

Volume XIV ♦ Issue 1

2023

Logos & Episteme

an international journal
of epistemology

**Romanian Academy
Iasi Branch**



**“Gheorghe Zane” Institute
for Economic and Social
Research**

Founding Editor

Teodor Dima (1939-2019)

Editorial Board

Editor-in-Chief

Eugen Huzum

Executive Editors

Vasile Pleșca

Cătălina-Daniela Răducu

Assistant Editors

Irina Frasin

Bogdan Ștefanachi

Ioan Alexandru Tofan

Web&Graphics

Codrin Dinu Vasiliu

Virgil-Constantin Fătu

Simona-Roxana Ulman

Contact address:

Institutul de Cercetări

Economice și Sociale „Gh.Zane”

Iași, str.T.Codrescu, nr.2, cod 700481

Tel/Fax: 004 0332 408922

Email: logosandepisteme@yahoo.com

<http://logos-and-episteme.acadiasi.ro/>

https://www.pdcnet.org/pdc/bvdb.nsf/journal?openform&journal=pdc_logos-episteme

Advisory Board

Frederick R. Adams

University of Delaware, USA

Scott F. Aikin

Vanderbilt University, USA

Daniel Andler

Université Paris-Sorbonne, Paris IV, France

Panayot Butchvarov

University of Iowa, USA

Mircea Dumitru

Universitatea din București, România

Sanford Goldberg

Northwestern University, Evanston, USA

Alvin I. Goldman

Rutgers, The State University of New Jersey, USA

Susan Haack

University of Miami, USA

Stephen Hetherington

The University of New South Wales, Sydney, Australia

Paul Humphreys

University of Virginia, USA

Jonathan L. Kvanvig

Baylor University, USA

Thierry Martin

Université de Franche-Comté, Besançon, France

Jürgen Mittelstrab

Universität Konstanz, Germany

Christian Möckel

Humboldt-Universität zu Berlin, Germany

Maryvonne Perrot

Université de Bourgogne, Dijon, France

Olga Maria Pombo-Martins

Universidade de Lisboa, Portugal

Duncan Pritchard

University of Edinburgh, United Kingdom

Nicolas Rescher

University of Pittsburgh, USA

Rahman Shahid

Université Lille 3, France

Ernest Sosa

Rutgers University, USA

John F. Symons

University of Texas at El Paso, USA

TABLE OF CONTENTS

RESEARCH ARTICLES

Leandro DE BRASI, Jack WARMAN, Deliberative Democracy, Epistemic Injustice, and Epistemic Disenfranchisement.....	7
Rauf ORAN, Lie <i>for the Other</i> : A Socio-Analytic Approach to Telling Lies.....	29
Timothy PERRINE, Prejudice, Harming Knowers, and Testimonial Injustice.....	53
Balder Edmund Ask ZAAR, Dispositional Reliabilism and Its Merits.....	75

DISCUSSION NOTES/DEBATE

Eric RAIDL, Neutralization, Lewis' Doctored Conditional, or Another Note on "A Connexive Conditional"	101
Erratum Notice.....	119
Notes on the Contributors.....	121
Notes to Contributors.....	123
<i>Logos and Episteme</i> . Aims and Scope.....	127

RESEARCH ARTICLES

DELIBERATIVE DEMOCRACY, EPISTEMIC INJUSTICE, AND EPISTEMIC DISENFRANCHISEMENT¹

Leandro De BRASI, Jack WARMAN

ABSTRACT: In this paper, we explore some links between deliberative democracy, natural testimony, and epistemic injustice. We hope to highlight the exclusionary effects of some cases of testimony-related epistemic injustice within the deliberative democratic framework and, in particular, two subtle ways of epistemic injustice that are not often highlighted in the political domain. In other words, we hope to highlight two specific mechanisms of epistemic exclusion within the democratic deliberative process that are not explicitly noticed in the relevant literature. In section 1, we present the deliberative model of democracy and the deliberative process. We then introduce the notion of epistemic (dis)enfranchisement, which we distinguish from formal enfranchisement, and explain the role that natural testimony plays in establishing citizens' epistemic enfranchisement. In section 2, we introduce Fricker's notion of testimonial injustice and two further testimony-related forms of epistemic injustice which seem to have been largely neglected in the debate so far, namely, discursive injustice and testimonial void. We also point out negative epistemic consequences of positive identity-prejudicial stereotypes. In section 3, we argue that these testimony-related forms of epistemic injustice can lead to epistemic disenfranchisement, which, we note, is an obstacle to deliberative democracy that warrants serious consideration.

KEYWORDS: deliberative democracy, epistemic injustice, disenfranchisement, testimony, stereotypes

Introduction

In this paper, we explore some links between deliberative democracy, natural testimony, and epistemic injustice. We hope to highlight the exclusionary effects of some cases of testimony-related epistemic injustice within the deliberative democratic framework and, in particular, two subtle ways of epistemic injustice that are not often highlighted in the political domain. In other words, we hope to highlight two specific mechanisms of epistemic exclusion within the democratic

¹ This research was funded by Agencia Nacional de Investigación y Desarrollo, Chile, FONDECYT Regular No. 1210724 (PI: Leandro De Brasi) and Agencia Nacional de Investigación y Desarrollo (ANID), Chile, FONDECYT Postdoctorado No. 3200770 (PI: Jack Warman).

deliberative process that are not explicitly noticed in the relevant literature. In section 1, we present the deliberative model of democracy and the deliberative process. We then introduce the notion of *epistemic disenfranchisement*, which we distinguish from formal enfranchisement, and explain the role that natural testimony plays in establishing citizens' epistemic enfranchisement. In section 2, we introduce Fricker's notion of *testimonial injustice* and two further testimony-related forms of epistemic injustice which seem to have been largely overlooked in the debate about deliberative democracy so far. The first of these is *discursive injustice* (Kukla 2014). The central thought here is that certain speakers can be excluded from the testimonial practice when their testimony *qua* speech act is not recognised as such. While this is typically explained as a consequence of negative identity prejudicial stereotypes, we argue that this can occur as a consequence of positive identity-prejudicial stereotypes too. The second of these is *testimonial void* (Carmona 2021). The idea here is that a person can suffer a kind of epistemic injustice when another withholds their testimony from them as a consequence of the identity-prejudicial stereotypes they, the would-be speaker, hold about their audience. Finally, in section 3 we argue that these testimony-related forms of epistemic injustice can lead to epistemic disenfranchisement, which, we explain, is an obstacle to deliberative democracy that warrants serious consideration.

1.1 Deliberative Democracy and the Public Sphere

Since the second part of the 20th Century, democracy has been regarded as the only legitimate political arrangement (Dunn 2005). On a normative level, democracy now represents an undisputed background. Moreover, democracy *qua* collective self-government embodies the aspiration to be guided by the *demos'* better reasons (Young 2000). After all, a government can commit grave forms of injustice (the sort of injustices that can affect many people and in systematic ways) and we do not want to make these high-cost mistakes (Aikin and Talisse 2019). We want democratic decision-making to be able to recognize good reasons and reject bad ones. Indeed, there may or may not be a uniquely best policy on some issues but there are many bad ones, and we want the political decision-making process to rule those out (Neblo 2015; Steinberger 2018). Now, although it is common to think of modern representative democracy in terms of regular and fair elections, it is much more than that. It involves a great variety of collective activities. For instance, voting is preceded by electoral campaigns where candidates, journalists, experts, and ordinary citizens interact in the attempt to exchange information and reasons (Jacobs et al. 2009; Page 1996). And after voting, citizens, experts, and journalists are to hold elected authorities accountable for their decisions. Indeed, it is a basic

commitment of modern democracy that people can participate in acts of protest and dissent and many of the freedoms protected by it, such as the freedoms of expression, of press and of association, are directly linked to that (Pettit 2013; Whelan 2019). Dissenting citizens, even if they are a minority, can, in principle, deliberate and critique a given political decision and bring about social change, which shows our social aspiration for our collective lives to be guided by our better reasons (cf. Habermas 1996, 306).

Thus, a pivotal component of democracy is the free exchange of reasons and information in an attempt to argue with each other about what we collectively should do (Bohman 1996; Fishkin 2018; Gutmann and Thompson 2004; Landemore 2013). Democracy can then be thought of as consisting in the attempt to collectively determine via public deliberation the policies and actions that enjoy the support of our better reasons. In fact, one can take this deliberation to be the source of legitimacy of political decisions (Cohen 1989; Estlund 2008; Manin 1987; Peter 2009). Indeed, three decades ago, democratic theory took a “deliberative turn” (Dryzek 2000, v; Hansen 2012) as a mixed group of theorists challenged models of democracy focusing on voting and turned their attention to the role played by public deliberation in political decision-making. Regarded as one of the most promising approaches in democratic theory and the predominant framework (Bächtiger et al. 2018; Talisse 2019), deliberative democracy sees the communicative processes in which decision-making procedures are embedded as the primary source of political legitimacy. This normative framework puts an emphasis on the notion of the public sphere and the discourse by which it is constituted, as well as highlighting the utmost importance of such political discourse being adequate.

The political public sphere is a vital part of democratic society. It is constituted by complex, communicative networks, “where information, ideas and debate can circulate in society, and where political opinion can be formed,” which connects scattered people, sometimes across large geographical areas (Dahlgren 1995, ix; see also Fraser 1990, 57; Habermas 1996: 360, 373-4). It promotes the shaping of opinion on political issues and two central communicative processes within it (but of course not the only ones; see e.g., Young 2000) are the transmission of information relevant to those issues and the deliberative argumentation concerning them (Cohen 1989; Habermas 1996; Estlund 2008). These communicative processes, like much public communication, have a general cooperative orientation: we share information and collaboratively search for the better position. Ideally, in the public sphere, information is shared, different

perspectives are presented, the reasons behind them exchanged and, in the long run, the better reasons prevail.

1.2 Natural Testimony and Epistemic Disenfranchisement

Of course, there are different (sometimes vague) definitions of deliberation and public sphere within different fields of research and even within the sub-field of deliberative democracy (Bächtiger et al. 2018; Gripsrud et al. 2010; McKee 2005; Wodak and Koller 2008). Having said that, for present purposes, a useful outline of the deliberative procedure, and by extension the public sphere, which is consistent with the above crucial features, is provided by Habermas (1996, 305-306), following Cohen (1989). Among other things, the procedure is understood as: (a) an argumentative exchange of reasons and information among people who introduce and critically test proposals; (b) which is inclusive and public and where all the affected by the issue have equal chances to participate; (c) which is free of external constraints and the participants are only bound by the presuppositions of communication and rules of argumentation; and (d) which is equally free from internal constraints to the extent that every participant has the same opportunity to be heard, and indeed, be spoken to, when taking part in the debate (see also Estlund 2008; Bernstein 2012).

This is of course an ideal and might (often) not be realized in the real world. But utopian as it may be, this ideal can have a real-world effect and, certainly, if it is not impossible to achieve (even if it is very unlikely that it will be), there is no reason to reject it (Estlund 2008, 2020). Minimally, the ideal “serves as a template against which to judge reality in order to identify and deal with deviations” (Estlund 2008, 199), even if the result ends up not being exactly the ideal situation (2008, 200-201). The aim of this paper is to identify some such deviations in the deliberative process; more particularly, regarding one of the communicative processes that occur within it: namely, the transmission of information. Before introducing some aspects of our testimonial practice, we should point out that the deviations that we are interested in are cases of, as we shall say, *epistemic disenfranchisement*. Note that we are observing a technical distinction between formal and epistemic (dis)enfranchisement. The former refers to the right to vote as it is afforded to citizens by the law. The latter refers to a kind of informal (dis)enfranchisement relating to the realities and practicalities which determine citizens’ ability to participate freely and fully in the epistemic practices essential for democratic deliberation.

Universal suffrage is nowadays typically taken for granted in democracies and the franchise is widely regarded as a basic individual right. Of course, there are

electoral exclusions: for example, in most democracies, minors and persons with mental impairments, and in some democracies, noncitizen-residents and criminal offenders are formally disenfranchised (Blais, Massicotte and Yoshinaka 2003). However, most democrats would find it impermissible to exclude persons because of race, gender, religion, weight, and sexual preferences, among other things. The limits of the legal right to vote are certainly not set in stone but presently, say, a gender criterion for formal disenfranchisement is unacceptable (although before the 20th Century women were denied the right to vote [Dahlerup 2018]). Our aim is to emphasise that, within the framework of deliberative democracy, where participation in the deliberative process is as crucial to democratic decision-making as voting, some persons can and are likely to be excluded from the epistemic practices that take place within the deliberative process due to systematically held prejudices against them. We shall refer to cases where persons are so excluded as cases of epistemic disenfranchisement and although such disenfranchisement can occur in relation to all the epistemic practices involved in the deliberative process, here we shall focus on the testimonial ones to illustrate the phenomenon of epistemic disenfranchisement.

The mechanism whereby so many citizens are epistemically enfranchised in deliberative democracy is what C.A.J. Coady (1992) calls natural testimony, which is encountered in everyday circumstances, as opposed to formal testimony, of which a paradigmatic example is the testimony of a witness in a court of law. The ability to learn from other people's testimony and share one's own beliefs and experiences by offering one's own testimony is essential for the good function of the democratic deliberative process. Now, it is beyond the scope of this paper to rehearse the key positions in the epistemology of testimony.² It is sufficient for our purposes to assume that knowledge can be transferred *via* testimony and that this mechanism is essential for deliberative democracy. In the following sections, we will consider how epistemic injustice nevertheless limits the power of testimony to enfranchise citizens.

² Epistemologists of testimony typically distinguish between Humean or reductionist positions, on the one hand, and Reidian or anti-reductionist positions, on the other. According to the Humean view, it is epistemically reasonable to believe that *p* based on someone's testimony that *p* if you have independent, non-testimonial reasons to think that they are a reliable source of knowledge. In contrast, on the Reidian view, it is epistemically reasonable to believe that *p* based on someone's testimony that *p* unless you possess independent epistemic reasons to believe that they are not a reliable source of knowledge. Key defences of anti-reductionism include those of Coady (1992) and Burge (1993). A prominent defence of reductionism is due to Elizabeth Fricker (see, for instance, Fricker 1987, 1994). See Lackey 2008 for a thorough critique of this debate.

2.1 Some Varieties of Testimony-related Epistemic Injustice

In this section, we introduce and explore some existing work on the topic of epistemic injustice. Fricker's *Epistemic Injustice: Power and the Ethics of Knowing* (2007) has been incredibly influential in establishing the topic of epistemic injustice as a mainstream research programme in contemporary epistemology. This will also help us to contextualise the varieties of testimonial injustice that we introduce in the following section. It is natural, then, that we take Fricker's account of epistemic injustice as our starting point.³

Fricker identifies two distinctive ways in which we can be wronged specifically in our capacity as epistemic agents and posits two varieties of epistemic injustice to explain them: *testimonial injustice* and *hermeneutical injustice*. Since the testimonial strain of epistemic injustice is more relevant to our project, we will focus on that aspect of Fricker's view.⁴ In Fricker's words, 'testimonial injustice occurs when prejudice causes a hearer to give a deflated level of credibility to a speaker's word' (2007, 1). Broadly speaking, testimonial injustice occurs when one person, the speaker, tries to tell another person, the hearer, that *p*, but the hearer does not accept the speaker's testimony, and in particular, because they, the hearer, possess prejudicial stereotypes about the speaker's social identity according to which the speaker is not a credible source of testimony. How does this come about? It is a fact of our epistemic lives that we are dependent on one another. We rely for a great deal of our beliefs, among other things, on the testimony of others. In a perfect world, this would make things straightforward. But this is not a perfect world: some people are incompetent, some are insincere, and some are both. We need to be able to determine who is a good informant, that is, someone who is both competent and sincere. According to Fricker, this role is performed by indicator

³ Dotson points out that, in treating recent work on epistemic injustice as a starting point, we should be careful not to overlook the fact that many thinkers have addressed very similar phenomena (2012). For instance, Patricia Williams describes how her testimony regarding her experiences of racial prejudice while shopping in New York City was met with a credibility deficit when one of articles discussing these experiences was reviewed by a Stanford Law School class. She was afforded an unwarranted credibility deficit on the basis of negative stereotypes that identify African Americans as, among other things, dishonest and paranoid (Williams 1991; discussed in Dotson 2012, 26-28).

⁴ By focusing on testimonial injustice, we do not mean to suggest that we believe that hermeneutical injustices make no difference to the health of democracy and the success of deliberation. We would agree, following Medina (2013), that epistemic injustices related to testimony and hermeneutical injustice are intimately related and feed each other. However, in this paper, we will focus on the testimonial side of this relation, leaving questions about hermeneutical injustice to one side for now.

properties, the visible, external signs that someone is a good informant (2007, 114-120). Many of these indicator properties themselves rely on stereotypes about social identity. Sometimes indicator properties reliably indicate whether a speaker is credible. However, these indicator properties—or rather, our reliance upon them—also leave us vulnerable to several types of error. Sometimes people possess indicator properties but lack competence or sincerity. In other words, some people who are not good informants are nevertheless regarded as if they were. This can lead us to form beliefs on the basis of the testimony of people who are ill-informed or insincere. This is undesirable.

But perhaps the more pernicious variety of error is that which occurs when someone who is both competent and sincere is not recognised as such. In cases of testimonial injustice, a speaker's testimony is not accepted by her hearer because, according to negative identity-prejudicial stereotypes held by the hearer, the speaker is either incompetent or insincere. 'A negative identity-prejudicial stereotype is,' according to Fricker,

[a] widely held disparaging association between a social group and one or more attributes, where this association embodies a generalization that displays some (typically, epistemically culpable) resistance to counter-evidence owing to an ethically bad affective investment. (2007, 35)

Fricker provides the example of Marge Sherwood from Minghella's screenplay for *The Talented Mr. Ripley* (2007, 86-91). In short: Marge recently got engaged to Dickie Greenleaf, the renegade son of a wealthy industrialist, Herbert Greenleaf. Dickie has gone missing and Marge thinks—with good reason—that his friend, the eponymous Mr. Ripley, is involved. But when Marge explains her suspicions to Herbert, he dismisses them: 'Marge, there's female intuition, and then there are facts' (cited in Fricker 2007, 88). Marge is constructed as a hysterical woman who cannot be relied upon to form true beliefs about the world. Moreover, it is worth noting that while Ripley coordinates this construction of Marge's social identity, the other men are certainly complicit in it too. In any case, the result is that nobody accepts Marge's testimony because they judge her not to be a credible source of knowledge on the basis of the prejudices they hold about women. Of course, other social identities beyond gender can be a source of the kinds of negative identity-prejudicial stereotypes that lead to testimonial injustice: class, gender, sexuality, and others bring with them the risk of social identity-based stereotypes, which can intersect.^{5 6}

⁵ Many other groups often suffer from negative identity-prejudicial stereotypes about credibility. For example, old people are sometimes subject to the negative stereotype that they are not capable (Jackson 2020) and African Americans can be subject to such negative stereotypes even

In summary, according to Fricker, testimonial injustice occurs when one person, the speaker, tries to tell another person, the hearer, that p , but the hearer does not accept the speaker's testimony because they, the hearer, possess prejudicial stereotypes about the speaker's social identity according to which the speaker is not a credible source of testimony.

2.2 Two Under-discussed Types of Testimony-related Epistemic Injustice

Fricker's concept of testimonial injustice helps us to understand a range of cases in which people are wronged in their capacity as epistemic agents. This account does not explain what has gone wrong in cases of epistemic disenfranchisement. So, in this section, we will draw attention to two further varieties of epistemic injustice related to testimony which we believe to contribute significantly to epistemic disenfranchisement, but which are sometimes overlooked. These are *discursive injustice* and *testimonial void* (Kukla 2014; Carmona 2021).

Discursive injustice. In standard cases of testimonial injustice as described by Fricker, the speaker testifies that p but while the hearer understands what the speaker is trying to communicate to them, they do not find the speaker credible because of negative identity-prejudicial stereotypes about the speaker's identity. This, of course, leads to communication failures. But not all communication failures have this structure. In the standard cases, the speaker achieved one part of what they set out to do: they were recognised as testifying that p . The problem in this case is that they were not recognised as a credible source. Reflecting on such failures of communication, Kukla introduces the concept of *discursive injustice*. Kukla argues that:

Sometimes a speaker's membership in an already disadvantaged social group makes it difficult or impossible for her to deploy discursive conventions in the normal way, with the result that the performative force of her utterances is distorted in ways that enhance disadvantage (Kukla 2014, 441).

Grammatical structure and semantic content are not sufficient for fixing the performative force and pragmatic structure of a given speech act, Kukla explains. *Discursive conventions* also play an important role in the fixing and interpretation of speech acts. These conventions 'determine when a speaker is entitled to issue a

by those who are politically and socially liberal and believe that they are not prejudiced (Dovidio, Gaertner & Pearson 2017).

⁶ Relatedly, Peet (2017) identifies a variety of epistemic injustice that occurs at the level of utterance interpretation, though Peet is concerned with how stereotypes influence our understanding of the content of utterances.

speech act of type A in context C' (Kukla 2014, 444). They determine whether a speech act gets the uptake in its audience that was intended by the speaker, where uptake is a matter of recognising the normative status changes that the utterance makes.⁷ Kukla argues that social identity can disrupt the working of these discursive conventions, such that a speaker who is entitled to make a speech act of type A in context C nevertheless fails to get the correct uptake for that speech act, because of stereotypes about the kinds of speech acts that members of that marginalised group tend to (or ought to) make (2014, 445).

Let's consider the speech act of testimony. In cases of discursive injustice, a speaker from a marginalised social group testifies that *p* but their speech act (i.e., testimony) is not recognised as such by the hearer, but rather, is intercepted as a different kind of speech act, and this is because of stereotypes about the speaker's social identity held by their audience. Women's emotional speech acts may be especially vulnerable to this kind of discursive injustice. Discussing Scheman's work, Kukla writes:

Women's emotional speech acts are often interpreted (including self-interpreted) as incapable of bearing cognitive content that is accountable to external facts about how things are; they are taken as mere *expressions* of emotion rather than as claims. (Kukla 2014, 451).⁸

In these sorts of cases, the audience mistakes the speaker's intended speech act (namely, testimony) for another one (for instance, opining, joking, or emoting). Consider Fricker's example of Marge from *The Talented Mr. Ripley* (Fricker 2007). In Fricker's reading of this case, the problem is that Herbert Greenleaf believes that Marge is incompetent and therefore that her testimony regarding the identity of the murderer should not be believed. In this case, Greenleaf recognises her testimony as such, but then rejects it because he judges her not to be a good informant. Another way we can understand Greenleaf's apparent rejection of Marge's testimony is that he does not even realise that she is testifying. Rather, we might speculate, Greenleaf interprets Marge's utterance (offered as a piece of testimony) as an entirely different kind of speech act. Maybe he thinks she is simply gossiping or playing along with a conversation she does not really understand or care about. Perhaps he holds that the purpose of testimony is to provide facts, but someone equipped with 'female intuition' is unlikely or even unable to have this objective. After all, as Greenleaf says, 'Marge, there's female intuition, and then there are facts' (Fricker 2007, 9). The thought is that testimony

⁷ Kukla diverges here from the Austinian notion of uptake, which is that the audience correctly recognises the speaker's intention. (See Hornsby & Langton 1998.)

⁸ See Scheman 1993.

based on women's intuition literally fails to count as testimony and instead, at best, qualifies as 'mere opinion.' And indeed, when Marge becomes understandably upset at her mistreatment, her speech will be recast as hysterical expressive outbursts rather than assertions of fact.

Moreover, the exclusion of people from the epistemic practice need not exclusively be due to a negative identity-prejudicial stereotype, but also to positive ones. As Davis argues, positively valenced or 'benevolent' stereotypes can also lead to credibility excesses which in turn cause epistemic injustice (2016). Davis provides the following example of what she calls identity-prejudicial credibility excess:

A group of American high-school students struggle to complete a difficult algebra question during their lunch period. After several failed attempts to solve the problem among themselves, the students decide to seek outside help. The students have heard that Asian-Americans are particularly good at math, so they ask an Asian-American student seated nearby for help with the problem. (2016, 487).

According to Davis, the identity-prejudicial credibility excess afforded to—or perhaps, imposed upon—the Asian American student in the example above is an instance of epistemic injustice because it involves compulsory representation, a form of epistemic exclusion whereby 'marginalized knowers are invited to participate in epistemic exchanges [but] the invitation is extended to the individual only insofar as the individual satisfies a certain description' (Davis 2016, 490). This, Davis explains, is harmful because it is a form of tokenism whereby marginalized individuals are unjustly treated as representatives of an exotic group (Davis 491). Tokenism is a special case of identity-prejudicial credibility excess where specific members of disadvantaged groups are singled out by members of the dominant class and obliged to represent that group regardless of their own wishes or abilities.⁹

Davis's examples show how positive stereotypes can lead to credibility excesses which cause epistemic injustice. Interestingly, though, there is an interesting disconnect between the overall valence of a stereotype (positive or negative) and its epistemic consequences for members of the group to whom it applies. In contrast to the examples discussed by Davis, positively valenced stereotypes can also lead to credibility deficits and exclusion. For an example of the

⁹ Thus, not all cases of identity-prejudicial credibility excess are harmful. If a member of the British aristocracy is afforded a credibility excess with respect to the topic of horse breeding, they may be embarrassed if it turns out they are not knowledgeable on this topic, but it seems implausible that they will have been substantially harmed in any way. That's because members of the British aristocracy are not marginalised in British society.

latter, consider how overweight as well as people perceived as camp are sometimes subject to positively valenced stereotypes, among other negative ones, regarding their playfulness and funniness (Diedrichs and Puhl 2017; Jackson 2020). But such positive attributes can bear negative effects in conversations since people subject to this stereotype can wrongly be thought of attempting to make a funny remark as opposed to a serious contribution. Regardless of its positive valence, the stereotype can nevertheless be harmful in the sense that, for example, some intended piece of testimony might instead be regarded as joking.

Testimonial void. So far, all the cases of testimonial injustice we have considered are related to how an audience reacts to a speaker's testimony, and in particular how their reaction wrongs that speaker. Another kind of epistemic injustice which is both highly relevant for epistemic disenfranchisement and almost completely undiscussed in the literature can be found in what Carmona labels *testimonial void* (2021). Carmona's argument begins in a reflection on Dotson's concept of testimonial smothering (2011). Dotson considers how speakers sometimes truncate or outright withhold their testimony because they reasonably believe that their testimony is likely to be misunderstood by the hearer in ways that have harmful consequences for the speaker, because of pernicious situated ignorance on the part of the would-be audience (2011, 244). She calls this *testimonial smothering*. Consider the following example, which is due to Dotson: Some African American women withhold from testifying about domestic violence committed by African American men because, while they recognise the harm of domestic violence, they fear that their testimony will help to justify harmful beliefs about African Americans (Dotson 2011, 245). An important insight of this work is that it is not just in how testimony is received that epistemic injustice may occur, but also in how, and indeed whether, it is offered. But where Dotson focuses on how the withholder may be the victim of injustice, Carmona focuses on cases where a person is wronged because a speaker withholds testimony from them.

Sometimes people withhold their testimony from a potential audience because they hold negative identity-prejudicial views about the social group to which the audience appears to belong, according to which they, the audience, would be unable or unwilling to respond appropriately to their testimony or not deserving of the testimony. To put it simply, there are cases in which a speaker who believes that *p* does not tell her audience that *p* because she thinks that they are either too incompetent to understand *p* or too dishonest or immature to respond appropriately to the testimony that *p*. With this in mind, let's consider one

of Carmona's illuminating examples of testimonial void. It may help to illustrate this further. Carmona shares the striking example of her own grandmother,

[who] was deprived of the epistemic resources to handle a man's everyday life and relied on my grandfather for everything that involved the world outside the home, with the exception of going to church. To this day, she continues to tell me today how much she misses my grandfather, who died a few years ago, because 'he used to take care of everything.' It is only after writing this paper that I feel that I am beginning to give full weight to another utterance that is typical of her, 'Girl, I don't understand', by which she often expresses her bewilderment regarding issues having to do with her finances or other *worldly* things. No less significant is my mother's complaint: 'I have become your grandfather for her.' (2021, 8)

Carmona emphasises the long-term consequences of testimonial void that occurs when someone is denied access to important information. When a substantial body of knowledge is systematically withheld from a group of people, they are put at an epistemic disadvantage. While we are all epistemically dependent on others to some extent, people who are systematically denied access to large swathes of knowledge pass from being dependent to being excluded. We will return to this thought in the following section.

We have presented a variety of examples of how negative identity-prejudicial stereotypes can cause a particular kind of epistemic injustice whereby a speaker withholds some or all of their testimony from an audience because they believe, on the basis of prejudice, that the audience is either incompetent or dishonest or not entitled to be part of the political conversation. These examples as well as the previous ones presented in this section have a distinctly political air to them and they should have already helped the reader connect these cases of epistemic injustice to the phenomenon of epistemic disenfranchisement. In the following section and to conclude, we will make the connection explicit.

3. Epistemic Injustice and Disenfranchisement

In the previous section, we considered how members of groups that are subject to negative and positive identity-prejudicial stereotypes can suffer some epistemic injustices related to the testimonial practice. In this section, we will explain how these forms of testimony-related epistemic injustice can lead to epistemic disenfranchisement.

On the one hand, the testimony of some members of these groups may not be recognised as such by the intended audience due to either negative or positive identity-prejudicial stereotypes, which make the audience mistake the speech act of testifying for another one, such as opining or joking. This is discursive injustice.

Let's consider a case where negative identity-prejudicial stereotypes lead to epistemic disenfranchisement via discursive injustice. Politics is still regarded as a "man's game" (Burns, Schlozman and Verba 2001; Clavero and Galligan 2005; Koenig et al. 2011; Mendez and Osborn 2010). When Nancy Astor, the first woman Member of Parliament, took her seat in the House of Commons, Winston Churchill is famous for allegedly remarking that it was as if she had interrupted him in the bathroom.¹⁰ It is common, when it comes to politics, that women's views are often ignored, interrupted, or dismissed. This notion of politics as a man's game, which many have internalised and which makes the space of conversation the domain of the men, is certainly still problematic for its exclusion of women.¹¹ Although most women nowadays have the right to vote and the right to stand for election (although, as one might have expected, still not sufficiently many women are given the opportunity to do so; Dahlerup 2018; Htun 2016), the idea that women do not belong in politics, partly because they are not expected to be as competent as men (Karakowsky, McBey and Miller 2004) but also because it is sometimes deemed to be simply inappropriate or unfitting (given that political discussion is a man's domain), can clearly affect whether, say, someone recognises a woman's attempted testimonial contribution to a political discussion as testimony at all. So, a woman's politically charged testimony, when not deliberately ignored or chided for being political, might instead be regarded by someone who holds that politics is not for women as a misplaced attempt to express their emotions.

Positively-valenced identity-prejudicial stereotypes may also lead to epistemic disenfranchisement via discursive injustice. As seen, due to fatphobic and homophobic stereotypes, overweight and people perceived as camp are sometimes subject to some such stereotypes regarding their playfulness and funniness. So, their serious contributions to some political discussion might wrongly be taken to be funny remarks by someone holding these stereotypes and a disagreeing view. This person, given that it takes the view to be wrong and the speaker to be playful and funny might naturally regard the speaker's testimony as a joke.

In both these cases, the testimony of the speaker is not being regarded as such by the audience, who instead understand it as some other speech act which does not allow the speaker's contribution to have an epistemic impact in the deliberation.

¹⁰ These comments may be apocryphal.

¹¹ Indeed, some have accused Habermas of positioning women outside the public sphere and not seeing them as making significant contributions (Brooks 2019).

On the other hand, testimony may be withheld from some of the members of these groups because the potential testifier holds negative identity-prejudicial stereotypes about them, according to which they are either incompetent or dishonest or ineligible for political discourse. This leaves these people in a testimonial void. Carmona identifies a politically charged case of testimonial void in the educational system of Franco's dictatorship. She writes:

The wider educational system during the Franco regime structurally disesteemed the intellectual abilities of women. Reforms in 1945 segregated education by sex, and women's education focused on preparing them to be (house)wives and mothers. Consequently, there was a specific curriculum for girls, which included housework, sewing, and childcare. In addition, in the subjects studied that were also taught to men, the curriculums differed significantly. For example, in History classes, women's education focused on the feminine qualities (mostly concerning self-sacrifice) of queens and other Catholic heroines. In this manner, the Franco regime controlled who was (not) told what in the educational context (2021, 2).

In Spain during the Francoist dictatorship, certain kinds of knowledge were withheld from women. This is a kind of testimonial void. We might attempt to explain this because of stereotyping. The thought would be that because women were held, according to the prevalent stereotypes of the time, to be unsuited intellectually and emotionally for the topic of politics. That being said, it is not clear that this instance of testimonial void is a consequence of negative identity-prejudicial gender stereotypes. Rather, this case of testimonial void may have been part of an ideological project designed to subjugate women. On this reading, the creation of a testimonial void for women is based not on the dictatorship's beliefs about how women are but rather on its beliefs about how women should be. Indeed, these explanations are not mutually exclusive. In any case, while Carmona focuses on an explicitly educational setting here, the same thought applies in the public sphere in democratic deliberation.

Given what Carmona has shown about testimonial void in other contexts, it is easy to imagine how it would look in political deliberations. We have already seen how politics is regarded as a male territory. This is usually understood as a matter of excluding women from the physical space of deliberation or diminishing their contributions to the deliberations which they take part in, but we can also imagine how these kinds of exclusion are accompanied by testimonial void. In much-discussed situations where women's contributions are ignored, their hearer additionally—albeit indirectly—deprives those women of information that they, the audience, would have shared with a speaker they respected as worthy participants in the deliberation. This could take the form of explicitly refusing to communicate with someone, but it is perhaps more likely that involves humouring

the speaker by engaging only superficially with what they say. It is easy to imagine a case where a woman's testimony in some political discussion is not engaged by the male audience who might anyway have information relevant to the assertion made. In this case, where the audience does not put forward some relevant information, there is testimonial void.

Another case of epistemic disenfranchisement via testimonial void involves those who, given their prejudiced association with some other group which is normally disenfranchised (i.e., not given the right to vote), are not taken to be legitimate interlocutors. For example, in his presidential campaign, Donald Trump is well known for having referred to Mexican immigrants as criminals (especially drug dealers and rapists; Lee 2015):

When Mexico sends its people, they're not sending their best. They're not sending you. They're not sending you. They're sending people that have lots of problems, and they're bringing those problems with us. They're bringing drugs. They're bringing crime. They're rapists.

Endorsement of such stereotypes can be consequential: criminals, as mentioned before, are disenfranchised in many democracies, including the USA (in fact, they are permanently denied the right to vote in the USA), and so people who stereotype Mexican people in that way can believe that they should not be included in political conversation. Given this, these people may not testify to Mexicans simply because they are prejudicially associated with a disenfranchised group.¹²

We have argued that there are some ways in which people can be excluded from the testimonial practice. With the help of some contemporary work by Kukla and Carmona, we have described two varieties of testimony-related epistemic injustice that are particularly overlooked in the literature on deliberative democracy, namely, discursive injustice and testimonial void.

As already seen, a crucial component of deliberative democracy is the exchange of information in the attempt to argue with each other about what we

¹² Similarly, since the 9/11 attack on the World Trade Center, negative identity-prejudicial stereotypes about Muslims, which portray them as violent criminals and, and, in particular, as terrorists, are prevalent (Sides & Gross 2013). In this way too, in the political arena, people may also withhold their testimony from Muslims given their perceptions that Muslims are, in some unspecified way, ineligible for inclusion in political discourse. Feminist and LGBT activists may suffer from the stereotype that they are not acting in good faith, but rather, they are seeking offence in order to further their political project. People may then refrain from engaging in conversation with people whom they perceive to be interested in feminist and LGBT activism (cf. Barvosa 2018).

collectively should do. Testimony then is one central communicative process within deliberative democracy (but certainly not the only one, e.g., Young 2000). More generally, deliberative models of democracy focus their attention mainly on the role played by public deliberation in political decision-making rather than the casting and counting of votes. This being so, the key democratic value of inclusion is also to be considered in the deliberative process (as highlighted in the above schematic characterization of the process; §1.2). Accordingly, the formal enfranchisement that citizens gain when given the right to vote is not enough for these models: voting is not enough. Citizens should not only participate in the voting process but also in the deliberative process and on equal terms (Young 2000). So, deliberative models require citizens to be epistemically enfranchised: namely, to be included in the epistemic practices of the deliberative process. The exclusion from the testimonial practice, as in the above cases of epistemic injustice, represents a sort of epistemic disenfranchisement.¹³ So, although when we talk about enfranchisement, we usually think about the right to vote, we must not lose sight of the following important fact within deliberative models of democracy: even if the state grants a given social group the right to vote, the members of that group are nevertheless epistemically disenfranchised if they are unable to participate in the deliberative aspects of democracy.

It is important to emphasize this difference between the phenomenon we have identified in this work and testimonial injustice as understood by Fricker. Crucially, cases of testimonial injustice are not cases of epistemic disenfranchisement. This is because the victims of testimonial injustice are included in the testimonial practice, even though they are not treated equally, as the deliberative model further requires. Where testimonial injustice presents an obstacle to equality among formally and epistemically enfranchised citizens, the injustices identified in this paper present obstacles to inclusion in the deliberative process.¹⁴

Of course, many have pointed out the various inequalities prevailing in society that are likely to be amplified (rather than mitigated) in the public sphere

¹³ This is one variety of informal disenfranchisement, we have suggested, but there may be others. For example, another sort of epistemic disenfranchisement concerns exclusion from argumentation. Although we do not have space to discuss it here, we think that a rather similar case can be made for this sort of disenfranchisement.

¹⁴ Inclusion and equality are both baseline normative ideals in democratic theory generally. However, within deliberative democracy, these ideals promote deliberative dimensions concerning the inclusion and equal treatment of all viewpoints (and so their experiences, reasons, and arguments; Bächtiger & Parkinson 2019; Christiano 1996). It is these dimensions that we have in mind here.

and even the Habermasian account of it has been criticized for ignoring that fact (Fraser 1990; see also Young 2000). There are many subtle forms of political oppression and control that may prevail in inclusive arenas and do not permit fairness in participation. People's deliberative contributions should be considered equally on their merits but prejudices, as in Fricker's testimonial injustice cases, can create inequalities within the deliberative process with regard to the credibility that people ought to be attributed. So, the viewpoints of some groups may play a disproportionate role in various parts of the deliberative process.

However, here we have been interested in cases where people are excluded from the deliberative process, rather than treated unequally within it. Now, although many have also been interested in the informal disenfranchisement of people, they focus on cases where people with greater power and resources may purposely leave others out of the political discussion; as Young (2000) would say, they focus on "external exclusion:" namely, when people do not have access to the fora for discussion.¹⁵ But Young (2000) is mainly interested in cases of "internal exclusion," which she takes to be less noticed than the cases of external exclusion. The former are cases where people, often unconsciously, ignore or dismiss or patronize others' contributions in political discussion (2000, 55). Given this, she suggests that some forms of communication, which indicate recognition, such as greetings, rhetoric, and narratives, are essential to inclusive deliberation (2000, 57ff.). Where Young focuses on explaining and ameliorating the broader phenomenon of internal exclusion, we have sought to refine one way in which epistemic disenfranchisement as a kind of internal exclusion can occur using the theoretical tools provided by recent work in social epistemology on epistemic injustice. So, we have attempted to specify the mechanisms behind two very particular and under-discussed ways in which people are internally excluded in relation to one key epistemic practice of the deliberative process. In particular, two sorts of epistemic injustices that internally exclude people from testimony that are not normally considered within the political domain. These epistemic injustices then generate the epistemic disenfranchisement of people that are formally enfranchised.

We have argued that much epistemic disenfranchisement can occur unintentionally due to a series of negative and positive identity-prejudicial stereotypes to which different marginalized groups are subject. Some of these groups, such as women and Black people, have only recently gained formal enfranchisement in some democracies. However, within the deliberative democracy framework, we may still be failing these groups inadvertently and in

¹⁵ This seems to be related to what Hookway calls *participatory injustice* (2010).

subtle ways that nevertheless bear the burden of certain identity-prejudicial stereotypes that cause individuals to exclude members of these groups from the epistemic practices involved in the deliberative process and so face informal epistemic disenfranchisement even if not formally disenfranchised. Moreover, within the deliberative democracy framework, we may also inadvertently be failing other groups who are subject to such identity-prejudicial stereotypes and were never thought to be disenfranchised. Certainly, most of us find it impermissible to exclude persons from political decision-making because of race, gender, religion, weight, age, and sexual preferences, among other things. Having said that, if we are to live up to our own ideals and those set by the deliberative model of democracy, we need to start paying more attention to the specific mechanisms behind the epistemic disenfranchisement that is likely to go on in political deliberation.

References

- Aikin, Scott, and Robert Talisse. 2019. *Why We Argue (And How We Should)*. 2nd Edition. New York: Routledge.
- Bächtiger, André, John Dryzek, Jane Mansbridge and Mark Warren. 2018. *The Oxford Handbook of Deliberative Democracy*. Oxford: Oxford University Press.
- Bächtiger, André and John Parkinson. 2019. *Mapping and Measuring Deliberation: Towards a New Deliberative Quality*. Oxford: Oxford University Press.
- Barvosa, Edvina. 2018. *Deliberative Democracy Now: LGBT Equality and the Emergence of Large-Scale Deliberative Systems*. Cambridge: Cambridge University Press.
- Bernstein, Richard. 2012. "The Normative Core of the Public Sphere". *Political Theory* 40: 767-78.
- Blais, Andre, Louis Massicotte and Antoine Yoshinaka. 2003. *Establishing the Rules of the Game*. 2nd ed. Toronto: University of Toronto Press.
- Bohman, James. 1996. *Public deliberation. Pluralism, complexity and democracy*. Massachusetts: MIT Press.
- Brooks, Ann. 2019. *Women, Politics and the Public Sphere*. Bristol: Policy Press.
- Burge, Tyler. 1993. "Content Preservation." *Philosophical Review* 102: 457-48.
- Burns, Nancy, Key Schlozman and Sidney Verba. 2001. *The Private Roots of Public Action: Gender, Equality and Political Participation*. Cambridge: Harvard University Press.
- Byford, Jovan. 2011. *Conspiracy Theories: A Critical Introduction*. London: Palgrave Macmillan.

- Carmona, Carla. 2021. "Silencing by Not Telling: Testimonial Void as a New Kind of Testimonial Injustice." *Social Epistemology* 35: 577-592.
- Christiano, Thomas. 1996. *The Rule of the Many: Fundamental Issues in Democratic Theory*. Boulder: Westview Press.
- Clavero, Sara, and Yvonne Galligan. 2005. "'A Job in Politics Is Not for Women': Analysing Barriers to Women's Political Representation in CEE." *Czech Sociological Review* 41: 979-1004.
- Coady, C.A.J. 1992. *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
- Cohen, J. 1989. "Deliberation and Democratic Legitimacy". In *The Good Polity*, edited by Alan Hamlin and Philip Pettit, 21-42. Oxford: Blackwell.
- Dahlerup, Drude. 2018. *Has Democracy Failed Women?* Cambridge: Polity.
- Dahlgren, Peter. 1995. "Introduction." In *Communication and Citizenship: journalism and the public sphere in the new media age*, edited by Peter Dahlgren and Colin Sparks, 1-24. London: Routledge.
- Davis, Emmalon. 2016. "Typecasts, Tokens, and Spokespersons: A Case for Credibility Excess as Testimonial Injustice." *Hypatia* 31: 485-501.
- Diedrichs, Phillippa and Rebecca Puhl. 2017. "Weight Bias: Prejudice and Discrimination toward Overweight and Obese People". In *The Cambridge Handbook of the Psychology of Prejudice*, edited by Chris Sibley and Fiona Barlow, 392-412. Cambridge: Cambridge University Press.
- Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26: 236-257.
- . 2012. "A Cautionary Tale: On Limiting Epistemic Oppression." *Frontiers: A Journal of Women Studies* 33(1): 24-47.
- Dovidio, John, Samuel Gaertner and Adam Pearson. 2017. "Aversive Racism and Contemporary Bias." In *The Cambridge Handbook of the Psychology of Prejudice*, edited by Chris Sibley and Fiona Barlow, 267-294. Cambridge: Cambridge University Press.
- Dryzek, John. 2000. *Deliberative Democracy and Beyond*. Oxford: Oxford University Press.
- Dunn, John. 2005. *Setting the People Free: The Story of Democracy*. 2nd Edition. New Haven: Princeton University Press.
- Estlund, David. 2008. *Democratic Authority*. New Haven: Princeton University Press.
- . 2020. *Utopophobia*. New Haven: Princeton University Press.
- Fine, Cordelia. 2017. *Testosterone Rex*. London: Faber & Faber.
- Fishkin, James. 2018. *Democracy When the People are Thinking*. Oxford: Oxford University Press.

Leandro De Brasi, Jack Warman

- Fraser, Nancy. 1990. "Rethinking the Public Sphere." *Social Text* 25/26: 56–80.
- Fricker, Elizabeth, 1987. "The Epistemology of Testimony." *Proceedings of the Aristotelian Society Supplement* 61: 57–83.
- Fricker, Elizabeth. 1994. "Against Gullibility." In *Knowing From Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, edited by B. Matilal and A. Chakrabarti. Dordrecht: Kluwer Academic Publishers.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Gripsrud, Jostein, Moe Hallvard, Anders Molander and Graham Murdock. 2010. *The Idea of the Public Sphere*. Plymouth: Lexington Books.
- Gutmann, Amy and Dennis Thompson. 2004. *Why Deliberative Democracy?* New Haven: Princeton University Press.
- Habermas, Jürgen. 1996. *Between Facts and Norms*. Massachusetts: MIT Press.
- Hansen, Kasper. 2012. "Deliberative Democracy: Mapping Out the Deliberative Turn in Democratic Theory." In *Rhetorical Citizenship and Public Deliberation*, edited by Christian Kock and Lisa Villadsen, 13–27. Pennsylvania: Penn State University Press.
- Hookway, Christopher. 2010. "Some Varieties of Epistemic Injustice: Reflections on Fricker." *Episteme* 7 (2): 151–163.
- Htun, Mala. 2016. *Inclusion without Representation in Latin America*. Cambridge: Cambridge University Press.
- Jackson, Lynne. 2020. *The Psychology of Prejudice*. 2nd edition. Washington: American Psychological Association.
- Jacobs, Lawrence, Fay Cook and Michael Delli Caprini. 2009. *Talking Together*. Chicago: University of Chicago Press.
- Karakowsky, Leonard, Kenneth McBey, and Diane Miller. 2004. "Gender, Perceived Competence, and Power Displays Examining Verbal Interruptions in a Group Context." *Small Group Research* 35: 407–39.
- Koenig, Anne, Alice Eagly, Abigail Mitchell and Tiina Ristikari. 2011. "Are Leader Stereotypes Masculine? A Meta-Analysis of Three Research Paradigms". *Psychological Bulletin* 137: 616–642.
- Kukla, Rebecca. 2014. "Performative Force, Convention, and Discursive Injustice." *Hypatia* 29(2): 440–457.
- Lackey, Jennifer. 2008. *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.
- Landemore, Helene. 2013. *Democratic Reason*. New Haven: Princeton University Press.
- Lee, Harper. 1960. *To Kill A Mockingbird*. (Various editions.)

- Lee, Michelle Ye Hee. 2015. "Donald Trump's false comments connecting Mexican immigrants and crime." *The Washington Post*. <https://www.washingtonpost.com/news/fact-checker/wp/2015/07/08/donald-trumps-false-comments-connecting-mexican-immigrants-and-crime/> Accessed: 1 December 2020.
- Manin, Bernard. 1987. "On Legitimacy and Political Deliberation." *Political Theory* 15: 338-68.
- McKee, Alan. 2005. *The Public Sphere*. Cambridge: Cambridge University Press.
- Medina, José. 2011. "The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice: Differential Epistemic Authority and the Social Imaginary." *Social Epistemology* 25(1): 15-35.
- . 2013. *The Epistemology of Resistance*. Oxford: Oxford University Press.
- Mendez, Jeanette, and Tracy Osborn. 2010. "Gender and the Perception of Knowledge in Political Discussion." *Political Research Quarterly* 63: 269-279.
- Neblo, Michael. 2015. *Deliberative Democracy between Theory and Practice*. Cambridge: Cambridge University Press.
- Page, Benjamin. 1996. *Who Deliberates?* Chicago: University of Chicago Press.
- Peet, Andrew. 2017. "Epistemic Injustice in Utterance Interpretation." *Synthese* 194(9): 3421-3443.
- Peter, Fabienne. 2009. *Democratic Legitimacy*. London: Routledge.
- Pettit, Philip. 2013. *On the people's terms*. Cambridge: Cambridge University Press.
- Scheman, Naomi. 1993. "Anger and the Politics of Naming." In *Engenderings: Construction of knowledge, authority, and privilege*, 22-35. New York: Routledge.
- Sides, John, and Kimberly Gross. 2013. "Stereotypes of Muslims and Support for the War on Terror." *Journal of Politics* 75: 583-98.
- Steinberger, Peter. 2018. *Political Judgment*. London: Polity.
- Talisso, Robert. 2019. *Overdoing Democracy: Why We Must Put Politics in Its Place*. Oxford: Oxford University Press.
- Whelan, Frederick. 2019. *Democracy in Theory and Practice*. London: Routledge.
- Williams, Patricia. 1991. *The Alchemy of Race and Rights: A Diary of a Law Professor*. Cambridge: Harvard University Press.
- Wodak, Ruth and Veronika Koller. 2008. *Handbook of Communication in the Public Sphere*. Berlin: de Gruyter.
- Young, Iris M. 2000. *Inclusion and Democracy*. Oxford: Oxford University Press.

LIE *FOR THE OTHER*: A SOCIO-ANALYTIC APPROACH TO TELLING LIES

Rauf ORAN

ABSTRACT It is a widely held view that lying is defined in the traditional tripartite model as the conjunction of a statement, the false belief, and the intended deception. Much of the criticisms have been levelled at the third condition—intended deception—with contemporary counterexamples. My main criticism of the traditional and contemporary model of lying centres on that philosophers discard the social existence of the hearer. Schutz's phenomenological sociology gives a sheer inspiration to redefine the third condition by taking the hearer as a *consciously social being* into account. Lying should be an intersubjective action *for the Other* from the perspective of the liar; it might be, thus, reasonable to assume that there should be *commonsense awareness* between the speaker and the hearer. This paper, by focusing on this commonsenseness and its *typifications*, introduces a new approach to the third condition: S must intend that H be induced to believe that *p*, where *p* is false. In this regard, once you lie, by being subjected to the *taken-for-granted* commonsenseness in our daily life, you must try *as hard as possible to succeed in deceiving* the hearer by stating that *p*. You, as a *typical* person, tell a *typical* lie in *typical* contexts for *typical* Others. The focus of attention, therefore, is on the hearer and it is the key to understanding that mere *intent to deceive* is too broad and unpragmatic for a social human being who always intends to flee the negative consequences of the context in which she has to lie. Making the extension narrower necessitates a new term, *anti-social bullshit* generally being replied rhetorically as “how can you expect me to believe that?” comprises the excluded cases.

KEYWORDS: lying, Schutz. commonsense-world, anti-social
bullshit, induce-to-believe

1. Introduction

As Nietzsche (1998, 7) said, the lie is “a condition of life.” There is no denying the fact that lying is as much a part of our social life as any other use of our language. Lying is a social action that involves at least two people interacting linguistically with one another; if the speaker intends to lie, then manifestly there must be a hearer for whom the lie is intended.

The traditional definition of lying is broken down into three conditions, 'to make a statement, 'to believe that the statement is false,' and 'to intend to deceive.' More exactly, S lies to H, if and only if,

C₁- S states that *p* to H,

C₂- S believes that *p* is false,

C₃- S intends to deceive H by stating that *p*.

Although this had been considered as the universal definition of lying for ages, it has come to be seen as debatable by some contemporary philosophers—specifically C₃ (hereafter trad-C₃) of which emphasis is firmly on the speaker's 'intention to deceive.' Yet contemporary analytic philosophers make no attempt to consider the social norms of lying. Thus, as in many other analytic philosophers, I also claim that trad-C₃ should be open to criticism; however, by being different from others, I claim that it is problematic in that it ignores the participants' social awareness in daily life. Lying is an intersubjective action between the speaker's and the hearer's social existence, so, by taking the hearer's social existence into account, lying should be redefined as on par with the binary—speaker and hearer—relationship.

To better understand how this relationship operates on lying, it might be helpful to briefly mention Alfred Schutz's phenomenological sociology which chiefly elaborates on the significance of social action in terms of the commonsense experience regarded as an everyday world. Consistent with it, roughly, the typical human being—the speaker or the hearer, in this case—is integrated into her social world which is taken for granted. The significance of why such a perspective is chosen resides in his philosophy, which links the human, *qua* social being, and the taken-for-granted social actions together hint that the traditional or any contemporary model of lying is incompatible with the commonsensical social world and also what a new model should be. Since lying is an intersubjective action that depends on the speaker-hearer relationship, a mere hearer-insensitive analytic perspective may be insufficient to define what lying is and thus socio-analytic model should be taken into consideration to underline the dependency of lying on the hearer's social existence. To formulate a new model, I should associate analytic philosophy with Schutzian phenomenological sociology to make lying more rational, commonsensical, and pragmatic as a speaker-hearer-sensitive social action. Given this situation, it is hardly surprising that the new model entirely agrees with the statement and false belief conditions, namely that it is in keeping with C₁ and C₂. As a result, in this paper, special attention is mainly given to the hearer-sensitive party, and the significance of the new-C₃ lies in having the intention to induce to believe that *p*. More precisely,

C₁ - S states that *p* to H,

C₂ - S believes that *p* is false,

new-C₃ - S must intend that H be induced to believe that *p*.

Central to the definition of the new-C₃ is the term, *induce to believe*: unlike *intend to deceive*, the new component alludes to trying to be successful, namely that S must try as hard as possible to succeed in leading H to believe that *p*. Put another way, S must *try as hard as possible to succeed in deceiving* H by stating that *p*. One idiosyncrasy with this condition, which is further considered, however, is that it reduces the new model's scope but makes the new model more rational and commonsensical.

Section 2 attempts to provide a brief introduction to a few preliminary Schutzian terms and perspectives. It does not carry out an in-depth analysis of it to avoid digressing from the main topic. It is presented concisely so that the reader can get a picture of the new model.

Section 3 associates Schutzian terms with the new-C₃ and describes the logic used in this new model. Furthermore, in combination with concrete examples as well as in comparison with the traditional model, the new model becomes more rational and daily-life-friendly. At the end, it introduces a new term, *anti-social bullshit*, which encompasses the ruled-out cases of the new model.

Section 4 provides an overview of Chisholm and Feehan's (hereafter abbreviated C&F) model of lying in their notable paper, *The Intent to Deceive*. It then goes on to discuss and compare with the new model. A little consideration will show that C&F also imply that the lie should be successful to deceive someone; however, they do not take social norms and existence into account. Despite a resemblance, thus, the salient discrepancy in the method cannot be ignored.

Section 5 introduces the non-deceptive lie definitions of Thomas Carson and Jennifer Lackey with the analysing of, from the viewpoint of the new model, the two prominent objections that have been raised against the traditional model at large by them. This section is divided into three subsections: 'inveterate liar', which is deceptive and not a counterexample to the traditional, but rather to the C&F and the new model; 'bald-faced lies' and 'coercion lies,' on the other hand, as non-deceptive cases, are examined to stress the distinction between the new model and its contemporary rivals.

2. Socio-Phenomenological Remarks for the New-C₃

Phenomenology has influenced sociology in many ways and some social scientists have approached phenomenology as an alternative perspective to understand social

processes. Alfred Schutz is one of the most important key figures, who was particularly interested in Edmund Husserl's philosophy and focused on analysing the structure of daily life interactions in the social world.

The purpose of this section is briefly to introduce some ideas of Alfred Schutz which will be helpful to comprehend the new model of lying. Again, I do not intend to carry out an in-depth analysis of his philosophical works, what I only present is the essential points of some of his terminologies, namely, commonsense world, typification, Thou-orientation as well as We-relationship, to understand the sense underlying the new model. These underlie the new model of lying which focuses on the intersubjectivity between the speaker's and the hearer's social existence in our everyday life. Analysing these concepts allows us to understand the new model without hindrance.

2.1. Commonsense World and Typification

The *commonsense world*, variably as 'world of daily life' or 'every-day world,' is the domain of social interactions where people come into contact and have a relationship with one another by being *taken for granted*. The taken-for-granted (*das Fraglos-gegeben*), wrote Schutz (1967, 74), "is always that particular level of experience which presents itself as not in need of further analysis." In short, the taken-for-granted commonsense world is a kind of immediate experience that is familiar to all of us. For instance, the existence of other people, meaningful communication and collaboration with others, socially accepted rules and principles for everyday life, etc. are all in our commonsense world and taken-for-granted.

We are entirely aware that this world already existed and it was understood, interpreted, and experienced by others before us. Now, we, with our contemporaries and consociates¹, are experiencing and interpreting it with the help of the stock of our previous experiences which is called *knowledge at hand*. It encompasses all of the knowledge coming from our world and the others, i.e. parents, teachers, friends, members of society, etc. and it functions as a reference to us for our daily life. The constituents of the stock of knowledge at hand are not individual; all of them are categorized under the related classes, which are called *typification*. For better understanding, as Schutz (1962, 8) notes,

The outer world is not experienced as an arrangement of individual unique objects, dispersed in space and time, but as 'mountains,' 'trees,' 'animals,'

¹ Consociates are the people in face-to-face situations directly and simultaneously experienced. Contemporaries, on the other hand, are the people not directly in contact with.

'fellowmen.' I may have never seen an Irish setter but if I see one, I know that it is an animal and in particular a dog, showing all the familiar features and the typical behavior of a dog and not, say of a cat. I may reasonably ask: "What kind of dog is this?"

In other words, we possess a sort of fundamental knowledge concerning our world. For instance, in the previous example, I can identify that it is a *typical* dog even though I cannot identify its genus. The chief source of this knowledge is from all of my previous experiences of seeing and typifying dogs.

The rest of the paper, however, only concentrates on the socially derived—or accepted—stock of knowledge at hand, namely, typical anticipations, characters, behaviours, etc. that have been piling up from our early life without our control. As Maurice Natanson (Schutz 1962, XXIX) states,

This 'stockpiling' of typifications is endemic to common-sense life. From childhood on, the individual continues to amass a vast number of 'recipes' which then serve as techniques for understanding or at least controlling aspects of his experience. The thousands of concrete problematic situations that arise in the course of daily affairs and have to be handled in some form are perceived and even initially formulated in terms of the individual's stock of knowledge at hand. The typifications which comprise the stock of knowledge are generated out of a social structure.

One of the sub-argument of that paper is that lying, as a social action, has also typifications in our stock of knowledge at hand. More precisely, we lie *typically* for any typical problematic situations that have been experienced in the past. Typifications that have been formed from childhood are applied to the current situation and both past and current experiences are relevant to the formulation of plans of action for the future. In the light of this information, this paper attempts to show that lying, as all of our intersubjective social actions, postulates typifications grounded in the commonsense world, or as Schutz (1962, 20) puts it, "the Husserlian idealization, 'I-can-do-it-again,' that is the assumption that I may under typically similar circumstances act in the typically similar way that I did before in order to bring about a typically similar state of affairs." More exactly, you, as a *typical* person, tell a *typical* lie in *typical* contexts for *typical* Others.

Although you can lie via pigeon post, mail, message, phone, etc., this paper analyses only the most complex, intersubjective, and familiar method, namely the face-to-face lie. It does not mean that one cannot employ the new model of lying to the non-face-to-face methods; it is rational for any kind of method, but the analysis will run on the face-to-face method. Thus, the following subsection gives some preliminary remarks on Schutz's approach to face-to-face intersubjectivity.

2.2. The Face-to-Face Situation and the We-Relationship

If we take intersubjectivity into account for lying, we should focus on *face-to-face* relationships in the social world. When the speaker encounters her hearer face-to-face, she shares a spatio-temporal domain within both of them reach in which she interprets the Other's acts. According to Schutz (1967, 163), "the face-to-face situation presupposes ... an actual simultaneity with each other of two separate streams of consciousness." This is the thesis that for the act of lying, the hearer is a conscious being as same as the speaker, both of whom are aware of one another in a psychophysical sense, and of the context where they experience together.

When the speaker lies to her hearer in the face-to-face context, she is conscious of the hearer and, thus, her conscious, attention, etc. oriented toward the hearer, and this attitude is called *Thou-orientation*. The Thou-orientation can be either one-sided or reciprocal. One-sided Thou-orientation is that only one of the parties is aware of the Other. For the new model of lying, however, I only focus on the reciprocal Thou-orientation, namely, the speaker and the hearer are mutually aware of one another, that is, the speaker is Thou-oriented toward the hearer, and at the same time, the hearer is also Thou-oriented toward the speaker. In that kind of relationship where the partners are aware of each other and "sympathetically participate in each other's lives for however short a time we shall call the *pure We-relationship*" (Schutz 1967, 164). The pure We-relationship, according to Schutz (1967, 168), "involves our awareness of each other's presence and also the knowledge of each that the Other is aware of him." The pure We-relationship, in other words, is merely the reciprocal form of the Thou-orientation. Schutz separated the pure We-relationship from simply 'the We-relationship' that is "a *close attentive awareness* of the Other, wherein the two interact and share their experiences with each other" (Cox 1973, 122, italics mine). This simultaneous, reciprocal as well as close attentive awareness places the speaker and the hearer in a We-relationship in telling a lie.

In conclusion, the abovementioned terms and remarks are sufficient to comprehend the new model. First of all, typification gives us a clue about how to build commonsense or socially accepted actions. In the commonsense world, we always choose one of the relevant types of lies for relevant context. Metaphorically, it may be said that human is by nature not only social animal but also a socially-accepted-behaved animal. Secondly, if the speaker-hearer mutually interacts with one another with the social awareness, that reciprocal intersubjectivity places them in the We-relationship. Once it occurs, the speaker is subject to a shared commonsense world, social norms, rules, etc. In short, when

you lie, your statement is based on a commonly-typified lie and you are in a sort of social coexistence with your hearer.

3. Definition of the New Model of Lying

Thus far, this paper has focused on a brief explanation of some Schutzian concepts from phenomenological sociology to gain a clear idea of the new model. We now move on to consider the comparison with the trad-C₃ with a useful example making its analysis easier and then to turn to analyse more clearly what has been asserted so far.

3.1. Difference between new-C₃ and trad-C₃

Suppose that you overslept and missed an important meeting. When this happens and you interact with your supervisor, you anticipate that, for instance, she notices your absence, she wants to know where you have been, she might accept your excuse, etc. In more general terms, you anticipate that she behaves typically in line with the commonsense world and its typifications; as you and many others behave—act, lie, etc.—under typically similar circumstances. In the face of such a situation, therefore, you state, to your supervisor, a plausible-to-believe, *typical* lie such as “I felt very ill” or “I missed the bus” that anyone might believe in our commonsense world. As Schutz (1962, 27) states, “If I, if we, if ‘anybody belonging to us’ found himself in typically similar circumstances he would act in a similar way.” Unlike the trad-C₃ for which only *the intention to deceive* is rational, you would not, in all likelihood, dare to state that “a hippopotamus held up the traffic” or “I was abducted by aliens” as an excuse for missing the meeting. You have to opt for *the most successful-to-be* or *the most commonsensical* anecdote to induce your hearer to believe that it can be true. To understand how we specifically concoct *p* to deceive the hearer, special attention should be given to the commonsense typifications for relevant context. Consequently, as already mentioned, redoing the ‘similar’ actions presuppose an *I-can-do-it-again* idealization: the speaker experienced the sort of similar situations in her past², hence, she acted sort of a similar way in the current situation.

To come back to the example, it is important to notice that the process of stating that *p* can be divided into two subprocesses, i.e. *thinking on* and *opting for*.

² Even if she has not experienced the relevant situation, she still might possess a relevant typification from the observations of the other people or her faculty of inference. For instance, I have never been pulled over by any traffic police but if it occurs and I have to tell a lie in this first-time situation, I still refer to close-relevant typifications of lies in my stock of knowledge at hand to be as commonsensical as possible.

Strictly speaking, the social context for lying might be called a *disjunctive syllogistic* situation: once you decide to lie, you may *think on* at least two different statements, say p , and q , and then you *opt for* the most commonsensical alternative for the hearer; with logical notation, $p \vee q, \sim q, \therefore p$. Alternatively, as Dewey (1930, 190) asserted, “deliberation is a dramatic rehearsal (in imagination) of various competing possible lines of action.” This phenomenon can best be illuminated by an analysis of the previous example: you overslept and missed an important meeting and have to concoct a valid excuse to protect yourself from the negative consequences of being tardy. You may concoct a considerable amount of statements as an excuse, i.e. “my car broke down,” “my cat ran away,” “I felt very ill,” “my house burned down,” etc. For the sake of simplicity, let us say you reduce your set of excuses into two alternatives: “broken-down car” (p) and “burned-down house” (q). You reasonably think on that p is more commonsensical than q for the hearer in question since if you state that q , then you have to feign that you lost your house, all your belongings, etc. and certainly you must seem upset about the incident. What is more, it would be quite absurd to come to the office on such a hard morning. In the end, stating that q would be costly as well as quite unmaintainable and thus, a moment’s reflection is sufficient for you to realize that q is a pathetic excuse for the hearer and to opt for p as a plausible excuse.

This scenario is useful for any typical supervisor whom you do not know anything about. More accurately, the situation is identified as context-sensitive but hearer-insensitive; it was assumed a typical hearer in a missed-meeting context. Putting forward the new-C₃, however, gives rise to a further problem that resides in how you can be sure that the hearer can be induced to believe that p , that is, how you could try as hard as possible to succeed in deceiving. There has been no such a problem with the trad-C₃ for it has generally been concerned with the speaker only and mostly ignored the hearer’s social existence. In definitional terms, ‘to intend to deceive’ is sufficient for it. To get the discussion on a concrete footing, let us consider that S with the intention of deceiving states that p to H₁, H₂, H₃, etc. where p is false. The traditional model implies that p is a lie for anyone—H₁, H₂, H₃, etc.; for it only depends on S; the new model, on the other hand, implies that p does not only depend on S but also on H—H₁, H₂, H₃, etc., namely that p might be a lie for H₁ but not for H₂ owing to the hearer-sensitive factor. To provide a clear picture, the following two cases will make the set of hearers narrower by transcending from the typical supervisor to the *subtypical* ones and they show how p might be a lie for H₁ but not for H₂.

Take the previous case as an example again: you overslept and missed your meeting.

Case-1: Your boss (H_g) is a firm believer in supernatural beings such as ghosts, evil spirits, etc. Knowing your boss' superstitious beliefs, you might state that p_g , "my house was haunted by ghosts in this morning and I dealt with them." as a lie for the 'ghost-believer' Other since you are aware that she has a high potential to believe that p_g .

Case-2: Your boss (H_a) is a rational person with a scientific perspective and she is an avid animal lover. Knowing all that you most probably would not say 'ghost anecdote' as a lie; even if it provides all the necessary conditions of the traditional model, you by no means dare to say that for any rational person. Instead, you might perhaps say that p_a , "my cat looked a bit in low spirits this morning and I dealt with her." as a lie for the 'avid-animal-lover' Other since you are aware that she always gives high importance to animals.

In the above cases, the attentive reader will notice that the hearers— H_g and H_a are not *typical* supervisors or persons for the speaker this time. Knowing something concerning the hearer reduces the scope of her typicality: 'avid-animal-lover Other' and 'ghost-believer Other' for the cases in question. If the speaker is acquainted with the hearer, then she does not have to categorize the hearer as anonymous in the broadest sense. If we assume, on the other hand, that the speaker has no knowledge concerning the hearers, then she *commonsensically* assumes that the hearer is a *typical person* and she would not prefer to state that neither p_g nor p_a . In other words, p_g and p_a are not to be opted for by a typical hearer(H_t) and thus, any typical speaker would not prefer them as an instance of lying. Technically speaking, wrote Schutz (1962, 18),

The more anonymous the typifying construct is, the more detached is it from the uniqueness of the individual fellowman involved ... In complete anonymization the individuals are supposed to be interchangeable and the course-of-action type refers to the behavior of 'whomsoever' acting in the way defined as typical by the construct.

In conclusion, in comparison with the traditional model, however, the two cases considered, p_g and p_a can both be lies for H_g , H_a , and H_t in trad-C₃ since it does not give particular importance to the hearer's social existence. In other words, trad-C₃ is a 'whomsoever' action. In the new-C₃, on the other hand, p_g can be a lie for H_g but not H_a or H_t , and p_a can be a lie for H_a but not H_g or H_t . Needless to say, neither of them can be a lie for H_t . An analogy can be drawn here: if the traditional—or any hearer-insensitive— model of lying is a factory-product, then the new model of lying is tailor-made. As a result, little thought is required to see that mere trad-C₃ is too broad, unpragmatic, and uncommonsensical for a social human being who always intends to flee the negative consequences of the context in which she has to lie.

3.2. Analysis of the New-C₃

Having discussed how to construct the lie *for the Other*, we now move on to explain that once you lie to the hearer, you share the same commonsense world that both of you are part of *intersubjectively*. This intersubjectivity, however, is not the rejection of subjectivity: you *decide to lie* subjectively, but you decide *what you state* as a lie intersubjectively. It is intersubjective because you live as an individual person among other people, “bound to them through common influence and work, understanding others and being understood by them” (Schutz 1962, XXX). Hence, even though the individual defines her world from her own perspective, she is nevertheless “a social being, rooted in an intersubjective reality” (Schutz 1962, XXX). More precisely, although we all are different individually, when we live together and constitute a social world, then we all are constituents of society and lose our individuality under intersubjectivity necessitating commonsense awareness by being taken for granted. That is, commonsenseness comprehends and conducts its relations with the Other without recognizing it. And the new-C₃ renders that unrecognized yet epistemically given part of lying, which is deeply rooted in our daily life. As commonsense people, *qua* deceivers and deceives,

We are all born into the same world, grow up as children guided by parents and other adults, learn a language, come into contact with others, receive an education, move into some phase of the business of life, and go through the infinitely detailed catalogue of human activity: we play, love, create, suffer, and die. But throughout all of the routine elements and forms of existence, we simply assume, presuppose, and take it for granted that the daily world in which all of these activities go on is there... Thus, the essential foundation of mundane existence remains unrecognized by commonsense men whose lives are nevertheless structured by and built upon the matrix of daily life (Schutz 1962, XXV).

Therefore, the new model of lying is to render the socially taken-for-granted yet unrecognized constituent of the definition of lying that was ignored by the traditional as well as the contemporary rivals. If the speaker genuinely wishes to lie to the hearer *in order to* get what she expects, she has to imagine both herself and the hearer as a typical person in the We-relationship under the awareness of commonsenseness and make herself a typical liar for the relevant context. The speaker understands herself, the context, the Other, etc. to the extent permitted by the stock of knowledge and previous experiences. Once the speaker places herself in the We-relationship, she should not determine her actions independently of the commonsense world; all of her intersubjective actions, as already stated, are determined by the social existence of the Other and commonsenseness, namely,

she is subject to be typified by the social world. As Schutz (1962, 11) emphasized, “in common-sense thinking if we take into account that this world is not my private world but an intersubjective one and that, therefore, my knowledge of it is not my private affair but from the outset intersubjective or socialized.” Accordingly, being a typical person implies that the speaker or the hearer approximately guesses what the Other states, how the Other behaves, etc. in line with the commonsense world. And both represent similar urges towards commonsense, namely, the evasion of uncommon and the adaption of dominant of related typification. The phrase that *the hearer be induced to believe the false statement* in the new-C₃ resides in that *commonsenseness*: once you lie to the Other, you are aware of the hearer’s state of mind, context, etc., and you state a lie in a most *commonsensical for the Other* by avoiding any unusual, atypical or implausible one. Thus, *lie to the other* should be transformed into *lie for the Other* in our everyday life.

In the missed-meeting context in which both participants are actively engaged with one another, telling a lie would be like that: first, you want to get rid of the negative consequences of being tardy, so you decide that your excuse has to be persuasive. And, as a typical person—you have been late and found an excuse couple of times in your past—therefore, you have a stock of knowledge at hand for that typical missed-important-thing. Second, even if you do not know anything about the supervisor, you, *qua* rational person, assume that she is a typical supervisor who expects to hear something typical as an excuse. The collection of all assumptions, typifications, and knowledge is based on commonsense awareness which has been built from your childhood. As constituents of the same commonsense world, you and the supervisor merged into a single and typical commonsense world citizen. As Cox (1973, 123) states,

My experience of the other weaves a network of interconnecting meanings, formed in presence to me and which I follow as it builds. The reality of the other overlaps my reality, and the two become merged into a single co-subjective here and now. I experience the other’s experiences, though not directly. I am aware of what he is thinking, that he believes this or that, and that he thinks such and such is true of me.

Put another way, once you lie to your hearer, you put yourself into her shoes and analyse whether your lie is commonsensical or not for your hearer. Hence, there is now no doubt that if you intend that your hearer be induced to believe something false, you have to opt for the most successful-to-be *p* for the hearer in question. Yet, it is crucial to note that successful-to-be does not have to entail that the speaker *must be successful* to induce to believe that is false. She lies even if she is

unsuccessful as well; the key point is that the speaker intends to try *as hard as possible* to obtain success in inducing what to state.

Before finishing this subsection, I now want to turn an analytic eye to the discrepancy between *to deceive* and *to induce*. The rough definition of deceiving is that you deliberately cause someone to believe something you know to be false by changing her epistemic status. The new-C₃, however, does not consist of *deceiving*; instead, it consists of *inducing*, here, which refers to *succeeding in* causing someone to do something. Semantically speaking, *to induce to believe that is false* is a subset of *to deceive*.

Inasmuch as *to induce to believe* comprises both the speaker and the hearer, the new-C₃ is, thence, a subset of trad-C₃. More exactly, all lie-N for the new-C₃ is also a lie for the trad-C₃, whereas all lie-T for the trad-C₃ is not a lie for the new-C₃; in technical notation, $N \subseteq T^3$. It is self-evident that the scope of the new model is narrower than the scope of the traditional rival and it is no coincidence that this narrowness will be thought of as a caveat for the new model. The scope of the new model is narrower than of the traditional rival, certainly; but this narrowness leads to the new one being more commonsensical and rational, as has been exemplified. Owing to its social characteristics, the definition of lying should treat both the speaker and the hearer as being of equal importance at the expense of being narrower. The other instances belonging to the scope of the traditional model—or of some contemporary models—but not of the new model will be called *anti-social bullshit* and it is to this we now turn.

3.3. Out of the Scope of New-C₃: *Anti-Social Bullshit*

In his essay *On Bullshit*, Harry Frankfurt (2005) states that the distinctive feature of bullshit is that the bullshitter is indifferent toward the truth or falsity of what she says. According to him, “her statement is grounded neither in a belief that it is true nor, as a lie must be, in a belief that it is not true. It is just this lack of connection to a concern with truth— this indifference to how things really are—that I regard as of the essence of bullshit” (Frankfurt 2005, 33–34). In other words, the distinction between a liar and a bullshitter is the fact that a liar must concern about whether what she says is true or false, whilst the bullshitter need not: she just says things without regard to their truth value. Technically speaking, the liar must employ C₂ to her statement.

There are, undoubtedly, some objections to Frankfurt’s definition of bullshit. However, my purpose is not to discuss what bullshit should be. Contrariwise, it is

³ Blackboard bold typeface denotes the all of the entries of the lie sets of relevant definitions.

to construct a new term to coin the instances which fall within the scope of trad-C₃ but beyond the scope of new-C₃ by adopting a similar perspective of Frankfurt's bullshit.

As argued above that the distinguishing characteristic of bullshit is the lack of concern with the truth which is the key difference between a lie and bullshit. The new term, anti-social bullshit, however, is essentially different from both of them. The anti-social bullshit, in contrast to classical bullshit, does care for the truth or falsity of the statement, namely, as if lying, it concerns C₂. The anti-social bullshitter, as a result, cares about whether what she says is true or false. The distinctive feature of anti-social bullshit is the fact that *it does not care for the hearer's state of mind*. Dissecting the term, the 'anti-social' part denotes that the term is *against the social norms* and commonsense world; the 'bullshit' part, on the other hand, denotes that the term *lacks concern* with the hearer's state of mind. Properly speaking, anti-social bullshit is a statement that can be replied to as "how can you expect me to believe that?" in daily life. In the extreme cases, the hearer may feel treated like dirt owing to the lacking of concern for herself. Whereas the speaker is aware of the hearer psycho-physically, she ignores her state of mind. Thus, anti-social bullshit is the taking no notice of the hearer by stating a false statement.

Related to deception, there is also a distinction between traditional lies, Frankfurt's bullshit and anti-social bullshit that is worth mentioning. As Frankfurt (2005, 54) suggests, "the bullshitter may not deceive us, or even intend to do so, either about the facts or about what he takes the facts to be." From this point, Frankfurt's bullshit resembles the traditional lie with respect to intentional deception. Conversely, anti-social bullshit *does not necessitate* an intentional deception condition—namely trad-C₃ or new-C₃— like some forms of the contemporary definition of lies which will be introduced in section 5; instead, it is characterized by the fact that *the speaker is fully aware that the hearer regards her as a dishonest person*. As a result, you, as an anti-social bullshitter, are regarded as a dishonest person from the perspective of your hearer and you know that, even if you do not intend to deceive her.

The attentive reader might ask why it is called 'bullshit' despite the contrast to classical bullshit. The reason why that word is chosen is to emphasize the 'lacking of concern' factor. If it could have been named as an 'anti-social lie,' it may cause confusion, since I would have asserted that anti-social lies are not lies. It should be further stressed that the principal characteristic of the anti-social bullshit, 'lacking concern with' the hearer would be lost. For that reason, anti-

social bullshit, as a term, is appropriate to emphasize ‘lack of concern’ and ‘against the society’ at the same time.

With the introduction of the new model of lying, this sort of neologism, anti-social bullshit, is indispensable to be defined in order to incorporate the ruled-out cases of the traditional or contemporary definition of lying. By the virtue of this neologism, the narrow scope of the new model does not pose an analytic problem since the anti-social bullshit encompasses, as already noted, the excluded instances which fall within the trad-C₃ scope but beyond of the new-C₃ scope. Nevertheless, the new model is not, fortunately, the only model having a relatively narrow scope in the literature. The following section moves on to describe in detail the C&F model of lying.

4. Chisholm and Feehan’s Model of Lying

As stated by C&F (1977, 149), if S lies to H, there should be two conditions:

cf-C₁- S says *p* to H for the purpose of causing H to believe that *p*;

cf-C₂- S believes that *p* is not true or she believes it to be false.

On the authority of C&F (1977, 149), “... in telling the lie, the liar ‘gives an indication that he is expressing his own opinion.’ And he does this in a special way—by getting his victim to place his faith in him. The sense of ‘say,’ therefore, in which the liar may be said to ‘intend to say what is false,’ is that of ‘to assert’.” Here, ‘to assert’ means ‘to be taken seriously;’ in the traditional model, stating is not asserting. If one states something as a joke, for example, then the statement is not an assertion. Consequently, pursuant to C&F, the *seriousness* that assertion involves resides in this fact: “the concept of assertion is essentially normative. We can explicate it only by reference to justification. And the justification in question is epistemic, the type of justification that is implied by knowledge and evidence”(Chisholm & Feehan 1977, 152). More precisely, once S asserts something to H, then S believes H to be justified in assuming not only cf-C₁ that S believes that *p*, but also cf-C₂ that she intends to cause H to believe that S believes that *p*. In the opinion of C&F, the point of asserting *p* is that of causing justified belief in the propositions cf-C₁ that the speaker accepts the assertion and cf-C₂ that she intends to convey her acceptance of the assertion. Strictly speaking, ‘asserting a proposition’ is:

S asserts *p* to H =*df* S states *p* to H and does so under conditions which, he believes, justify H in believing that he, S, not only accepts *p* but also intends to contribute causally to H’s believing that he, S, accepts *p*. (Chisholm & Feehan 1977, 152)

And the definition of lying is,

S lies to H \Rightarrow There is a proposition p such that (i) either S believes that p is not true or S believes that p is false and (ii) S asserts p to H. (Chisholm and Feehan 1977, 152)

To understand better what they mean, two well-known cases, which have been proposed by Augustine (1952, 57), may be raised:

Case-1: We have a person who knows or thinks that he is speaking falsely, yet speaks in this way without the intention of deceiving. Such would be the case of a man who, knowing that a certain road is besieged by bandits and fearing that a friend for whose safety he is concerned will take that road, tells that friend that there are no bandits there. He makes this assertion, realizing that his friend does not trust him, and, because of the statement to the contrary by the person, in whom he has no faith, will therefore believe that the bandits are there and will not go by the road.

C&F propose that S does not lie to H; even though S believes that the statement is false and acts with the intention of deceiving, S does not assert a proposition p , because he does not believe that the conditions under which he states p are conditions that justify H in believing that S accepts p .

In the same way that the new model also claims that S does not lie to H. S states that "there are no bandits on the road" by being aware that H cannot be induced to believe that statement. Having the We-relationship with the hearer, S is entirely aware that H *by no means* believes that p ; therefore p is not counted as a lie from the viewpoint of the new model.

Case-2: There is the case of the person who, knowing or thinking what he says is true, nevertheless says it to deceive. This would happen if the man mentioned above were to tell his mistrustful acquaintance that there are bandits on the road, knowing that they are there and telling it so that his hearer, because of his distrust of the speaker, may proceed to take that road and so fall into the hands of the bandits. (Augustine 1952, 57)

C&F(1977) claim that S does not lie to H. Even though S believes that there are bandits on the road, S intends to cause H to believe that S believes that there are no bandits on the road. But S does not believe the statement to be false. Hence his assertion of that statement is not a lie.

Similarly, the new model argues that S does not lie to H here either. Although S intends that H be induced to believe his statement, S does not believe the statement to be false. This example has shown the speaker-sensitive party of the new model. That is, whereas new-C₃ has been satisfied, C₂ has not been. The importance of this example lies in the fact that it makes us recall that even if you could have induced your hearer to believe that something is false, it is still not a lie as long as you do not believe it to be false.

It goes without saying that both C&F and the new model have a characteristic aspect in common:⁴ you lie only if you expect that you will be successful in deceiving the hearer with your false statement. In spite of close resemblance, the ground of the new model deviates considerably from the C&F model. By contrast with C&F, this paper's thesis is based on the human being as a social being who relates with one another intersubjectively in the commonsense world in which lying occurs and, unlike being seriousness, the new model asserts that lying, as an intersubjective action, should be based on the awareness of the commonsense world and of its typification. In the following paragraph, this disparity is conveniently exemplified.

Here is a substantial discrepancy between C&F and the new model is that the hearer is an animal being able to understand some basic words and acting based on these words: suppose that I say to my cat "Look out! There is a bird over there!" by knowing that she completely understands what I mean. Do I lie to her? According to C&F (1977), if I make my statement to cause her to believe that there is a bird over there, then I lie to her. As per the new model, on the other hand, I do not and cannot lie to her. The reason is that it cannot be treated as a social being to my cat. More accurately, in the new model of lying, the hearer must be a conscious human being. What is more, aside from animals, the new model also asserts, as opposed to C&F (1977), that you cannot lie in any case to a polygraph or artificial intelligence since, needless to say, there is no such other-party social being.

In conclusion, it is hardly an exaggeration to say that the new model is the narrowest version among all of the definitions of lying. For that reason, there are good grounds for doubting that it will be criticized owing to ruling out some notable instances of lying. The following section, thus, examines some types of lying from the viewpoint of the new model.

5. Objections to the New Model of Lying

While the new model is more rational and commonsensical than its rivals for our social world, it is self-evident that it will be open to criticism. For the time being, however, I will concentrate of the objections raised to traditional and C&F lying. There are, admittedly, many objections and objectors, yet to keep this paper

⁴ Fried (1978, 55) also states that "A person lies when he asserts a proposition he believes to be false... Their [Chisholm and Feehan's] central emphasis on assertion is identical to mine, which is not necessarily remarkable given the fact that the authors are heavily influenced, as am I, by Augustine's and Kant's discussion of lying. We differ principally in that they find a way to treat as not lying at all some cases which seem to me to be cases of justified lying. But my reasons and theirs are close and the difference is largely one of form"

concise and focused, I mainly adhere to the objections of Thomas Carson and Jennifer Lackey. Before I present the instances, it would be useful to introduce their lie definitions which are distinct from the traditional view.

As Carson (2010, 30) defined lying:

A person S tells a lie to another person S1 iff: 1. S makes a false statement X to S1, 2. S believes that X is false or probably false (or, alternatively, S does not believe that X is true), 3. S states X in a context in which S thereby warrants the truth of X to S1, and 4. S does not take herself to be not warranting the truth of what she says to S1.

“To lie, on my view, is to invite others to trust and rely on what one says by warranting its truth, but, at the same time, to betray that trust by making false statements that one does not believe” (Carson 2010, 34). In other words, Carson thinks that the liar betrays trust when she lies. His main argument is that when you lie, you betray trust and lying does not be with intending to deceive.

Lackey (2013) also argues that lying does not involve an intention ‘to deceive.’ Instead, it involves an intention ‘to be deceptive.’ She proposes that there is a distinction between the intention ‘to deceive’ and ‘to be deceptive.’ More precisely,

Deceiving: A deceives B with respect to whether *p* if and only if A aims to bring about a false belief in B regarding whether *p*.

Being deceptive: A is deceptive to B with respect to whether *p* if A aims to conceal information from B regarding whether *p* (Lackey 2013, 241).

According to Lackey (2013, 237), therefore, the three conditions of lying are, “(i) A states that *p* to B, (ii) A believes that *p* is false and (iii) A intends to be deceptive to B in stating that *p*.”

The following subsections present the three major objection-to-be cases, i.e. ‘inveterate liar’, ‘bald-faced lies’, and ‘coercion lies.’ Of these, the bald-faced lie is “an undisguised lie, one where a speaker states that *p* where she believes that *p* is false and it is common knowledge that what is being stated does not reflect what the speaker actually believes” (Lackey 2013, 237-238). Another objection is called ‘coercion lies’ which occurs “when a speaker believes that *p* is false, states that *p*, and does so, not with the intention to deceive, but because she is coerced or frightened into doing so” (Lackey 2013, 239).

5.1. Case-1: Inveterate Liar

Carson (2006, 292) argues that

Chisholm and Feehan’s definition has the very odd and unacceptable result that a

person who is notoriously dishonest couldn't tell lies to those he knows distrust him. Their definition implies that it is self-contradictory to say that I lie when I know that others know that I am lying (and thus are not justified in believing that I believe (accept) what I say).

As seen above, Carson claims that the C&F model is problematic since it is self-contradictory.⁵

The new model is eager to share this objection fully with C&F. It is certain that it is not an instance of lying from the viewpoint of the new model either since by ignoring the hearer's state of mind, the action that the speaker does cannot be called even a real conversation or speech, much less lying; or it might be a typical example of anti-social bullshit. More precisely, if H knows that S is an inveterate liar and S also has this knowledge—H knows that S is an inveterate liar—, then all of S's attempts to induce H to believe that something false will be in vain. As a famous for being an inveterate liar, nobody takes her opinion, testimony, etc. seriously. Hence it is not an overstatement to say that being known as an inveterate liar is equal to being *socially inaudible*.

To put this into perspective, imagine an infant talking barely, a parrot mimicking human speech, and a person, *qua* hearer, all in a room. And suppose that the infant and the parrot start to use profanity towards the hearer. However, if the hearer is a sensible adult, then she would not take them seriously for she is acutely aware that they do not and cannot intend to intimidate, offend or otherwise give rise to emotional harm. Technically speaking, contrary to any adult human being, the infant or the parrot lack intention to offend as well as commonsense awareness; and this is not surprising, considering the lack of awareness with not acting *for the Other*.

The case of the inveterate liar greatly resembles the abovementioned case, the swearer infant—or parrot in that both of their hearers are aware that the speaker lacks concern with the awareness of the Other's states of mind. As a result, unlike the traditional rival, the new model claims that even if she has intended deception, an inveterate liar cannot lie to her hearer. It goes without saying that it seems quite absurd that an inveterate liar cannot lie. A little consideration, however, will show that the word 'liar'—of inveterate liar— refers to the traditional model since the ground of English vocabulary is based on the most

⁵ As Fallis (2009, 46) puts it, "...when someone who is known to be an inveterate liar makes a statement, there is no reason for anyone to believe that she believes that the statement is true. So, if she knows that she is known to be an inveterate liar, the conditions of CFL will not be satisfied. But presumably, someone who is known to be an inveterate liar can still lie. Thus, CFL is still too narrow" (CFL stands for Chisholm Feehan Lying).

prevailing—traditional—definition of lying. The phrase ‘cannot lie,’ on the other hand, refers to the new model of lying. In short, an inveterate liar can only lie in the traditional sense. From the perspective of the new model, the inveterate liar just bullshits anti-socially.

Case-2: Bald-faced Lies

A bald-faced lie is when the speaker states that p where p is false and both the speaker and hearer are aware that the Other knows this. Let us cite an example:

A student is caught flagrantly cheating on an exam for the fourth time this term, all of the conclusive evidence for which is passed on to the dean of Academic Affairs. Both the student and the dean know that he cheated on the exam, and they each know that the other knows this, but the student is also aware of the fact that the dean punishes students for academic dishonesty only when there is a confession. Given this, when the student is called to the dean’s office, he states, “I did not cheat on the exam” (Lackey 2013, 238).

It deserves mention that this case might evoke the case of the inveterate liar; however, by denying the guilty of cheating, the student is here not trying to deceive the dean into thinking otherwise, rather he protects himself from sanction. Hence, the student’s false statement does not satisfy the traditional model, pursuant to which lying is qualified by the intent to deceive.

As argued by Carson (2006) and Lackey (2013), the student is clearly lying. Carson (2006, 295) asserts that if the student “plays it straight and looks grave and serious, then his statements are warranted to be true and count as lies according to my definition.” Lackey (2013, 237), on the other hand, asserts, as already stated, that “A intends to be deceptive to B in stating that p .” As specified by her, although the student does not intend to bring about any false beliefs in the dean, he is clearly lying with the intention of being deceptive. Even though Carson and Lackey are distinct from one another, both agree that bald-faced lies are an example of lying.⁶

As the reader may easily guess that bald-faced lies are not an instance of lying for the new-C₃ since it lacks intentional deception. From this point, she is totally right about that. However, I would like to analyse why it lacks intentional

⁶ Some philosophers argue that bald-faced lies are not lies. For example, Meibauer (2014, 140) argues that bald-faced lies are not lies because the bald-faced liar does not really present p as true in the context since he lets shine through that p is false. He would not feel committed to the truth of p , and he would not be ready to provide further evidence. Keiser (2016, 464) also thinks that bald-faced lies are not genuine instances of lying because they are not genuine instances of assertion.

deception from the perspective of the new model. Put differently, I quite the contrary argue that the student is not lying in conformity with the new model insofar as he does not concern with the dean's state of mind and does not try as hard as possible to succeed in deceiving the dean about his false statement: whatever the student says, the dean continues to believe that he is a cheater and the student knows this. He does not and cannot attempt to manipulate the dean's state of mind; the action is ineffective, taking no notice of the hearer and not *for the Other*. What he states, in this case, seems entirely independent of the hearer's state of mind. Thus, by knowing that the dean absolutely knows that he cheated, what the student states entirely fits the definition of anti-social bullshit. To round off this picture, a concrete example should be given: if the student says something in Mandarin by knowing that the dean cannot understand any Mandarin, then the dean's belief remains unchanged since any statement is *for the dean* since the student takes no notice of the dean and unsurprisingly, cannot affect his state of mind.

5.2. Case-3: Coercion Lies

A typical example of a coercion lie is as follows:

I witness a crime and clearly see that a particular individual committed the crime. Later, the same person is accused of the crime, and, as a witness in court; I am asked whether or not I saw the defendant commit the crime. I make the false statement that I did not see the defendant commit the crime, for fear of being harmed or killed by him. It does not necessarily follow that I intend that my false statements deceive anyone. (I might hope that no one believes my testimony and that he is convicted in spite of it.) Deceiving the jury is not a means to preserving my life. Giving false testimony is necessary to save my life, but deceiving others is not; the deception is merely an unintended 'side effect'. I do not intend to deceive the jury in this case, but it seems clear that my false testimony would constitute a lie. (Carson 2006, 289)

As seen above, Carson, as well as Lackey (2013), asserts that the witness clearly lies, because the witness knows that the jury and the judge will not be justified in believing that he believes what he says.

For this case, how the witness acts for the hearers in question is significant. The witness will be aware of the fact that any judge might be in all likelihood remarkably experienced in spotting fictitious testimony and the fact that everybody knows about this, so does the defendant (supposing that the defendant is in the court at the time of the hearing). Therefore, once the witness states that "I did not see the defendant commit the crime," the statement is *for* two different *types* of Others, namely 'the judge or the jury' and 'the defendant.' And if the

witness really fears being harmed or getting killed by the defendant, he has to immerse himself in the role of a perjurer and looks earnest and assertive concerning his intention that both the judge and the jury be induced to believe his false statement; otherwise, he might still be in trouble. The defendant might say that “you have just droned something out, you have deliberately acted like that so that the judge could spot your perjury and punish me!” In case of such a probable bad consequence, the witness must envision the judge, the jury, the defendant—typifications of a typical courtroom—and himself as in the *We*-relationship. If the witness feigns in this direction, then it is called a lie from the point of view of the new model. Conversely, if the witness only claims the defendant’s innocence, in an atonic manner without looking assertive with taking no notice of anyone’s state of mind he is not lying; he just acts perfunctorily just because he is coerced to do that and does not try as hard as possible for to succeed in deceiving. Put differently, instead of the action of lying, it might look as same as a kind of performing art done compulsorily from the perspective of the hearers by ignoring their social existence and states of mind, namely in short, it would be a coerced anti-social bullshit.

Last, it may be of interest to add that even though contemporary definitions of Lackey and Carson differ from one another, it can roughly be said that they generally possess the broadest extension. Put differently, the traditional definition is a subset of contemporary rivals, that is, all lie-T for the trad-C₃ is also a lie for the contemporary rivals, whereas all lie-C for the contemporaries is not a lie for trad-C₃. The set of all instances of lies used in this paper in technical notation can be summarized as $N \subseteq F \subseteq T \subseteq C$, where the set of F refers to Chisholm and Feehan’s lies.

6. Conclusion

A good deal of progress has been made towards giving a new way of considering lying in which not only the speaker-sensitive but also the hearer-sensitive despite having clearly become more disputable. Granted that by ruling out some admitted cases of lying, the scope of the new model is much narrower than its rivals. Despite being narrower, however, it might be more commonsensical and rational for our intersubjective social world in which lying occurs. And the ruled-out cases are classified as anti-social bullshit which can be replied to with that rhetorical question, “how can you expect me to believe that?” in our daily life. Nevertheless, the reader might think and ask that “the new model only claims that the lie must be plausible-to-believe. So what?” You can undoubtedly define the new model as plausible-to-believe. However, it is not what this paper intended. Instead, this

paper aims to show why most of the prevalent attitudes toward lying are defective. Now that lying is an intersubjective social action, we should give more importance to the hearer's social existence as well as the social norms and phenomenological sociology was ideally suited for this new attitude. Therefore, being plausible-to-believe is not what to look for; it is only the result of the socio-analytic attitude which implies that lying makes you—*qua* liar—a typical *the Other*. The lie is not only the statement that comes out of the mouth but also that goes into the ear. As a result, if you, *qua* constituent of the commonsense world, would like to obtain the desired output for your action, you must give proper input to the hearer.

References:

- Augustine, Saint. 1952. "Lying." *Treatises on Various Subjects*, edited by R.J. Deferrari, Vol. 16, 53–120. New York: Fathers of Church.
- Carson, Thomas. 2006. "The Definition of Lying." *Noûs* 40(2): 284–306. <http://www.jstor.org/stable/3506133>.
- . 2010. *Lying and Deception: Theory and Practice*. Oxford: Oxford University Press.
- Chisholm, Roderick and Feehan, Thomas. 1977. "The Intent to Deceive." *The Journal of Philosophy* 74(3): 143–159. <https://doi.org/10.2307/2025605>.
- Cox, Ronald R. 1973. "Schutz's Theory of Relevance and the We-relation." *Research in Phenomenology* 3: 121–145. <http://www.jstor.org/stable/24654260>.
- Dewey, John. 1930. *Human Nature and Conduct*. New York: The Modern Library.
- Fallis, Don. 2009. "What Is Lying." *The Journal of Philosophy* 106(1): 29–56. <http://www.jstor.org/stable/20620149>.
- Frankfurt, Harry G. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.
- Fried, Charles. 1978. *Right and Wrong*. Harvard University Press.
- Keiser, Jessica. 2016. "Bald-faced lies: how to make a move in a language game without making a move in a conversation." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 173(2): 461–477. <http://www.jstor.org/stable/24703894>.
- Lackey, Jennifer. 2013. "Lies and deception: an unhappy divorce". *Analysis* 73(2): 236–248. <http://www.jstor.org/stable/24671096>.
- Meibauer, Jörg. 2014. "Bald faced-lies as acts of verbal aggression". *Journal of Language Aggression and Conflict* 20 (1): 127–150. <https://doi.org/10.1075/jlac.2.1.05mei>.
- Nietzsche, Friedrich. 1998. *Beyond Good and Evil*, translated by Marion Faber, Oxford World's Classics. Oxford Paperbacks.

- Schutz, Alfred. 1962. *Collected Papers I: The Problem of Social Reality*, edited and introduced by Maurice Natanson. The Hague: Martinus Nijhoff Publishers.
- . 1967. *The Phenomenology of the Social World*, translated and with an introduction by George Walsh and Frederick Lehnert. Northwestern University Press.

PREJUDICE, HARMING KNOWERS, AND TESTIMONIAL INJUSTICE

Timothy PERRINE

ABSTRACT: Fricker's *Epistemic Injustice* discusses the idea of testimonial injustice, specifically, being harmed in one's capacity as a knower. Fricker's own theory of testimonial injustice emphasizes the role of prejudice. She argues that prejudice is necessary for testimonial injustice and that when hearers use a prejudice to give a deficit to the credibility of speakers hearers intrinsically harm speakers in their capacity as a knower. This paper rethinks the connections between prejudice and testimonial injustice. I argue that many cases of prejudicial credibility deficits do not intrinsically harm speakers. Further, I suggest that prejudice is not necessary for harming speakers. I provide my own proposal on which testimonial injustice occurs when speaker's capacity as a giver of knowledge is interfered with in important ways. My proposal does not give prejudice any essential role.

KEYWORDS: epistemic injustice, testimonial injustice, prejudicial deficits; Miranda Fricker

As thousands of citations attest, Fricker's *Epistemic Injustice* is a landmark contribution to ethics and epistemology. She highlights a phenomenon of "epistemic injustice" where people are harmed in their capacity as knowers. She focuses mostly on a particular kind of epistemic injustice—testimonial injustice, where people are harmed in their capacity as givers of knowledge.

On her theory of testimonial injustice, morally objectionable attitudes—specifically prejudices—are necessary and perhaps even constitutive of harms in cases of testimonial injustice. Indeed, for her, the central cases of testimonial injustice are cases of "identity prejudicial credibility deficits." Roughly, these are cases where hearers' prejudice against a social identity of speakers causes the hearers to give less credibility to speakers. She claims that these prejudices are necessary for harming speakers. And she later develops this idea by claiming that such hearers treat speakers as mere sources of information, thereby epistemically objectifying them. Fricker's theory thus sees important connections between morally problematic attitudes and harming speakers

This paper rethinks those connections. Specifically, I argue that morally problematic attitudes, such as prejudices, are neither necessary nor constitutive for harming people in their capacity as givers of knowledge. Rather, I propose, people

are harmed in their capacity as givers of knowledge when that capacity is interfered with in important ways. While prejudice can cause such interference, so can other things as well.

In sections I-IV, I critically examine Fricker's theory about the harms of testimonial injustice. In section I, I provide Fricker's basic theory, which I criticize in section II. In section III, I provide Fricker's more developed theory involving claims about epistemic objectification, which I criticize in section IV. In section V, I suggest, *contra* Fricker and others, that morally problematic attitudes such as prejudice are neither necessary nor constitutive of harming people in their capacity as givers of knowledge. Finally, in section VI, I propose a theory on which being harmed in one's capacity as a giver of knowledge occurs when that capacity is interfered with in important ways. I develop this proposal before showing ways in which it differs from Fricker's theory.

I. Fricker On Testimonial Injustice

Fricker's book produced a groundswell of work.¹ In it, she focuses on a kind of injustice that she regards as distinctively epistemic (Fricker 2007, 1). Fricker understands injustice through the concept of harm as opposed to other potential concepts (e.g. fairness, rights, hypothetical agreements, etc.); I'll follow her lead here. Consequentially, Fricker and others suppose that *epistemic* justice is being harmed in one's capacity *as a knower*. She focuses on "testimonial injustice," a type of epistemic injustice where one is harmed in one's capacity as a *giver* of knowledge. Here I will focus on this type of injustice.²

Fricker describes the "central case" of testimonial injustice as cases of "*identity-prejudicial credibility deficit*" (Fricker 2007, 4, 28), hereafter cases of IPCDs. She characterizes these as cases where a speaker "receives a credibility deficit owing to identity prejudice in the hearer" (Fricker 2007, 28; 2012b, 292-3;

¹ While Fricker's work is well-known, others have also discussed the phenomenon of epistemic injustice. For a discussion of other authors that discuss similar topics, sometimes predating Fricker, see Ivy (2016, 438-9) and Pohlhaus (2017). Other recent relevant discussions include Medina (2013), Kidd, Medina, & Pohlhaus (2017), Sherman (2019). Fricker latter labels this type of injustice "discriminatory" as opposed to "distributive" (Fricker 2013; 2017b). I won't use this label because it is unnecessary here—I won't be discussing distributive injustice, but only a type of discriminatory epistemic injustice, testimonial injustice.

² It is not clear that what matters is being harmed as a knower *per se* as opposed to other things closely associated with knowledge. (For instance, almost all of the cases people discuss could be subsumed under the less pithy category 'being harmed in a capacity as a reliable belief former and/or testifier') But I set this issue aside here. For a discussion that assume it is knowledge *per se* that matters, see Luzzi (2016).

2013, 1319; 2017, 161; 2017b, 53-4). Let's unpack this characterization. A credibility deficit is where a hearer attributes less credibility to a speaker than is warranted by the evidence the hearer has. A prejudicial credibility deficit is a credibility deficit caused by a prejudice (Fricker 2007, 17). Prejudices for Fricker are stereotypes where a person has some resistance to counter-evidence (Fricker 2007, 35); stereotypes themselves being "widely held associations between a given social group and one or more attributes" (Fricker 2007, 30). Finally, an identity prejudice is a prejudice against a person in virtue of one of their social identities (Fricker 2007, 4, 27) Thus, we can unpack this characterization as: in the central cases of testimonial injustice, a hearer has a prejudice against a social identity of a speaker which causes the hearer to judge the credibility of the speaker less than the hearer would have if the hearer lacked the prejudice.

Fricker focuses on cases of IPCDs because, as she argues (Fricker 2007, 41-3), testimonial injustice only occurs if hearers bear problematic attitudes—specifically prejudices—towards speakers. As she memorably puts it, "the ethical poison of testimonial injustice must derive from some ethical poison in the judgment of the hearer, ...The proposal I am heading for is that the ethical poison in question is that of *prejudice*" (Fricker 2007, 22). I'll call this claim:

Hearer's Attitudes, Speaker's Harms: A person commits testimonial injustice towards another—harms that person in their capacity as a giver of knowledge—only if the former bears morally problematic attitudes towards the later.

Fricker identifies the problematic attitude as prejudice, but cases of IPCD may also include other problematic attitudes, as we'll see in sections III and V below.

Turning to harm, Fricker distinguishes between two types of harm in cases of testimonial injustice, a "primary" and a "secondary." The main difference is whether the harm is intrinsic and essential to the testimonial exchange or extrinsic and contingent. She writes (Fricker 2007 44):

The harm that concerns us here is ... the immediate wrong that the hearer does to the speaker who is on the receiving end of a testimonial injustice.

We should distinguish a primary from a secondary aspect of the harm. The primary harm is a form of the essential harm that is definitive of epistemic injustice in the broad. In all such injustices the subject is wronged in her capacity as a knower... When one is undermined or otherwise wronged in a capacity essential to human value, one suffers an intrinsic injustice. The form that this intrinsic injustice takes specifically in the case of testimonial injustice that the subject is wronged in her capacity as a giver of knowledge...

She elaborates on secondary harms (Fricker 2007, 46):

Turning now to the secondary aspect of the harm, we see that it is composed of a

Timothy Perrine

range of possible follow-on disadvantages, extrinsic to the primary injustice in that they are caused by it rather than being a proper part of it.

Secondary harms might include things like loss of personal or professional opportunity or even loss of life, as in one of Fricker's chief examples, Tom Robinson in *To Kill a Mockingbird*.

Fricker understands these primary harms in a non-aggregative way (cf. (Fricker 2007, 20-21)). Each occasion harms the speaker on its own. It might also be that a sufficient number of IPCDs will also have an aggregating or cumulative effect. But Fricker sees that as a distinct phenomenon from the primary harms that occur at each IPCDs on their own.

Thus, Fricker's theory of testimonial injustice endorses:

Intrinsic Harm: In cases of IPCDs, there is a harm that is intrinsic and essential to the case.

Capacity Harm: In cases of IPCDs, hearers are harmed in their capacity as knowers, specifically, as givers of knowledge.

These claims are proposed as true of the primary harms, but not necessarily true of secondary harms. Indeed, the term 'primary harm' could just be defined as those harms, whatever they are, that both *Intrinsic Harm* and *Capacity Harm* are true of.

II. Intrinsic Harm and Capacity Harm

Initially Fricker does not analyze *Capacity Harm*, instead relying on intuitions about being harmed in a capacity as a giver of knowledge. However, in (Fricker 2007, chp. 5), she develops an analysis of that harm in terms of epistemic objectification. This section examines the intuitiveness of *Capacity Harm*, while section IV will examine her developed analysis. My argumentative strategy in this section is to provide a range of types of cases of IPCDs. Relying on intuition, I'll suggest that these types of cases do not essentially and intrinsically harm speakers. That is, *Intrinsic Harm* and *Capacity Harm* are not both true.³

Case Type 1. A hearer gives an IPCD to a speaker. But the deficit is not great. As a result, the hearer does end up believing what the speaker says, though they would have believed more readily if the hearer lacked their prejudice.

Case Type 2. A hearer gives an IPCD and, as a result, the hearer does not believe the speaker. But the speaker does not know any of this. (Perhaps the hearer

³ I assume in all cases of IPCDs hearer see speakers as engaging in speech acts like saying, asserting, reporting, etc. For hearers are making attributions of credibility. So I won't consider cases where the hearer fails to identify the correct illocutionary act that speakers are in engaged in. That is an important, but distinct, phenomenon.

merely nods along as if they believe or quickly changes the subject.)

Case Type 3. A hearer gives an IPCD and, as a result, the hearer does not believe the speaker. The speaker knows that the hearers does not believe the speaker, but not that the reason for this is an identity prejudice. (Perhaps the speaker thinks the hearer is merely not convinced by the evidence she adduced or is just being odd.)

Case Type 4. A hearer gives an IPCD and, as a result, the hearer does not believe the speaker. The speaker knows the hearer does not believe the speaker and this is because of an identity prejudice or stereotype. Nonetheless, the speaker does not take this exchange to question her own competency. (One can imagine a possible speaker in such cases responding, “You don’t believe me because I’m a such-and-such; ha, I always thought you were a bastard!”)

Case Type 5. A hearer gives an IPCD and, as a result, the hearer does not believe the speaker. The speaker knows the hearer does not believe the speaker and this is because of an identity prejudice or stereotype. As a result of the exchange, the speaker questions her own competency either on the subject matter or more generally.

In each case, hearers give IPCDs. Thus, given *Intrinsic Harm* and *Capacity Harm*, in each of these types of cases, speakers are harmed in their capacity as givers of knowledge. However, my own intuition is that such a result is not uniformly plausible. Specifically, it is not plausible at all in *Case Type 1* and is much more plausible for *Case Type 5*. My intuitions about being harmed in a capacity as a giver of knowledge do not track cases of IPCDs. To be sure, any of these cases might also have extrinsic “secondary” harms. And it might be that if we focus on an aggregation of these cases there is also some sort of harm. But *Intrinsic Harm* and *Capacity Harm* imply that there is a further harm from any of those.⁴

Maitra also argues that not all cases of IPCDs harm speakers (Maitra 2010, 197-9). However, my argument may avoid some complex issues that her argument may get embroiled in. Her arguments are based on a variety of cases, but her chief case involves a person, Zara, who has a stereotype that members of the “tea party” are ill-informed, despite not knowing much about them. On the basis of this stereotype, Zara dismisses a piece of news written by a “committed tea-partier.”

⁴ There’s a tension in Fricker’s writing. At one point, Fricker claims that the harms of testimonial injustice might be “very little” even “trivial” (Fricker 2007, 43). On the next page, she describes these harms using harsh language like “degrading.” I find it hard to reconcile these passages. Further, Fricker clearly wants testimonial injustice to constitute a relatively uniform class. But if some harms are trivial while others are deeply degrading, they are not a unified class. So it is not plausible to respond to these cases by claiming that there is a “trivial” harm in each case type.

Timothy Perrine

Maitra claims, plausibly, that Zara does not harm the writer. This case is purportedly inconsistent with *Intrinsic Harm*.

While I am sympathetic with Maitra's conclusion, her argument may get embroiled in disputes about proper regard for evidence. Specifically, some philosophers draw a sharp distinction between evidence agents have and evidence agents could have; they further argue that agent's epistemic obligations are determined only by the former, not the latter (see, e.g., (Feldman 2000, 688-70)). Given such views, Zara is not showing improper regard for the evidence; she merely does not have much evidence. Thus, Zara lacks a prejudice and this is not a case of an IPCD. However, Maitra might reject this position (cf. Maitra 2010, 200)) and others have argued against it (e.g. Kornblith 1983; Goldberg 2017; Weiland 2017). On an alternative position, since Zara does not acquire readily available evidence, she does have a prejudice and thus this is a case of an IPCD. Thus, whether Maitra's counterexample succeeds may depend upon antecedent views about proper regard for evidence and prejudice. My objection side steps these issues by merely *stipulating* that these cases involving IPCDs.

III. Fricker on Primary Harm as Objectification

In (Fricker 2007, chp. 5), Fricker develops *Capacity Harm*. Her development of this claim makes *Intrinsic Harm* true, because on the developed theory the harm is intrinsic to the prejudice of the hearer. The way she develops *Capacity Harm* has two parts. First, utilizing a distinction from Edward Craig—between informants and sources of information—she claims hearers treat speakers as mere sources of information in cases of IPCDs. Second, drawing on Kant and Nussbaum, she claims treating someone as a mere source of information is a kind of morally problematic objectification.

Fricker utilizes Craig's distinction between an "informant" and a "source of information." She writes, "Broadly speaking, informants are epistemic agents who convey information, whereas sources of information are states of affairs from which the inquirer may be in a position to glean information" (Fricker 2007, 132). She then suggests that in cases of testimonial injustice people are not treated as informants or sources of information but rather *mere* sources of information (Fricker 2007, 132). As she writes (Fricker 2007, 132-3):

The moment of testimonial injustice wrongfully denies someone their capacity as an informant, and in confining them to their entirely passive capacity as a source of information, it relegates them to the same epistemic status as a felled tree whose age one might glean from the number of rings.

Let's call this idea:

Prejudice, Harming Knowers, and Testimonial Injustice

Mere Source of Information: In cases of IPCDs, hearers treat speakers as mere sources of information.

Additionally, treating someone as a mere source of information is a kind of “objectification” (Fricker 2007, 133):

...testimonial injustice demotes the speaker from informant to source of information, from subject to object. This reveals the intrinsic harm of testimonial injustice as *epistemic objectification*: when a hearer undermines a speaker in her capacity as a giver of knowledge, the speaker is epistemically objectified.

So when a hearer treats a speaker as a mere source of information, she epistemically objectifies the hearer.

She develops this idea by utilizing the work of Nussbaum and Kant (Fricker 2007, 133-5). Just as Nussbaum claims there’s no problem with treating a person as an object—people are objects!—so too treating someone as a source of information isn’t problematic—people are sources of information! The problem is treating someone as a *mere* source of information. Just as Kant claims that treating someone as a mere means is problematic because it is inconsistent with treating them as an end in themselves, so too treating a person as a mere source of information is problematic because it is inconsistent with treating them as a subject of knowledge, a knower. I’ll distill these points as:

Epistemic Objectifying Claim: In cases of IPCDs, the speaker is epistemically objectified, that is, the speaker is denied a general status as knower by the hearer.⁵

It is unclear how Fricker sees the relationship between the prejudice hearers have and their epistemically objectifying speakers. Presumably these are additional, further attitudes that are caused by the prejudice, but technically speaking distinct from it. This level of subtlety will not matter for my objections.

IV. Sources of Information and Epistemic Objectification

In this section I object to Fricker’s way of developing *Capacity Harm*. In short, cases of IPCDs neither essentially involve treating speakers as mere sources of information (contra *Mere Source of Information*) nor essentially involve denying speakers a general status as a knower (contra *Epistemic Objectifying Claim*).

A. Credibility Deficits and Mere Sources of Information

⁵ At one point, Fricker claims that a hearer’s prejudice will rarely have them claim that speakers are not knowers at all (Fricker 2007, 134-5). She then shifts to talking about speaker’s “general status” as a knower being denied. Fricker does not explain the distinction between a “status as a knower” and “general status as a knower.”

Mere Source of Information is false. First, recall instances of *Case Type 1*. In these cases, a hearer believes what a speaker asserts because the speaker asserts it. To be sure, they also have an IPCD that deflates the speaker's credibility, but not to a sufficient degree to disbelieve them. Clearly, in this case the speaker is being treated as an informant. So not all cases of IPCDs are cases of a hearer treating a speaker as a mere source of information.

Let's turn to any case from *Case Type 2-5*. In each, a hearer has an IPCD and, as a result of this, does not believe the speaker. But in these cases the hearer is *not* treating the speaker as a source of information—merely or otherwise. For the hearer does not believe what the speaker says!

The way we normally respond to the content of assertions does not allow for a response where we treat the content of people's assertions as a mere source of information. For we normally respond to the content of an assertion by either believing it or not. But if we believe the content of a speaker's assertion, then we are treating them as an informant. And if we do not believe the content of a speaker's assertion, then we normally do not form any belief on the basis of what they say and are thus not treating them as a source of information—merely or otherwise.

To be clear, hearers *could* treat speakers' assertions as a source of information. However, such cases are rarer and unlike cases of IPCDs. To see this, it is useful to recall why Craig thought this distinction was important. As he saw it, informants are more *useful* than sources of information (Craig 1990, 36-7). For gleaning information for sources of information may require some additional beliefs or specialized knowledge (Craig 1990, 36). For instance, one can glean information from the fact that a red litmus paper turned blue—but only if one has additional beliefs about that fact.⁶

It is possible to use the content of a speaker's assertion as a mere source of information, but this requires some further belief about the connection between their assertion and some other fact. To give some illustrations, perhaps I utter a certain sentence in Chinese using the universal measure word instead of a more specific one that fits the noun. A native speaker might glean from my choice of measure word that I am not fluent in Chinese (or my relative fluency). Or perhaps there is a well-known "pickup artist" in my extended friend group. One night at the bar, he sidles up to me and says some coy and flattering things about me. I

⁶ Craig indicates a second reason based on the idea of "the special psychology of team-work in a community," but he regards this reason as "more questionable" and "far harder to pin down" and is uncomfortable with resting the distinction just on it (Craig 1990, 36). Interestingly, in her discussion of Craig, Fricker emphasizes this feature (Fricker 2007, 131; 2012: 252ff.).

won't believe what he says, and he is thus not an informant for me, but I can glean information about his intentions given facts about what he said. In these cases, hearers have a further belief that they use to glean information from facts about the content of speaker's assertions. But these kinds of cases are not only rarer but quite different from standard cases of IPCDs.

In criticizing *Mere Source of Information*, I have focused on whether or not hearers treat speaker's *assertions* as a mere source of information. However, obviously we can treat other features of an agent as a source of information. Seeing a person soaked in water can tell me that it is raining outside. Likewise, what a person is carrying or wearing can also be used as a source of information so that I can reasonably infer other things about them. However, treating these other features of a person as a source of information are independent of testimonial exchanges and IPCDs. That is, one can treat these other features of person as a source of information, even if there is no testimonial exchange. And one can treat these other features of a person as a source of information, regardless of whether one gives an IPCD. Thus, in criticizing *Mere Sources of Information* I have focused on responding to the assertion or its content as opposed to these other features.

My criticisms are similar to Pohlhaus' criticisms (Pohlhaus 2014, 103-4). She claims that when we glean information from states of affairs, they are "objects" that make demands on us which we cannot resist. When I see a person come in from the rain, there is a demand to believe that it is raining that I cannot normally resist. But in cases of testimonial injustice hearers *do* resist what they are being told. Additionally, cases of testimonial injustice occur when "engaging in ordinary epistemic practices for ascertaining truth from another epistemic agent based on testimony" (Pohlhaus 2014, 103); but we do not try to elicit information from objects. In this way, cases of IPCDs are unlike cases where we treat something as an object to glean information from it.

Pohlhaus and I agree that cases of IPCDs are not always or necessarily cases of treating a person as a mere source of information; that is, *Mere Sources of Information* is false. However, Pohlhaus is arguing that if *Mere Sources of Information* is true, then in cases of IPCDs, hearers treat speakers as objects, and she rejects that claim. By contrast, I object that when we attend to the distinction between informant/source of information and how we normally react to assertions, it is implausible to think that speakers are treated as mere sources of information in cases of IPCDs.

B. Mere Sources of Information and Epistemic Objectification

Fricker also defends *Epistemic Objectifying Claim*: that cases of IPCDs are cases where hearers deny speaker's general status as a subject of knowledge (recall, "it relegates them to the same epistemic status as a felled tree whose age one might glean from the number of rings"). Additionally, she believes that hearers deny speaker's general status as a subject of knowledge *because* hearers treat speakers as mere sources of information. However, neither this explanatory claim nor *Epistemic Objectifying Claim* are plausible for all cases of IPCDs.

Epistemic Objectifying Claim is not plausible, as can be seen by revisiting the standard reading of Kant. On that reading, when one treats a person as a mere means, one is treating them in a way that is inconsistent with them being an "end in themselves." Analogously, for Fricker, when one treats a person as a mere source of information, one is treating them in a way that is inconsistent with them being a "subject of knowledge" (Fricker 2007, 135). However, the analogy breaks down. A singular action that treats a person as a mere means is sufficient for not treating that person as an end in themselves. By contrast, treating a person with a singular IPCD is *not* sufficient for denying that person general status as a subject of knowledge.

An example. Suppose I frequently take my car to the repair shop. But my mechanic—who only has a high school diploma—frequently wants to talk about the stock market. Perhaps on the basis of a prejudicial stereotype about people with a high school diploma, I routinely give a credibility deficit to what my car mechanic says about the stock market. Here I give a persistent IPCD to my car mechanic. Nonetheless, I see him as a knower in general. Indeed, when it comes to the workings of my car, I see him as the expert and much more knowledgeable than I. So treating someone with a IPCD—even a persistent one—is consistent with also treating them as a subject of knowledge in general.⁷

Additionally, and for essentially the same reason, the explanatory claim is not true: treating someone as a mere source of information need not be inconsistent with treating them as a general subject of knowledge. For instance, Craig gives an example of a person who is "systematically wrong about what day of the week it is: he is always a day behind" (Craig 1990, 37) Craig notes one could use their assertions to figure out what day it is by asking them and adding a day. Such a person is not an informant—they don't know and one doesn't believe what they say. But there's no trace of denying the person's general status as a knower. In

⁷ Congdon (2017, 247) gives a similar kind of example, though more extreme than mine. I hope my example indicate how to identify additional cases beyond the ones I or Congdon give.

short, cases of IPCDs don't essentially involve treating speakers as mere sources of information or as mere objects, lacking a general status as knowers.

V. Hearer's Attitudes and Speaker's Harms

I have argued that cases of IPCD do not intrinsically or essentially harm speakers. Specifically, if being harmed implies being treated as a mere source of information or being denied a general status as a knower, then cases of IPCD do not intrinsically or essentially harm speakers. Some might think that Fricker's specific proposals are wrong and what we need is just some "fine-tuning." In this section, I identify a deeper disagreement over IPCDs and harm.

While others have criticized *Epistemic Objectifying Claim*, I take their objections to be of the "fine-tuning" kind. They *agree* that IPCDs contain morally problematic attitudes that harm speakers; they merely maintain that Fricker has identified the wrong attitudes. Some illustrations. Pohlhaus claims that in cases of IPCDs people are not treated as objects. Rather, they are treated as truncated subjects who have knowledge but in a derivatized way (Pohlhaus 2014, 105ff.). Those are morally problematic attitudes to adopt towards speakers. Similarly, Davis criticizes Fricker for claiming that cases of identity prejudicial credibility excesses cannot harm speakers. She follows Pohlhaus' alternative model, but extends it to cover cases of credibility excess (Davis 2016). McGlynn also objects to *Epistemic Objectifying Claim*, but in a different way. McGlynn retains the idea that cases of testimonial injustice involve epistemic objectification but rejects the idea that epistemic objectification should be understood as exclusively denying someone's general status as a subject of knowledge. Rather, McGlynn—following Nussbaum—suggests that there are different ways we can objectify a person, with denial of knowledge or agency being just one among many (McGlynn 2021, 169ff.).

While these authors are not explicit on this point, I see them as agreeing with Fricker in accepting:

Hearer's Attitudes, Speaker's Harms. A speaker commits testimonial injustice towards a hearer—harms that hearer in their capacity as a giver of knowledge—only if the speaker bears morally problematic attitudes towards the speaker.

The dispute between them is primarily over the nature and content of these morally problematic attitudes.

While *Hearer's Attitudes, Speaker's Harm* only states a necessary condition between attitudes and harm, some may desire a stronger condition. Specifically, one might claim that the hearer's attitude partly *constitutes* the harm to the speaker. Indeed, one might claim that the "primary harm" of testimonial injustice *just is* the hearer's problematic attitude towards the speaker. Fricker herself never

makes these stronger constitutive claims. But they do naturally fit with the way she develops her theory of testimonial injustice. Further, they would also explain the appeal of *Intrinsic Harm*: of course there is a harm that is intrinsic to cases of IPCDs, the prejudice (or negative attitude more generally) *itself* constitutes a harm.

However, I reject *Hearer's Attitudes, Speaker's Harms* for a simple reason. In general, in order to harm others, one does not need to have morally problematic attitudes towards them, prejudicial or otherwise. A doctor harms a person by accidentally giving them medicine they are allergic to, even if the doctor lacks any morally problematic attitudes about them. A company may dump chemicals at the federally required level and, as a result, cause infertility in a nearby population. The company harms the nearby population, even if it bears no morally problematic attitudes towards them. So I reject *Hearer's Attitudes, Speaker's Harms* because it is not true, in general, that a person is harmed by another only if the former bears a morally problematic attitude towards the latter.

I am also skeptical of the constitutive claim. For the harm under discussion is being harmed in one's capacity as a giver of knowledge. But a hearer's attitude *by itself* need not have any causal influence on a speaker's capacity as a giver of knowledge. But if the hearer's attitude *by itself* has no influence on a speaker's capacity as a giver of knowledge, I am unsure how it can *constitute* harming a speaker's capacity as a giver of knowledge. To be sure, having a prejudicial—or otherwise morally problematic—attitude towards another can be a bad thing. But I am unsure how such an attitude on its own, independent of its causal influence on the world, harms another.

Thus, the deeper disagreement I have with these authors is over whether morally problematic attitudes, including prejudices, are necessary or constitutive of harming people in their capacity as knowers. Thus, I doubt that the right way to analyze testimonial injustice is in terms of such attitudes. And while I cannot claim to have refuted such views, I hope to have motivated interest in an alternative proposal, which I'll provide in the next section.⁸

VI. Harming Speakers—An Alternative Account

This section provides an alternative proposal for testimonial injustice. It retains the idea that testimonial injustice harms others in their capacity as givers of knowledge. But I propose that one is harmed in one's capacity as a giver of knowledge when that capacity is interfered with in important ways. I begin by

⁸ Li argues that prejudice is not necessary for testimonial injustice in cases involving testifiers with cognitive or psychological impairments (Li 2016). I think Li is correct—but that there is no reason to restrict to cases involving testifiers with cognitive or psychological impairments.

describing the idea of being harmed in a capacity, before turning to being harmed in a capacity as a giver of knowledge, and ending on some differences between my proposal and Fricker's.

A. Being Harmed in a Capacity

The guiding idea of my proposal is this:

Harmed in a Capacity. A person is harmed in a capacity, when their ability to utilize that capacity is interfered with in important ways.

Let me say a little bit more about utilizing, interference, and importance.⁹

Utilizing. A simple way to utilize a capacity is to *manifest* or *use* it. Capacities are clusters of dispositions and we can use them when we manifest them. Some capacities have a discrete effect—like how, for most people, hitting their knee patella results in their leg extending. But many capacities have more open-ended effects. For instance, people have a capacity to hit a home run, write a philosophical paper, sing a song, paint in watercolor, and give birth to a child.

Capacities can be used by developing them or modifying them over time. A person may start with a capacity to run a kilometer. But they can develop that capacity to run faster, with better form. A person can develop their capacity to paint with watercolor by doing it more. Other capacities can be utilized primarily by manifesting them without modifying them very much. For instance, most people have the capacity to be in ketosis—a state where their bodies are creating more ketone bodies as a result of low bioavailability of glucose. But this is a “hard wired” capacity that can only be utilized by manifesting it (by, e.g., going on a diet restricting glucose availability).

Interference. A person's capacity can be interfered with. It is useful to divide the interference into two types. First, a person's capacity can be interfered with by modifying the capacity. An extreme form of modifying the capacity is removing it altogether. For instance, some criminals and mental health patients were forced into sterilization in the late 19th, early 20th century in the US (see (Largent 2008)). Their reproductive capacities were removed altogether. But a person's capacity might not be permitted to develop or positively worsen. Intentionally depriving a child of adequate nutrition is a serious offense for exactly this reason. Second, a person's capacity can be interfered with, not by modifying the capacity, but by

⁹ Understanding justice and injustice in terms of interfering with capacities is not new to me. It is the cornerstone of approaches to welfare pioneered by Amartya Sen and Martha Nussbaum (Sen 1987, 2009; Nussbaum 2000, 2006). However, my proposal here is independent of the political aspirations of their approaches.

keeping people from utilizing that capacity. For instance, prior to 1947, many African American baseball players were not admitted into the Major Baseball League. Many African American players had the capacities to play baseball well at that level—they simply were not allowed to utilize those capacities in the Major Baseball League.

Importance. People can have their capacities interfered with, but not in important ways. Unfortunately, I have no criteria for what constitutes important and unimportant ways. However, some capacities might be more important than others. For instance, developing and manifesting some capacities may be important for one's own well-being, maintaining practical identities, or bringing about the good. All else being equal, interfering with those capacities may be more important than other capacities.

Whether an interference is important may depend upon the availability of utilizing that capacity on other occasions. To illustrate, if the library is closing, this interferes with my ability to finish the book I am reading. But I have many other opportunities to finish reading that book by returning on some other day. Thus, the library closing right now is not an important interference. By contrast, if I am barred from the library in virtue of my political or ethnic status, such barring would constitute an important interference. Indeed, it may be that normally *sequences* or *sets* of interferences constitute an important interference.

B. Being Harmed in a Capacity as a Giver of Knowledge

Testimonial injustice harms someone as a giver of knowledge. A giver of knowledge is not only someone who knows something—has information—but someone who has the capacity to *give* that information, normally in an intentional act of communication. Given *Harmed in a Capacity*, my proposal is this:

Interfering Harm: One experiences testimonial injustice when one's capacity to give knowledge is interfered with in an important way.

I will briefly describe ways this capacity can be interfered with. As we'll see, this proposal will subsume a number of cases others have discussed.

One's capacity can be interfered with by either modifying the capacity or by keeping people from utilizing the capacity. First, let us focus on being interfered with in utilizing that capacity. That interference can be either internal or external. For instance, an external interference of manifesting that capacity includes situations where people are not asked to testify or are barred or otherwise prevented from testifying. To illustrate, consider countries that have laws barring literacy for certain groups—effectively keeping them from testifying through a primary method of testimony, writing. Or consider medical care providers who

either interrupt patients who are testifying or simply don't request information from them, assuming that they won't have anything medically relevant to say (cf. (Kidd & Carel 2017)). Or, lastly, consider government censorship of individual citizens—like, for instance, when the government removes posts by individuals describing their experiences on social media and scrubs any mentions of those posts. Here, as well, there is an external interference of manifesting a capacity.

Interference with utilizing a capacity may come through internal means. Specifically, speakers might silence themselves because of poor track records of testimonial exchanges involving themselves or those similar to them. If there is a track record of the speaker not being believed, the speaker may not see the point in testifying, in manifesting their capacity as a giver of knowledge. Dotson calls this “testimony smothering,” as speakers smother their own testimony (Dotson 2011, 244). Fricker describes a woman who stops proposing things in meetings, and simply passes her ideas to a male coworker for him to propose because she was rarely believed in the past (Fricker 2007, 47). Another way this interference may occur is if a person internalizes a social norm to *not* testify because doing so is not appropriate for someone with one of their social identities. Nussbaum describes an Indian mother who was highly critical of male authority with her daughter privately, not publicly, yet nevertheless taught her daughter to be “submissive, silent, and innocent” (Nussbaum 2000, 42)¹⁰

One's capacity can be interfered with by having the capacity itself be modified. The capacity to be a giver of knowledge requires two parts. First, it requires having knowledge, relevant information. Second, it requires the ability to *provide* or *give* that information, primarily though not exhaustively through written or spoken word. For most adults, having knowledge is normally sufficient for having a capacity to give that knowledge. Thus, if one's capacity to give knowledge is interfered with by modifying that capacity, this would normally be by interfering with the capacity to know itself, and not necessarily the second capacity to provide or give that information. At the very least, I will focus on the first capacity here.¹¹

¹⁰ See also Lee (2021a, b), who describes cases of “anticipatory epistemic injustice.” I'm not sure what Lee describes is distinct from testimonial smothering (compare (García 2021)). But, unlike Lee, I would categorize the cases she describes as cases of testimonial injustice, since they involve interfering with the capacity to give knowledge.

¹¹ One important exception may be people and communities who do not primarily use spoken word to communicate. For instance, those who primarily use a signed language to communicate may have their ability to give knowledge interfered with in important ways even if their ability to know is not at the same time interfered with. Thus, governments may be obligated to provide signed language interpreters in various contexts—to not interfere with the manifesting of those

Once again, we can distinguish between external and internal interference. A person can be interfered with in developing their capacity to know through external sources. For instance, denial to education is one such way. Such cases will *ipso facto* also be cases where agents are interfered with in developing their capacity as givers of knowledge (cf. Anderson 2012, 169ff.). Similarly, a parent who is unable to do the research they intend to do because of inadequate access to childcare is having their capacity to know interfered with through external sources (cf. Hookway 2010, 154). However, agents can also fail to develop their capacity to know, and thus be a giver of knowledge, through internal means. Specifically, if agents acquire sufficient cases where they are not believed in testimony, they may come to doubt their own intellectual abilities. That is, they might *internalize* the doubt and lack of trusts others have. In such a case, this internalization will interfere with them developing their capacity to know and be a giver of knowledge. (Cf. Fricker 2007, 47-51, where she describes cases like this.)

I have highlighted four ways that speakers can be interfered with in their capacity as knowers. They can be interfered with internally or externally in giving information—in testifying. Additionally, they can be interfered with internally or externally in gaining knowledge and information—knowing. However, I have said little about the role of hearer’s *disbelief* in interfering. If a hearer does not believe what a speaker says, does this disbelief constitute an interference of the speaker’s capacity as a giver of knowledge? No. Giving knowledge is fundamentally a matter of having information and being able to intentionally communicate that information. It is not normally a requirement of having information or intentionally communicating that information that others receive that information. To be clear, a lot of really important activities require that hearers listen to speakers. But when hearers do not listen to speakers, we should not classify this as a failure on the part of the hearer to manifest a capacity.

These interferences must also be *important* interferences. As I said, I lack a good method for sorting which kinds of interferences are important or not. However, presumably some rules of thumb are correct. Systematic and persistent interferences are more likely to reduce the number of occasions for utilizing a capacity than localized and non-persistent interferences. Interferences regarding issues of one’s self-identity or well-being might be more important than some that are not.

capacities to give knowledge. This is a rich and important topic I cannot hope to adequately explore here.

C. Departures from Fricker's View

As a way of further developing my proposal *Interfering Harm*, I will briefly highlight and defend some further departures from Fricker's theory. I do take my theory here to conflict with Fricker's, for I take us to both be describing the same issue: being harmed in one's capacity as a giver of knowledge. Further, as indicated above, both of theories agree on a range of cases that they are cases of testimonial injustice. Thus, a concluding comparison may help see where our two theories diverge.¹²

First, and most obviously, it is not part of *Interfering Harm* that testimonial injustice occurs only in cases of IPCDs. In fact, they might not occur in testimonial exchanges at all. Being barred from educational opportunities is a kind of testimonial injustice on my proposal because it interferes with one's capacity to know and thus, *ipso facto*, be a giver of knowledge. But such barring is not a testimonial exchange.

Nonetheless, cases of IPCDs can cause and exasperate testimonial injustice. Specifically, testimonial injustice can be caused in cases where speakers are aware that the credibility deficit they receive is because of their social identity (*Case 4* and *Case 5* from above). For when speakers are aware that they are not believed because of their social identity, this *external* fact can cause an *internal* response whereby the speaker interferes with his own capacity as a giver of knowledge.

Second, and relatedly, given *Interfering Harm*, the claim *Hearer's Attitudes, Speaker's Harm* is false. Otherwise put, people can harm speakers even if they do not bear any negative attitudes—like prejudice, objectifying, etc.—to speakers. For one's capacity as a giver of knowledge can be interfered with even if speakers do not harbor negative attitudes. As noted above, Fricker rejects this. I will briefly consider her reason.

Fricker claims that if prejudice is not required for testimonial injustice than it would be “too easy” to harm others (Fricker 2007, 42; 2012b, 290-1). Specifically,

¹² Dotson suggests that there cannot be a “catch-all” theory of epistemic injustice, and this may lead one to think that Fricker and I are merely talking past each other and offering theories that could stand side-by-side (Dotson 2012, 41). However, as I see it, Dotson is led to this conclusion by claiming—correctly—that addressing different types of epistemic injustice require different solutions (Dotson 2012, 41, 36; 2014, 117). But I don't see how it would follow from the fact that there are distinct responses to distinct types of epistemic injustice that there could be no common or unifying feature. Further, my discussion here is focused on a particular type of epistemic injustice—testimonial injustice—so I am doubtful that types of considerations Dotson raises for epistemic injustice *in general* would indicate that Fricker and I are giving theories of distinct things *in a particular case*.

her worry is that people might use non-prejudicial reliable stereotypes to give credibility deficits that harm. She gives an example of people being skeptical of a used-cars salesman. Her thought is when we use a reliable, non-prejudicial stereotype to give a credibility deficit to him, this should not count as a case of testimonial injustice. One proposal that excludes this as a case of testimonial injustice is *Hearer's Attitudes, Speaker's Harm*.¹³

However, credibility deficits—and hearers' attitudes more generally—do not essentially interfere with speaker's capacity as a giver of knowledge. Thus, on *Interfering Harm*, credibility deficits are not essentially cases of testimonial injustice. And, indeed, in most cases where one is skeptical of a cars salesman because he is a car salesman, this does not interfere with his capacity as a giver of knowledge.

Third, Fricker claims that testimonial injustice only occurs in cases of credibility *deficits* not *excesses*. Several authors have criticized this claim (e.g., Davis 2016; Medina 2013, chp. 2). Davis gives an example of an Asian-American high school student who is enlisted by peers to help with a math problem in virtue of a stereotype that Asian-Americans are especially good at math. We can easily imagine a version of this case that is included under my proposal. For instance, suppose the student is *not* especially good at math and thus his attempt to aid his peers with math problems frequently fails. Having failed to reach the lofty social standards set for him, this student may simply despair that he has any great competence here at all. In such a case, the credibility *excess* sets standards too high for the student and when he fails them he *internalizes* doubts about his own capacities for knowledge in this area. He is thereby harmed in a capacity for knowledge in virtue of a credibility excess.

Fourth, Fricker uses the term 'primary harm' to refer to harms that are intrinsic and essential to cases of IPCDs and 'secondary harms' to refer to harms that are extrinsic. Suppose we take these usages as a stipulative definition. Given *Interfering Harm*, plausibly, there are no primary harms. Cases of IPCDs do not *essentially* interfere with speaker's capacity as a giver of knowledge—even if some, or even most, do. Given *Interfering Harm*, there are only so-called 'secondary harms.'

Finally, and as a way of summing up, we might put the differences between my approach and others like this. What makes testimonial injustice *testimonial* injustice? For Fricker and others, testimonial injustice is testimonial partly because of where the harm occurs—in the testimonial exchange itself because of the

¹³ Fricker develops her ideas of testimonial injustice using some of Nussbaum's work on capacities; but she still sees prejudice as central to testimonial injustice (Fricker 2015, 79).

attitudes of speakers.¹⁴ However, on my proposal testimonial injustice is *testimonial* partly because of its effects—what it interferes with—namely, the practice of testimony. Indeed, as Fricker and others have pointed out, testimonial injustice can have entirely deleterious effects in the lives of peoples and communities. My proposal accepts, and underscores, this point.¹⁵

References

- Anderson, Elizabeth. 2012. “Epistemic Justice as a Virtue of Social Institutions.” *Social Epistemology* 26, 2: 163-73.
- Congdon, Matthew. 2017. “What’s Wrong with Epistemic Injustice? Harm, Vice, Objectification, and Misrecognition.” In Kidd, Medina, and Pohlhaus (2017).
- Craig, Edward. 1990. *Knowledge and the State of Nature*. Oxford: Clarendon Press.
- Davis, Emmalon. 2016. “Typecasts, tokens, and spokespersons.” *Hypatia* 31, 3: 485-501.
- Dotson, Kristie. 2011. “Tracking Epistemic Violence, Tracking Practices of Silencing.” *Hypatia* 26, 2: 236-256.
- . 2012. “A Cautionary Tale: On Limiting Epistemic Oppression.” *Frontiers*. 33, 1: 24-47.
- . 2014. “Conceptualizing Epistemic Oppression.” *Social Epistemology* 28, 2: 115-138.
- Feldman, Richard. 2000. “The Ethics of Belief.” *Philosophy and Phenomenological Research* 60, 3: 667-695.
- Fricker, Miranda. 2007. *Epistemic Injustice*. Oxford: Oxford University Press.
- . 2012a. “Group Testimony? The Making of a Collective Good Informant.” *Philosophy and Phenomenological Research* 84, 2: 249-276.
- . 2012b. “Silence and Institutional Prejudice.” In *Out from the Shadows*, edited by Sharon Crasnow and Anita Superson. Oxford: Oxford University Press.
- . 2013. “Epistemic Justice as a Condition of Political Freedom?” *Synthese* 190: 1317-32.

¹⁴ Pynn (2021) provides an account of the harm of testimonial injustice in terms of degradation. However, the harm of degradation still occurs in the testimonial exchange—when a person publically does not accept the testimony of a knower. (Though, note, Pynn doesn’t think that degradation requires prejudice (Pynn 2021, 166).)

¹⁵ This paper was written during my course on Contemporary Philosophical Issue: Epistemic Injustice. Thanks to my students from their participation and lively discussion. For helpful feedback on the paper, I thank Ben Cross, Harrison Waldo, Steve Wykstra, and an anonymous reviewer.

Timothy Perrine

- . 2015. "Epistemic Contribution as a Central Human Capability." In *The Equal Society*, edited by George Hall. Lanham: Lexington Books.
- . 2016. "Fault and No-fault Responsibility for Epistemic Prejudice." In *The Epistemic Life of Groups*, edited by Michael Brady and Miranda Fricker. Oxford: Oxford University Press.
- . 2017a. "Epistemic Injustice and the Preservation of Ignorance." In *The Epistemic Dimensions of Ignorance*, edited by Rik Peels and Martijn Blaauw. Oxford: Oxford University Press.
- . 2017b. "Evolving Concepts of Epistemic Justice." In Kidd, Medina, and Pohlhaus (2017).
- García, Eric Bayruns. 2021. "On Anticipatory-Epistemic Injustice and the Distinctness of Epistemic-Injustice Phenomena." *Social Epistemology Review and Reply Collective* 10, 7: 48-57.
- Goldberg, Sanford. 2017. "Should Have Known." *Synthese* 194, 8: 2863-2894.
- Hookway, Christopher. 2010. "Some Varieties of Epistemic Injustice." *Episteme* 7, 2: 151-163.
- Kidd, Ian James and Havi Carel. 2017. "Epistemic Injustice and Illness." *Journal of Applied Philosophy* 34, 2: 172-190.
- Kidd, Ian James, José Medina, and Gaile Pohlhaus Jr. eds. 2017. *The Routledge Handbook of Epistemic Injustice*. Routledge: Routledge Press.
- Kornblith, Hilary. 1983. "Justified belief and epistemically responsible action." *The Philosophical Review* 92, 1: 33-48.
- Largent, Mark. 2008. *Breeding Contempt*. New Brunswick: Rutgers University Press.
- Lee, J. Y. 2021a. "Anticipatory Epistemic Injustice." *Social Epistemology* 35, 6: 564-576.
- . 2021b. "On Anticipatory Epistemic Injustice: Replies to Eric Bayruns García and Trystan S. Goetze." *Social Epistemology Review and Reply Collective* 10, 10: 39-42.
- Li, Yi. 2016. "Testimonial Injustice without Prejudice: Considering Cases of Cognitive or Psychological Impairment." *Journal of Social Philosophy* 47, 4: 457-469.
- Maitra, Ishani. 2010. "The Nature of Epistemic Injustice." *Analytic Philosophy* 51, 4: 195-211.
- Ivy, Veronica. 2016. "Epistemic Injustice." *Philosophy Compass* 11, 8: 437-446.
- McGlynn, Aidan. 2021. "Epistemic Objectification as the Primary Harm of Testimonial Injustice." *Episteme* 18, 2: 160-176.

- Medina, José. 2013. *The Epistemology of Resistance*. Oxford: Oxford University Press.
- Nussbaum, Martha. 2000. *Women and Human Development*. Cambridge: Cambridge University Press.
- . 2006. *Frontiers of Justice*. Cambridge: Harvard University Press.
- Pohlhaus Jr., Gaile. 2014. "Discerning the Primary Epistemic Harm in Cases of Testimonial Injustice." *Social Epistemology* 28, 2: 99-114.
- . 2017. "Varieties of Epistemic Injustice." In Kidd, Medina, and Pohlhaus (2017).
- Pynn, Geoff. 2021. "Epistemic Degradation and Testimonial Injustice." In *Applied Epistemology*, edited by Jennifer Lackey. Oxford: Oxford University Press.
- Sen, Amartya. 1987. *Commodities and Capabilities*. Oxford: Oxford University Press.
- . 2009. *The Idea of Justice*. Cambridge: Belknap Press.
- Sherman, Benjamin and Stacey Goguen, eds. 2019. *Overcoming Epistemic Injustice*. Lanham: Rowman & Littlefield.
- Weiland, Jan Willem. 2017. "Evidence One Does Not Possess." *Ergo*. 4, 26: 739-757.

DISPOSITIONAL RELIABILISM AND ITS MERITS

Balder Edmund Ask ZAAR

ABSTRACT: In this article I discuss two counterexamples (the New Evil Demon Problem and Norman's Clairvoyance) to reliabilism and a potential solution: dispositional reliabilism. The latter is a recent addition to the many already-existing varieties of reliabilism and faces some serious problems of its own. I argue here that these problems are surmountable. The resulting central argument of the article aims to demonstrate how viewing reliabilism as an intrinsic dispositional property solves many of the issues facing reliabilism to date.

KEYWORDS: dispositional reliabilism, counterexamples to reliabilism, new evil demon problem, accidental reliability

1. Introduction

Reliabilism can easily be said to capture what is essential about epistemic justification. It does so by emphasizing that justification is about a certain kind of reliable relation to truth such that if one is justified in holding a belief that P, then P is also very likely to be true. Were one to make a stronger claim, for instance by saying that the relation between epistemic justification and truth is such that if one has a belief that P, then P is invariably true, then one is also incapable of handling the very plausible scenario wherein one is justified in holding a belief that P without P being true. Reliabilism effectively solves this problem by offering the next best thing: epistemic justification is when one's true belief is reliably (but not infallibly) produced. This still leaves open the possibility that one's belief is produced with a reliable method, yet that belief is false. Reliabilism thus offers the closest possible relation that a justified belief may be said to hold to truth without thereby making the relation logical or nomological. Reliability alone, however, has also been shown to be neither necessary nor sufficient to account for justification, and so the trouble begins.

This article amounts essentially to a defense of reliabilism by defending 'dispositional reliabilism.' I will begin by introducing in a bit more detail what I call 'standard reliabilism' and two of its more prominent counterexamples in order to then turn to a longer discussion of dispositional reliabilism and attempt to show how the theory is capable of facing the purported counterexamples head on.

The standard reliabilist theory of justification states that a belief is justified if and only if it has been formed by way of a reliable process. Along the same line, reliabilist theories of *knowledge* necessitate that a true belief is the result of a reliable process if it is to count as a state of knowledge (Goldman 2021). For a belief-forming process to be reliable means that it is truth-conducive. That a process is truth-conducive, in turn, means that the process has to have a high probability of producing true beliefs or, put differently, it has to have a high truth ratio; the process has to produce a higher ratio of true beliefs compared to false beliefs. For many, this amounts to the essence of whatever it is that takes us from mere true belief to genuine knowledge. That is to say, the reliabilist wants us to think that regardless of how one conceives of epistemic justification, it is essential that states of justification are truth-conducive, and thus reliable, or else we would have no reason to view having a justified true belief as more valuable than having a mere true belief.

There are two types of direct counterexamples to the reliabilist conception of justification.¹ One of them is the so-called New Evil Demon (NED) Problem, which first appeared in Lehrer and Cohen (1983) and Cohen (1984). The counterexample hinges on what I call the ‘internalist intuition.’ The intuition arises when we consider a world just like ours except for the fact that there is an evil demon that ensures that all the normally reliable processes only engender false beliefs. Their belief-acquiring processes, in other words, are no longer reliable. Yet, the internalist would say, the inhabitants of the demon world are nonetheless justified in holding their beliefs insofar as they are doing things such as appealing to the best available evidence, and so they are still being maximally epistemically responsible, and are, in a sense, still justified in holding various beliefs about their world. What counts as being out of the control of the NED-worlders simply cannot be used to undermine their status of being justified. Now, if the NED-worlders are as justified as we are, then the following *reductio* argument can be constructed to undermine reliabilism:

1. The NED-world inhabitants cannot acquire beliefs reliably (NED-world Stipulation).
2. A belief is justified if and only if it has been formed by way of a reliable process. (Reliabilist Assumption)
3. The NED-world inhabitants’ perceptual beliefs are as justified as our own. (Internalist Intuition)

¹ To my knowledge at the time of writing this.

4. Therefore, the perceptual beliefs² of the NED-world inhabitants have been produced by reliable processes. (1, 4; \perp)

The other counterexample is arrived at through considering worlds where a type of reliable process is merely reliable for seemingly accidental reasons. Consider the following quote by Bonjour (1980, 62):

Norman, under certain conditions that usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power, or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power, under circumstances in which it is completely reliable.

Now what kind of trouble does this cause for the reliabilist? It shows that accidentally reliable processes, the ones that are also highly irresponsible processes of belief-formation, may nonetheless amount to states of being justified simply in virtue of being truth-conducive processes. One may thus use reliable processes to acquire a belief that P without thereby having good reasons to believe that P. This is far from ideal.

Both counterexamples indicate that reliabilism alone is neither necessary nor sufficient to account for whatever it is that brings us from true belief to knowledge. The NED-problem shows us that reliability is not necessary in order for a person to be justified. The problem of mysterious or non-normal reliabilism shows us that reliabilism is not sufficient to bring us to a state of justification or knowledge. As long as there is something accidental or seemingly irresponsible about one's reliably formed belief, that belief cannot be seen as justified. Facing such powerful counterexamples, one would not be amiss to think the reliabilist project to be rather hopeless.

In the proceeding article I am going to attempt to show how adopting dispositional reliabilism undercuts both types of counterexamples and as a result preserves a form of standard reliabilism, albeit a more specified version of it. Let us now make clear what dispositional reliabilism is and how it purports to fix the problems of standard reliabilism.

² Perceptual beliefs are normally taken to be justified, which is why I use them here, but any other kind of belief that we paradigmatically take to be justified or reliably acquired could be used (so instead of a perceptual belief, it could be one arrived at through using sound reasoning, and so on).

2. Why Consider Dispositional Reliabilism?

Dispositional reliabilism arises out of a rather plausible analogy argument, which will be presented shortly. But let us first introduce the general notion and how it serves as an improvement on standard reliabilism and other prominent varieties of reliabilism.

The idea, in short, is to view reliable processes as possessing particular kinds of dispositional properties. Dispositional reliabilism then posits that to use a reliable process is to use a process *disposed* to produce a high ratio of true beliefs. So, for example, to be engaged in a reliable perceptual process resulting in a true belief is to use a process that has the dispositional property to produce a high ratio of true beliefs. This minor modification of standard reliabilism means, according to Baysan (2017), that we need not ‘weaken, relativize, or indexicalize’ (paraphrasing Baysan 2017, 42) standard reliabilism in order to solve the NED-problem (as well as the problems surrounding how to view accidental reliability’s relation to being justified). The kind of relativization and weakening here involves views such as ‘home-world’ or ‘actual-world reliabilism’ (Majors & Sawyer 2005), indexical reliabilism (Comesaña 2002) and normal-world reliabilism (Goldman 1986, 107). Under the category of ‘weakened’ versions of reliabilism I would also include two-concepts responses, such as Goldman’s strong and weak justification (1988) and Sosa’s (2003) apt and adroit justification. A single-concept response would be preferable simply via considering something like Grice’s razor, but the two-concepts responses also fail in their own right. Goldman’s two-concepts response involves something like the following (this exact formulation can be found in Majors & Sawyer 2005, 270):

(S/W) Strong/Weak Justification: Justification consists either in reliability in the world the subject happens to inhabit (strong justification), or in unreliable but cognitively responsible belief (weak justification).

While this view accounts for the internalist intuition, it can be said to fail to satisfy the crucial desideratum of externalist epistemology. Perhaps by calling the form of justification that the internalist intuition appeals to ‘weak,’ Goldman still captures the fact that truth-conduciveness in the world a subject happens to inhabit is what epistemic justification primarily aims for. But in allowing cognitively responsible beliefs (which can exist without any relation to the truth whatsoever) to count as justified beliefs, we seem to have conceded too much. Standing in a particular relation to the way things are is no longer essential to our notion of epistemic justification. We thus end up with two concepts of justification. One (the strong) which leads to a contradiction (see the NED-argument above), another which, to an externalist, is no theory of justification at all, since complete cognitive

responsibility can be in place (as in the NED-world) without an epistemic subject having any sort of relation to the way things are. That is to say, weak justification lacks the requisite truth relation, and as such, is in an externalist framework, no theory of epistemic justification at all. While this kind of conciliatory response has many merits in its own right and might be acceptable given more extended consideration, I take this kind of approach to be, at best, a last resort. As long as we can remain within single-concept theories of justification, we should do precisely that.

With Sosa's two-concepts theory things are not looking any better. The case against it will be presented in brief (it has been convincingly undermined already, cf. Graham 2016 and Majors & Sawyer 2005). Sosa calls the two ways in which one can be justified apt and adroit justification. The former is the justification one has when using intellectual virtues to arrive at beliefs where using the virtues yields a high ratio of true beliefs in the world of usage; the latter is the kind of justification one has when using intellectual virtues that yield a high ratio of true beliefs in the actual world. Consider again the case of a user of clairvoyance in a world where clairvoyance happens to be reliable for accidental reasons. Does Sosa's theory account for this scenario? Adroit justification clearly does not work since clairvoyance is not reliable in the actual world (and actual world reliabilism generally cannot judge whether someone is justified in a non-actual world since they could never in fact be justified merely in virtue of being in the wrong world). Apt justification, on the other hand, is accounted for, but it seems to face the exact same issues that standard reliabilism faces. Accidentally or contingently reliable processes are just not what we typically take to be responsible ways to acquire beliefs. Moreover, a highly irresponsible way of forming a belief cannot be used to justify a belief. Actual-world reliabilism (either adroit or apt) is not able to account for this problem.

Now these types of responses have been brought up in part to show that plurality of concepts does not necessarily lead to a working theory of justification, but also to illustrate that if a single-concept response was indeed possible, it would be doubly preferable; not only in virtue of Grice's razor, but because these two-concepts responses seemingly do not work. Let us now turn to the question of the plausibility of dispositional reliabilism and see how it faces the problems that other varieties of reliabilism are seemingly unable to handle.

3. The Problems Facing Dispositional Reliabilism and How to Move Forward in Spite of Them

Why should we think that dispositional reliabilism is plausible? It is meant to serve as a solution to the NED-problem. But the NED-worlders, by stipulation, are not able to produce true beliefs when using their perceptual processes, so how can a process be disposed to produce true beliefs without being capable of producing beliefs? The underlying assumption supporting dispositional reliabilism is that one can be a bearer of a dispositional property without manifesting the property for entirely contingent reasons, even when these reasons are systematically present. This puts into question what reliabilism actually consists of, and how it arises. Is it an internal state of a person which is central or an internal state mixed with a particular environment? The latter is also the point of tension that Madison (2021) picks up on, which will drive the coming discussion. But let us first go through the analogy argument in favor of dispositional reliabilism to explain how it can overcome the immediate problem noted above regarding how one can be disposed to produce true beliefs but never doing so.

The argument runs roughly like this (ibid., Baysan 2017, 44-45):

- (1) A vase can be fragile without ever breaking simply by never being struck (or going through any other event which could break it).
- (2) Further, there could be vases which never break despite being struck, let us say if there is some kind of magic spell on it which prevents it from breaking when it otherwise would have.
- (3) The vase is nonetheless fragile, since if it were not for the protective spell (a contingent fact about the vase), it would have manifested its fragility by breaking on being struck.
- (4) Similarly, if (3) is a possible state of affairs then the following state of affairs is also possible: say *a* is a reliable belief-forming process; *a* is used by a subject *S*; *a* nonetheless fails to produce true beliefs for *S* and does so systematically.

If the vase is fragile without being breakable, then, similarly, our perceptual processes can be reliable without outputting true beliefs. If this is plausible, it could then solve the NED-problem by using standard reliabilism understood dispositionally. The NED-worlders can then truly be said to be using processes that have the dispositional property of being reliable with the caveat that they are being systematically prevented from manifesting this property by the evil demon. If epistemic justification consists in dispositional reliability, then it also explains why the NED-worlder is as justified as their actual-world counterpart: their perceptual processes share a certain kind of dispositional property. We also solve

the issue of accidental reliability. The person using clairvoyance is in fact not using a reliable process insofar as the process being used does not have the dispositional property of being reliable. Its manifest reliability is entirely contingent. Alternatively, we could view the case of clairvoyant reliability as a case where a world has acquired clairvoyance-waves and receptor or similar strange phenomena that could indeed posit the existence of a non-contingent form of reliability in that world. But in that case, we move further and further away from the accidental nature of such a scenario, and we can no longer claim that such processes are irresponsible to use. Instead, we can now claim that the accidental nature of the reliability is based on the fact that the process lacks a vital dispositional property. So far so good.

But what are we to make of the idea that someone in the NED-world is using reliable processes? That is, how do we specify the notion of reliability so that a process can be said to be reliable without thereby producing a high ratio of true beliefs in certain scenarios? Madison (2021, 197-198) gives two suggestions.

One suggestion is to view reliability as a property ascribable to processes that have a track record of producing true beliefs. The frequency with which the process produces true beliefs compared to false beliefs must then meet some threshold in order to be seen as reliable. Now this does not work since the NED-problem, as Madison points out, ensures that the track record of the processes used by the NED-worlders is such that it has not produced a high ratio of true beliefs. The point is broader, also, in that we cannot take producing a high proportion of true beliefs as either necessary or sufficient for reliability. It is conceivable that a process exists that is reliable, were it to be used, but that nonetheless is never used, therefore having used the process cannot be necessary for ascribing the property of being reliable. Along the same line, Madison (2021, 198) points out that one can use processes that are accidentally or ‘luckily’ truth-conducive – i.e., a student may guess all the answers on a test, through sheer luck be correct in those guesses, and be deemed to use a reliable process (although this is of course based on the type-process one is considering – guessing is not exactly a process that is reliable in general). These arguments may not be impossible to respond to, but together they at least make the plausibility of a functioning frequentist conception of reliability more problematic than its alternative.

These arguments suggest that reliability is better viewed in modal terms. It does not have to be the case that a process has already been shown to produce a high ratio of true beliefs (compared to false beliefs), but it is enough that the process *would* produce a high ratio of true beliefs if it were used.³ A modal

³ Goldman (1976, 771) himself, importantly, early on saw the importance of the counterfactual

conception of reliability, then, states that were one to use a reliable process, then one would yield a high ratio of true to false beliefs. Of course, if there is a posited demon which invariably blocks any perceptual belief from being true, a standard modal conception of reliability cannot work either. Since the NED-worlders are justified, that would mean they are using processes that, if they were used, would yield a high proportion of true beliefs (compared to false beliefs). But, seeing as it is stipulated in the NED-problem that such a favorable ratio of true to false beliefs cannot arise, this amounts to a clear contradiction. Although, with further specification, it is going to be argued here that modal reliabilism, interpreted through a realist dispositional framework with focus on intrinsic dispositional properties, is the best path forward. But let us first continue with Madinson's argument.

It is indeed impossible for our NED-world counterparts to yield true beliefs from their supposedly reliable processes. So what Madison, I believe correctly, points out is that Baysan fails to consider whether the dispositional properties under consideration are extrinsic or intrinsic properties (or a mix of the two). For the vase, the fragility can be said to be an *intrinsic* property of the vase; fragility is a property which comes from the vase's microstructure (as Madison puts it, *ibid.*, 200).⁴ The relational aspects of the vase as it is put under a protective spell, however, make it non-fragile in virtue of its *extrinsic* properties. The assumption Madison operates under here is that dispositional properties stem from the intrinsic

side of reliabilism: "a cognitive mechanism or process is reliable if it not only produces true beliefs in actual situations, but would produce true beliefs, or at least inhibit false beliefs, in relevant counterfactual situations. The theory of knowledge I envisage, then, would contain an important counterfactual component."

⁴ Cf. something like Armstrong's (1993, 87-90) argument against the phenomenalist conception of dispositions. He reaches the conclusion that dispositions need support or arise from non-dispositional states. The argument for this is roughly that if a disposition is not due to intrinsic properties of an entity with a certain disposition, in that from the behaviorist/phenomenalist point of view one does not accept unobservables, then the disposition must be explained by its extrinsic properties (or by contingent connections between 'categorical properties and dispositional properties'). The extrinsic properties, however, are not observable either, which forces the phenomenalist to reject dispositions altogether – a position Armstrong takes to be too extreme. Thus, if we want to accept that things have dispositions, we also have to be realists about them. Being a realist about dispositions, in turn, seems to lead to a view where a thing's dispositional property is determined by its intrinsic properties. As Armstrong puts it (*ibid.*, 88): "Dispositions are seen to be states that actually stand behind their manifestations". It is not an uncontroversial account, see for instance Mellor (1974, 164-5), but it nonetheless is at least plausible that dispositions in an object arise from that object's non-dispositional properties.

properties of a given thing, and that the intrinsic dispositional properties of our cognitive processes are not enough to yield genuine reliability.

Taking this distinction into account, we begin to see a disanalogy between the vase case and the case with the cognitive faculties – or so Madison claims. That is, supposedly two vases (one under a protective spell, one not) can share intrinsic properties and simultaneously have different conditions for manifesting their dispositions depending on the extrinsic properties of each vase. For the case of cognitive processes, Madison writes (*ibid.*):

If two subjects are exact intrinsic duplicates, and have the same belief forming processes, these processes need not be equally reliable – for instance, the subjects might be in radically different environments, as the NED cases make vivid. Whether a process produces true beliefs is partly determined relationally. This means that whether a process is reliable necessarily depends on the environment in which it is used. Baysan seems to implicitly recognize this, as reliable belief-forming processes are described as tending to produce true beliefs, in the right circumstances.

What Madison stresses here is that reliability is not a wholly intrinsic property, and so the relationship between the vase example and the cognitive faculties example is not analogous, and so the argument Baysan posited fails. The result of this is that Baysan seemingly has to argue for a kind of *Modal Reliabilism*, akin to the varieties of reliabilism discussed in section 2 that are basically modified versions of standard reliabilism, relativized to special kinds of possible worlds. As Madison writes (*ibid.*, 201): “In short, reliability is determined not only by the relevant belief forming process, but also the relevant environment, and the Dispositionalist response to the NED problem overlooks this.” And so, the reliabilist has to type-individuate environments, as well as processes, in order to have genuine reliability (or in order to manifest a high ratio of true beliefs as acquired on the basis of using a reliable process).

But here we can question a few of the assumptions made by Madison.⁵ First, let us consider the idea that the reliabilist has to type-individuate environments in order for reliabilism to be plausible. On this point there is a distinction to

⁵ Another approach to Madison’s argument which aims to uphold the analogy between vases and reliability would be to argue that fragility is not a property ascribable only on the basis of the intrinsic properties of an object. In a possible world where ceramic is the hardest material, for example, it seems that it would not be deemed to be fragile in the same sense a cognitive process would not be deemed reliable if there was an evil demon influencing things. Fragility seems to also be dependent on extrinsic factors such as forces and objects that are capable of instantiating the breaking of object disposed to breaking. Therefore, the analogy holds by viewing fragility as a mix of intrinsic and extrinsic factors as well.

highlight, which is the notion of reliability as compared to the notion of having a fully specified relation between a cognitive faculty and the environment it attempts to model, predict, understand, perceive, etc. Part of the virtue of a reliabilist notion of justification is that it is not an infallibilist conception of justification. The relationship between justification and truth need not be one-to-one. All we have to do in order to be justified, according to the reliabilist, is to use cognitive faculties, processes, or methods, that yield a high ratio of true beliefs (in this case, using the modal account, the method *would* yield a high ratio of true beliefs if used). But the requirement that we have to specify the entire process to the point where there is no room for failure, that we have to guarantee that if one uses a certain process, then one is also going to yield a true belief, is not in the spirit of reliabilism as I understand it. What makes reliabilism attractive in part is that it is fallibilistic. That is, it is attractive because it does not require that we specify the full set of extrinsic properties that need to be in place in order to actually yield a true belief in each token use of the process. That a type of process yields a high ratio of true beliefs comes from the fact that the process allows, when the right circumstances are in place, one to gain knowledge of a state of affairs – not that it invariably does so in every conceivable situation. We are not always in favorable circumstances, and so there is no guarantee of gaining knowledge in each token use of a perceptual process (there are many ways in which to hinder a perceptual process from performing its proper function). Can this not be explained by the fact that reliability (of the modal variety) is an *intrinsic* dispositional property? The remainder of this section will attempt to provide reasons to answer this question in the affirmative.

Part of the problem with Madison's account is the assumption that dispositional reliabilism, in order to count as a form of justification, has to involve both intrinsic *and* extrinsic properties. Such a requirement seems to fly in the face of reliabilism insofar as it is normally taken to be a form of fallibilism. Reliabilism allows for error in the following sense: we might be in a situation where we use a type of process that normally produces true beliefs and acquire a false belief, but we need not worry about this as long as the process produces true beliefs in most cases. If we have to guarantee, in any knowledge-seeking activity, that both relevant environmental/relational factors are present as well as making sure that the process involved has the desired dispositional property of being reliable, then it seems that we no longer have a notion of justification which allows for error. If, within the analysis of knowledge or justification, we systematically remove all environmental factors that may lead to error as well as use only processes that have a dispositional property of being reliable, then knowledge presumably cannot be

fallible. Furthermore, one may not ever be able to be in a state wholly devoid of the possibility of doubt and therefore it is hard to see how there could be such a thing as knowledge instantiated in everyday epistemic situations. For a frequentist kind of reliabilism to arise, of course it is true that certain external conditions need to be in place in order for a process to produce mostly true belief. But perhaps what is important in order to be justified, is that one is engaging in the best possible processes one has at hand, i.e., those processes that are disposed to produce true beliefs, even if they end up never producing true beliefs, due to entirely extrinsic factors.

For if we on the other hand view justification as using processes that have the intrinsic dispositional property of being reliable, we seem to reach a theory of justification that explains the fallibility of knowledge as well as the internalist intuition. Reliability, I take it, does not have to be guaranteed by relational properties, it only has to contain the broader possibility of being reliable (by 'broader' here I mean that even if there are worlds in which some processes that are normally reliable are not reliable in those worlds, there nonetheless remains a possibility that they could turn reliable, were the relational properties that make the production of a high ratio of true beliefs impossible to disappear). Whether one has knowledge when one has a justified belief, then, could be said to be a contingent fact both in that it depends on whether the belief is true, but also on whether the intrinsic dispositional property of being reliable is in favorable conditions (that is, in conditions devoid of manifestation-blockers⁶ such as evil demons or blindfolds). These external conditions, however, do not have to be taken as elements in the analysis of knowledge; they seem to come with the fact that epistemic agents inhabit mostly favorable epistemic conditions. So, while Madison is correct that extrinsic and intrinsic properties both need to be in place in order for there to be an observably truth-conducive process, this need not imply that reliabilism as a form of epistemic justification has to be understood as a mix of intrinsic and extrinsic factors. It may just as well be understood as consisting of primarily intrinsic factors; as properties of various cognitive processes or other methodological procedures (properties of various instruments used in experiments, and so on). If we view reliabilism in intrinsic and modal terms, it seems that Baysan's analogy nonetheless works.

There is also an added bonus with intrinsic or internal reliability in that one can analyze the dispositional property conditionally by including anti-masker and anti-mimic clauses (the latter being what happens in the clairvoyance scenario where a process is used but is not really disposed to produce true beliefs in virtue of

⁶ Called a "masker" by Johnston (2012), "antidote" by Bird (1998).

the intrinsic properties of the process, but for seemingly entirely accidental reasons that mimic what a genuinely reliable process does). The NED-problem could then be viewed as a case where a disposition is masked by a demon's interference with the belief-acquiring processes, whereas the clairvoyance problem would then be a case where the disposition is mimicked (but where we can say that the dispositional property is not really there, and so cannot afford justification to a subject using such a process, thereby solving the problem of accidental reliability). To make the formulation of a conditional analysis of dispositional reliability a bit more precise: S is disposed to produce a high ratio of true beliefs when using process X if and only if S produces a high ratio of true beliefs given the use of X and there is no antidote or mimic present. This kind of understanding of reliability seems to preserve the fallibilism of standard reliabilism while undermining both purported counterexamples. It also avoids the criticism by Madison by taking the justification-conferring aspect of a reliable state to be intrinsic factors alone, thus plausibly upholding the analogy to Baysan's vase.

I will propose, then, that reliabilism is best viewed as an intrinsic dispositional property that a process has if it has the capacity to yield a high ratio of true beliefs in virtue of its intrinsic properties. The process, in some sense, has to be shown to be capable to provide a subject with information about the outside world in a way that is not accidental, in order to be viewed as a reliable process; there has to be proof of receptivity. So, we can say that even in the NED-world, the perceptual processes do have the capacity to yield a high ratio of true beliefs, although they cannot exercise this capacity due to the extrinsic properties that the world imposes on them. In virtue of their intrinsic properties, however, the processes are still reliable, but masked. We are now in a position to explain the internalist intuition with the help of an externalist framework, since we can say that the justified status of the NED-worlders is conferrable in virtue of the dispositional properties of their perceptual processes. Whether one is justified is still about factors that need not be present to the mind, and there is still an emphasis on the relation between justification and truth, only now focused more on the use of processes whose intrinsic properties are truth-conducive (in a modal sense).

So, to clarify, if we permit that justification consists in the intrinsic propositional property of being reliable, then the NED-problem is solved. For the same way the vase maintains its intrinsic fragility despite the sorcerer's protective spell, then, we could say that perceptual processes maintain their intrinsic reliability despite being in extraordinarily strange environments that mask their manifestation. On the flipside, we can also say that in the same way that the *fragile*

vase is impossible to break, the reliable processes are incapable of yielding true beliefs (in that possible world). In both cases, these seemingly contradicting facts are merely due to the extrinsic factors involved, and so have no bearing on the dispositional status of the intrinsic properties. If this much can be permitted, we can now ascribe NED-worlders with intrinsic dispositional reliability without contradiction. Let us formulate intrinsic dispositional reliability (IDR):

IDR Epistemic justification consists in using processes that are intrinsically disposed to yield a high ratio of true beliefs.

Let us now input this version into the NED-argument:

1. The NED-world inhabitant cannot acquire beliefs reliably. (NED-world Stipulation).
2. A belief is justified if and only if it was acquired via processes that are intrinsically disposed to yield a high ratio of true beliefs. (IDR)
3. The NED-world inhabitants' beliefs are as justified as our own (Internalist Intuition)
4. Therefore, the justified beliefs of the NED-world inhabitants have been acquired via processes that are intrinsically disposed to yield a high ratio of true beliefs. (3, 2)

Now there is no contradiction – being disposed to manifest a certain property does not mean one invariably does so (especially not in worlds that are epistemically unfavorable). We are also now in a position to better deal with the clairvoyance counterexample, in two separate senses. One problem with the possible world wherein clairvoyance is a reliable belief-forming process is that we would not view using such processes as being justified or responsible. Whereas it is a problem for standard reliabilism that there is reliability without justification, this is not a problem for IDR, since clairvoyance, as a state of mind, is not disposed towards yielding a high ratio of true beliefs in virtue of its intrinsic properties. So, even if clairvoyance happens to be reliable in this world, this cannot be in virtue of the dispositional property, and so we cannot say that the inhabitant of such a world is justified in using clairvoyance as a belief-forming process. With IDR we have a better idea of why it is that the clairvoyance reliability is accidental – it is only in virtue of unspecified extrinsic properties that Norman's clairvoyance is truth-conducive.

Alternatively, in the scenario suggested by Goldman (1988)⁷ where the feeling of clairvoyance is coupled with new natural phenomena (clairvoyance

⁷ To quote him directly (ibid., 62):

waves and clairvoyance wave receptors, for instance, which involves a real causal connection between the feeling and the phenomena in the world). Using clairvoyance in such a world would afford a subject the status of being justified in that the feeling of clairvoyance would indeed be disposed to produce true beliefs in virtue of its intrinsic properties. And so, it seems, the desideratum of having a non-accidental relation between justification and truth can also be maintained.

I would then make the case that intrinsic reliabilism is enough for justification. One is rarely in a position to verify the complete causal relationship between one's cognitive states and the environment; such a requirement would be too demanding. The conjecture here, then, is that in order to be justified it is enough to use a type of process whose intrinsic dispositional property allows the acquiring of true beliefs. While knowledge may consist in having a justified true belief, a justified belief need not be true, and this can be explained by having justification be tantamount to an intrinsic dispositional property of our cognitive faculties. The same way our cognitive faculties are disposed to produce a high ratio of true beliefs even if there is an evil demon systematically deceiving us, sugar is disposed to dissolve in water even if all water in a possible world is at an absolute zero.

The intrinsic dispositional property can now be said to be ascribable to the NED-worlders, yet nonetheless it cannot manifest itself due to the strange circumstances. If one were to remove the relational property of being 'influenced by an evil demon' for the NED-worlder, the intrinsic dispositional property would again manifest itself. In the same way, the vase would break if struck if the protective spell was removed. The analogy seems to hold, all that was needed was to heed the distinction between intrinsic and extrinsic properties (and perhaps reframe the NED-problem as a masking problem for a conditional analysis of

Consider a possible non-normal world *W*, significantly different from ours. In *W* people commonly form beliefs by a process that has a very high truth-ratio in *W*, but would not have a high truth-ratio in normal worlds. Couldn't the beliefs formed by the process in *W* qualify as justified?

To be concrete, let the process be that of forming beliefs in accord with feelings of clairvoyance. Such a process presumably does not have a high truth ratio in the actual world; nor would it have a high truth ratio in normal worlds. But suppose *W* contains clairvoyance waves, analogous to sound or light waves. By means of clairvoyance waves people in *W* accurately detect features of their environment just as we detect features of our environment by light and sound. Surely, the clairvoyance belief-forming processes of people in world *W* can yield justified beliefs.

dispositions and reframe the clairvoyance problem as a mimicking problem). While Madison's criticism was that genuine reliability required some kind of extrinsic property in order to guarantee reliability, the view espoused here is that a frequentist kind of reliability need not be guaranteed in order for one to be justified, instead one only needs to use processes that hold the dispositional property of being reliable, seeing as such a process would be what actually ends up making a frequentist conception of reliability possible. The cognitive faculties we have are reliable intrinsically, but not infallibly. They can even systematically be manipulated. Luckily, in normal conditions in our actual world, as far as we know, this is not the case, and so something like knowledge with all likelihood exists and we need not wade into skeptical waters.

Before concluding this article, I would like to add some comments in support of the analogy between the vase's fragility and our perceptual processes' reliability and formulate an argument in favor of dispositional reliabilism.

4. The Argument for Perpetual Dispositional Masking

In Mellor's *In Defense of Dispositions* (1978) there is relevant distinction between a thing being mortal and a thing being fragile. In the former case, there are implications regarding the future; the thing will die. In the latter, it is not necessary that the thing either has been broken or will break. Dispositions can thus be said to be different to other kinds of properties in a very clear way, i.e. (*ibid.*, 159) in a way that also gives support to Baysan's analogy:

[B]eing forty or mortal now has past or future consequences where being fragile or soluble does not. His past birth being what makes a man forty now, it must have him thirty ten years ago; similarly a man who is mortal now is bound to be mortal until he dies. We draw no such consequences from the present ascription of dispositions. A fragile glass may (or may not) be toughened by heat treatment at any time.

Moreover (*ibid.*, 173):

The safety precautions at our nuclear power station [...] are intended to prevent an explosion by making impossible the conditions in which fuel would explode. It is ridiculous to say that their success robs the fuel of its explosive disposition and thus the precautions of their point.

This seems to add some support to the plausibility that one can have a dispositional property without it ever manifesting. If we accept that stimulus conditions for a certain disposition can be in place without the related disposition manifesting itself, we should also be able to accept that the reason for the failure of the disposition to manifest itself could be there in perpetuity. If we accept this, the

following (similar to Baysan's) argument may be plausibly posited to sum up the discussion:

- (1) Having a disposition is compatible with it failing to manifest despite having met some stimulus condition (the situation may involve some kind of interference with the stimulus conditions, i.e., an evil demon).
- (2) There must be some reason as to why a disposition failed to manifest itself (Realist assumption about dispositions).
- (3) There is a possible world where the reason (the condition) for the disposition's failure to manifest itself is present in perpetuity.
- (4) Thus, there can be a possible world in which one can have a disposition despite its manifestation being impossible in that world.

While it would not be possible for NED-worlders to identify the underlying reliability of their perceptual states (which is stipulated anyway) due to them never manifesting their capacity to yield a high ratio of true beliefs, we can nonetheless know that the NED-worlders are using processes that are intrinsically reliable. We are also able to identify with fairly high precision which of the processes among our cognitive faculties that are disposed to produce a high ratio of true beliefs based on how we experience our own use of them. It seems that regardless of how we approach the metaphysics of dispositions, we should be able to confer the dispositional property of being reliable to the NED-worlders' processes insofar as they are using the same processes that we are using to acquire beliefs about the world. Seeing as the internalist posits that there must be something that confers justification on our beliefs as well as the NED-worlder's beliefs and that whatever does so must be identical in both circumstances, the explanation for this could be that their belief-acquiring processes possess the same type of intrinsic dispositional property.

Similarly, we are often able to identify the maskers of the intrinsic dispositional property of being reliable when, despite using processes with this property, we are left confounded. Maskers of our perceptual processes that are also unidentifiable on the other hand are very rare and need not necessarily factor into the analysis of knowledge (but of course this can be done). Perhaps we simply *should not* include such factors into the analysis of justification or knowledge, on pain of a far too demanding and unrealistic infallibilistic notion of knowledge seeing as we obviously cannot gain knowledge about per definition unknowable deceivers.

So, in order to be epistemically justified, it may very well be enough to use whatever cognitive processes are available to you. Importantly, this is not incompatible with the idea that this is in virtue of the intrinsic dispositional

reliability these processes possess. The contention here is that this is precisely the reason for why an appeal to one's internal states can be associated with epistemic justification to begin with. Were these internal states not disposed to yield a high ratio of true beliefs, they would not afford a subject the status of being justified simply in virtue of being accessible to a cognizer.

If we only consider the internal aspects of a person (their perceptual organs), we run the risk of being in unsuitable or deceptive environments that could threaten the overall truth-conduciveness of the external and internal state we are in. But is this a problem? Can internal reliabilism be enough to account for both justification and knowledge (when it is coupled with a true belief)? Ultimately this seems to come to down whether one can acquire a true belief by using processes that are intrinsically disposed to generate a high ratio of true beliefs without being in a state of knowing. Normally, environments that are highly deceptive or unfavorable only engender false beliefs, and so such examples cannot serve as counterexamples to any standard JTB analysis of knowledge involving intrinsic dispositional properties. But what about barn examples, where one acquires true beliefs through sheer luck by forming the belief 'this is a barn' about the only real barn in a field of barn-façades? The types of scenarios where the process used is clearly generally reliable yet produces an accidentally true belief due to a deceptive environment are quite difficult to handle with this conception of justification. While the accidental reliability of clairvoyance is problematic insofar as it shows that a process is reliable for reasons that have nothing to do with the intrinsic properties of the process itself, the accidental nature of the knowledge one gains of the fact that there is a real barn among many barn facsimiles is due to broader safety considerations. Consider Williamson's (2000, 128) formulation of a safety condition for knowledge (for heuristic purposes understood topologically where α and β are situations similar enough to each other, as it is in the barn scenario where the percepts are similar enough): For all cases α and β , if β is close to α and in α one knows that C obtains, then in β one does not falsely believe that C obtains. Knowledge analyzed into a JTB theory where justification takes the form of intrinsic dispositional properties of being reliable are seemingly unable to handle the barn scenario. If reliability were a mix of extrinsic and intrinsic factors, the barn scenario would simply be ruled out as a case of knowing since visual perception in fake barn county is not exactly reliable. But visual perception is intrinsically disposed to be reliable, and so we have a case of knowledge without meeting the safety condition.

A potential way forward here would be to consider Goldman's (1976) solution where a necessary condition for knowledge, over and above using a

perceptual mechanism (or any intrinsically reliable process), would be to add a condition stating that there cannot be any (ibid., 786) “relevant counterfactual situations in which the same belief would be produced via an equivalent percept and in which the belief would be false.” Such a condition would rule out fake barn counties and stopped clock cases of knowing. But seeing as reliability understood as a mix of extrinsic and intrinsic factors eliminates such cases by type-individuating the environment along with the perceptual process, in the process ruling out perception-in-barn-county or telling-the-time-on-a-stopped-clock-scenarios as cases of knowledge in that they lack reliability, it may be hard to see how an intrinsic dispositional property view of reliabilism with a ‘no relevant counterfactual situations’ clause would be preferable to standard reliabilism. If they are judging equivalently in relevant problematic scenarios, as they seem to be doing, the dispositional view of reliabilism is perhaps still preferable as it avoids the NED-problem and it has the ability to solve certain cases of accidental reliability. As always, as some problems are solved, others arise, and so perhaps it is best to leave a more thorough discussion of the potential problems of intrinsic dispositional reliabilism for another paper. It is nonetheless important to note that if we focus too much on the subject-internal in the notion of justification, problematic results may arise that need to be addressed.

I will now proceed to conclude this paper with some clarifying remarks regarding the notion of an ‘internally reliable’ process and how such a position relates to externalism and internalism of justification more generally, as well as make some comments on the value of justification conceived as an intrinsic dispositional property, and, finally, its relation to naturalist conceptions of knowledge.

5. Concluding Remarks

While this account of justification also faces some problems, it is an interesting result for the following reason. If nothing else, it shows that we can accept the internalist intuition wholesale while remaining externalists about justification. The NED-problem is not a knockdown argument against reliabilist conceptions of knowledge and justification. But what can we say about the suggestion that reliability is a form of truth-conducive capability ascribable to the intrinsic properties of belief-acquiring processes?

To say that reliability is an intrinsic dispositional property does not make it an internalist notion. Whether an intrinsic dispositional property of some cognitive faculty is truth-conducive is ultimately only evaluable based on factors external to the mental content of that cognizer. In extreme cases of cognitive

decline, for instance, one is not in a position to evaluate whether one's faculties are still truth-conducive. This does not mean that in most cases, in normal or ideal conditions, one is not in a position to evaluate the state of one's cognitive faculties with adequate degree of accuracy (for example, people notice if they get something in their eye, precluding them from seeing clearly, and so on). In any case, it is hopefully clear that justification in this view is not internalist, but merely a *subject-internal property*.⁸

Yet this view has some conciliatory value. The central internalist epistemic desideratum (accessibility) can likely, if not be completely accounted for, at least be appealed to, and be afforded plausibility despite insisting on truth-conduciveness as essential for epistemic justification. For instance, one can still maintain that whether we are justified is largely accessible to us (even implicationally so, with the right kind of caveats). Only in this case, it is in virtue of reliability being a property of cognitive faculties which we happen to be consciously aware of in the process of using them (in a broad sense, we know, or are in a position to know, that we are using our visual system when reading, our olfactory system when noticing a scent, etc., and we also know that these are normally reliable). If we maintain that justification is a subject-internal property, this explains the notion of accessibility as an epistemic desideratum in that we normally have some access to our subject-internal states.

Externalism need not imply that justification depends on factors external to the cognizer in a way that by necessity factors the environment into our analysis of justification. Instead, we could say that justification depends on whether one's faculties have certain epistemically valuable properties, such as being disposed to produce a high ratio of true beliefs. Madison (2021) questions the value of reliability as an intrinsic dispositional property. Justification has to have an instrumental value, in the sense that it leads us to truth. The intrinsic dispositional property of being reliable, however, is not a perfect path to truth – the path may be obstructed through various means. So, why would it be epistemically valuable? Here is my answer: Reliable processes do not necessarily yield true beliefs (the reliabilist does not demand a perfect truth ratio) but they nonetheless do so contingently⁹ as long as we are in good cognitive health and use the processes in

⁸ Similar views are expressed by Mulnix (2013), who notes that invoking 'mental states' or the internal properties of a subject is not a rejection of externalism. Similarly, being an internalist does not necessarily mean that the accessibility desideratum pertains to subject-internal properties (e.g., one could be said to be accessing universals or sense data; see Mulnix 2013, 37, also Fumerton 1995, 60-66).

⁹ I separate "contingently" from "accidentally" here, even though they are often used

the kinds of environments we normally inhabit. We only need to include the latter in the analysis of justification if we are aiming for the guarantee that if one is justified, one is tracking truth infallibly. But, as we likely have to accept, there are no such external guarantees – at least not in the actual world.¹⁰ As epistemically valuable states or virtues go, then, it seems that reliability as an intrinsic property of our cognitive faculties may be about as good as it gets. In any case, the fact that there are possible worlds wherein certain dispositional properties are perpetually blocked from manifesting does not mean that these are less valuable for inhabitants of worlds wherein they are not systematically precluded from manifesting themselves. A process which allows us to produce true beliefs in virtue of its intrinsic properties is valuable for precisely that reason.

Something can now be said about responsibility and its relation to epistemic justification, as well. We can follow Mulnix (2013, 47) in denying that responsibility fully exhausts the concept of justification. The similarities between the NED-worlders and us do not have to be that the acquisition of beliefs proceeds responsibly in both worlds. It may lie in the properties of the types of processes being used, therefore we do not have to invoke the notion of responsibility at all or accuse the internalist of conflating responsibility, blamelessness, and epistemic justification.

A desideratum of externalism is that a theory justification should be amenable to naturalization. Is intrinsic dispositional reliability compatible with the naturalization of knowledge? It is not entirely obvious. Insofar as justification here is a dispositional property, we should ask, can this dispositional property be a natural kind? It seems plausible enough. Kornblith's (2002, 61f.) take is that natural kinds are stable homeostatic clusters of properties and as such can factor into causal explanations or inferences based on natural laws. He claims that knowledge is precisely this type of well-behaved category (*ibid.*, 62-63):

The knowledge that members of a species embody is the locus of a homeostatic cluster of properties: true beliefs that are reliably produced, that are instrumental in the production of behavior successful in meeting biological needs and thereby

interchangeably. Whether one is privy to the truth is not only a matter of one's internal states, but it also depends contingently on whether the environment is perceivable and whether one is not precluded from exercising one's perceptual capacities by external influences (intoxication, brain damage, blindfolds, systematic deception by evil demons or barn-builders, etc.). The relation between justification and truth is not accidental, the property is let's say 'designed' to, or has a 'proper function,' to produce a high ratio of true beliefs.

¹⁰ As some argue, "fitness beats truth" (Prakash, et al, 2021). Meaning we are not evolved to know, but to survive, and so we cannot take our cognitive faculties as genuinely truth-conducive.

implicated in the Darwinian explanation of the selective retention of traits. The various information-processing capacities and information-gathering abilities that animals possess are attuned to the animals' environment by natural selection, and it is thus that the category of beliefs that manifest such attunement-cases of knowledge-are rightly seen as a natural category, a natural kind.

Kornblith, however, says knowledge is an ecological kind, consisting of a certain fit between organism and environment. While I agree with this, I think IDR can specify the way in which knowledge can be taken as an ecological kind by separating the internal justificatory aspect from the external (truth) aspect of knowledge. If knowledge is a fit between environment and organism, I take justification to be the internal aspect of this fit. Specifically, justification is equivalent to the properties of organisms that allow them to receive information about their environment. The external part is simply "truth," or the state of the environment at the time of using a reliable process. Given the full analysis of knowledge as a true justified belief, we seem to have the two parts that make up the fit between organism and world in the way Kornblith aims for. On the one end, the organism is using cognitive processes that are disposed to produce a high ratio of true beliefs, on the other, the environment lays before the organism using this process. It seems to me that an organism using such a process – without overt deception going on – would indeed acquire knowledge about its environment. Justification, one could say, is an openness to the world (as Merleau-Ponty often notes regarding perception, cf. 2012, 17). Knowledge arises when the world is not such that it would block or manipulate this openness to the world.

Seeing as there is at least one version of reliabilism that solves the NED-problem as well as the clairvoyance problem, it can at least be concluded that hope is not lost for the externalist. With this view of justification, it becomes possible to avoid relativizing reliabilism to specific types of worlds or conditions. It amounts to a notion of justification simpliciter; a notion of justification applicable in all possible worlds, one that arguably maintains an externalist spirit while heeding the internalist intuition.

This approach is also in line with Graham's (2014) arguments against transglobal reliabilism. He argues that reliabilism need not hold in all, or even most, possible environments in order to amount to justification, or (*ibid.*, 533): "Organisms with more stable predictable natural environments can get by without such learning mechanisms; organisms do not always need transglobally reliable processes to successfully navigate their normal environments." Maybe non-accidental local reliability is good enough and transglobal reliabilism may be too much to ask since regardless of the type of cognitive process we conceive of, we can always conceive of a world in which such a process is not reliable. I believe

this can be taken as further support for the view that intrinsic dispositional reliabilism is what confers justification to a subject or belief. Even if cognitive processes are sometimes systematically blocked from manifesting themselves, these scenarios are not often faced in the actual world, and so should not deter us from viewing cognitive processes that give us information about our environment as ways of acquiring justified, and in most cases, true beliefs.

References

- Armstrong, D. M. 1993. *A Materialist Theory of the Mind* (Rev. ed). Routledge.
- Baysan, U. 2017. "A New Response to the New Evil Demon Problem." *Logos & Episteme* 8(1): 41-45.
- Bird, A. 1998. "Dispositions and Antidotes." *The Philosophical Quarterly* 48: 227–234.
- Cohen, S. 1984. "Justification and Truth." *Philosophical Studies* 46(3): 279-95.
- Comesaña, J. 2002. "The Diagonal and the Demon." *Philosophical Studies* 110(3): 249-266.
- Fumerton, R. A. 1995. *Metaepistemology and Skepticism*. Rowman & Littlefield Publishers.
- Goldman, A. 1976. "Discrimination and Perceptual Knowledge." *Causal Theories of Mind* 174.
- .1988. "Strong and Weak Justification." *Philosophical perspectives* 2 : 51-69.
- .2021. "Reliabilist Epistemology." *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2021/entries/reliabilism/>>.
- Graham, P. 2014. "Against Transglobal Reliabilism." *Philosophical Studies* 169(3): 525-535.
- .2016. "Against Actual-World Reliabilism." In *Performance Epistemology: Foundations and Applications*. Oxford University Press.
- Johnston, M. 1992. "How to Speak of the Colors." *Philosophical studies* 68(3): 221-263.
- Kornblith, H. 2002. *Knowledge and Its Place in Nature*. Oxford University Press.
- Lehrer, K, & S. Cohen. 1983. "Justification, Truth, and Coherence". *Synthese* 55(2): 191-207.
- Madison, B. J. C. 2021. "Reliabilists Should Still Fear the Demon." *Logos & Episteme* 12(2): 193-202.
- Majors, B. & S. Sawyer. 2005. "The Epistemological Argument for Content Externalism." *Philosophical Perspectives* 19: 257-280.

- Mellor, D. H. 1974. "In Defense of Dispositions." *The Philosophical Review* 83(2): 157–181.
- Merleau-Ponty, M. 2012. *Phenomenology of Perception*. Routledge.
- Mulnix, J. W. 2013. "Reliabilism and Demon World Victims." *Tópicos (México)*, (44): 35-82.
- Prakash, C, K. Stephens, D. Hoffman, M. Singh, M, and C. Fields. 2021. "Fitness Beats Truth in the Evolution of Perception." *Acta Biotheoretica* 69(3): 319-341.
- Williamson, T. 2002. *Knowledge and Its Limits*. Oxford University Press.

DISCUSSION NOTES/ DEBATE

NEUTRALIZATION, LEWIS’ DOCTORED CONDITIONAL, OR ANOTHER NOTE ON “A CONNEXIVE CONDITIONAL”

Eric RAIDL¹

ABSTRACT: Günther recently suggested a ‘new’ conditional. This conditional is not new, as already remarked by Wansing and Omori. It is just David Lewis’ forgotten alternative ‘doctored’ conditional and part of a larger class termed neutral conditionals. In this paper, I answer some questions raised by Wansing and Omori, concerning the motivation, the logic, the connexive flavor and contra-classicality of such neutralized conditionals. The main message being: Neutralizing a vacuist conditional avoids (some) paradoxes of strict implication, changes the logic essentially only by Aristotle’s Thesis, makes strong connexivity impossible, and remains in the realm of non-contra-classical logics.

KEYWORDS: neutral conditional, paradoxes of strict implication, paradoxes of material implication, definable conditional, vacuism, connexivity, super-strict Implication, contra-classicality

Wansing and Omori (2022) recently provided some historic and logical context to a proposal by Günther (2022) to define a ‘new’ conditional. The purpose of this note is to add more context and address some of their questions.

Günther proposes to define a conditional $A \Box \Rightarrow B$ by augmenting a Lewisian conditional $A \Box \rightarrow B$ by the possibility of the antecedent. Semantically, the proposal amounts to saying that $A \Box \Rightarrow B$ is true at world w iff the most similar A -worlds are B -worlds *and* there is a most similar A -world. As Wansing and Omori remark, and Günther partly acknowledges, this proposal is not new.

Wansing and Omori trace the account back to Priest (1999, 145). An earlier proposal was made by Burks (1955) (cf. Pizzi 1977, 289-90). In these accounts, the underlying conditional is not a Lewisian conditional but a strict conditional. Following Gherardi and Orlandelli (2021, 2022), I call the resulting conditional (weak) super-strict implication and denote it by \Rightarrow .² The semantic definition here

¹ Eric Raidl’s work was funded by Germany’s Excellence Strategy – EXCNumber 2064/1 – Project number 390727645 and the Baden-Württemberg Foundation.

² Priest also suggested the stronger alternative to add the possibility of the negated consequent.

amounts to saying that $A \Rightarrow B$ is true at world w iff all accessible A -worlds are B -worlds and there is an accessible A -world. From this perspective, “it seems that Günther simply repeats for the Lewis-Stalnaker conditional what Priest suggested for a strict conditional” (Wansing & Omori 2022, 327). But Lewis (1973a, 24–6) himself already suggested to consider $A \Box \Rightarrow C$ as an alternative to his counterfactual $A \Box \rightarrow C$, more than two decades prior to Priest. He called it ‘doctored counterfactual’ (Lewis 1973b, 438). Thus Günther really studies Lewis’ forgotten alternative doctored conditional.³ The same idea was investigated in the related possibilistic and ranking semantics (Benferhat, Dubois, & Prade 1997; Dubois & Prade 1994; Huber 2014; Raidl 2019). Furthermore, the underlying construction is quite general: Add the assumption that the antecedent is possible to your preferred conditional. I will call the result *neutralized conditional*.

Such a general approach was conducted by Raidl (2020). Slightly modifying my previous terminology, let us call *neutralized conditional* \rightarrow any conditional definable from a basic conditional $>$ in the following way

$$A \rightarrow B := (A > B) \wedge \Diamond A,$$

where $\Diamond A := \neg(A > \perp)$ is the so-called outer possibility of $>$.⁴ This is a more general syntactic definition, englobing all previous proposals. The basic conditional $>$ is arbitrary. It need neither be a strict conditional nor a Lewisian conditional, it can be, more generally, some kind of variably strict conditional (as studied by Raidl) or a relevance conditional (as imagined by Priest).

The semantics of a neutralized conditional is as follows: $A \rightarrow B$ is true (or accepted) at world w iff the defining clause for $A > B$ holds at w and the defining clause for $\neg(A > \perp)$ holds at w . The semantics for \rightarrow is only fixed, once the semantics for $>$ is fixed. In a very weak neighborhood (sentence) selection semantics, the defining clause becomes: B is in the A -neighborhood and \perp is not in that neighborhood. A belief reformulation, where the A -neighborhood is interpreted as the set of sentences believed given A , would be: B is believed given A , but \perp is not. If we add some further constraints on neighborhoods or conditional beliefs, a closeness reformulation becomes available: closest A -worlds are B -worlds, and there are closest A -worlds. If closeness is analyzed in a Lewisian sphere semantics, we obtain Lewis’ alternative doctored conditional (as studied by Günther). Possibilistic and ranking theoretic versions can be embedded into such

This was called *strong super-strict implication* by Gherardi and Orlandelli, and *implicative conditional* by Gomes (2020), and Raidl and Gomes (2023).

³ Although Günther does not fix the semantics, he speaks in terms of Lewisian similarity.

⁴ Günther considers the alternative $\Diamond' A := \neg(A > \neg A)$. In his ‘semantics,’ the two are equivalent.

semantics, and if we suppose that there is only one sphere around each world, we obtain a semantics for (reflexive) normal weak super-strict implication. If additionally, the unique sphere is the same for each world, we obtain Priest's (S5-based) proposal. Thus all mentioned proposals are neutralized conditionals. Their underlying conditionals are just of different type or strength.

The main point in Günther (2022), however, is that neutralization is a natural way to 'connexivize' the original conditional. A similar point was made by Priest (1999, §2.5-6). However, Günther's conditional is not connexive, as Wansing and Omori remark, neither is Priest's conditional, nor any neutralized conditional, as I will show. Neutralized conditionals are rather motivated by nullifying vacuism. Instead of making an impossible antecedent conditional vacuously true, as vacuism, the neutralization makes it false. The connexive flavor is a side-effect.

The following sections echo some of the questions raised by Wansing and Omori, and provide some answers. Section 1 motivates neutralization. Section 2 presents logics for neutralizations, in particular for the neutralized weakly centered Lewisian conditional. Section 3 compares the latter to super-strict implication. Section 4 proves that connexivity is impossible for neutralizations, and Section 5 discusses contra-classicality. Non-obvious proofs are collected in the Appendix A.

1. Motivation

What is the motivation behind strengthening a conditional by the possibility of the antecedent?

Günther argues that conditionals with a contradictory antecedent are 'unintelligible' (2022, 58). Wansing and Omori rightly contest. We can very well utter and understand

- (1) If it snows and it does not snow, I am the queen of England.
- (2) If it snows and it does not snow, it snows.

We also reason from a contradiction without complaining about the unintelligibility of that contradiction. The problem of contradictory antecedent conditionals, and more generally, impossible antecedent conditionals, is not so much that we do not use them or that we do not understand them or their antecedents, but that our intuitions with respect to their truth or falsity, as with respect to their logical behavior are less clear than for possible antecedent conditionals.

Consider the following conditionals

- (3) If $1 + 1 = 3$, I'm the queen of England.

(4) If $1 + 1 = 3$, $1 + 1 + 1 = 4$.

According to a relevance-based view, (1) and (3) should be false, since there is no connection between the antecedent and the consequent. But (2) is relevantly judged true. And maybe (4) should be judged true as well. After all, if $1 + 1 = 3$ and $3 + 1 = 4$, then $1 + 1 + 1 = 4$, by adding +1 to each side, so that the (wrong) antecedent equality seems to be relevant to the (equally wrong) consequent equality.

Another view is that impossible antecedent conditionals carry another message than their cousins with possible antecedents. The meaning conveyed by (3) is not that normally or relevantly $1+1=3$ implies that I am the queen of England. Besides mockery, such a conditional rather states that $1+1=3$ is impossible. Let's call this the reductive view. If this were the only meaning, impossible antecedent conditionals like (3) could (and maybe should) be rephrased as simple modal statements, without loss of meaning. But some content seems lost when we rephrase any of the above (1)–(4) by ' $1+1=3$ is impossible', as the relevance's analysis suggests. The consequent contributes to the meaning. But how? Maybe the conditional has an additional performative meaning. The conditional (rather than the modal) statement is used to illustrate the antecedent impossibility by another, often more intuitive impossibility in the consequent. Combining the reductive with the performative reading we obtain that an impossible antecedent conditional expresses the impossibility of the antecedent by illustrating it with another often more intuitive impossibility in the consequent. According to this view, it is (3) which is true (or acceptable), and rather (4) which should be false (or rejected), since in the latter, the consequent impossibility is not more intuitive than the antecedent impossibility. (Similarly (1) is true and (2) is false.)

The above are only two views for impossible antecedent conditionals. The point to present them side-by-side was merely to show that they diverge in their truth evaluation of (3) and (4). Whereas the relevance view judges the first as false and the second as true, the reductive-performative view makes the opposite judgment.

The deviance of impossible antecedent conditionals also concerns their inference behavior. For possible antecedent conditionals, many conditional accounts usually accept the following two laws:

ID	$A > A$	Identity
RW	If $\vdash B \supset C$ then $\vdash (A > B) \supset (A > C)$	Right Weakening

That is, possible antecedents imply themselves and are closed under logical implication. But it is unclear whether these laws transfer to impossible antecedent conditionals. According to relevance, ID holds but RW needs to be drastically restricted. From the reductive-performative perspective, it is ID which fails, but maybe parts of RW can be retained.

We may agree that the meaning and reasoning behavior of impossible antecedent conditionals deviates from their cousins with possible antecedents. But we may disagree on what this deviance is and how to formalize it. There are different options. We might want to judge all impossible antecedent conditionals as true – a position called *vacuism* (Williamson 2007). Conversely, we might want to judge them all as false – called *neutralism* (Raidl 2019, 2020). Hybrid options fall in between: we could suspend judgment and attribute a third truth value (for 'indeterminate'), or we might want to discriminate between some true and some false impossible antecedent conditionals (as in impossible world semantics or in relevance logic). Suitable restrictions of ID and RW will be correlated with such semantic choices. Impossible world semantics, vacuism and relevance logic all agree that impossible and possible antecedent conditionals can be treated in *the same* semantics. But they disagree whether they can be treated in the same way. Impossible world semantics treats impossible antecedent conditionals in a radically different way than possible antecedent conditionals – the former follow almost no law at all (apart from ID). Vacuism and relevance logic, on the other hand, treat both kinds in exactly the same way, the laws in vacuism being inspired by possible antecedent conditionals, whereas the laws in relevance logic are rather inspired by impossible antecedent conditionals. By contrast, I take neutralism to be a proposal for possible antecedent conditionals only, which is either in wait of completion by a suitable extension to impossible antecedent conditionals (if one thinks that the two kinds interact), or which needs to be considered as strictly separated from a theory for the latter (if one thinks that the two kinds don't interact).

Priest (1999) argued for neutralization by the 'cancellation view' of negation. Affirming a sentence and then its negation cancels both affirmations. That is, a sentence joined with its negation ($A \wedge \neg A$) should not entail everything, as in vacuism, nor should it entail something (A and $\neg A$), as in relevance logic, but it should entail nothing. But this restricted 'null view' only motivates neutralism half way. What about other contradictions, and impossibilities? We extend the null view from conjunctive contradictions to classical contradictions if we endorse a form of Left Logical Equivalence. The possibilistic framework based a form of neutralization on this more general null view: classical contradictions should entail

nothing.⁵ But neutralization rests on a much stronger claim which is just neutralism: *Impossible antecedents entail no consequent*. Lewis (1973a, 25) motivated neutralization from neutralism and although adopting vacuism, admitted that he had no decisive argument for choosing the latter.⁶ A similar motivation, based on doxastic considerations, can be found in Raidl (2019).

Neutralism stands in contrast to *Vacuism*. Vacuism treats all impossible antecedent conditionals as true. For a conditional to be vacuist it suffices that it validates ID and RW (and that \supset behaves classically). Let's call such a conditional *pure*. Thus pure conditionals are vacuist. But the reverse need not hold, since similar results can be proven for slightly weaker conditionals, for example where $>$ validates ID and the following deductive version of RW

dRW If $B \vdash C$ then $A > B \vdash A > C$	deductive Right Weakening
--	---------------------------

Most conditionals are pure and hence vacuist, including the material and strict conditional, Lewisian-Stalnaker conditionals and many much weaker variably strict conditionals. Other conditionals are almost pure in that they validate ID and restrict RW (or dRW). Relevance conditionals are almost pure in this sense.

The problem with vacuist and pure conditionals is that they inherit two central paradoxes from strict implication:

AA $\perp > C$	Antilogical Antecedent
IA $\neg \Diamond A \supset (A > C)$	Impossible Antecedent

Almost pure conditionals may validate restricted versions of these.

The neutralization of a pure conditional avoids these paradoxes: it invalidates AA since it validates the negation NAA, and it invalidates IA, since it invalidates the inner scope negation NIA:

⁵ The view is presented by the authors as if it applied to all impossibilities. But in their language, only boolean impossibilities are considered, that is classical contradictions. This is due to the fact that the authors interpret impossibility as having possibility measure 0, where the impossibility measure ranges over a boolean algebra and where additionally only (boolean) contradictions receive possibility 0.

⁶ Lewis (1973b, §9) also highlighted that the doctored conditional is better suited than its vacuist cousin for analyzing conditional obligation (*Given A, it ought C*), temporal conditionals (*When next A, it will C*; *When last A, it was C*), Prior's egocentric relation (*The A is C*).

NAA	$\neg(\perp \rightarrow C)$	No Antilogical Antecedent
NIA	$\neg\Diamond A \supset \neg(A \rightarrow C)$	No Impossible Antecedent

where now the possibility needs to be expressed by $\Diamond A := (A \rightarrow \top)$.

Thus neutralization neutralizes the paradoxes of vacuist conditionals. However, since NIA entails NAA if the modality is normal,⁷ the core axiom here is NIA. Yet NIA is nothing else than an object language expression of neutralism: *impossible antecedent conditionals are false*. And thus, the avoidance of the paradox IA by endorsing NIA is tantamount to adopting neutralism. In this sense, neutralization is the minimal and maybe most natural way to adopt neutralism and avoid the mentioned paradoxes of material and strict implication.

2. The Logic

It remains to be seen, what are the particular implications when we combine the Lewis-Stalnaker conditional with Priest's framework? (Wansing & Omori 2022, 327)

The logical side of this question has been partly answered. Indeed, Raidl (2020) provided a detailed analysis, completeness results included, of neutralized conditionals in various semantics, starting from a very weak neighborhood set-selection semantics all the way up to a Lewisian (non-centered) semantics. Extending the results of that paper, we obtain that

Theorem 1. The following logic, NW, is sound and complete for the neutralized conditional in weakly centered Lewisian models:⁸

MP	If $\Gamma \vdash A$ and $\Gamma \vdash A \supset B$ then $\Gamma \vdash B$	Modus Ponens
LLE	If $\vdash A \equiv B$ then $\vdash (A \rightarrow C) \supset (B \rightarrow C)$	Left Logical Equivalence
RW	If $\vdash A \supset B$ then $\vdash (C \rightarrow A) \supset (C \rightarrow B)$	Right Weakening
<hr/>		
PT	Substitutions of classical tautologies	

⁷ It suffices that $\neg\Diamond\perp$ is valid.

⁸ For a strongly centered semantics, we need to add the debatable law of Conjunctive Sufficiency (CS). If we want to drop \supset from the language, we need to replace MP by the rules for \wedge and \neg , and restate any axiom $X \supset Y$ in rule form $X \vdash Y$, and the rules LLE, RW in deductive form (e.g. RW becomes dRW).

AND	$(A \rightarrow B) \wedge (A \rightarrow C) \supset (A \rightarrow B \wedge C)$	Consequent Conjunction
\Diamond ID	$\Diamond A \supset (A \rightarrow A)$	Possible Identity
AT	$\neg(A \rightarrow \neg A)$	Aristotle's Thesis
OR	$(A \rightarrow C) \wedge (B \rightarrow C) \supset (A \vee B \rightarrow C)$	Antecedent Disjunction
IOR	$(A \rightarrow C) \wedge \neg \Diamond B \supset (A \vee B \rightarrow C)$	Impossible Disjunct
RM	$(A \rightarrow C) \wedge \neg(A \rightarrow \neg B) \supset (A \wedge B \rightarrow C)$	Rational Monotonicity
TID	$T \rightarrow T$	Tautological Identity
MI	$(A \rightarrow C) \supset (A \supset C)$	Material Implication

In this logic, one can further derive:

wBT	$(A \rightarrow B) \supset \neg(A \rightarrow \neg B)$	weak Boethian Thesis
NAA	$\neg(\perp \rightarrow C)$	No Antilogical Antecedent
NAC	$\neg(A \rightarrow \perp)$	No Antilogical Consequent
PA	$(A \rightarrow B) \supset \Diamond A$	Possible Antecedent
N	If $\vdash A$ then $\vdash \Box A$	Necessitation
CM	$(A \rightarrow C) \wedge (A \rightarrow B) \supset (A \wedge B \rightarrow C)$	Cautious Monotonicity

The law wBT follows from AND, RW and AT. NAA follows from RW, \Diamond ID and AT. NAC follows from RW and AT. PA follows from RW. N follows from AT and LLE. CM follows from RM and wBT.

Note that the above neutralized conditional is really Lewis' alternative conditional $\Box \Rightarrow$ in a weakly centered semantics. And as long as we interpret Günther's intuitive talk of similarity in the Lewisian sense, the above is a logic for the Lewisian doctored conditional considered by Günther. To carve out the difference between $\Box \Rightarrow$ and $\Box \rightarrow$, note that Lewis' weakly centered conditional can be axiomatized by replacing \Diamond ID + TID by ID, removing AT [and IOR], but adding CM. AT is invalid for $\Box \rightarrow$, whereas ID is invalid for $\Box \Rightarrow$. Thus the neutralization differs from the original Lewisian conditional in that identity is restricted to tautological and possible antecedents, AT holds, CM is not required, and OR needs the additional help of IOR to make the logic complete.

By the same method, we can analyze neutralizations of weaker conditionals. For example, let's say that $>$ is an *orthodox conditional* if it is ID normal, that is, it validates ID together with the first five principles (MP)–(AND) above.⁹ As corollary to Theorems 6 and 7 from Raidl (2020), we obtain:

⁹ A normal conditional has a *normal conditional logic* in the sense of Chellas (1975), i.e. (MP)–(AND) together with $A > T$, which in the presence of ID becomes redundant due to RW.

Theorem 2. The complete logic of the neutralization of an orthodox $>$ is given by the first 7 principles (MP)–(AT). And (wBT)–(N) remain derivable.

Thus the neutralization differs from the underlying conditional only in adopting AT and restricting ID. In this context, we can equivalently replace AT by wBT or by NAC.¹⁰ And thus AT and wBT are equally at the heart of neutralizing vacuous conditionals. Further strengthenings of the logic for $>$ result in corresponding strengthenings of the logic for \rightarrow . For example, adding OR for $>$ results in adding OR+IOR for \rightarrow , adding RM for $>$ results in adding RM for \rightarrow , adding $\neg(T > \perp)$ for $>$ results in adding TID for \rightarrow , and adding MI for $>$ results in adding MI for \rightarrow . The weakest neutralized logic, E, analyzed by Raidl (2020, p. 148) is given by the first four principles (MP)–(PT) together with NAC. It is the neutralized companion of the (non-normal conditional) logic given by the first four principles together with $A > T$.

3. Comparing Neutralizations

There might be something revealing in working with a Lewis-Stalnaker conditional instead of a strict one, but that is at least not made clear in (Günther 2022). (Wansing & Omori 2022, 327)

What is the difference between neutralizing a strict conditional or a variably strict conditional? To simplify, consider a strict conditional in reflexive normal models (with the modal logic KT). How does its neutralization (the super-strict implication) differ from the neutralization of the previous Lewisian conditional? An axiomatization of super-strict implication with proof of completeness is presented by Gerhardt, Orlandelli and Raidl (2022).¹¹ They use the inner modality $\Box A := (T \rightarrow A)$. An alternative axiomatization consists in simply augmenting the logic from Theorem 1 by the single axiom

IO $\Box A \supset \Box A$

Inner to Outer modality

Theorem 3. The logic NW (from Theorem 1) augmented by IO is sound and complete for the super-strict conditional in reflexive Kripke models.

¹⁰ AT implies NAC by RW. NAC implies wBT by AND. And wBT implies AT by RW and \Diamond ID. Raidl (2020) chose NAC to formalize his neutral conditional logics.

¹¹ These authors also axiomatize neutralizations of some non-normal strict implications.

IO is invalid for the Lewisian neutralization, but the reverse Outer to Inner modality (OI) is valid. Thus both neutralizations just differ by a single axiom.¹²

There are further differences. For example, super-strict implication validates a version of Transitivity, and restricted versions of Contraposition and Strengthening the Antecedent:

wTR	$(A \rightarrow B) \wedge (B \rightarrow C) \supset (A \rightarrow C)$	weak Transitivity
PC	$\Diamond \neg B \wedge (A \rightarrow B) \supset (\neg B \rightarrow \neg A)$	Possibilistic Contraposition
PM	$\Diamond(A \wedge B) \wedge (A \rightarrow C) \supset (A \wedge B \rightarrow C)$	Possibilistic Monotonicity

These are invalid for the neutralized Lewisian conditional.¹³ Simply by construction, super-strict implication is ‘closer’ to strict implication than the neutralized Lewisian conditional, which in turn is closer to its underlying conditional.

4. Impossible Connexivity

Günther’s conditional is *not* connexive. It does, however, have some connexive flavour” (Wansing & Omori 2022, 325)

A conditional is called *connexive*,¹⁴ if it invalidates Symmetry

$$S \quad (A \rightarrow B) \rightarrow (B \rightarrow A),$$

and validates AT and

$$BT \quad (A \rightarrow B) \rightarrow \neg(A \rightarrow \neg B). \quad \text{Boethius Thesis}$$

It is called *Kapsner strong* if the following hold

Unsat1. In no model is $A \rightarrow \neg A$ satisfiable,

Unsat2. In no model are $A \rightarrow B$ and $A \rightarrow \neg B$ satisfiable.

It is *strongly connexive* if it is connexive and Kapsner strong. If negation and \supset are classical, then Unsat1 and Unsat2 are respectively equivalent to AT and wBT. Let’s

¹² This difference really boils down to the underlying conditionals – strict or Lewisian. The inner and outer modality of a Lewisian conditional are distinct: $\Box A = (T > A)$ and $\Box A = (\neg A > \perp)$. These are equivalent for the strict conditional. But otherwise, the latter validates the same principles as a weakly-centered Lewisian conditional.

¹³ An essential difference between (weak) super-strict implication and strong super-strict implication, is that the latter validates Aristotle’s second Thesis (AT2) $(A \rightarrow B) \supset \neg(\neg A \rightarrow B)$, which is invalid for (weak) super-strict implication. For an axiomatization of strong super-strict implication in reflexive Kripke models, see (Raidl & Gomes 2023).

¹⁴ McCall (1963, 1966) and Wansing (2022).

call a conditional *pseudo-connexive* if it invalidates S and validates AT and wBT. It is *strongly pseudo-connexive* if additionally it is Kapsner strong.

Günther's (Lewis' doctored) conditional is not connexive, since it invalidates Boethius' thesis, as noted by Wansing and Omori. However, it is pseudo-connexive and due to classicality of \neg and \supset it is strongly pseudo-connexive.¹⁵

This will hold for many neutralizations of conditionals with a consistent logic. For Unsat2 it suffices that the underlying conditional $>$ validates the deductive version dAND of AND (built in a similar way from AND as dRW from RW) and dRW applied to $B \wedge \neg B \vdash \perp$. For Unsat1, it suffices that $>$ additionally validates ID.¹⁶ For AT it then suffices that additionally \neg is classical, and for wBT it suffices that \supset is also classical. For invalidity of S it suffices that the underlying $>$ validates ID and dRW applied again to $B \wedge \neg B \vdash \perp$.¹⁷ Let's say that $>$ is *conjunctive*, if it validates ID, dAND, and dRW applied to $B \wedge \neg B \vdash \perp$.

Then we obviously have:

Theorem 4. Let \rightarrow be the neutralization of $>$.

- If $>$ is conjunctive, then \rightarrow is Kapsner-strong and invalidates S.
- If additionally \neg, \supset are classical, then \rightarrow is (strongly) pseudo-connexive.

From this perspective, the distinction between pseudo-connexivity and strong pseudo-connexivity (by adding 'Kapsner strong') does not make much sense, since as soon as pseudo-connexivity is ensured by classicality of \neg and \supset , the conditional is automatically Kapsner strong. Thus, from the perspective of neutralizations, one rather approximates connexivity by the following steps: first ensure Unsat2 (by dAND and dRW for $>$), then Unsat1 (by ID for $>$), and thereby invalidity of S. Classicality of \neg, \supset then ensures AT and wBT. Hence rather than being a strengthening of pseudo-connexivity, being 'Kapsner strong' is a precondition of pseudo-connexivity.

From the above result, it follows that the 'connexive flavor' of neutralizations of orthodox conditionals is that they are strongly pseudo-connexive. One might think that we then only have one step to go to obtain a connexive conditional: add Boethius' thesis. However this is impossible:

Theorem 5. Adding BT to a pure neutralized conditional logic is inconsistent.

¹⁵ That \Rightarrow validates AT, the deductive version of wBT, and some other principles was noted by Priest (1999).

¹⁶ If one takes the alternative outer modality, Unsat1 follows by definition, but Unsat2 requires dRW additionally.

¹⁷ The special case $((\perp > \perp) \wedge \neg(\perp > \perp)) > ((\perp > \perp) \wedge \neg(\perp > \perp))$ of ID suffices.

Proof. A pure conditional is given by MP, PT, RW, ID. The neutralization of a pure conditional still validates AT, PA, and N. BT implies $\Diamond (A \rightarrow C)$ for any A , C , by PA. Thus $\Diamond (T \rightarrow \perp)$. But $\neg(T \rightarrow \perp)$ by AT. Hence $\Box \neg(T \rightarrow \perp)$ by N. That is $\neg\Diamond(T \rightarrow \perp)$. QED.

It's not just that neutralization does not give us new insights into connexivity, connexivity is incompatible with neutralization. BT is not only invalid, but strongly invalid, since any BT extension of a pure neutralized conditional logic is inconsistent. For the same reason, neutralized conditionals will (strongly) invalidate any nested law of the form $(A \rightarrow B) \rightarrow C$. The strong invalidity of S and BT fall into the same basket. The problem concerns a vast class of neutralized conditionals. Only neutralizations of impure conditionals (non-ID or non-RW) escape. But impure conditionals don't create the vacuist problems (AA, IA) for the avoidance of which neutralization was conceived in the first place! The only comfort we may take in neutralized conditionals (apart from being pseudo-connexive), is maybe that they validate the outer-scope version of BT

$$\text{oBT. } \neg((A \rightarrow B) \rightarrow (A \rightarrow \neg B)) \quad \text{outer scope Boethian Thesis}$$

For this \Diamond ID and wBT [i.e. AT, AND, RW] suffice.

The more intricate worry about connexivity is as follows. The combination of the standard principles RW and ID is incompatible with AT and also with wBT. Indeed, if ID would hold, $\perp \rightarrow \perp$ would hold and by RW $\perp \rightarrow T$ would hold. But this contradicts AT (it also contradicts wBT). Thus upholding ID and RW together is not compatible with AT (nor with wBT). Hence either ID or RW need to go, for a connexive conditional. Neutralization restricts ID but keeps RW, the result being that it makes connexivization impossible (Theorem 5). Thus, maybe if we have learned something it is that neutralization will not help in the study of connexive logic, and that ultimately, we should better explore the route where we keep ID but drop or restrict RW. This is basically the relevantist route.

5. Contra-Classicality?

If neutralization does not lead to (strong) connexivity, then at least, it may be one way of exploring contra-classical logics, as Wansing and Omori suggest.

[...] a simple variant of Lewis conditional will bring us to the realm of contra-classical logics (cf. (Humberstone 2000)). The same applies to the variants of strict implications explored by Gherardi and Orlandelli, and this seems to be a simple and interesting route to contra-classicality. (Wansing & Omori 2022, 326-7)

I will argue that this is only true in a very restricted sense, and that contra-classicality is not the appropriate notion to characterize logics of neutralizations (or related constructions).

In general, a neutralized logic, say NL, arises from a companion conditional logic L for some underlying conditional \succ . The neutralized logics considered here are extensions of classical propositional logic CL (since L extends classical logic), thus they are *not* contra-classical in the sense of being incompatible with classical logic. They verify if $\vdash_{\text{CL}} \alpha$ then $\vdash_{\text{NL}} \alpha$ and also the converse for α a classical sentence. However, the neutralized logics are contra-classical in another, very strict sense: Call t *the literal translation* if the conditional \rightarrow is translated into the material conditional \supset and t preserves Booleans and propositional variables. A propositional logic S with a new conditional-like connective \rightarrow is *literally contra-classical* iff the literal translation t does not satisfy

$$\text{If } \vdash_S \alpha \text{ then } \vdash_{\text{CL}} t(\alpha) \quad (5.1)$$

The neutralized logics are literally contra-classical, since AT (or wBT) is literally translation resistant, i.e., it is derivable in the neutralized logic, but classically invalid under the literal translation, and thus not classically derivable. Thus \rightarrow cannot receive the classical material conditional interpretation. But literal contra-classicality is not the notion Wansing and Omori had in mind.

A propositional logic is *contra-classical* iff it is not a sublogic of classical propositional logic, not even modulo a translation which preserves propositional variables. Yet contra-classicality without some restriction (called 'profound') is too restrictive since it reduces to the notion of inconsistency (Humberstone 2000, Proposition 1.1). But we can require the translation to preserve Booleans (\neg , \wedge , \vee , \supset , \top , \perp), and speak of contra-classicality *modulo Booleans*. Literal contra-classicality is a special case, and Humberstone's notion of contra-classicality (modulo Booleans) simply extends literal contra-classicality by testing (5.1) for other translations than the literal one. What is really being tested thereby is whether \rightarrow can receive any classical interpretation at all. But the neutralized logics are not contra-classical in this sense either, as we will now see.

Neutralizations are definable conditional constructions from some basic conditional \succ (Raidl 2020, 2021). This is to say that there is a translation \circ from the language of \rightarrow to the language of \succ , preserving Booleans and propositional variables and such that scheme (5.1) holds from NL to L, modulo \circ . The translation of neutralizations arises naturally by using the semantic definition. It is induced from

$$(A \rightarrow B)^\circ := (A^\circ \succ B^\circ) \wedge \neg(A^\circ \succ \perp)$$

meaning that all standard connectives are normally translated, and propositional variables remain untranslated. Thus the translation preserves Booleans. Furthermore, one can prove that if $\vdash_{\text{NL}} \alpha$ then $\vdash_{\text{L}} \alpha^\circ$. In the above terminology: NL is not contra-L modulo Booleans.

Whether NL is contra-classical modulo Booleans reduces to the question of whether L is contra-classical modulo Booleans. But neither the Lewisean weakly centered logic (VW), nor the logic of normal strict implication are contra-classical modulo Booleans. We can indeed translate the Lewisean $>$ into \supset – denote the translation $\#$ – and satisfy (5.1) for $S = \text{VW}$ and $t = \#$.

That is, $>$ can be interpreted classically and in fact literally (although this is not the intended interpretation). (Similarly for a normal strict implication.) Chaining \circ and $\#$, we then obtain that \rightarrow translates into \wedge and $t = \circ\#$ still respects (5.1) for $S = \text{NL}$. Hence \rightarrow can also be interpreted classically, but not literally, and the \wedge -interpretation is of course not the intended one.¹⁸ Hence the neutralized logics are *not* contra-classical (modulo Booleans), either. A similar remark holds in general for other conditional constructions out of normal conditionals.¹⁹

Overall, neutralization does not generate contra-classical logics out of logics which are not contra-classical. Contra-classicality of the conditional construction may at best be inherited from the underlying conditional, not from the construction. If at all, neutralization allows to construct new contra-classical logics from already existing contra-classical logics.

An example is the neutralization of an S6 strict implication. The modal logic S6 can be seen as S2 augmented by the axiom $\neg\Box\Box A$. Gherardi, Orlandelli, and Raidl (2022) present a complete axiomatization (ST2) of the neutralization of S2 strict implication. The neutralization of S6 strict implication only requires to add the axiom $\neg\Box\Box A$ (that is $\neg(T \rightarrow (T \rightarrow A))$) to ST2. Since S6 is a consistent contra-classical modal logic (Humberstone 2000, Proposition 2.1), the neutralization is also consistent and contra-classical. The reason here is the backtranslation \bullet of \Box into super-strict implication, induced by $(\Box A)^\bullet = \Box A^\bullet$. We have: if $\vdash_{\text{S6}} A$ then $\vdash_{\text{ST6}} A^\bullet$, analogously to Lemma 2 of Gherardi, Orlandelli, and Raidl (2022) for S2 and ST2. Thus if ST6 were not contra-classical, then we would have a translation T , such that $\vdash_{\text{ST6}} B$ implies $\vdash_{\text{CL}} T(B)$, and hence a translation $t' = \bullet T$, such that (5.1) holds for $t = t'$ and $S = \text{S6}$. But then S6 would not be contra-classical, contrary to

¹⁸ The \wedge -interpretation can however be used to show that NL is consistent (has a model), and to find non-derivable formulas.

¹⁹ This is analogous to Humberstone's remark that there are no consistent normal modal logics which are contra-classical modulo Booleans.

Humberstone's result. Contra-classicality is here due to the non-congruentiality, which is transferred from S6 to its neutralization. In short, the neutralization of an S6 strict conditional has no classical truth-functional interpretation whatsoever.

Finally, if we slightly stretch the notion of classicality and count first order logic as classical, then we lose contra-classicality altogether. As long as the underlying conditional is first order translatable, the neutralized conditional is as well. One then obtains that if $\vdash_{NL} \alpha$ then $\Gamma \vdash_{FOL} \forall x \alpha^*$, under suitable assumptions Γ on the relations used for the first-order translation.²⁰ In particular, since the Lewisian conditional and strict implication are first order translatable, the conditional construct is also first order translatable. Thus these conditional constructions are not contra-classical in the first order sense either.

For these reasons, I see definable conditional constructions rather as a way to explore semantic strengthenings (or weakenings) or mixtures of existing conditionals. The conditional construction comes immediately with a proper axiom for the definable construction. For the neutralized conditional, the proper axiom is AT, or wBT, or NAC (depending on how one sees it). In view of this and Theorem 4, neutralization is essentially pseudo-connexivization, but nothing more on the connexive hierarchy, by Theorem 5.

A. Proofs

Proof of Theorem 1. Raidl (2020, Corollary 1) proved that the logic, say NV, given by MP, PT, LLE, RW, AND, NAC, \Diamond ID, \Box M, OR, IOR, RM is sound and complete for the neutralized conditional in Lewisian models (where \Box M is the monotonicity axiom $\Diamond A \supset \Diamond(A \vee B)$). By the same proof procedure, we can obtain a complete logic for weakly centered Lewisian models. For this it suffices to recall that (1) the weakly centered Lewisian conditional has the logic VW and extends the logic V of the Lewisian conditional by the axiom MI, and that (2) the backtranslate of MI is of the form $((A \rightarrow B) \vee \neg(A \rightarrow T)) \supset (A \supset B)$ and can be decomposed into MI and $\neg(A \rightarrow T) \supset \neg A$, the contraposed of which is $A \supset (A \rightarrow T)$. From this TID follows. Conversely TID and MI together with the remaining axioms imply $A \supset (A \rightarrow T)$: Assume A . Thus $\neg(T \supset \neg A)$. Hence $\neg(T \rightarrow \neg A)$ by MI. But $T \rightarrow T$ by TID. Thus $A \rightarrow$

²⁰ For the first order translation of a KT strict conditional we need to assume that R is reflexive. For a first order translation of a (weakly centered) Lewisian conditional, we need to encode the semantic assumptions on the accessibility relation R and the similarity relation ($R'xyz$ iff $y \lesssim_x z$) in first order language – the binary relation R is reflexive, and the ternary R' when restricted to its first component $R'x$ is a total preorder over R -accessible points from x , such that Rwv implies $R'wv$. All these constraints are first order definable.

Eric Raidl

T by RM and LLE. Hence NV+MI+TID is sound and complete for \rightarrow in weakly centered Lewisian models.

It now suffices to show that NV+MI+TID is equivalent to our NW.

First we show that we can derive PA and AT from NV+MI+TID.

PA. Suppose $A \rightarrow B$. Hence $A \rightarrow T$ by RW. This is $\Diamond A$.

AT. Suppose $A \rightarrow \neg A$. Then $\Diamond A$ by PA. Thus $A \rightarrow A$ by \Diamond ID. Hence $A \rightarrow \perp$ by AND. This contradicts NAC. Therefore $\neg(A \rightarrow \neg A)$.

Second let us conversely show that our NW derives NAC and \Box M.

NAC. Suppose $A \rightarrow \perp$. Hence $A \rightarrow \neg A$ by RW. This contradicts AT. Hence $\neg(A \rightarrow \perp)$.

\Box M. Suppose $A \rightarrow T$. If $\neg(B \rightarrow T)$, that is $\neg\Diamond B$, then $A \vee B \rightarrow T$ by IOR. If on the other hand $B \rightarrow T$, then $A \vee B \rightarrow T$ by OR. QED.

Proof of Theorem 3. Gherardi, Orlandelli, and Raidl (2022, Theorem 18) proved that the following logic, SST, for super-strict implication is sound and complete in reflexive Kripke models: MP, PT, LLE, RW, AT, \Diamond PA, INC, AND, TID, SPRES, \Box T, where

$(A \rightarrow B) \supset \Diamond A$	\Diamond PA
$(A \rightarrow B) \supset \Box(A \supset B)$	INC
$\Box(A \supset B) \wedge \Diamond A \supset (A \rightarrow B)$	SPRES
$\Box A \supset A$	\Box T

We show that SST is equivalent to NW+IO (i.e. replacing \Diamond PA, INC, SPRES, \Box T by \Diamond ID, OR, IOR, RM, MI, IO).

First we show that \Diamond ID, OR, IOR, RM, MI, IO are derivable in SST. \Diamond ID, OR, RM, MI were shown derivable (Gherardi et al., 2022, Lemma 11). It remains to derive \Diamond ID, IOR, IO, and OI.

IO. We show the contraposed $\Diamond A \supset \Diamond A$. Assume $\Diamond A$. That is $A \rightarrow T$. Hence $\Diamond A$ by \Diamond PA.

OI. We show the contraposed $\Diamond A \supset \Diamond A$. Assume $\Diamond A$. That is $\neg(T \rightarrow \neg A)$. But $T \rightarrow T$ by TID. Thus $A \rightarrow T$ by RM. This is $\Diamond A$.

\Diamond ID. Assume $\Diamond A$. Thus $\Diamond A$ by IO. Hence $A \rightarrow A$ by \Diamond ID.

IOR. Suppose $A \rightarrow C$ and $\neg\Diamond B$. Thus $\Box(A \supset C)$ by INC, $\neg\Diamond B$ by OI, and $\Diamond A$ by \Diamond PA. From $\neg\Diamond B$ we obtain $\Box\neg B$ and hence $\Box(B \supset C)$, by standard reasoning with \Box (a KT necessity). Thus also $\Box(A \vee B \supset C)$, and $\Diamond(A \vee B)$, again by standard reasoning with \Box . Hence $A \vee B \rightarrow C$ by SPRES.

Second, and conversely, let us derive \Diamond PA, INC, SPRES, \Box T from NW+IO.

\Diamond PA. Suppose $A \rightarrow B$. Thus $\Diamond A$ by PA (i.e. RW). Hence $\Diamond A$, contraposing IO.

INC. Suppose $A \rightarrow B$. Thus $A \rightarrow (A \supset B)$ by RW. If $\neg\Diamond\neg A$, then $T \rightarrow (A \supset B)$ by IOR and LLE. If $\Diamond\neg A$, then $\neg A \rightarrow \neg A$ by \Diamond ID. Hence $\neg A \rightarrow (A \supset B)$ by RW. Therefore $T \rightarrow (A \supset B)$ by OR. Thus overall $\Box(A \supset B)$.

\Box T. Suppose $T \rightarrow A$. Hence $T \supset A$ by MI. That is A .

SPRES. Suppose $T \rightarrow (A \supset B)$ and $\neg(T \rightarrow \neg A)$. Then $A \rightarrow (A \supset B)$ by RM. Hence $\Diamond A$ by PA. Thus $A \rightarrow A$ by \Diamond ID. Therefore $A \rightarrow (A \wedge B)$ by AND and RW. Hence $A \rightarrow B$ by RW again. QED.

References

- Benferhat, S., Didier Dubois, and Henri Prade. 1997. "Nonmonotonic reasoning, conditional objects and possibility theory." *Artificial Intelligence* 92: 259–276.
- Burks, Arthur W. 1955. "Dispositional statements." *Philosophy of Science* 22(3): 175–193.
- Chellas, Brian F. 1975. "Basic conditional logic." *Journal of Philosophical Logic* 4(2): 133–153.
- Dubois, Didier, and Henri Prade. 1994. "Conditional objects as non-monotonic consequence relations." In *Principles of knowledge representation and reasoning*, edited by Jon Doyle, Erik Sandewall, and Pietro Torasso, 170–177. San Francisco: Morgan Kaufmann.
- Gherardi, Guido, and Eugenio Orlandelli. 2021. "Super-strict implications." *Bulletin of the Section of Logic* 50(1): 1–34.
- (2022). Non-normal super-strict implications. In *Proceedings of the 10th international conference on non-classical logics. Theory and applications*, edited by Andrzej Indrzejczak and Michał Zawidzki, 1–11. Sydney: EPTCS.
- Gherardi, Guido, Eugenio Orlandelli and Eric Raidl. 2022. "Proof systems for super-strict implication." Forthcoming in *Studia Logica*.
- Gomes, Gilberto. 2020. "Concessive conditionals without *Even if* and nonconcessive conditionals with *Even if*." *Acta Analytica* 35:1–21.
- Günther, Mario. 2022. "A connexive conditional." *Logos & Episteme* 13(1): 55–63.
- Huber, Franz. 2014. "New Foundations for Counterfactuals." *Synthese* 191: 2167–2193.
- Humberstone, Lloyd. 2000. "Contra-classical logics." *Australasian Journal of Philosophy* 78(4): 438–474.
- Lewis, David. 1973a. *Counterfactuals*. Oxford: Blackwell.

Eric Raidl

- Lewis, David. 1973b. "Counterfactuals and Comparative Possibility." *Journal of Philosophical Logic* 2(4): 418–446.
- McCall, Storrs. 1963. *Non-classical propositional calculi*. Ph. D. thesis. Oxford University.
- . 1966. "Connexive implication." *The Journal of Symbolic Logic* 31(3): 415–433.
- Pizzi, Claudio. 1977. "Boethius' thesis and conditional logic." *Journal of Philosophical Logic* 6(1): 283–302.
- Priest, Graham. 1999. "Negation as cancellation, and connexive logic." *Topoi* 18: 141–148.
- Raidl, Eric. 2019. "Completeness for counter-doxa conditionals – using ranking semantics." *The Review of Symbolic Logic* 12(4): 861–891.
- . 2020. "Strengthened conditionals." In *Context, conflict and reasoning. Logic in Asia Series*, edited by Beishui Liao and Yi N. Wang, 139–155. Singapore: Springer.
- . 2021. "Definable Conditionals." *Topoi* 40: 87–105.
- Raidl, Eric, and Gilberto Gomes. 2023. "The implicative conditional." *Journal of Philosophical Logic*, forthcoming.
- Wansing, Heinrich. 2022. "Connexive Logic." In *The Stanford encyclopedia of philosophy* (Summer 2022 ed.), edited by E. N. Zalta. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2022/entries/logic-connexive/>
- Wansing, Heinrich, and Hitoshi Omori. 2022. "A note on "a connexive conditional"." *Logos & Episteme* 13 (3): 325–328.
- Williamson, Timothy. 2007. *The philosophy of philosophy*. Oxford: Oxford University Press.

ERRATUM NOTICE

There is an error in the article “Rational Decision-Making in a Complex World: Towards an Instrumental, Yet Embodied, Account,” authored by Ragnar van der Merwe and published in the previous issue of *Logos & Episteme*. The second last sentence of section 2.2 on page 393 reads “It does apply to generic agents (including AI systems and aliens perhaps) engaged in (Turing machine-like) decision-making simpliciter.” The sentence should read “It does *not* apply to generic agents (including AI systems and aliens perhaps) engaged in (Turing machine-like) decision-making simpliciter.”

NOTES ON THE CONTRIBUTORS

Leandro De Brasi obtained his PhD at King's College London and is Associate Professor of Philosophy at the Universidad de La Frontera (Chile). His research interests include social, legal and political epistemology and normative theory of democracy. He is currently directing the State-funded Fondecyt project "Epistemic Pathologies of the Public Sphere" (1210724). Contact: Leandro.debrasi@ufrontera.cl.

Rauf Oran has a Bachelor's degree in physics from Middle East Technical University, Turkey, and a Master's degree in Chinese philosophy from Xiamen University, China. Currently, he is a Ph.D. student in Western philosophy at Xiamen University. His current research is mainly focused on the epistemology of testimony. Aside from that, he is interested in the philosophy of science, phenomenology, philosophy of language and Chinese philosophy. Contact: narofuar@163.com.

Timothy Perrine is currently a Post-doctoral Researcher at Rutgers University. He works primarily in Epistemology, Value Theory, and Philosophy of Religion. He has published over two dozen papers in peer-reviewed academic journals. His current research focuses on the experience of "Divine Presence"—experiences whereby people feel a divine being present to them. Working with an interdisciplinary team of anthropologists and cognitive science, he hopes to explore the philosophical significance of these experiences for religious belief. Contact: timperrine@gmail.com.

Eric Raidl is a research fellow at the University of Tübingen (Germany) within the Excellence Cluster "Machine Learning for Science" and project leader of AI, Trustworthiness and Explainability. He works on non-classical logics, formal epistemology and philosophy of AI. He wrote his PhD on objective probability and his habilitation on conditional logics, including explanatory and justificatory relations. He uses these to investigate issues in Machine Learning. Contact: Eric.Raidl@uni-tuebingen.de.

Jack Warman is a FONDECYT Postdoctoral Researcher in the Department of Social Sciences at the University of La Frontera in Temuco, Chile. He was awarded his PhD in Philosophy at the University of York, UK in 2020. He is primarily

Logos and Episteme

interested in social epistemology and the ethics of belief. He is also interested in the philosophy of mental health. Contact: jack.warman@alumni.york.ac.uk. URL: <https://philpeople.org/profiles/jack-warman>.

Balder Edmund Ask Zaar is a Lund University graduate specializing in Theoretical Philosophy. His research interests are primarily within epistemology and philosophy of language, with a particular interest in externalist theories of knowledge and justification. The paper in the current issue was first written as part of a master's thesis written at Lund under the supervision of Professor Erik J. Olsson. Contact: balderaskzaar@gmail.com.

NOTES TO CONTRIBUTORS

1. Accepted Submissions

The journal accepts for publication articles, discussion notes and book reviews.

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

There are no submission fees or page charges for our journal.

2. Publication Ethics

The journal accepts for publication papers submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The submitted papers should be the author's own work. All (and only) persons who have a reasonable claim to authorship must be named as co-authors.

The papers suspected of plagiarism, self-plagiarism, redundant publications, unwarranted ('honorary') authorship, unwarranted citations, omitting relevant citations, citing sources that were not read, participation in citation groups (and/or other forms of scholarly misconduct) or the papers containing racist and sexist (or any other kind of offensive, abusive, defamatory, obscene or fraudulent) opinions will be rejected. The authors will be informed about the reasons of the rejection. The editors of *Logos & Episteme* reserve the right to take any other legitimate sanctions against the authors proven of scholarly misconduct (such as refusing all future submissions belonging to these authors).

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*.rtf) or Microsoft Word document (*.docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author's affiliation and professional postal and email address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles which pass the editorial review, with the exception of articles from the invited contributors, will be subject to a strict double anonymous-review process. Therefore the authors should avoid in

their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it.

The submissions will be sent to at least two reviewers recognized as specialists in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

The review process is intended to be as quick as possible and to take no more than three months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors.

The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.

The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations.

11. Checking Proofs

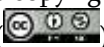
Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice și Sociale „Gh. Zane” Academia Română, Filiala Iași, Str. Teodor Codrescu, Nr. 2, 700481, Iași, România.

The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Copyright & Publishing Rights

The journal holds copyright and publishing rights under the terms listed by the CC BY-NC License (). Authors have the right to use, reuse and build upon their papers for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.

No manuscripts will be returned to their authors. The journal does not pay royalties.

14. Electronic Archives

The journal is archived on the Romanian Academy, Iasi Branch web page. The electronic archives of *Logos & Episteme* are also freely available on Philosophy Documentation Center web page.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears at the end of March, June, September, and December. Its fundamental mission is to support philosophical research on human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal publishes articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal promotes all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the Continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

Logos & Episteme is published and financed by the "Gheorghe Zane" Institute for Economic and Social Research of The Romanian Academy, Iasi Branch. The publication is free of any fees or charges.

For further information, please see the **Notes to Contributors**.

Contact: logosandepisteme@yahoo.com.