature on descriptive

---

[26] Recent studies suggest, however, that people tendency to follow descriptive norms might in some cases even be stronger than their motivated cognition in favor of (e.g.) their political in-group: Pryor *et al.* ([2019], p. 1) found that, '[c]ounter 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'. For discussion, see Peters ([2020a]).

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' (i.e., messages directed at the public, produced by governments or private institutions to raise awareness of, and alter attitudes and behavior 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 papers 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-fulfilling 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 it, 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 too have so far not paid much attention to the problem of PST self-fulfilling effects related to descriptive norms. To be sure, in 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 behavioral 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' (Hacking [2007], p. 286). Hacking holds that the behavioral and cognitive changes triggered by the classifications may in turn trigger revisions in the classifications in order to ensure they reflect these changes, a phenomenon he calls 'looping effects'. While Hacking's ideas are connected and congenial to the points made here about self-fulfilling effects, he has not yet considered the role of descriptive norms in the context of how social scientific classifications might 'make people up' (Hacking [2007]). And he has not yet discussed the related implications specifically when it comes to 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-fulfilling 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' (Fine [2012], p. 286). Similarly, in her argument that certain scientific research and claims 'should be banned' (*inter alia*) because of their harmful self-fulfilling effects, Kourany ([2016], p. 779) concentrates only on research and testimony related to cognitive differences between gender and race groups. Neither Fine nor Kourany consider whether the kind of problem of a self-fulfillment of scientific claims might in

fact be highly general, potentially arising even for *prima facie* unproblematic claims (e.g., about meat eating, voting, social media use, etc.).

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 behavior 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 is linked to self-fulfilling effects: by leading subjects to underperform in stereotype-related tasks, it may contribute to a confirmation of stereotypes (e.g., 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-fulfilling effects.

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

Notice too that the empirical basis of the argument introduced above is more robust than that of their argument. 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]; Flores *et al.* [2019]). In contrast, meta-analyses of studies on descriptive norms repeatedly confirmed the reality of the type of self-fulfilling effects discussed above (e.g., 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, if any, are social scientists responsible for the pernicious self-fulfilling effects that might arise from their PST?

4.3 Scientists' responsibility for self-fulfilling effects

It is widely accepted that scientists have a moral duty toward society to avoid causing harms to society and prevent wrongful outcomes and omissions that they are in a position to foresee (Douglas [2009], p. 83; Resnik and Elliot [2016], p. 36). 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. And 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; for an explicit statement of this requirement, see, e.g., the European Commission for Research and Innovation ([2018], p. 20).

However, scientists are clearly 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-fulfilling effects qualify as reasonably foreseeable?

To be sure, many social scientists or their peers might not be aware of the connection between their PST and its potential self-fulfilling effects. But consider, for instance, a member of hiring committees. Suppose they lack of 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 committee members *should be aware of* implicit bias even if they do not in fact do so (for further support, see Washington and Kelly [2014]).[27]

Similarly, notice that the data on descriptive norms and their impact on behavior and cognition, too, have been known for about 30 years now (Cialdini *et al.* [1990]). Given this point and the fact that institutional boards (i.e., 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 too *should be aware* of PST self-fulfilling effects too. After all, it is hard to see why the social role specific obligation on scientists to assess and limit risks of social harm that ethics committees confer on them *prior* to conducting an experiment should ceases 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[28] have the following responsibility:

> *Communicative Risk Anticipation*
> Scientists should assess whether their PST is likely to have detrimental self-fulfilling effects and consider adopting strategies to mitigate the risks of them.

Communicative Risk Anticipation is modest. While it requires scientists to analyze the risks of their

---

[27] Thanks to Dan Kelly here for drawing my attention to this externalist line of thought.

[28] 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. A separate argument is needed; for a helpful thought experiment to developed one, see Washington and Kelly [2014]. To keep the discussion focused, I shall set science reporters aside here.

PST and consider implementing ways of reducing them, it does not yet ask them to in fact adopt strategies to avoid problematic self-fulfilling effects related to descriptive norms. This matters because, as acknowledged above, there could be cases where the negative consequences of PST self-fulfilling 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)[29] complexities, a blanket requirement on scientists to adopt ways of avoiding descriptive norms related negative self-fulfilling effects would be too strong.

Still, the preceding sections suggest that these effects are likely to arise in a wide range of cases. Scientists should thus factor them in. *Communicative Risk Anticipation* captures this point. Suppose, then, that in line with *Communicative Risk Anticipation*, scientists have assessed the self-fulfilling 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-fulfilling 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 [2020a]). 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.

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. For these norms refer to features or behavior common (rather than infrequent) among people (Cialdini *et al.* [1990]). Now, there is evidence that social scientific results expressed with generic language 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-fulfilling 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.

---

[29] Adding further complications, the effects of PST depend partly on who the PST audience is (e.g., 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 (e.g., 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 *Communicative Risk Anticipation*, scientists ought to assess these effects when thinking about their PST. Thanks to an anonymous reviewer here.

Granted, if the quantified expressions to replace the generics and other broad generalizations in the PST could only be 'all' or 'most' so as to preserve accuracy, then even with this rephrasing, PST would remain problematic for the same reason as before. However, notice that, as Henrich *et al.* ([2010)] found,

> 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. (p. 61)

Henrich *et al.* went on to show that these assumptions are not justified. By reviewing the comparative database from across the behavioral 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' (Henrich *et al.* [2010)], p. 61).

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

Indeed, in a study already mentioned above, DeJesus *et al.* ([2019]) analyzed 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' (p. 18375). In fact, scientists often used generics despite omitting information on sample features (e.g. socio-economic status) or having only small or unrepresentative samples (e.g., Western, white, middle-class etc.) as a basis. Strikingly, authors failing to mention, for instance, socio-economic status tended to use *more* generics than those who did mention it (DeJesus *et al.* ([2019]), 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 articles freely accessible online can be replaced with claims involving quantifiers other than 'all' or 'most' in the description of the samples. Since doing so helps reduce the risk that the public is exposed to socially harmful self-fulfilling effects related to descriptive norms, there is (given *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-fulfilling 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 socio-economic 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 require it (Resnik [2005], p. 48; Forge [2008]). Relatedly, focusing specifically on *intra*-scientific communication, Simons *et al.* ([2017]) 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. (p. 1123)

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 1% 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 norm of scientific accuracy.

However, this would overlook the following. Cimpian et al. ([2010]) found that while generic claims of the form '*P*s (e.g., men etc.) are *f* (e.g., resist green behavior, etc.)' are generally accepted on relatively weak evidence, when they are unfamiliar with *P*, subjects construe such claims strongly as conveying that almost all *P*s are *f*. Generics in PST thus involve an ambiguity that leads different people (e.g., scientific experts vs. non-experts) to different interpretations depending on their expertise concerning *P*.[30] This ambiguity or openness to different interpretations itself is in tension with scientific accuracy. Because if a particular empirical claim has been experimentally supported only with a specific sample, then even though using a generic might not involve making a false claim, it would still be inaccurate in that it leaves the social reference class of the claim more indeterminate than the experimental data warrant. It inaccurately suggests generalizability or 'projectability' (Munton [2020]) where the data may not support it. Replacing broad generic generalizations with specific and explicitly quantified claims thus helps improve scientific accuracy.

It is worth noting too that a number of philosophers have argued that hearing and using generics makes subjects significantly more likely to 'essentialize' social kinds such as, for instance, racial or gender groups (Langton *et al.* [2012]; Leslie [2017]): generics are thought to lead people to

---

[30] Thanks to Olivier Lemeire for comments on this point.

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 (e.g., 'Latinos are temperamental', 'women are nurturing'). Langton *et al.* ([2012]) 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 [e.g.] racial [gender etc.] essences' (p. 765). To counteract such essentialization, Langton *et al.* ([2012]) plausibly propose that we should convey the content of generics differently, namely by an 'explicitly quantified statement' involving 'some', 'most', or 'all' (765). Leslie ([2017]) adds that hearing generics inclines individuals not only to essentialize those classified by the generics but also to more readily apply stereotypes to them (p. 416).

5.3 Combining ethical and epistemic considerations

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

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

Following *PST Generality Constraint* has significant benefits. Given the points from the previous section, it helps (1) decrease the risk of negative self-fulfilling effects tied to PST, (2) reduce scientific inaccuracy related to over-generalizations, and (3) counteract social essentialization and stereotyping.

Moreover, *PST Generality Constraint* 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, and 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 (e.g., when scientists or science reporters have little time/space to give an opinion online, on TV, etc.; Gerken [2020]).

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. For instance, with respect to several examples of claims (1)-(10), there are unlikely to be ethical, epistemic, or feasibility concerns overriding the benefits of rephrasing these claims with explicitly quantifications and relativization to specific samples. For example, assuming that the evidential basis permits it, 'Men resist green behavior as unmanly', 'Americans eat too much processed meat', and 'People on social media are quick to repeat something that's wrong' could be easily rephrased thus: 'Many men in the U.S. resist green behavior as unmanly', 'Many White, middle-class Americans eat too much processed meat', and 'Many people using Facebook

are quick to repeat something that's wrong'. In fact, even if there are overriding conditions against such rephrasing, following *PST Generality Constraint* would still help improve the current provision of PST. For social scientists would then first need to consider the potential pernicious descriptive norms related effects of their PST and check them against other potential harms/benefits before testifying.[31]

Having said that, following *PST Generality Constraint* does not eliminate all potential detrimental self-fulfilling 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 the emergence 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 *PST Generality Constraint* can 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 an intriguing one of them: social scientific claims about negative features of individuals, groups, society, or social structures can cause cognitive and behavioral changes contributing to the self-fulfillment of these claims. I provided reasons to believe that this is likely to happen frequently and result in a wide range of harmful upshots (e.g., environmentally unfriendly behavior, low voter turnout, proliferation of 'fake news', bias against women, etc.). I argued that this problem is 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-fulfilling effects of PST related to descriptive norms can be decreased, namely to ensure that the generalizations in the PST (a) contain an explicitly quantified relativization to the population(s) to which they pertain, and (b) are not broader than warranted by the evidence. This strategy might in some cases be overridden by other considerations and may not eliminate all the harmful effects related to people's conformist responding to descriptive norms. Still, it can help counteract them, and it is independently well motivated by the fact that it increases scientific accuracy.

Acknowledgements

---

[31] *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, e.g., Gerken [2018]; [2019]; [2020]).

References

Alexandrova, A. [2018]: 'Can the Science of Well-Being Be Objective?', *The British Journal for the Philosophy of Science*, **69**, 2, pp. 421–445.

Anderson, E. [2011]: 'Democracy, Public Policy, and Lay Assessments of Public Scientific Testimony', *Episteme*, **8(**2), pp. 144–164.

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

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. 158–166.

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, NY: Oxford University Press.

Bird, A. [2018]: 'Understanding the Replication Crisis as a Base Rate Fallacy', *The British Journal for the Philosophy of Science*, axy051, https://doi.org/10.1093/bjps/axy051

Blanken, I., van de Ven, N., and Zeelenberg, M. [2015]: 'A Meta-analytic Review of Moral Licensing', *Personality and Social Psychology Bulletin*, **41**(4), pp. 540–558.

Borsari, B., & Carey, K. B. [2003]: 'Descriptive and Injunctive Norms in College Drinking: A Meta-analytic Integration', *Journal of Studies on Alcohol*, **64**(3), pp. 331–341.

Brewer, P., and Ley, B. [2013]: 'Whose Science Do You Believe? Explaining Trust in Sources of Scientific Information about the Environment', *Science Communication*, **35**(1), pp. 115–137.

Cacciatore, M., Browning, N., Scheufele, D., Brossard, D., Xenos, M., Corley, E. [2018]: 'Opposing Ends of the Spectrum: Exploring Trust in Scientific and Religious Authorities', *Public Understanding of Science*, **27**(1), pp. 11–28.

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. 201–233.

Cialdini, R. B. [2003]: 'Crafting Normative Messages to Protect the Environment', *Current Directions in Psychological Science*, **12**(4), pp. 105–109.

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

Cimpian, A., Brandone, A., & Gelman, S. [2010]: 'Generic Statements Require Little Evidence for Acceptance but Have Powerful Implications', *Cognitive Science*, **34**, pp. 1452–1482.

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*, **116** (37), pp. 18370–18377.

Duguid, M., and Thomas-Hunt, M. [2015]: 'Condoning Stereotyping? How Awareness of Stereotyping Prevalence Impacts Expression of Stereotypes', *Journal of Applied Psychology*, **100**, pp. 343–359.

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

European Commission for Research and Innovation [2020]: *Ethics in Social Science and Humanities*. Available online, URL:
https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020_ethics-soc-science-humanities_en.pdf

Fine, C. [2012]: 'Explaining, or Sustaining, the Status Quo? The Potentially Self-Fulfilling Effects of "Hardwired" Accounts of Sex Differences', *Neuroethics*, **5**(3), pp. 285–294.

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., & 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. 140–174.

Funk, C., and Rainie, L. [2015]: 'Public and Scientists' Views on Science and Society', Available at: http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/

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

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. [2019]: 'Public Public Scientific Testimony in the Scientific Image', *Studies in History and Philosophy of Science A*.
https://www.sciencedirect.com/science/article/pii/S0039368119300640

Gerken, M. [2020]: 'How to Balance Balanced Reporting and Reliable Reporting', *Philosophical Studies,* **177**, pp. 3117–3142.

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. 472–482.

Guyll, M., Madon, S., Prieto, L., & 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**(1), pp. 113–130.

Hacking, I. [1995]: 'The Looping Effects of Human Kinds', in Sperber, D., Premack, D. and Premack, A. (eds), *Causal Cognition*, Oxford: Clarendon Press, pp. 351–383.

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

Iyengar, S., and Massey, D.S. [2019]: 'Scientific Communication in a Post-truth Society', *Proceedings of the National Academy of Sciences*, **116**(16), pp. 7656–7661.

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

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

Keohane, R. O., Lane, M., and Oppenheimer, M. [2014]: 'The Ethics of Scientific Communication Under Uncertainty', *Politics, Philosophy & Economics*, **13**, pp. 343–368.

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**(5), pp. 781–792.

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

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., & Van Lange, P. A. [2015]: '"Who Doesn't?" – The Impact of Descriptive Norms on Corruption', *PloS one*, **10(**6), e0131830. https://doi.org/10.1371/journal.pone.0131830

Kormos, C., Gifford, R., & Brown, E. [2015]: 'The Influence of Descriptive Social Norm Information on Sustainable Transportation Behavior: A Field Experiment', *Environment and Behavior*, **47**(5), pp. 479–501.

Kourany, J. [2016]: 'Should Some Knowledge be Forbidden? The Case of Cognitive Differences Research', *Philosophy of Science*, **83**(5), pp. 779–790.

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. 753–767.

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**(1): e562

Leslie, S.J. [2017]: 'The Original Sin of Cognition: Fear, Prejudice, and Generalization', *Journal of Philosophy*, **114**(8), pp. 393–421.

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

Liu, J., & 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**(1), 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**(1), pp. 4–17.

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**, 2, pp. 201–210.

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

Muldoon, R., Lisciandra, C., Bicchieri, C., Hartmann, S., & Sprenger, J. [2014]: 'On the Emergence of Descriptive Norms', *Politics, Philosophy & Economics*, **13**(1), pp. 3–22.

Munton, J. [2019]: 'Beyond Accuracy: Epistemic Flaws with Statistical Generalizations', *Philosophical Issues*, **29**, pp. 228–240.

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

Nan, X., & 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**(1), pp. 56–66.

Neff, A., Edwards, D., Palmer, A., Ramsing, R., Righter, A., and Wolfson, J. [2018]: 'Reducing Meat Consumption in the USA: A Nationally Representative Survey of Attitudes and Behaviours', *Public Health Nutrition*, **21**, 10, pp. 1835–1844.

Nguyen, H. H. D., & 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**(6), pp. 1314–1334.

Nisbet, E. C., Cooper, K. E., & Garrett, R. K. [2015]: 'The Partisan Brain: How Dissonant Science Messages Lead Conservatives and Liberals to (Dis)Trust Science', *The ANNALS of the American Academy of Political and Social Science*, **658**(1), pp. 36–66.

O'Connor, C., & Joffe, H. [2014]: 'Social Representations of Brain Research: Exploring Public (Dis)engagement With Contemporary Neuroscience', *Science Communication*, **36**(5), pp. 617–645.

Paryavi, M., Bohnet, I., & van Geen, A. [2019]: 'Descriptive Norms and Gender Diversity: Reactance from Men', *Journal of Behavioral Public Administration*, **2**(1). https://doi.org/10.30636/jbpa.21.51

Peters, U. [2020a]: 'How (Many) Descriptive Claims about Political Polarization Exacerbate Polarization', *Journal of Social and Political Psychology*. Preprint available online: https://philpapers.org/archive/PETHMD.pdf

Peters, U. [2020b]: 'What is the Function of Confirmation Bias? *Erkenntnis*. Online First. https://link.springer.com/article/10.1007/s11229-020-02846-2

Posetti, J. and Bontcheva, K. [2020]: *Disinfodemic: Deciphering COVID-19 disinformation*. United Nations Educational, Scientific and Cultural Organization.

Prentice, D. A. [2007]: 'Prescriptive vs. Descriptive Norms', in R. Baumeister & K. Vohs (eds), *Encyclopedia of Social Psychology*, California: Sage Publications,(pp. 629–630).

Pryor, C., Perfors, A., & Howe, P. [2019]: 'Conformity to the Descriptive Norms of People with Opposing Political or Social Beliefs', *PloS one*, **14**(7), e0219464. https://doi.org/10.1371/journal.pone.0219464

Raymond, L., Weldon, L. Kelly, D. Arriaga, X., and Clark, A. [2014]: 'Making Change: Norm-Based Strategies for Institutional Change to Address Intractable Problems', *Political Research Quarterly*, **67**, pp. 197–211.

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', *Account Res*, **23**, pp. 31–46.

Richter, I., Thøgersen, J., & Klöckner, C.A. [2018]: 'A Social Norms Intervention Going Wrong: Boomerang Effects from Descriptive Norms Information', *Sustainability*, **10**, 2848. https://www.mdpi.com/2071-1050/10/8/2848#cite

Ritchie, K. [2019]: 'Should We Use Racial and Gender Generics?' *Thought: A Journal of Philosophy*, **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. 218–233.

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. 374–387.

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**(5), pp. 534–553.

Saul, J. [2018]: '(How) Should We Tell Implicit Bias Stories?', *Disputatio*, **10**(50), pp. 217–244.

Schmader, T., M. Johns, and C. Forbes. [2008]: 'An Integrated Process Model of Stereotype Threat Effects on Performance', *Psychological Review*, **115**(2), pp. 336–356.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. [2007]: 'The Constructive, Destructive, and Reconstructive Power of Social Norms', *Psychological Science*, **18**(5), pp. 429–434.

Schwartz, B. [1997]: 'Psychology, Idea Technology, and Ideology', *Psychological Science*, **8**(1), pp. 21–27.

Simons, D. J., Shoda, Y., & Lindsay, D. S. [2017]: 'Constraints on Generality (COG): A Proposed Addition to All Empirical Papers', *Perspectives on Psychological Science*, **12**(6), pp. 1123–1128.

Sparkman, G. and Walton, G. [2017]: 'Dynamic Norms Promote Sustainable Behavior, Even If it is Counternormative', *Psychological Science*, **28**, pp. 1663–1674.

Staunton, M., Louis, W. R., Smith, J. R., Terry, D. J., & McDonald, R. I. [2014]: 'How Negative Descriptive Norms for Healthy Eating Undermine the Effects of Positive Injunctive Norms', *Journal of Applied Social Psychology*, **44**(4), pp. 319–330.

Steele, C.M. [1997]: 'A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance', *American Psychologist*, **52**(6), pp. 613–629.

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

Turnwald, B. P., Goyer, J. P., Boles, D. Z., Silder, A., Delp, S. L., & Crum, A. J. [2019]: 'Learning One's Genetic Risk Changes Physiology Independent of Actual Genetic Risk', *Nature human behaviour*, **3**(1), pp. 48–56.

Wassenaar, DR., and, Mamotte, N. [2012]: 'Ethical Issues and Ethics Reviews in Social Science Research', in Leach, M., Stevens, M., Lindsay, G., Ferrero, A., Korkut, Y., (eds), *The Oxford Handbook of International Psychological Ethics*. New York: Oxford University Press, pp. 268–282.

Simons, D., Shoda, Y., and Lindsay, D. [2017]: 'Constraints on Generality (COG): A Proposed Addition to All Empirical Papers', *Perspectives on Psychological Science*, **12**, pp. 1123–1128.

Steele, C.M. [1997]: 'A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance', *American Psychologist*, **52**(6), pp. 613–629.

von Wagner, C., Hirst, Y., Waller, J., Ghanouni, A., McGregor, L. M., Kerrison, R. S., et al. [2019]: 'The Impact of Descriptive Norms on Motivation to Participate in Cancer Screening—

Evidence   from   Online   Experiments',   *Patient   Education   and   Counseling.*
https://doi.org/10.1016/j.pec.2019.04.001

Walton, G. M., & Wilson, T. D. [2018]: 'Wise Interventions: Psychological Remedies for Social and Personal Problems', *Psychological review*, **125**(5), pp. 617–655.

Washington, N. & Kelly, D. [2016]: 'Who's responsible for this? Moral responsibility, externalism and knowledge about implicit bias', in M. Brownstein & J. Saul (*eds*), *Implicit bias and philosophy: Moral responsibility, structural injustice, and ethics* (Vol. 2). Oxford: Oxford University Press, pp. 11-37.

Wolpe, P. [2006]: 'Reasons Scientists Avoid Thinking About Ethics', *Cell*, **125**(6), pp. 1023–1025.

Zigerell, L. J. [2017]: 'Potential Publication Bias in the Stereotype Threat Literature: Comment on Nguyen and Ryan (2008)' *Journal of Applied Psychology*, **102**(8), pp. 1159–1168.