# Science and Values: A Philosophical Perspective on the Justifiability of Evidence based Policymaking

O. Çağlar Dede

Dede, .O. Ç.

*Science and Values: A philosophical perspective on the justifiability of evidence based policymaking*



Cover design by Ender Yıldızhan, www.behance.net/enderyildizhan

Printed by Print Service Ede

Science and Values:

A philosophical perspective on the justifiability of evidence based policymaking

Wetenschap en Waarden:

Een filosofisch perspectief op de rechtvaardiging van *evidence based policy*

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the

Rector Magnificus

**Prof. dr. A.L. Bredenoord**

and in accordance with the decision of the Doctorate Board.

**The public defence shall be held on**

**Wednesday the 8th of December 2021 at 15:30 hrs**.

by

**Osman Çağlar Dede**

born in Istanbul, Turkey

**Erasmus University Rotterdam**

**Doctoral Committee:**

**Promotor:** Prof. dr. J.J. Vromen

**Other members:** Prof. dr. R.E. Backhouse

Prof. dr. I.P. van Staveren

Prof. dr. G. Irzık

**Copromotor:** Dr. H.C.K. Heilmann

*To Deniz, Sema, Seda, Rabia and Fikret*

# Table of Contents

# Preface

I pursued a bachelor's degree in economics and had my fair share of excelling in the mathematical modeling of macroeconomic dynamics, individual and public choice and welfare. Though I secretly enjoyed the fancy and analytical wonders of mathematical economics, I often was discontented with the curriculum of my programme. My sentiment was that we-the-students had little to contribute positively to society with the knowledge we had been acquiring. As a curious and critically-minded twenty-something, I was hoping to intervene in what I then perceived as the wrongs in the world, understand and change the institutions that shape people's economic behavior. Back then, I could not grasp how to do things with economics, so I entertained the possibility of following elective courses that offer alternative perspectives to studying economies and economics. I followed courses in social policy, political economy, sociology of markets, and the history of thought with the vague hope that I would acquire some less abstract and more applicable pieces of knowledge about societal change. Intriguingly, these courses were all about criticizing economics rather than offering alternatives to it. Contrary to my sentiment regarding the impracticality of economics to make a positive contribution to society, the professors I have met during these courses often held the opinion that economics is too powerful as a science, and the experts of this supposedly "dismal" science entertain a great deal of influence in shaping the social world we live in.

This background has shaped my early academic interests. I was curious about how economics does things in the world and why and how one would be critical about it. I was initially inspired by the sociologists of economics (the participants of the so-called "performativity of economics" literature) who claim that economists do not only study markets but also actively create them (MacKenzie 2006). Back then, experimental and behavioral economics was on the rise, and some commentators claimed that experiments in economics were not only used for testing economic theories and but also for building markets and creating agents suitable for applying

economic theories to the practice (Guala 2005, 2007). I started to develop some philosophical taste in these subjects and was wondering whether economists were really engaging in creating a sort of moral agency; something like the infamous *homo economicus*; the perfectly rational, utility-maximizer man. Indeed, the behavioral policies such as nudges and auction design practices, which seemed to go in this direction, were also becoming popular and catching philosophers' attention. And, I was interested in understanding whether the influence of economists over our lives is somewhat objectionable with respect to freedom and democracy. With the encouragement of my academic mentors at Boğaziçi University (Yahya Mete Madra, Fikret Adaman, and Emrah Aydınonat), I have moved to the Netherlands to pursue research in these subjects at the very research-center of philosophy of economics; *Erasmus Institute for Philosophy and Economics* (EIPE).

At EIPE, I had written an MA thesis on the methodological and ethical presuppositions of the nudge policies, which are based on the pieces of evidence assembled by behavioral economists to change people's behaviors. I was then kindly offered a bursary to keep philosophizing about economists', and more generally, scientists' relation to society.

During these formative years, researching behavioral policies from a philosophical perspective has shaped the main theme of this dissertation: *the conditions for the justifiability of evidence based policy*. I was gradually convinced that a healthy and fruitful way to explore sciences' impact on society would be to focus on the use of scientific knowledge in policymaking. If behavioral economists are doing things in the world and shape our behaviors, they do so by applying their findings to the policy arena and by being able to shape the policy agendas and frameworks. Though this may sound like a straightforward premise, it was not initially obvious to me or to many humanities scholars who may be unfamiliar with the field of philosophy of economics and hold a negative political judgment about the hegemonic influence of economists (e.g., "economics functions like an ideology"). After years of studies in philosophy of science, I have endorsed the opinion that an ideology-centered analysis of how sciences influence people does some injustice to its subject matter and is

devoid of suitable tools and perspectives to critically reflect on the nature of science and the pragmatic roles it plays in society. Tracking sciences' influence in the world as instances of evidence based policymaking; however, equips one to look closely at the content of scientific evidence and how they are used for developing and justifying policy interventions. It also puts one in a better position to examine what roles scientists' interests, viewpoints, or social positions may play during the episodes where scientific practices influence the management of current affairs.

By the time I started my doctoral research, evidence based policy was a hot topic among the philosophers of economics and biomedical sciences. Philosophers I have met who were working on evidence based policy (such as Julian Reiss, Federica Russo, Nancy Cartwright, Catarina Marchionni) were mainly interested in the discovery of causal mechanisms applicable to policymaking (e.g. Reiss 2008); or in Nancy Cartwright's terms, "hunting causes and using them" (Cartwright 2007). Accordingly, the philosophical approach to evidence based policy was rather a narrow and technical one focusing predominantly on the methodological problem of "external validity" of scientific evidence (particularly those gathered through the lab and field experiments). However illuminating and important these philosophical reflections on evidence based policy might be, they do not respond to the broader and more classical philosophical puzzles regarding interactions between science and society that I have been setting myself to reflect upon: that is, the evaluation of evidence based policies with respect to moral, ethical, and political (non-epistemic) desiderata such as enhancement of the freedom or welfare. The target of this doctoral thesis has thus been to offer a perspective from the philosophy of science to the analysis of evidence based policy's broader success.

Pursuing this line of doctoral research has been transformative and truly rewarding: I have learned a great deal from the scholarship on the interface between science and policymaking. The "values in science" literature in philosophy of science has supplied the key philosophical perspective and the conceptual tools needed for the development of this project. Thanks to the academic events hosted in EIPE or elsewhere in the Netherlands, I have had the unique chance to meet some of the most inspiring

philosophers who have contributed to our understanding of values in science, such as *Anna Alexandrova, Alison Wylie, Heather Douglas, and Philip Kitcher*. I have also been quite lucky to have some travel budget attached to my bursary for participating in the relevant academic events around Europe as a young scholar. The international conferences and workshops have supplied me with the invaluable academic environment required to develop my thinking and deepen my research on the subject matter of this PhD thesis. These types of scholarly experiences have not only educated me in the field of philosophy of science but also helped me develop published academic articles, book chapters, and publishable manuscripts contributing to the ongoing discussions on science, values, and empirically informed decision-making. These research outputs have, in turn, formed the content of this PhD thesis.

To conclude, I do wish to express how grateful I am to the many individuals and institutions who have contributed to the successful completion of this project and the vitalization of my ongoing intellectual journey. Let me undertake this task in the *Acknowledgment*.

## References


Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge University Press.

Guala, F. (2007). How to do things with experimental economics. In Donald MacKenzie, Fabian Muniesa, & Lucia Siu (Eds.). *Do economists make markets? On the performativity of economics*. Princeton University Press.

MacKenzie, D. A. (2006). *An engine, not a camera: How financial models shape markets.* MIT Press.

Reiss, J. (2008). *Error in economics: Towards a more evidence-based methodology*. Routledge.

# Acknowledgments

It takes a village to raise a child properly. It takes a complex web of interactions and collaborations to complete a doctoral dissertation. It takes some individual, interpersonal, and institutional experience, dedication, passion, perseverance, and care to prepare one to be an independent researcher in a worthy but challenging field of human inquiry. I feel lucky to be the carrier of these collective efforts and the receiver of my peers' criticisms, insights, and experiences; of the relevant service providers' care and support; and of the family and friends' companionship, kindness, patience, and love. I feel honored to finally express my gratitude to everyone who has contributed to the development of my academic studies in one way or another. In customary fashion, this acknowledgment is devoted to thanking some of those special individuals and highlighting their contributions.

I should start with my early academic mentors: I am indebted to *Yahya Mete Madra* for broadening my intellectual horizon and encouraging me to follow my curiosity and pursue a career in research. I also would like to thank *Fikret Adaman* for offering me excellent and comprehensive advice on the graduate programs suitable for reflecting critically on economics and the pros and cons of moving abroad for the purpose of obtaining a PhD. Finally, I am grateful to *Emrah Aydınonat* for introducing me with the field of history and philosophy of economics, supporting me in joining EIPE for graduate studies, inviting me to give talks in Istanbul and Helsinki, and still being an insightful, supportive, and cool colleague!

Of course, this thesis could not have come together without the invaluable support and advice of my co-supervisors: *Jack Vromen* and *Conrad Heilmann*. *Jack* has been the voice of critical reason and academic excellence throughout the process. He has taught me that there is no satisfactory endpoint in one's philosophical thinking and writing: the deeper you dive, the more interesting -but more challenging- it could get. Apart from his humbling feedback and advice on my research and manuscripts, I have learned a great deal from his wise realism, quest for sharpness, wide-ranging

knowledge in philosophy, and the history of economic thought. I also very much enjoyed our friendly chats and discussions after research seminars. *Conrad* has not only been my supervisor but also served as my academic mentor since the beginning of my master's in philosophy of economics. He has immensely contributed to the development of my philosophical thinking, writing, and presentation skills. He has encouraged me to give talks in a variety of academic contexts, publish articles, write referee reports, and aim very high. He has patiently read my early and advanced drafts and provided me with useful, considerate, and quick feedback. I have benefitted greatly from *Conrad*'s professionalism, meticulousness, and conscientiousness. Our conversations have educated me a great deal. Thanks, *Jack* and *Conrad,* for your contributions, feedback, support, and patience with me!

I am also grateful to all the scholars who commented on my work throughout the years at various academic events. Thanks to the audiences at the EIPE PhD seminars, including research master's students and the PhD researchers of EIPE. I also thank to the audiences at various Helsinki, EPSA, INEM, ENPOSS, NordicPSA, DEMOCRASCI workshops and conferences, as well as graduate workshops at OZSW, LSE, UvA, and VU. These events have been inspiring and instrumental for me to learn more about how to engage with the world of philosophers as a young scholar and contribute to publishing some of the chapters in this thesis. I particularly would like to thank *Anna Alexandrova*, *Erik Angner*, *Christopher Clarke*, *John B. Davis*, *Ben Baumberg Geiger*, *Daniel Hausman, Lisa Herzog*, *Till Grüne-Yanoff*, *Francesco Guala*, *Philip Kitcher*, *Inkeri Koskinen*, *Magdelena Malecka*, *Michiru Nagatsu*, *Julian Reiss, Federica Russo, Eric Schliessler,* and *Johanna Thoma* for their responsiveness, encouragement, and feedback. Special thanks to *Alessandra Arcuri* and *Florin Coman-Kund* for offering me the opportunity to discuss the philosophical values-in-science perspective with the international socio-legal scholars of science and technology and contribute to their edited volume. I am also grateful to the anonymous referees and the editors of the *British Journal for Philosophy of Science*, *European Journal for Philosophy of Science*, *Journal of Economic Methodology*, and *Perspectives on Science.*

Special thanks to the members of the examination committee, *Roger Backhouse*, *Irene van Staveren*, *Gürol Irzık*, *Heather Douglas*, *Anna Alexandrova*, and *Christopher Clarke*, for agreeing to evaluate my doctoral thesis. I appreciate their time and effort, and interest in my research.

I also would like to express my thankfulness to my colleagues at EIPE. *Constanze Binder*, it has always been a pleasure and inspiration to work and chat with you! *Christopher Clarke*, thanks for your advice and feedback and for being always cheerful, attentive, and enthusiastic philosopher of science. *Erwin Dekker*, many thanks for your kind willingness to keep me "in the conversation." It has always been instructive to collaborate with you on several projects, including the in-person interview with *Bob Sugden*.

A subset of my colleagues at EIPE were the co-editors of the *Erasmus Journal for Philosophy and Economics* (*EJPE*). I have served as an editor of this serious PhD-run journal for a few years. During this time, I have collaborated with *Tom Wells*, *Luis Flores-Mireles*, *Philippe Verreault-Julien*, *Willem van der Deijl*, *Huub Brouwer*, and *Erwin Dekker*. Thanks to them, I have learned a great deal about the editorial processes for publishing academic articles, as well as communication, collaboration, and decision-making practices for managing an academic journal. Special thanks to *Tom* for mentoring me (and his dedication to the journal!) and *Philippe* for his guidance, collegiality, and hospitality.

Some of my colleagues at the *EJPE* were also my PhD peers on the 4th floor of the (old) H building. We have co-experienced the PhD candidature in different times with *Willem*, *Philippe*, *Jasper*, *Melissa*, *Huub*, *James*, *Attilia*, *Sine*, *Daphne*, *Akshath*, *Vaios* (and later with *Lennart*, *Måns*, *Annalisa*, *Erica* at the new building). I thank each one of them for their companionship during this journey.

Some of my PhD peers have become my sisters: Thanks, *Melissa Vergara Fernández*, for the genuine exchanges regarding our shared academic and personal challenges; *Sine Bağatur* for always being kind, helpful, guiding, and welcoming; *Attilia Ruzzene*

for being remarkably caring, attentive and inspiring. Perhaps I should also name it here: *Attilia* is the *honorary paranymph* of all the international PhDs of EIPE!

Some have become my brothers: *Vaios Koliofotis* has contributed a great deal to my intellectual and personal development. He has helped me find my way around since my early days in the Netherlands and been supportive immensely all these years. Thank you so much for the sustained and reliable brotherhood *Vaios*!

Some PhD peers have become my friends and comrades. *James Grayot* has been my officemate and "wowed" me a lot of times with his linguistic skills, humor, creativity, and intellectual passion. It has been a complex joyful experience to go through this journey together with *James* as it has been evolving into a genuine friendship: we have seen the best and worst in us and still been able to call each other "zutie pie." I hope we will stay friends! *Daphne Truijens* had played all these supportive roles when I needed them the most; hence her contribution is invaluable to me. Thank you, *Daphne*, for your generosity, support, and encouragement (also for promptly translating in Dutch my "*samenvatting*"). I wish our fellowship could have been full-time and started much earlier. Special thanks to *Daphne* and *James* for accepting to be my "*paranimpen*" and staying behind me one more time during my defense ceremony!

An indispensable but often under-acknowledged contribution to PhD theses is the service provided by the university staff and the professionals of care. I am grateful to the librarian *Judith Gulpers* and the staff of *Erasmus Graduate School of Social Sciences* and the *Erasmus School of Philosophy*. Special thanks to *Liesbet van Zoonen* for leading the university services dedicated to the PhD students in humanities. I have always found her creative, inclusive, and caring leadership remarkable and inspiring. Thanks to *Iris* and *Ellen* for the service they are offering for the well being and personal development of the early career researchers. Thanks to *Lena Schots* for always being kind and responsive to my practical questions regarding my teaching. Thanks to *Ticia Herold* for welcoming me to EIPE and EUR and helping me claim exemption from the mandatory military service in Turkey.

I also would like to thank my new colleagues at *Vrije University Amsterdam*. I feel very lucky and honored to join the research team of the ERC-funded project "*Social Epistemology of Argumentation*" as a postdoctoral researcher. It has already been a enormous pleasure and intellectually fulfilling to work with you, *Catarina Dutilh Novaes*, *Hein Duijf*, *Merel Talbi*, *Elias Anttila*, as well as the other associates of the project *César F. dos Santos, Colin Rittberg, Silvia Ivani, and Thirza Lagewaard*. I am indebted to you for giving me a genuine opportunity to do further research on the social epistemology of science and informed decision-making in the years to come!

An acknowledgment for a PhD dissertation naturally ends with expressing huge thanks to one's family and close friends. I would like to thank all those people who have contributed to the completion of this project through their friendship, companionship, kindness, patience, and love. *Barış*, *Pietro*, *Tej*: it has been wonderful to start with you but then become very challenging after you had left! I have been missing you so much in all these years. Dear members of the "Turkish diaspora": *Ayça*, *Berke*, *Cansu*, *Didem*, *Dilege*, *Efe*, *Erdem*, *Nesrin*, *Nur*, *Pelin*, *Sine*, *Umut*; I am grateful to each of you for your companionship and support throughout our shared journey. Thanks to all the other friends and fellows who have made my life richer in years past: *Akshath*, *Batu & Ece*, *Cor*, *Dinçer*, *Erinç & Kate*, *Hannie*, *Jeroen*, *Jolene*, *Marteen*, *Merve Dilara*, *Luis*, *Oh-Yoon*, *Öğüt family: Meltem, Özlem, Sabiş & Mehmet, Serap, Şafak Hocam, Uğur, Yıldız family: Ozan, Muzaffer & Gülseren.* Special thanks to *Efe Soyman*, who has not only sometimes proofread my drafts but also guided me with his experiences and insights being a wonderful friend. Thanks to BEYLORD (*Anıl, Ender, Sabun*) for accommodating my unavailability during the years of doctoral research and always feeding me with some healthy dosage of genuine and warming humor whenever I needed some.

Special thanks to my gifted friend *Ender Yıldızhan* for the beautiful cover art.

Last but not least: *Deniz* has been kind, loving, supportive, and patient from the start till the end. Deniz's contribution to the completion of this thesis is huge. *Deniz*, I very much appreciate your angelic presence in my life.

To my family (*Sema, Seda, Rabia, Fikret, and Deniz*): I hope this thesis and the resulting achievement could serve as a proper tribute to your awesomeness, altruism, and kindness. Would I even be able to move to the Netherlands and let alone complete a PhD had I not had such a self-sacrificing and loving family? This project is dedicated to you.

# Introduction

Science is widely regarded as the most reliable epistemic source of providing knowledge about the world. Policymakers intend to make purposeful changes in the world. The practice of policymakers relying on scientific experts to make informed decisions about which policies to implement is called *Evidence Based Policymaking*. This thesis provides a perspective from the philosophy of science in order to discuss the justifiability of Evidence Based Policymaking (EBP) with respect to broadly democratic and liberal values.

Justifying EBP with broadly democratic and liberal values entails that the practices of EBP promote, or at least are in harmony with, values such as democratic governance and enhancement of people's freedom and autonomy. Identifying the conditions under which practices of EBP meet such desiderata minimally requires an understanding of how sciences and scientific experts are instrumental in realizing the public's values, needs, interests, and pursuit of freedom. In order to approach this project, the thesis adopts a philosophical perspective to conceptualize how sciences are supposed to be guided by or promote society's values, needs, and interests. Specifically, it adopts a perspective from the philosophy of science that focuses on the relationship between science and (societal) values.

The kind of philosophy of science perspective on "values in science" that this thesis adopts here has two overarching pursuits relevant for the project of the thesis. Firstly, it seeks to inform the debates about which values and non-epistemic considerations are supposed to inform scientific research. For instance, it discusses the proper sources/owners of the non-epistemic desiderata that inform scientific research and the proper social mechanisms to identify these non-epistemic desiderata (e.g. Kourany, 2010; Kitcher, 2011). Second, it offers theories of the non-epistemic values' proper roles in scientific reasoning and research that specify how their involvement in science does not unduly compromise the epistemic pursuits of science. The values-in-science perspective thereby seeks to balance the instrumental value of science (i.e., its use to pursue certain societal projects and values) with its epistemic authority (i.e., its objectivity, non-dogmatism, and reliability). The thesis advances an understanding of EBP from the perspective of the values in science by addressing issues that come to

the fore when EBP is acknowledged as a value-laden practice of informed decision-making.

The thesis is structured in two parts which consist of self-standing research papers addressing the philosophical issues pertaining to both the theme of the respective part and the overall thesis. In Part I (entitled *Values in Science: Beyond the "Value Free Ideal" of Science*), the thesis engages with the philosophy of science literature on the values in science, illuminates how this literature relates to EBP, and discusses how the contemporary philosophical approaches describe the roles of values in scientific reasoning and how these values may change in relation to the social contexts within which they arise. In Part II (entitled *Philosophy of Evidence Based Policy from the Perspective of Values in Science*), the thesis focuses on the issues regarding the justifiability of EBP from the values-in-science perspective.

In the remainder of the introduction, I will describe and motivate the two parts of the thesis in further detail and offer a preview of the chapters constituting the thesis.

# Part I: Values in Science: Beyond the "Value Free Ideal" of Science

The traditional and the most commonsensical view on the values in science is the so-called "value free ideal of science" (VFI) (see Douglas, 2009 for an extensive review). The VFI suggests that scientists should not consider non-epistemic aspects of their research (e.g., values, societal consequences) when reaching scientific conclusions. Non-epistemic considerations could play plausible roles in science in "the context of discovery" (e.g., when choosing research questions and identifying which hypotheses deserve to be assessed scientifically), but not in "the context of justification." The latter includes methodological steps such as data characterization and assessment of claims/hypotheses based on evidence. In the face of the problems such as evidential uncertainty or underdetermination, values could play an acceptable role in making inferences, but these values are supposed by the proponents of the VFI to be "epistemic values" such as scope, simplicity, unificatory power, consistency, and

fruitfulness (as formulated in Kuhn, 1977). While epistemic values could legitimately shape scientists' conclusions, the involvement of the non-epistemic values such as ethical and moral values may corrupt the core phases of scientific thinking, or so the proponents of the VFI argue. As this brief description of the VFI suggests, the VFI provides a clear view of the proper limits to science's promotion of the public's certain values, needs, and interests.

The VFI is not just an account of values' legitimacy in science. The VFI also motivates a view on science's place in a democratic society. According to this view, science best contributes to society by seeking objective knowledge independently from societal concerns and assembling facts to inform normative policy deliberations. This view not only emphasizes the pragmatic significance of the public's reliance on scientific experts to make the right policy decisions but also promotes the idea that the judgments regarding normative matters (e.g., what is acceptable, what values, and whose should be promoted) should be made by democratically legitimate political actors and be engaged in the political domain (see Betz, 2013 for a review). Accordingly, the VFI also offers an approach to how scientists should properly identify the social values or needs that guide their research pursuits.

The VFI is a cogent *regulative* ideal, especially in the context where we reflect on how sciences contribute to policymaking, even though it may otherwise be an unattainable ideal in the actual practices of science. Indeed, the VFI has represented the dominant philosophical approach to the science-policy interface during the 20$^{th}$ century (Douglas, 2009). Nevertheless, the normative appeal of the VFI in the context of informed decision making has been challenged by prominent philosophers of science in recent decades (e.g. Anderson, 2004; Douglas, 2009, Kitcher, 2011). The main argument against the VFI emphasizes that science does not provide the decisionmakers with *certainty*. Even though science is the most reliable epistemic source into the state of affairs in the world (and how to make predictable changes in it), scientific judgments always face the risk of being wrong, and they are supposed to change over time through the discovery of new pieces of evidence and theories. Such non-dogmatic character of science is indeed one of the most important reasons

behind science's epistemic success, hence its reliability for making decisions (Douglas, 2009, Chapter 1). In the face of inductive risks, scientists who inform policy decisions are in a position to make morally consequential judgments (see Douglas 2009 for an extensive argument from the "inductive risk" to the VFI and Biddle & Kukla, 2017 for a review of "epistemic risks" that generalizes the arguments regarding inductive risk in science). For instance, when available evidence does not fully confirm a policy-relevant hypothesis, scientists could reach a judgment about the acceptability of the hypothesis only after they make a choice on what kind of evidence is relevant and how much of evidence is sufficient to accept or reject the hypothesis. Making this choice requires considering the non-epistemic context of the decision (e.g., how severe the moral and social consequences of the decision is, whether there is more time to work further on the question). Suppose the judgment is going to justify an urgent policy decision, and the consequences of rejecting the hypothesis are too severe and morally unacceptable. In that case, one may plausibly want to accept the hypothesis based on relatively lower evidential standards (Douglas, 2000). Consequently, in the face of such less-than-ideal circumstances for scientifically informed decision making, presupposing that the scientific experts provide the society with value free and universally true statements does not unambiguously lend itself to the intelligent use of science for policy making.

In the light of these kinds of philosophical considerations, it seems plausible to reject or at least transcend the VFI. However, transcending the VFI and adopting a less ideal (and a more practice-oriented) conception of science in society raises two main philosophical challenges for a discussion on the justifiability of the EBP in a democratic context. Firstly, the observation that scientific input for policymaking is uncertain and laden with (or incomplete without) making non-epistemic value judgments seems to challenge the very idea of the public's deference to scientific experts to make informed decisions. From the perspective of the so-called "science wars," for instance, going beyond the VFI may be understood as a challenge to the epistemic authority of science, but that is not a premise upon which one could justify the practices of EBP. By definition, the latter rests on the idea that scientific evidence should guide policy decisions because science is the most reliable epistemic resource

to make informed decisions and the value-ladenness of science may corrupt the reliability of science (e.g. makes it more like politics). Nevertheless, the philosophical approaches to values-in-science do not aim to challenge the epistemic authority of science for decision-making when they question the plausibility of the VFI. Instead, just like the VFI, they aim to define the proper roles of values in science (and for EBP) such that scientific judgments are epistemically authoritative (i.e., they are objective in a meaningful sense and reliable and trustworthy).

The second challenge of the alternative accounts of values in science concerns the appropriate content and the sources of the value judgments that inform scientific research. The scientific input for decision-making may presuppose certain value judgments, but it is unclear where these value judgments (should) come from. Yet, in a democratic context, it is essential to identify who the proper owners of the values that inform scientific research are and how the content of value judgments are determined, or why they change over time. The sources and the contents of the values that direct scientific research should be identifiable and transparent so that the society and the researchers could critically evaluate them and explore alternative social mechanisms, procedures, principles, and institutions for determining the values that shape EBP.

These two challenges are conceptualized in the thesis through a distinction between the *legitimacy problem of values* and *the authority problem of values* in science. The former problem is about defining the legitimate roles value presuppositions play in making and communicating value-laden scientific judgments. The latter problem is about defining the cogent sources and the content of value presuppositions.

The first part of the thesis is dedicated to introducing the particular philosophical perspective (i.e., the values in science perspective) to approach and understand the justifiability of the EBP in a democratic context. It introduces the values in science perspective to approach issues in EBP, reviews the philosophical landscape on the subject, and identifies the challenges and potentials of the contemporary philosophical approaches to the values in science in the light of the debate on the VFI (see below for a preview of **Chapter 1**). It also investigates how our best accounts of values can

apply to describe the values that shape paradigmatic evidence based policy practices (specifically through which social mechanisms and procedures scientists' evaluations of the inductive risks change over time) (see below for a preview of **Chapter 2**).

# Part 2: Philosophy of evidence based policy from the perspective of values in science

The second part of the thesis is devoted to analyzing philosophical issues in Evidence Based Policymaking (EBP) from the perspective of the values in science. Here, I briefly discuss how the values-in-science perspective contributes to the philosophical approaches to EBP and describe the philosophical issues the thesis addresses from this perspective.

The extant philosophical literature on the EBP mainly focuses on the question of "what works" in the light of a critical analysis of how scientific evidence should be interpreted and used in the right way for policymaking purposes (e.g., Cartwright & Hardie, 2012). Philosophers often conceptualize this as discovering causal relations applicable to policymaking and adequately applying them through extrapolation (i.e., "hunting causes and using them" in Cartwright's terms). The focus of this philosophical approach, therefore, is on the "external validity" of the scientific findings and the kind of evidence required for making externally valid (policy warranting) inferences. In the past decade, philosophers extensively questioned the limitations of the dominant methods used in the practices of evidence based medicine and social policies (see, for instance, Deaton & Cartwright, 2018 for an exhaustive analysis of the limitations of the randomized controlled trials for justifying policy claims). This philosophical perspective on EBP helps us understand how to produce and use scientific evidence *more effectively* (so that evidence based policies *work*).

The thesis complements and goes beyond the extant philosophical approach to EBP by offering an understanding of how to produce and use scientific evidence in ways that are *more responsive* to the values, interests, and needs of the members of the

public who are at the receiving end of the evidence based policies. This change of perspective shifts the focus of the philosophical discussion from the issue of "how to properly apply science for policy making purposes" to that of "how to properly inform and guide science for policy making purposes."

The second part of the thesis is dedicated to understand and address some of the philosophical issues that come to the fore when one asks "how to properly inform and guide science for policy making purposes." It discusses how to select the appropriate methods for evidence based policy analysis to be responsive to specific non-epistemic purposes (see below for a preview of **Chapter 3**). It reflects on how to treat value-laden scientific judgments in the context of informed decision-making such that these judgments are considered objective, reliable, trustworthy, *and* congruent with the requirements of democratic policymaking (see below for a preview of **Chapter 4**). Finally, the second part of the thesis explores how value judgments that inform EBP can be properly identified. It applies philosophical insights on the determination of values that guide scientific research, and discusses how the philosophical perspective offered can advance the normative justifiability of evidence based policy practices (see below for a preview of **Chapter 5**).

# Preview of the chapters

Five chapters of the thesis are written as self-standing research articles and meant as timely contributions to the relevant specialized discussions in philosophy of science. The arguments developed in each chapter are informed by socially relevant and philosophically interesting cases, historical episodes, or illustrative examples drawn from the scientific practice. This also reflects the thesis' ambition to contribute to the practice-oriented and socially engaging approaches to philosophy of science. In the following, the chapters are previewed for the reader who is conversant on the relevant literature.

In the first part of the dissertation (entitled "*Values in Science: Beyond the "Value Free Ideal of Science*"), the thesis investigates the consequences of abandoning the value free ideal of science. **Chapter 1** reviews the arguments for and against the value free ideal of science and illustrates these arguments by reference to the example of regulatory toxicology. The chapter describes the risks of abandoning the value free ideal. It argues that the prominent extant philosophical approaches that transcend the value free ideal of science do not reject but preserve the meta-values about science that the proponents of the value free ideal deem important. These values include the epistemic authority of science, scientific impartiality, and the avoidance of non-democratic (technocratic) policymaking.

**Chapter 2** reflects on how the contemporary philosophical frameworks that advance the most prominent value-laden conceptions of science describe and evaluate the actual episodes of scientific practice. Specifically, it focuses on Heather Douglas's inductive risk framework and discusses how the framework applies to a socially relevant and philosophically significant historical episode of regulatory toxicology. The chapter offers an empirically informed description of how regulatory toxicologists make methodological judgments about what the relevant evidence is for their purposes and choose between alternative evidence-gathering methods. It highlights some theoretically interesting aspects of this case (e.g., the roles of social and institutional processes in changing toxicologists' perceptions of inductive risks) that are puzzling or novel from the perspective of the inductive risk framework. It shows how the inductive risk framework can be extended to reflect on these aspects of the case.

In the second part, (entitled "*Philosophy of Evidence Based Policy from the Perspective of Values in Science*)," the thesis reflects on the philosophy of evidence based policy making from the perspective of the values in science).

**Chapter 3** investigates how an instance of behavioral interventions (Incentivized Smoking Cessation Policies) is evaluated by scientists. It demonstrates that the practice of evaluating these policies is attended by a plurality of researchers making use of different evidence gathering methods, representing different scientific

disciplines and policy perspectives regarding smoking as a public health issue. It argues that an empirical evaluation of whether the behavioral smoking policies are effective in reducing smoking-related health inequities (which is cited as the most pressing desideratum for many smoking control policies in Europe and the UK) requires integrating different methods (e.g., randomized controlled trials with qualitative methods used in social epidemiology). It discusses the implications of this pluralist perspective for the philosophical discussions over the evidential requirements of behavioral interventions.

**Chapter 4** focuses on the special cases where the claims made by social scientists include concepts that rest on unwarranted value presuppositions. Anna Alexandrova defines such claims as "mixed claims" because they mix normative content with empirical ones. Mixed claims are especially controversial when they are used for policymaking. Unless they are properly scrutinized, they implicitly impose some undetectably value-laden discursive perspectives on the policy deliberation. The chapter analyzes the contrast between Nagel and Alexandrova and offers a reconciliation that is argued to be agreeable to the both sides of the controversy. Specifically, based on the distinction between the authority and the legitimacy problem of values, the chapter demonstrates that Alexandrova's 2017 critique of Nagel (1961) is confined to the authority problem of values and leaves Nagelian conditionalization of mixed claims intact qua a strategy that addresses the legitimacy problem of values. It also shows that the contemporary Neo-Nagelian accounts (such as E. Anderson's and H. Douglas's accounts of objectivity) support the reconciliation the chapter proposes, and are thus compatible with both Nagel's and Alexandrova's insights regarding mixed claims.

**Chapter 5** focuses on an important instance of evidence based behavioral policies; nudges. Nudge policies are prominent and extensively applied examples of evidence based behavioral policies and raise significant ethical and political problems relevant to the philosophical issues explored in the thesis. Drawing on the literature on the democratization of science (particularly on Philip Kitcher's approach), the chapter proposes a conception of nudging, called well ordered nudge, that seeks justification

of these kinds of policies through deliberative democratic practices. It thereby offers a particular approach to address the authority problem of values in science. It seeks to identify practical and conceptual implications of this approach for an important instance of evidence based policy. The chapter compares the well ordered conception of nudges with the extant conceptions of nudges and elaborates on how the philosophical perspective offered can advance the ongoing discussions on the democratic justifiability of nudges and nudge-like evidence based policies.

# References

Alexandrova, A (2017). *A philosophy for the science of well-being*. Oxford University Press. https://doi.org/10.1093/oso/9780199300518.001.0001

Anderson, E. (2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia*, *19*(1), 1–24. https://doi.org/10.1111/j.1527-2001.2004.tb01266.x

Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, *3*(2), 207–220. https://doi.org/10.1007/s13194-012-0062-x

Biddle, J. B., & Kukla, R. (2017). The Geography of Epistemic Risk. In K. Elliott & T. Richards (Eds.), *Exploring inductive risk: case studies of values in science*. (pp. 215–237). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467715.003.0011

Cartwright, N, & Hardie, J. (2012). Evidence-based policy: A practical guide to doing it better. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001

Deaton, A, & Cartwright, N. (2018) Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine, 210,* 2-21.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, *67*(4), 559–579. https://doi.org/10.1086/392855

Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press. https://doi.org/10.2307/j.ctt6wrc78

Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.

Kourany, J. A. (2010). *Philosophy of science after feminism*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732623.001.0001

Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension: selected studies in scientific tradition and change* (pp. 320–339). University of Chicago Press.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. Harcourt, Brace & World, Inc.

# Part I

# Values in Science: Beyond the “Value Free Ideal” of Science

# Chapter 1

# Debating the Value Free Ideal of Science: the case of regulatory toxicology

## 1.1. Introduction

To be able to answer complex and technical questions, we defer to the judgment of scientific experts. And, doing so is wise. We need scientists' advice to be able to take high-stake policy actions. More generally, complicated and morally consequential decisions should better be relied on empirical evidence and scientifically supported arguments rather than uninformed opinions, doxastic preferences of a handful of decision-makers, or blind ideological commitments.

As the problems of our advanced post-industrial societies and globalized affairs are becoming more complicated, common governance and regulation mechanisms are ever more dependent on scientific advisory. But, with great power comes great responsibility. Further reliance on scientific advisory to manage the current affairs leads to more pressure on scientific experts. Concerned citizens and students of science expect ever more from scientific experts. We demand scientists to be as unbiased as possible in their assessments, but we also wish them to be responsive to,

and sometimes, protective of our values and ideals when they inform policies that will affect us. We expect scientific experts to be specialists. Yet, we also ask them to consider multiple dimensions of a decision problem and to take the interests of all the relevant stakeholders into account in doing so.

All these growing expectations concerning scientific advisers are in tension with our commonsensical understanding of scientists. The conventional image of a scientific expert is someone who is a specialist, policy-neutral and autonomous from societal affairs. If we need to abandon this old image of scientific experts and transcend it, we have to build a new conception of scientific expertise on solid foundations. These foundations will have to make sure that our reliance on expert judgments in the management of the current affairs is not spurious. We have to carefully and systematically ponder the grounds on which scientific experts can fulfil our pressing expectations. Social studies of scientific expertise provide us with helpful insights towards reaching this goal by conceptualizing the accountability of scientific experts and examine the cutting-edge cases where the role of scientific advisors is salient in policymaking. The present chapter contributes to these scholarship by offering a perspective from the philosophy of science.

In the following, I will review the debate about the most prominent account of science in society developed by the philosophers of science in the 20th century, which underpins the commonsensical image of scientists depicted above. This account is known as the *Value Free Ideal of Science*. The *Value Free Ideal* (VFI) suggests that non-epistemic (moral, political, environmental, economic and similar) values, commitments and aspirations have no legitimate role to play in scientific judgment-formation. This standard is quite stringent as it substantially limits what we could reasonably expect from scientific experts in the context of regulation and law. Yet, prominent contemporary philosophers of science have also proposed to abandon, or else amend, the VFI. I will introduce some of the key arguments for and against the VFI and discuss the implications of these arguments for conceptualizing the accountability of scientific experts in the context of evidence-based policy and

regulation. To make the discussion more accessible to the socio-legal scholars, I will illustrate these arguments by referring to regulatory toxicology and regulation of agricultural pesticides.

My aim is to demonstrate that the recent philosophical insights that contest the VFI do not debunk but rather revive and reconstruct the value of scientific expertise. Specifically, I argue that philosophers of science such as Heather Douglas, Philip Kitcher and Helen Longino appreciate and share the main worries of the proponents of the VFI, such as the importance of objectivity and neutrality, the need for rational argumentation in policy-making and the avoidance of technocracy. Nevertheless, they disagree with the proponents of the VFI about the premise that a genuine reliance on scientific expertise is sustainable within the boundaries the VFI defines. In the face of this problem, those philosophers who propose to transcend the VFI work towards developing an understanding of science as a reliable epistemic source for making complex and challenging (policy) decisions about the pressing and history-old problems. This emerging conception of science is not only responsive to urgent and significant non-epistemic values but also epistemically robust. Accordingly, I suggest, what contemporary philosophers of science have to say about the relationship between science and values is vital to develop any projects that attempt to critically assess the accountability of scientific experts in the contemporary contexts of law, governance and decision-making. My task in this chapter is to offer the reader a survey of some of the key recent philosophical insights for those socio-legal scholars who seek to strike the right balance between epistemic and non-epistemic dimensions of scientific expertise in developing a conception of the 'accountability of experts' in governance and law.

The rest of the chapter is structured as follows. In Section 2, I introduce an example concerning toxicological assessments of agricultural pesticides that I will use throughout the chapter. Based on this example, I review a number of contexts where scientific experts make non-epistemic value judgments in their assessments. In these contexts, the permeation of value judgments into their analysis is seen uncontroversial

by the philosophers of science, regardless of their position on the debate about the VFI. In Section 3, I describe the contexts where the inclusion of non-epistemic value judgments in science is impermissible according to the proponents of the VFI and discuss the main arguments for and against the VFI. In Section 4, I illustrate some of the key concepts in the contemporary philosophy of science that challenge the VFI and transcend it by allowing non-epistemic values a more expanded role to play in scientific judgment than the VFI permits. In Section 5, I conclude by elaborating on how the debate on the VFI and the recent insights from philosophy of science contribute to the socio-legal scholars' attempts to reconceptualize the accountability of scientific experts in law and governance.

## 1.2. Revisiting the Value Free Ideal of Science when examining the role of regulatory toxicologists in protecting environmental safety

### 1.2.1. The safety of pesticides and toxicological expertise

Consider environmental safety regulations in agriculture. Various kinds of pesticides are used for maximizing the productive capacities of the agricultural businesses, from which we all benefit as consumers. The overuse of toxic chemicals has adverse consequences for the environment and public health. The safety of the agricultural pesticides is, therefore, systematically regulated by institutions such as the US Environmental Protection Agency (EPA) and EFSA (European Food and Safety Authority). The processes of decision making in the regulatory institutions are far from being simple as it is not always a clear-cut judgment to decide whether a type of pesticide should be regulated more or less stringently. Making consequential regulatory decisions requires an estimation of relative costs and benefits of a particular regulatory action over a number of desiderata such as ecological safety, economic efficiency and the public interest.

It is, therefore, useful to conceive regulatory decision making about a type of pesticide as a multidimensional balancing process that goes beyond the binary judgment of banning or allowing the industrial use of a toxic substance. For instance, questions of the following type need to be answered, and the pertinent trade-offs must be drawn by the regulatory decision makers. What is the threshold (expressed in terms of dosages) on which a pesticide is safe to use in agricultural production? To which produces should the pesticide in question be applied to? Does the adversity of the pesticide depend on its proper application? If so, how should the substance be applied by its users? What is the severity of the costs on health and the environment? What are the consequences of imposing more or less stringent regulatory measures on the agricultural sector and the producers of the pesticides?

Answering such questions is likely to demand some technical knowledge and skills. Accordingly, regulators must rely on toxicological risk assessment to answer them as accurately as possible. Toxicological experts who perform required testing and who run the pertinent risk models, therefore, play some key roles in the making of regulatory decisions for environmental safety. Overall, it is desirable that the environmental safety regulations are evidence-based and informed by science.

The need to examine the accountability of scientific experts in regulatory decision-making processes arises in controversial cases. Regulatory mechanisms sometimes fail to ensure the safety of people's health and the environment and even lead to disastrous consequences. For instance, the use of clothianidin [1] in agricultural production has led to the death of millions of honey bees in Germany (Benjamin, 2008). Such cases are controversial not only because of the high environmental costs involved but also due to the severe uncertainties about what counts as the right course of regulatory action. In the case of the clothianidin, the details and the stringency of regulations have changed across time, and only after five years, its outdoor uses have been suspended in the EU. The gaps and incompleteness of the scientific advice over

[1] Clothianidin is a member of the class of insecticides known as neonicotinoids. Neonicotinoids are chemically similar to nicotine but more effective as a pesticide. It is applied on plants and attacks the nervous system of the targeted insects. It is first developed by Bayer BV.

the malignancy of neonicotinoids for the environment (the health of bees in particular) seem to play some role in the progression of the debate over the suspension of the neonicotinoids such as clothianidin (McDonald-Gibson, 2013).

In the face of such environmental failures caused by mis-regulation or under-regulation, it is appealing to question the role and the responsibilities of toxicological experts, and the regulatory scientists in general. More specifically, one feels compelled to ask whether regulatory toxicologists could have done better in preventing the environmental harms caused by the industrial use of the clothianidin. Perhaps, in the face of evidential gaps and uncertainties, the bits of advice for regulation could have been produced and communicated in a more precautionary manner, prioritizing the environmental safety over other concerns. For instance, regulatory toxicologists could have determined the parameters of their risk assessment models in a way that is more protective of environmental health.

But, to what extent is it justifiable to expect toxicological experts to be protective of *values such as environmental safety* in conducting their assessments? Should scientific advisors *qua* scientists endorse non-epistemic values or commitments and then let those values influence their analyses and conclusions? I suggest that philosophers of science have a lot of useful insights about these kinds of questions, and the VFI is a good place to start engaging with the philosophical perspectives on issues pertaining to the accountability of experts in governance and regulation.

The VFI is a deeply rooted analytical-philosophical view of how science relates to society. It defines clear boundaries under which non-epistemic values such as environmental health can legitimately influence scientific judgment. Specifically, according to the VFI, value judgments such as a commitment to maximizing environmental safety, may play *some* permissible roles in conducting scientific research. However, the VFI states, non-epistemic values should *not* influence scientists' inferences from the evidence, hence the results of their analysis (see, for instance, McMullin, 1982). In the context of science-based regulation, a proponent of the VFI would suggest that expecting scientific experts to let certain value

considerations to influence their results undermines the *raison d'être* of expert advice for making regulatory decisions. So, this view goes, the main reason for seeking scientific advice is to be informed about the facts of the world that are relevant for making morally consequential decisions; scientific advisors qua scientists are *not* regulators; hence scientists' judgments should not reflect any value judgments.

Accordingly, the VFI provides us with a clear benchmark that defines principles under which scientific experts can legitimately employ values such as environmental safety in their analysis. This ideal can, therefore, help us examine what we could justifiably expect from toxicologists to do in order to prevent regulatory failures, which may be partly made based on their judgments. I will now review a bit more closely what boundaries the VFI draws for the inclusion of values in science. We will then ask ourselves how those boundaries are justifiable and engage with the philosophical debates over whether and how the VFI should be transcended.

## 1.2.2. Values in regulatory toxicology according to the Value Free Ideal

In the face of regulatory failures in science-based regulation of agricultural production (manifested in the cases such as the regulation of clothianidin), we are interested in knowing whether non-epistemic values and commitments, such as maximizing the interest of public health or the environment, should have a more expanded role to play in the formation of scientific judgments made for regulatory action. I have suggested so far that the philosophical *Value Free Ideal of Science* provides us with a sound starting point to determine plausible ways in which values can play more or less expanded roles in science. My aim now is to offer a clear understanding of the principles of the VFI so as to proceed with a structured discussion about how regulatory toxicologists should employ value judgments in their analysis.

### 1.2.2.1. Scientific integrity and fraud

It is agreed by all the philosophers of science, irrespective of their position in the VFI debate, that scientists should avoid fraud. Scientific integrity is the most commonly

accepted (though also sometimes violated) moral principle in scientific practice. For instance, it is strictly forbidden to make up data for one's convenience. Similarly, scientists are not expected to cherry-pick a piece of evidence from a wide range of available evidential sources in a way that would favour the results they wish to obtain through their analysis. When regulatory toxicologists engage in these kinds of practices, they commit to *scientific fraud*.

Scientific experts who engage in fraudulent scientific practices causally contribute to the regulatory failures when the policy outcomes partly result from their untruthful judgments. The corruption of policy-informing experts is a genuine issue that needs to be discussed thoroughly. Indeed, many commonly known failures of expert-based regulatory advisory in environmental safety result from the violation of basic scientific integrity (see, for instance, Lacey, 2005; Gøtzsche, 2017). As extensively documented by the socio-legal scholars, the problem of fraud is most commonly observed among the so-called 'industry scientists' who are pressured to produce research results that serve the interests of large, profit-driven pharmaceutical and biomedical companies. The sources and the dynamics of indecent scientific practices need to be carefully studied by the socio-legal scholars and philosophers of science (see Bright, 2017 for a philosophical discussion).

#### 1.2.2.2. Non-epistemic values in the selection of research questions and tools

The VFI does not suggest that non-epistemic values have no legitimate function in science. There are two contexts where value judgments are needed, and it is legitimate for scientists to employ value judgments in their reasoning, according to the VFI. Firstly, the context in which scientists choose a research question to investigate and decide how to investigate that question. It is permissible that scientists attend to certain subjective value judgments, commitments or interests when they make decisions in such initial phases of their inquiry.

For instance, a toxicologist may want to study the effects of pesticides on the population of bees because she might deem it valuable to conserve the bee population's health. Her employment of values in this way does not conflict with the VFI, as values (ecological safety of bees) do not determine the results of her inquiry. Similarly, she can attend to certain (communal or personal) value considerations in determining the methods through which she will conduct her investigation. She may, for example, want to work only with molecular-genetic tools instead of exposed animal bio-assays because she may not prefer to harm or kill animals to conduct her scientific research. Again, the VFI does allow a role for non-epistemic values in this particular sense.

#### 1.2.2.3. No values in scientific inference

The VFI does limit the permeation of values into science when value judgments influence results obtained by an investigation or when values change scientists' judgments. Specifically, values should not influence certain methodological decisions in science. These are decisions such as how evidence is gathered, how hypotheses are justified and accepted, and how the results are interpreted. These methodological contexts are referred to by the proponents of the VFI as the core stages of the scientific reasoning; also known as the context of justification. The context of justification is commonly contrasted with the 'context of discovery' where the research questions and methods are chosen, in which values play a permissible role. According to the VFI, the context of justification should, in principle, be free from non-epistemic value considerations.

So, for instance, regulatory toxicologists could endorse the value of protecting environmental safety and choose to examine those hypotheses that would be protective of the environment (if their judgments will then be used for justifying regulatory processes). The regulatory toxicologists' endorsement of a non-epistemic value in this particular sense complies with the standard of the VFI. However, a commitment to protect environmental safety should not influence how the same

toxicologists justify the hypotheses they are assessing. In other words, values should not determine which judgments are accepted.

Moreover, scientific experts' role *qua scientist* must be distinguished from scientific experts *qua policy advisors* (see Steele, 2012; de Melo-Martín and Intemann, 2016 for elaborations). Qua policy advisors, scientists might make use of values in determining how they communicate their results (see John, 2015 for a philosophical analysis of this context). For instance, they could permissibly emphasize certain aspects of their results to motivate a particular regulatory response. However, qua scientists, policy advisors should not allow non-epistemic value judgments to influence how they obtain their results. In short, values should be ignored in scientific interference, according to the VFI.

#### 1.2.2.4. Except epistemic values

The so-called epistemic values are exceptional to the principle discussed above. The proponents of the VFI find it benign when various epistemic values of a hypothesis, such as its simplicity, explanatory or predictive power, accuracy, robustness and fruitfulness are used by scientists to assess that hypothesis. Thomas Kuhn, in particular, thought that scientific reasoning or hypothesis-confirmation should better be conceptualized as a process that appeals to some shared standards of epistemic values (which may differ across different communities of scientists) (Kuhn, 1977). Since Kuhn, it is commonly thought that the involvement of *epistemic* values in scientific inference does not harm the justification of scientific results or the objectivity of scientific judgment in general. The controversy has been over the status of the so-called 'non-epistemic values' in scientific justification. Kuhn's ideas about epistemic values in relation to scientific disagreement and objectivity were thought to be protective of the VFI (Douglas 2009, Chapter 3).

The VFI thereby provides us with a clear benchmark regarding how far we should expect scientific experts to be protective of certain non-epistemic values in producing their judgments for regulatory procedures. The principles of the VFI discussed so far

can be summarized as follows. Non-epistemic values such as environmental safety are permissible in scientific experts' reasoning only in certain contexts that are considered to be separate from the core stages of science, such as collecting evidence and making inferences from them. Only the so-called epistemic values can influence scientists' judgments over the confirmation of a hypothesis.

These principles of the VFI are supposed to be protective of scientific objectivity, hence scientific authority for informing the processes of governance and regulation. Given this conceptual background, we now need to ask ourselves: should the socio-legal scholars adopt or abandon the VFI? In the context of specific cases such as the assessment of clothianidin, what are the limits of scientific activism of the sort that is protective of values such as environmental safety and public health? I will now review arguments for and against the VFI in order to motivate relevant philosophical insights regarding whether and how the VFI could be transcended.

## 1.3. Debating the Value Free Ideal of Science

As I have discussed in Section 2, the VFI does not leave room for any policy desiderata to play a detrimental role during the formation of expert advice for regulatory processes. This implies that if the socio-legal scholars endorse the VFI as a benchmark model of how scientific experts make judgments about policy-relevant issues, they should not expect from scientific experts to adjust their findings or judgments in order to serve for the protection of certain values. For instance, toxicologists assessing clothianidin should not be expected to produce results that are protective of the financial interests of a patent-holder company or the economic efficiency of the agricultural industry. Similarly, if the VFI is endorsed, the dangers of clothianidin-use for ecological health should not influence the expert's assessment of clothianidin's toxicity. However, the VFI is not free from criticisms and challenges. At least starting from the 1980s, philosophers of science have questioned the descriptive adequacy and the normative appeal of the VFI. I will now review some of the influential arguments that are used for contesting or defending the VFI.

### 1.3.1. Contesting the Value Free Ideal

An important type of challenge against the VFI is directed at the attainability of it as an ideal (Douglas, 2009). Specifically, these criticisms are about the nature of scientific reasoning and methods and the core concepts of the scientific method, such as evidence, theories and explanation. Helen Longino, for instance, argues that inferences from the evidence to hypothesis often rely on background assumptions which may encode ontological assumptions about the world, value presuppositions about what deserves to be examined, or methodological assumptions about what types of evidence are relevant to given hypotheses (Longino, 1990). The problem, according to Longino, is that scientific communities' shared background assumptions may often be implicit, insufficiently examined, or not be subjected to critical scrutiny of alternative perspectives. This work has later motivated more recent studies in philosophy of science, which conceive of scientists as members of social groups and knowledge-production as a social activity characterized by a plurality of epistemic communities with varying values, interests and methods to pursue research. Given such an understanding of scientific judgment that is undetermined by evidence and laden with disagreements about background presuppositions, the VFI seems to be an unattainable ideal, if not an illusion (Kincaid, Dupré, & Wylie, 2009). In Section 4, I will also discuss how philosophers such as Longino make constructive suggestions about scientific objectivity while holding a descriptive view of science as a deeply pluralistic social practice.

However, the descriptive problems of the VFI are not of direct relevance to the problems examined in this edited volume. We are not interested in grand questions regarding the nature of science, scientific knowledge and objectivity. We are interested in the rules and principles that guide scientific experts when they need to attend to value judgments in their analysis for regulatory purposes. The critiques of the VFI that focus on the normative appeal of its principles are of more direct relevance to discuss the accountability of regulatory expertise. Some philosophers of science, such as Heather Douglas – whom we can see as a contemporary successor of Richard Rudner's approach to scientific reasoning – have argued against the VFI as

an ideal, irrespective of its feasibility (Rudner, 1953; Douglas, 2000, 2009). Heather Douglas' work has been particularly influential in recent years in pushing the argument against the VFI forward (Douglas, 2009).

Douglas argues against the normative appeal of the VFI on two grounds, which, when taken together, leads her to reject the VFI normatively. Firstly, Douglas emphasizes that scientific judgment often involves the risk of being wrong. That is to say, when a scientist makes an inference about a hypothesis, the available evidence does not fully confirm the hypothesis being true or false. And so, she argues, scientific confirmation actually takes place in degrees and commits to an error akin to the errors in basic statistical inferences: a false positive judgment or a false negative one (i.e. accepting p, when p is actually false; or rejecting p when p is actually true). This idea is known as the 'inductive risk' argument (see Brown, 2013 for a review and Elliot and Richards, 2017 for a collection of recent philosophical debates on the inductive risk argument). This is the first building block of Douglas's framework.

Douglas's second argument relies on the recognition of the pragmatic function of science in the last century. Scientific experts are increasingly playing a more authoritative role in resolving regulatory conflicts, making policies and legitimizing interventions to societies and people's lives (Douglas, 2009). Douglas suggests that it is better to conceive of scientific judgment-formation as a process that is not autonomous from the management of the current affairs and hence should not be governed by its own rules only.

The combination of these two ideas (the inductive risk argument and the [partial] causal dependence of policy outcomes to scientific expert judgments) compels Douglas to reject the normative appeal of the VFI as an ideal. Specifically, she claims that it is plausible for scientists to take into account the risks of being wrong in their judgments in the face of the inductive risk problem. Moreover, she argues, if the judgment made by a scientific expert influences policy outcomes (e.g. as it happens in the policy-relevant areas such as regulatory toxicology), the risks of being wrong bear on non-epistemic consequences of the policy action. For instance, as the

clothianidin case suggests, a regulatory toxicologist's erroneous judgment may have consequences for the health of the bee population and the economic productivity of the pesticide-producing sector. Accordingly, Douglas suggests, it should be normatively desirable that scientists make an evaluation of such non-epistemic consequences in order to decide how much evidence is sufficient to confirm a hypothesis given that there are inductive risks in scientific judgment-formation.

Douglas's conclusion goes against the normative appeal of the VFI. It does so because it leaves room for non-epistemic values to play a *permissible* role in the core stages of scientific inference. If we agree with Douglas that non-epistemic values can plausibly influence the results of scientists' judgments, we would then expect regulatory toxicologists to let values such as environmental protectionism to play a more prominent role during the formation of regulatory judgments. This idea has evaluative consequences for a case like the regulation of the clothianidin.

Suppose that a regulatory toxicologist collects data from the pathological examination of bee tissues, which are exposed to varying dosages of clothianidin in laboratory conditions. Based on the collected experimental observations, she chooses a model that describes how tissues respond to the different doses of clothianidin (typically in the form of a dose-response curve). Then, based on the predictions of the model, she makes a judgment about the hazardousness of clothianidin over a given bee tissue-type under varying dosages. This judgment is a probabilistic one such as 'clothianidin is likely/highly likely/unlikely to be malignant for the exposed tissue' (but typically with precise numbers) (Douglas, 2000).

Now, as discussed above, Douglas's idea is that because the available evidence is not *fully* confirmatory of the hypothesis that the clothianidin is harmful, the toxicologist's judgment will be an inductively risky one, consisting of one of the following errors. *Error I* is reporting clothianidin as benign when it is actually malignant. *Error II* is reporting clothianidin as malignant when it is actually benign. But, both errors have non-epistemic consequences because they can be used for legitimizing different regulatory agendas. Committing *Error I* could motivate regulatory actions that do *not*

strongly discourage the use of the clothianidin even though it should really be discouraged. This contributes to the development of adverse ecological or public health outcomes such as the collapse of the bee population. On the other hand, committing *Error II* provides justification for unnecessary over-regulation of clothianidin, leading to unwanted economic outcomes.

Douglas suggests that it should be legitimate when toxicologists consider the severity of these errors and pertinent outcomes in non-epistemic terms to decide how much evidence is sufficient to make one of the risky inferences. If failing to detect the malignancy of clothianidin has too severe environmental costs, it should be permissible for the toxicologist to confirm its malignancy *based on relatively weak evidential support*. In other words, non-epistemic values such as ecological health should be allowed in scientific reasoning to help scientists to determine *how much evidence is sufficient to confirm a given hypothesis*. If Douglas is right, then the ecological concerns should play a more extensive role in regulatory toxicologists' evaluation of clothianidin and similar toxic substances.

Douglas's normative argument against the VFI provides us with some reasons to depart from the VFI as a benchmark for evaluating experts' accountability in evidence-based law and governance. Perhaps, non-epistemic values or interests of the public or the environment should, after all, influence scientific experts' judgments more wittingly. But, notice that Douglas's argument only criticizes the VFI but does not offer an alternative benchmark that regulates values in science. If we do not know how to replace the ideal, isn't abandoning it too costly?

### 1.3.2. Why is the VFI normatively appealing?

Without a new ideal that is as useful and convincing as the VFI, the permeation of values to scientific inference may be undesirable because the scientific authority could be conceived as unduly politicized when there is no normative benchmark that regulates the permeation of values into scientific reasoning. Some philosophers of science have argued that there are good reasons for not abandoning the VFI, at least

until a better benchmark is proposed (see, for instance, Betz, 2013, 2018; Bright, 2018). The defences of the VFI are based on two main concerns.

The first and the most pressing worry against the idea of abandoning the VFI is the potential loss of scientific objectivity in the context of scientific advisory. Betz, for instance, argues that the best use of scientific knowledge is to hold it detached from social and political processes, as that is the best way to keep the reliability of science intact. Disbelief in the objectivity of the scientific advisory would undermine the very reason to defer to the experts in making complex regulatory decisions.

The inclusion of non-epistemic values into scientific reasoning may lead to dogmatism; that is, unwillingness to update or replace one's predispositions, values and judgment (see Anderson, 2004 for a compelling argument). Non-dogmatism is a characteristic commonly ascribed to scientists. Scientific experts are supposed to report what available evidence indicates regardless of whether her evidence-based inferences are in clash with her personal wishes, viewpoint or causes. A significant departure from the VFI depiction of science seems to spoil this basic feature of scientific reasoning.

Consider a regulatory toxicologist who endorses environmentalist values. She might be worried that the industrial use of clothianidin may harm biodiversity. She accordingly would wish that her assessment of clothianidin would conclude that clothianidin is extremely harmful to wild-life even when exposed in very small dosages. One could reasonably suspect that an expert endowed with such an activist mindset is likely to be more subjective in her judgments, and her conclusions are unreliable for informing regulatory processes. More generally, when a scientist makes a judgment that she prefers to be true even though there is no clear evidence supporting that judgment, she would commit to wishful thinking. Wishful thinking has no legitimate role to play in genuine scientific reasoning. When scientific experts engage in wishful thinking, they will lose their reliability for informing regulatory processes.

Accordingly, the VFI could be considered as protective of the public trust in scientific authority in regulatory processes. Philosophers who propose to transcend the VFI should then have an account of how scientific inference can remain non-dogmatic and trustworthy while being influenced by non-epistemic value judgments. As I will review in Section 4, philosophers such as Heather Douglas or Elizabeth Anderson have addressed this worry, but the controversy remains up-to-date (Betz, 2018).

The second argument for the normative appeal of the VFI emphasizes the inconvenience of scientists as agents of ethical and political decision-making (Betz, 2013). Letting scientific results be influenced by certain non-epistemic value considerations may lead to the expansion of experts' arbitrary political power in regulatory processes. Experts can influence the policy processes by adjusting the results in accordance with their preferred policy desiderata. However, scientific experts are not considered to be legitimate political actors, at least in a democratic context. In a democratic governance structure, regulatory decisions are not supposed to be made by experts but by democratically legitimized political subjects. The acceptance of values into the production of scientific judgments is in tension with the liberal/democratic values.

This argument is underpinned by an understanding of policy-making in which the roles of democratically legitimized decisionmakers and the scientific advisors are clearly divided. According to this understanding, scientific experts are tasked with a politically neutral, advisory role which is to report the available evidence as accurately as possible. The democratically and legally legitimate decisionmakers then bear the responsibility of making any politically and morally binding value judgment. This ideal resembles the traditional model of risk management in regulatory sciences, which was critically discussed by the science studies scholars such as Sheila Jasanoff (1990) as well as legal scholars who contributed to this volume (Arcuri and Simoncini, 2015; see also Arcuri 2021, Chapter 3 and Kanetake, 2021, Chapter 8). The model distinguishes scientists as *risk-assessors* and decisionmakers as *risk-managers*. While the decisionmakers are responsible for managing risks and taking risky decisions, the

scientific experts are tasked with assessing the risks without taking a stance on how those risks should be managed or letting that stance influence her assessment.

For instance, toxicologists conduct research on the nature of clothianidin's toxicity and assess the scale of environmental risks resulting from the use of clothianidin. Then, they communicate their estimations of risks to the decision-making agents. They make clear when the evidence is inconclusive or in how far certain policy actions are supported by the available evidence. Only then, the decisionmakers (e.g. board of directors of private companies or regulators) make a decision about how to regulate clothianidin use (e.g. selling it, regulating it, banning it) by taking into account the risks estimated by the experts and valuation of potential adverse or desirable consequences of their decisions (Jeffrey, 1956). If this model of evidence-based regulation would be transcended, the replacing accounts of expertise in governance should also be able to respond to the democratic defence of the VFI, either by effectively arguing against it or by arguing that the new alternatives to the VFI do not harm democratic legitimacy of regulatory decision-making processes.

In this section, I have reviewed some of the key arguments for and against the VFI. A complete and widely accepted philosophical alternative to the VFI is not yet offered, and how the VFI should be replaced is still discussed by philosophers of science (for a very recent attempt to replace the VFI, see Brown, 2020). In the rest of this chapter, I will review a number of prominent approaches in the contemporary philosophy of science which transcend the VFI. The conceptual insights from these approaches can be insightful for the socio-legal scholars' search for the right balance between epistemic and non-epistemic dimensions of scientific expertise in developing a conception of the accountability of experts in governance and law.

## 1.4. Going beyond the Value Free Ideal while keeping scientific objectivity and experts' reliability intact

The aim of this chapter has so far been to offer a perspective from the philosophy of science on the debates about the accountability of scientific experts in the context of regulation. I have questioned how far scientific experts can and should make non-epistemic value judgments in their assessments for regulatory purposes. In response to this question, I have introduced the philosophical notion of the VFI, which permits values in science in some restricted ways. I have then discussed why we may want to move beyond or else protect the VFI. Now, in this section, I will introduce some of the key ideas in the recent philosophy of science that transcend the VFI but nevertheless respect the normative reasons for endorsing the VFI such as objectivity, the authority of the scientific judgments, and democratic governance.

### 1.4.1. Does abandoning the VFI undermine science?

In contrast to the VFI depiction of value-free scientific experts who are distant from societal issues and neutral towards policy outcomes, the philosophers who are unconvinced by the VFI (such as Douglas, Longino, Anderson, Brown, Kitcher) portray a regulatory expert who is concerned with the policy outcomes and who is protective of relevant value judgments and commitments in producing her analysis. For some commentators, replacing the VFI with such a 'non-neutral' portrayal of scientific advisory might come across as a dangerous turn in the philosophy of science.

One could, for instance, forcefully argue that the VFI is too important to be dismissed and too risky to abandon because of its normative appeals. The VFI, the skeptic might argue, is an ideal that appreciates the authority of science and effectively regulates the science-society relationship in a regulatory context. Acknowledging and demanding a more expanded role to the non-epistemic considerations in scientific experts' reasoning opens the door to a lot of practical problems and may harm rational deference to experts in regulatory decision making. Similarly, one might be worried

that the quest for including general public's expectations and values into expert-based regulatory decision-making may lead to a 'politicization of expertise', (Douglas, 2009, p. 113, p. 134) 'tranny of ignorance or vulgar democracy' (Kitcher 2011, p. 113, p. 126, p. 140, p. 177), or democratically unjustifiable 'epistocracy' in governance structures (see Reiss, 2019 for a discussion). Moreover, leaving the VFI may trigger the erosion of the authority of science (see Longino, 2002; Kitcher, 2011 for elaborations of these worries), motivating unsubstantiated versions of post-truth politics and unconvincing adoptions of Paul Feyerabend's 'anything-goes' argument (see Russell, 1983 for a discussion). After all, given its broad normative appeal (reviewed in Section 3), the VFI could still be the best *normative* benchmark for evaluating experts' role and responsibilities in evidence-based decision-making (however untenable it might be as a descriptive account of science). If we allow the permeation of non-epistemic considerations into scientific reasoning and open the Pandora's box, so to say, how can we make sure that scientific experts are objective when they inform policy decisions? Does an expert who is driven to preserve values such as health-provision, environmentalism or inequality-aversion almost unavoidably lose her reliability as a neutral policy assessor? Is it not *too dangerous* to abandon the VFI as a normative benchmark for scientific reasoning? So, even if it might be an untenable ideal, why not keep the VFI as a guiding principle?

These important concerns need to be properly addressed if we are to fully replace the VFI. I do not claim that these concerns are already satisfactorily addressed by the philosophers of science. However, all I want to clarify is that the recent philosophical critics of the VFI, who seek to provide us with the foundations to go beyond it, endorse similar concerns but nevertheless believe that the best way to address these concerns, at least in the context of regulatory sciences such as regulatory toxicology, is to transcend the VFI. In their arguments against the VFI, philosophers of science such as Philip Kitcher, Helen Longino and Heather Douglas, have explicitly addressed these concerns.

In the remainder of this section, I will introduce some of the key concepts in the literature on the values in science that move beyond the boundaries of the VFI, which nevertheless are protective of the meta-values that the VFI subscribes to such as scientific objectivity, the importance of scientific authority, democratic quality of science-based decision-making, and the avoidance of imposing values to the citizenry.

## 1.4.2. Transcending the Value Free Ideal of Science

I will review four recent concepts in the contemporary philosophy of science that go beyond the VFI while keeping the authority of science for democratic governance intact. These are Douglas's objectivity as impartiality or detachment (Douglas, 2004), Steel's notion of epistemic priority (Steel, 2018), right values for science (see for example among others Hicks, 2014; Kourany, 2010) and Kitcher's account for the democratization of scientific expertise (Kitcher, 2011, pp. 86-113).

### 1.4.2.1. Objectivity as impartiality as opposed to value freedom

Heather Douglas argues that the exclusion of value judgments from scientific inference is *not* the only plausible rule to protect the objectivity of scientific experts' reasoning (Douglas, 2004, Douglas, 2009, pp. 115-132). She claims that values may, in principle, legitimately influence regulatory scientists' decisions about key methodological choices, which would then influence their results. However, she argues, permitting values in scientific inference does not necessarily imply that scientific objectivity qua impartiality is harmed. According to Douglas (2004, pp. 458-461) and Elizabeth Anderson (2004) the key to achieve objectivity in scientific reasoning while permitting non-epistemic values is that value judgments do not override scientists' evidence-based inferences, but rather *supplement them.* So, Douglas suggests, rather than banning the non-epistemic values in science as the VFI advises, we need to an account that distinguishes permissible and impermissible permeations of values into scientific reasoning. She, in turn, offers one such account.

Douglas introduces a distinction between the direct and the indirect use of values in scientific inference.

When values influence scientific judgment indirectly, they supplement evidential reasoning by helping scientists decide how much evidence is sufficient to confirm a given hypothesis, under inductive risks. The scientific reasoning is in the following form: if the consequences of me failing to accept p when, in fact, p is true would be very bad (ethically, socially), then I will accept that p on the basis of weaker evidence. When the role of values is direct, values go beyond this supplementary function (helping assess the sufficiency of evidence), and instead value considerations trump over evidential ones or replace them. It would then be in this form: if the consequences of me failing to accept p would be very bad (ethically, socially), I then will accept p (irrespective of the sufficiency of the evidence).

Now, Douglas argues that there is an important difference between the two uses of value judgments: when employing values indirectly, a scientist is prepared to be non-dogmatic. That is to say, she would reject p when evidential support against p would increase (as the severity of non-epistemic consequences of errors would diminish when there is more certain evidence). The same is not the case for the direct employment of value judgments in scientific inference: in that case, the scientists' reasoning disregards evidence. Therefore, only the indirect role of non-epistemic values should be permissible in scientific reasoning, according to Douglas.

The reader might be wondering: How would this rather conceptual-theoretical distinction work in practice such that it can inform our analysis of the real cases of expert accountability? Although the individual scientist's right reasoning attitudes, social norms of transparency and peer-review are supposed to filter out the impermissible, direct uses of values in scientific reasoning, the idea of detached objectivity should be conceived as a regulative ideal that may not be fully attainable in practice, just as the VFI.

The concept of *detached objectivity* transcends the VFI in that it allows values to play some (indirect) role at the core of scientific reasoning. It is, however, protective of the value of objectivity as impartiality or avoidance of dogmatism, which is at the central to the VFI.

#### 1.4.2.2. Epistemic priority

As noted in Section 2, the employment of epistemic values such as simplicity, internal consistency in making inferences in science does not conflict with the VFI but rather supports it. However, some philosophers of science, such as Rooney (2018), argue that the definitions and instances of epistemic values in science may sometimes be affected by hidden non-epistemic considerations as well. This kind of scepticism about the VFI leads to new approaches in philosophy of science that rethink the nature of the epistemic values in science and accept the idea that epistemic values may sometimes be context-dependent or not always serve for the attainment of truth (see, for instance, Longino, 1996).

Recently, Daniel Steel suggests that some of the epistemic values that help with scientific inference (such as simplicity, consistency with what is already known, openness to criticism, following agreed-upon methodological procedures) sometimes do not serve the attainment of true beliefs (Steel, 2010; 2018). Whether such epistemic values hinder or promote the attainment of truth depends on the context and the case at hand. Steel suggests that non-epistemic values may similarly help with the attainment of truth, so in some cases, their role is similar to the role of the benign employment of epistemic values. The studies on the values in regulatory sciences can make use of Steel's account to judge whether experts' employment of certain non-epistemic values harms the quality of scientific advisory in a certain case. For instance, the standardization of policy-relevant scientific inquiry (e.g. through establishing agreed-upon protocols and mechanical procedures) may minimize any particular community of experts' arbitrary influence on political decision-making in some contexts but could also lead to biases in different cases, depending on the structure and the quality of the procedures.

Steel's argument is another prominent example of a recent approach in philosophy of science that goes beyond the VFI, which nevertheless is clearly protective of the core meta-values that the proponents of the VFI have subscribed to. Notice that Steel's criterion for allowing non-epistemic values in scientific reasoning prioritizes the attainment of truth as the guiding principle to decide whether the permeation of non-epistemic values in science leads to any harm in the quality of scientific judgment in a given context. Similar to the idea of detached objectivity, Steel's argument for the inclusion of non-epistemic values in science is protective of the scientific authority in an important sense.

#### 1.4.2.3. Pluralism and the right values in science

Yet another remarkable and prominent philosophical approach to the role of values draws on feminism and the work of feminist philosophers of science who have grappled with the question: *Is a feminist science possible?* (see for instance Haraway, 1988; Anderson, 2004; Longino, 1987; Wylie, 2007). In advancing replies to this question, philosophers such as Alison Wylie, Elizabeth Anderson and Helen Longino, have offered compelling theories according to which scientific judgment can be value-laden but nevertheless deserve the label objective.

Famously, Helen Longino (1990) argues that the plurality of views, values and background assumptions in science is not a weakness but rather constitutive of scientific objectivity and rationality. For instance, Longino has established the idea that avenues for criticism and peer review, openness to new perspectives and participants, and similar social institutions of science that facilitate rational argumentation help science progress and establish itself as an objective inquiry. Longino's account, therefore, invites us to think that non-epistemic values in scientific reasoning are not to be refrained from but to be embraced, reflected upon and to be regulated through the critical social interactions among scientists.

While Longino's account is a general one, more recent approaches in the feminist philosophy of science, such as Janet Kourany's *Philosophy of Science after Feminism*

(2010), have addressed the problem in more specific terms by asking which non-epistemic values are conducive of good and responsible scientific practices. Note that venturing into an account of 'right values' is a controversial subject and may conflict with the ideals of democracy and neutrality. This kind of work is ongoing. However, Kourany's project is similar to Longino's in that she argues for a more extended role of non-epistemic values in science because she believes that cogent and responsible science cannot be realized without genuine engagement with values.

The feminist accounts of science, such as Longino's and Kourany's, invite us to assess the quality of the social dimensions of the regulatory sciences used for legitimizing policies. The relevant examples of these dimensions are pluralism, diversity of perspectives, transparency of non-epistemic values. Assessing these criteria for the quality of regulatory expertise can help socio-legal scholars detect whether and how failing to fulfil such feminist ideals in regulatory sciences hinders or obstructs the attainment of good regulatory outcomes (that are acceptable by the stakeholders and are protective of environmental safety and the public health). We can also use this kind of analysis to identify whether acceptable non-epistemic values are employed in regulatory sciences such as toxicology and pharmacology (see Hicks, 2014 for an exploration).

#### 1.4.2.4. Democratization and citizen participation in science

Some philosophers of science, notably Philip Kitcher (2011) and Stephen Turner (2001) have explored these issues from a different angle by asking under what conditions citizens' and governments' deference to expertise to manage current affairs is reconcilable with liberal and democratic values. In response to this theoretical problem, the science studies scholars and philosophers of science have explored to what extent and how scientific expertise can be democratized by establishing institutionalized avenues that facilitate a deep and meaningful interaction between experts and non-expert citizens. In turn, various models of inserting citizens into science, or more generally models of democratization of expertise, have been

proposed or investigated by the philosophers of science and science studies scholars (see, for a collection of contributions, Maasen &Weingart, 2005). This literature is also known as the democratization of science literature.

Some of the commentators in this literature, such as Kitcher and Douglas, have explicitly abandoned the VFI. But again, the point has not been to criticize the meta-normative values the proponents of the VFI subscribe to but to transcend them. Specifically, Kitcher addresses the second argument in favour of the VFI noted in Section 3. He proposes that the involvement of non-epistemic values and directions of scientific research should be decided by the informed citizens and deliberated in ideal settings. Kitcher's ideal demands that the heterogeneity of values in the public should be represented in scientific research projects. In order to avoid the arbitrary negative influence of value-laden experts over citizens, Kitcher argues in favour of reforming scientific institutions in such a way that science is properly situated in an advanced deliberative democratic settings. Trying to come closer to Kitcher's ideal of science ( the *Well Ordered Science*), he argues, is not only a requirement of democracy but also a vehicle to fulfil liberal values such as people's freedom to pursue meaningful life projects and equality in doing so.

Together with Kitcher's *Well Ordered Science*, the four concepts reviewed above show us that the worries of the proponents of the VFI can, in principle, be properly addressed (see Table 10.1 for a summary). For the examination of the accountability of regulatory experts, these philosophical insights highlight that departing from the VFI does not necessarily imply an unguarded politicization of scientific expertise in regulatory governance processes. Moreover, they form a helpful pool of conceptual resources that can be used for evaluating the quality of how non-epistemic issues are addressed by scientific experts across diverse regulatory contexts in practice.

**Table 1.1:** Four key concepts from the recent philosophy of science that transcend the VFI while reclaiming the value of scientific authority in democratic governance

| Concept | Description | Key Text |
|---|---|---|
| *Objectivity As Impartiality* | The permeation of pragmatic considerations into the core of scientific reasoning violates the sense of scientific objectivity as value-freedom. However, the violation of value-freedom does not imply that the sense of objectivity as 'detachment from values and evidence' is also violated. | *Science, Policy and the Value-Free Ideal* by Heather Douglas |
| *Objectivity through Pluralism and Dissent* | Proper social institutions of science (e.g. an inclusive and transparent peer-review community) may guard against the arbitrary dominance of any particular community of experts' subjective biases or hidden values over practical decision-making. Scientific objectivity is partly established inter-subjectively through learning from each other and discovering one's biases through critical peer-reviewing. | *Science as social knowledge: Values and objectivity in scientific inquiry* by Helen Longino |
| *Epistemic Priority* | Non-epistemic values can be included in scientific inference if we take the epistemic value of doing so as a guiding principle. | *Qualified Epistemic Priority* by Daniel Steel |
| *Well Ordered Science* | In order to keep the authority of science intact, we need to make sure that the heterogeneity of values in public is well represented in science. This would fulfil the instrumental value of science, which is to maximize societies' freedom to pursue meaningful life projects. In order to avoid the arbitrary negative influence of value-laden experts over citizens, we need to design the avenues for expert-citizen interaction in such a way that science is properly situated in an advanced democracy. | *Science in a Democratic Society* by Philip Kitcher |

# 1.5. Conclusion

As reviewed, the philosophical debate over the cogency of the VFI provides the socio-legal scholars who are interested in the accountability of regulatory scientific expertise with useful conceptual insights. These insights can be used for examining individual cases such as the case of the regulation of clothianidin. Based on the relevant conceptual grounds reviewed here, we could ask, for instance, whether the VFI is followed by the regulatory toxicologists or justifiably violated.

Recognizing the fact that value-free and objective scientific judgment is, at best, an ideal should not readily pave the way for the politicization of scientific expertise in the context of expert-based governance. After all, multi-level regulation and governance need fair amount evidence as opposed to opinions; rational arguments as opposed to politics-driven wishes. This however does not imply that we must stick to an old-fashioned positivist ideal that is hardly applicable to the complex contemporary multi-level governance mechanisms in which experts' judgments play crucial roles. Scientific judgments must be produced with the awareness that their conclusions might be misinterpreted, misused by the policymaking agents. It is also important that regulatory experts genuinely serve for public's interest. What is required is an understanding of science as a reliable epistemic source that does not refrain from attending to the relevant non-epistemic issues.

In this chapter, I have considered different ways in which scientific experts can be permissibly expected to let non-epistemic value judgments to influence their assessment. In particular, I have offered that the arguments for and against the philosophical Value Free Ideal of Science provide us with a suitable conceptual ground to examine the accountability of scientific experts in addressing the relevant value considerations such as economic interest and environmental safety in their analysis. I have emphasized that the contemporary philosophical debates on the role of values in science urge us to consider ways in which values may play roles in regulatory scientists' analysis, but within certain limits that hopefully protect the epistemic and democratic legitimacy of our reliance on scientific experts in regulatory decision-making.

# References

Anderson, E. (2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia*, *19*(1), 1–24. https://doi.org/10.1111/j.1527-2001.2004.tb01266.x

Arcuri, A., & Simoncini, M. (2015). Scientists and legal accountability: Lessons from the LLAquila Case. *EUI Department of Law Research Paper No. 2015/17.* https://doi.org/10.2139/ssrn.2600406

Arcuri, A. (2021). Three dimensions of accountability for global technocracy. In A.S. Arcuri & F. Coman-Kund (Eds.), *Technocracy and the law : Accountability, governance and expertise*. (Chapter 3). Routledge. https://doi.org/10.4324/9781003174769

Benjamin, A. (2008, May 23). Pesticides: Germany bans chemicals linked to honeybee devastation. *Guardian*. https://www.theguardian.com/environment/2008/may/23/wildlife.endangeredspecies

Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, *3*(2), 207–220. https://doi.org/10.1007/s13194-012-0062-x

Betz, G. (2018). Why the argument from inductive risk doesn't justify incorporating non-epistemic values in scientific reasoning. In K.C. Elliott & D. Steel (Eds.) *Current controversies in values and science* (pp. 94–110). Routledge. https://doi.org/10.4324/9781315639420-7

Bright, L. K. (2017). On fraud. *Philosophical Studies 174***,** 291–310. https://doi.org/10.1007/s11098-016-0682-7

Bright, L. K. (2018). Du Bois' democratic defence of the value free ideal. *Synthese, 195*(5), 2227–2245. https://doi.org/10.1007/s11229-017-1333-z

Brown, M. J. (2013). Values in science beyond underdetermination and inductive risk. *Philosophy of Science, 80*(5), 829–839. https://doi.org/10.1086/673720

Brown, M. J. (2020). *Science and moral imagination: A new ideal for values in science*. University of Pittsburg Press. doi:10.2307/j.ctv18b5d19

de Melo-Martín, I., & Intemann, K. (2016). The risk of using inductive risk to challenge the value-free ideal. *Philosophy of Science, 83*(4), 500–520. https://doi.org/10.1086/687259

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science, 67*(4), 559–579. https://doi.org/10.1086/392855

Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese, 138*(3), 453–473. https://doi.org/10.1023/B:SYNT.0000016451.18182.91

Douglas, H. (2009). *Science, policy, and the value-free ideal.* University of Pittsburgh Press. doi:10.2307/j.ctt6wrc78

Elliott, K. C., & Richards, T. (Eds.) (2017). *Exploring inductive risk: case studies of values in science*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467715.001.0001

Gøtzsche, P. (2017). *Deadly medicines and organised crime: how big pharma has corrupted healthcare*. CRC Press.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575. https://doi.org/10.2307/3178066

Hicks, D. J. (2014). A new direction for science and values. *Synthese*, *191*(14), 3271–3295. https://doi.org/10.1007/s11229-014-0447-9

Jasanoff, S. (1990). *The fifth branch: science advisers as policymakers*. Harvard University Press.

Jeffrey, R. C. (1956). Valuation and acceptance of scientific hypotheses. *Philosophy of Science*, *23*(3), 237–246. https://doi.org/10.1086/287489

John, S. (2015). Inductive risk and the contexts of communication. *Synthese*, *192*(1), 79–96. https://doi.org/10.1007/s11229-014-0554-7

Kanetake, M (2021). Transnational radiological standards in domestic contexts. In A.S. Arcuri & F. Coman-Kund (Eds.), *Technocracy and the law : accountability, governance and expertise*. (Chapter 8). Routledge. https://doi.org/10.4324/9781003174769

Kincaid, H., Dupré, J., & Wylie, A. (Eds.) (2009). *Value-free science?: ideals and illusions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195308969.001.0001

Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.

Kourany, J. A. (2010). *Philosophy of science after feminism*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732623.001.0001

Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension: selected studies in scientific tradition and change* (pp. 320–339). University of Chicago Press.

Lacey, H. (2005). *Values and objectivity in science: the current controversy about transgenic crops*. Lexington Books.

Longino, H. E. (1987). Can there be a feminist science? *Hypatia*, *2*(3), 51–64. https://doi.org/10.1111/j.1527-2001.1987.tb01341.x

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Longino, H. E. (1996). Cognitive and non-cognitive values in science: Rethinking the dichotomy. In J.Nelson (Eds.), *Feminism, science, and the philosophy of science* (pp. 39–58). Springer. https://doi.org/10.1007/978-94-009-1742-2_3

Longino, H. E. (2002). *The fate of knowledge*. Princeton University Press. https://doi.org/10.2307/j.ctv2tvzv0

Maasen, S., & Weingart, P. (Eds.) (2005). *Democratization of expertise?: Exploring novel forms of scientific advice in political decision-making*. Springer.

McDonald-Gibson, C. (2013, April). "Victory for bees" as European Union bans neonicotinoid pesticides blamed for destroying bee population. *Independent*. https://www.independent.co.uk/climate-change/news/victory-for-bees-as-european-union-bans-neonicotinoid-pesticides-blamed-for-destroying-bee-population-8595408.html

McMullin, E. (1982). Values in science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *2*. https://doi.org/10.1086/psaprocbienmeetp.1982.2.192409

Reiss, J. (2019). Expertise, agreement, and the nature of social scientific facts or: Against epistocracy. *Social Epistemology*, *33*(2), 183–192. https://doi.org/10.1080/02691728.2019.1577513

Rooney, P. (2018). The borderlands between epistemic and non-epistemic values. In D. Steel & K. C. Elliott (Eds.), *Current controversies in values and science*. Routledge.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science*, *20*(1), 1-6. https://doi.org/10.1086/287231

Russell, D. (1983). Anything goes. *Social Studies of Science*, *13*(3), 437–464. https://doi.org/10.1177/030631283013003005

Steel, D. (2010). Epistemic values and the argument from inductive risk. *Philosophy of Science*, *77*(1), 14–34. https://doi.org/10.1086/650206

Steel, D. (2018). Qualified epistemic priority. In *Current controversies in values and science* (pp. 49–63). Routledge. https://doi.org/10.4324/9781315639420-4

Steele, K. (2012). The scientist qua policy advisor makes value judgments. *Philosophy of Science*, *79*(5), 893–904. https://doi.org/10.1086/667842

Turner, S. (2001). What is the problem with experts? *Social Studies of Science*, *31*(1), 123–149. https://doi.org/10.1177/030631201031001007

Wylie, A. (2007). Doing archaeology as a feminist: Introduction. *Journal of Archaeological Method and Theory*, *14*, 209-216. https://doi.org/10.1007/s10816-007-9034-4

# Chapter 2

# Integrating Heather Douglas' Framework with an Account of Scientific Evidence: Why and How?

## 2.1. Introduction

Contemporary philosophers of science are increasingly interested in evidence-based policy assessment and the role of value considerations in scientific analysis. This has revitalized some of the traditional philosophical debates about the proper role of values in science, the norms of scientific reasoning, and the cogency of the traditional 'value-free ideal' (Betz, 2013; de Melo-Martín & Intemann, 2016; Douglas, 2009; Elliott, 2011; Hicks, 2018; Jeffrey, 1956; John, 2015; Longino, 1990; Rudner, 1953). Heather Douglas contributes to this debate by offering a normative account that distinguishes legitimate from illegitimate value-permeations in scientific reasoning (Douglas, 2000, 2015; Douglas, 2009).

Douglas conveys that scientific reasoning can legitimately be value-laden while its objectivity stays intact. This is possible, Douglas argues, if the role of non-epistemic

values is constrained to the assessment of evidential sufficiency in the face of uncertainty. More specifically, it is legitimate that non-epistemic values inform an evaluation of not-fully-confirmed hypotheses by providing researchers with information about pertinent risks of being wrong in one's inferences when the available evidence is inconclusive (i.e., inductive risks). Douglas argues that allowing non-epistemic values in science in such an "indirect" manner does not violate the objectivity of science, at least in a significant sense of objectivity that appeals to the virtue of assessing evidence in a detached way (Douglas, 2004). Douglas further suggests that this understanding helps us assess when scientific judgments get dogmatic or irrationally politicized and therefore untrustworthy (2009, pp. 112–114). The most prominent aspiration of Douglas' project is to replace the value-free ideal, which strictly prohibits value considerations in evidential reasoning during the so-called core (internal) stages of scientific justification.

Douglas' project has been found promising by philosophers of science who are interested in the role of values in science (see, for instance, Elliott & Richards, 2017; Kitcher, 2011 for prominent reviews and reflections). Many commentators involved in science-based policy discussions accept that scientific inputs should be sensitive to relevant moral and political considerations without losing their rationality, objectivity, and trustworthiness in some meaningful sense. I, too, am sympathetic to this pragmatic aim. Therefore I am interested in demonstrating in detail how Heather Douglas' rational account of scientific reasoning in the face of ethical and pragmatic considerations, and factual uncertainty, is and can be considered consistent with real-world contexts of evidential decision-making that scientists typically face.

In this chapter, I examine how Heather Douglas' inductive risk framework applies to a specific and carefully selected case of scientific practice. I analyze toxicologists' evidential judgments during a societally important historical episode of toxicological practice, the so-called molecularization of policy-relevant toxicology. Policy-relevant toxicology, which is also often aptly referred to as "regulatory toxicology", is practiced for the purposes of regulation of the industrial use of toxic components. In the U.S., the major research institutions in this area are the National Institute of

Environmental Health Sciences (NIEHS), which is also home to the National Toxicology Program (NTP), the Environmental Protection Agency (EPA), and the National Center for Toxicological Research (NCTR) based in the Food and Drug Administration (FDA). These institutions are tasked with protecting the public and environmental health through supplying scientific risk assessments about the potential adverse effects of the use of various toxic substances.

I shall focus on regulatory toxicologists' judgments about the acceptability of molecular evidential sources and evidence-gathering methods for their purposes. In analyzing this context, I draw on the evidence documented in the relevant historical studies of toxicology (such as (Frickel, 2004; Shostak, 2005, 2013) to offer a descriptive account of how regulatory toxicologists actually make and update methodological decisions about evidence. Based on my analysis, I demonstrate that toxicologists evaluate different kinds of evidence by assessing their epistemic advantages (such as further accuracy and precision achieved through the use of different kinds of evidence) and by considering relevant non-epistemic consequences of their inductively risky methodological decisions (such as the convenience of the use of different kinds of evidence for regulatory use, and for the broader purpose of protecting the public and environmental health). I highlight some philosophically interesting aspects of the case, such as the initial rigidity of toxicologists' inductive risk judgments and the role of social and institutional processes in the formation of their evidential decisions, which Douglas' account, and the broader inductive risk framework, is not designed to address. I argue that Douglas' account of scientific reasoning can be integrated with a suitable conception of evidence, such as Helen Longino's contextual empiricist one, in order to account for the highlighted aspects of the case. I then introduce such an integrated account and demonstrate how Longino's contextual empiricism and Douglas' inductive risk framework fruitfully complement each other in describing the case.

The chapter proceeds as follows. In section 2, I illustrate Douglas' theory for scientific reasoning, and how it can be used for representing and assessing the actual cases of scientific judgment-formation in the face of value considerations and uncertainty

about facts. In section 3, I examine how the inductive risk framework can be applied to understand toxicologists' acceptance of molecular methods and evidential sources over the conventional ones for the purposes of regulatory toxicology, and I highlight the exceptional and theoretically novel aspects of the toxicology case which the inductive risk framework is not designed to address. In section 4, I show how Douglas' inductive risk framework works in harmony with Longino's contextual empiricism to address the highlighted problems such as toxicologists' initial rigidity of the inductive risk judgments and the role of social context in mediating scientists' evidential judgments. In section 5, I conclude with a brief summary of the chapter.

## 2.2. Heather Douglas' Account as a Norm of Scientific Judgment in the Face of Inductive Risks

The so-called "value-free ideal" suggests that scientists' assessment of evidence should be free from non-epistemic values. As a norm of scientific reasoning, this suggestion prohibits appealing to contextual pragmatic or ethical considerations in making evidence-based judgments. The value-free ideal has been widely debated in the modern philosophy of science (Betz, 2013; Bright, 2018; de Melo-Martín & Intemann, 2016; Douglas, 2009; Jeffrey, 1956; Kincaid, Duprè, & Wylie, 2009; Kitcher, 2011; Rudner, 1953). Heather Douglas has contributed to the debate by re-examining the role of inductive risk in scientific reasoning in her *Science, Policy and the Value-Free Ideal* (Douglas, 2009). Once we acknowledge the extended pragmatic role of scientific knowledge in contemporary societies, such as the increasing prominence of scientific expert advisory in risk management and regulation. Douglas proposes an alternative norm for scientific reasoning that preserves the objectivity and rationality of science in a meaningful sense. In this section, I present a summary of Douglas's account, and motivate the project and the case I am going to analyze.

### 2.2.1. Distinguishing Roles of Values to Conceptualize an Alternative to the Value-Free Ideal

Heather Douglas proposes that scientists' judgments are often used for informing policy decisions, and thereby scientists may contribute to morally and politically significant outcomes in the world. In such cases, Douglas argues, scientists cannot and should not resist appealing to non-epistemic considerations in their decision-making. She argues that in many methodological decisions, scientists inescapably need to choose thresholds of evidential sufficiency, asking "how much evidence is sufficient for me to accept or reject a hypothesis?" because scientific inference and judgment often involve ineliminable uncertainties about what is actually true or right (Douglas, 2000, p. 559). Accordingly, Douglas argues that it is rational and morally desirable when scientists consider the non-epistemic consequences of their decisions in contexts where they accept a hypothesis, make methodological choices, or provide support for a course of policy action. She concludes that the value-free scientific reasoning is therefore "flawed" and "incomplete" as a normative and descriptive account of policy-relevant sciences because it precludes scientists from making decisions regarding evidential sufficiency in the face of factual uncertainty and the pertinent need to consider the moral and pragmatic consequences of scientific judgments (2009).

Douglas proposes an alternative norm for scientific reasoning, which conceives of non-epistemic considerations as information used for determining the seriousness of making inductive errors in evidence-based reasoning:

> The scientist will need to consider both the quantity of evidence or degree of confirmation to estimate the magnitude of inductive risk and the valuation of the consequences that would result from error to estimate the seriousness or desirability of the consequences. The weighing of these consequences, in combination with the perceived magnitude of the inductive risk (i.e., how likely one is to be wrong), determines which choice is more acceptable. Where inductive risks are involved and non-epistemic consequences follow from error, non-epistemic values are essential for

> deciding which inductive risks we should accept, or which choice we should make (2000, p. 565).

This suggests that non-epistemic values help researchers determine how much evidence is sufficient to accept or reject a hypothesis or to make a potentially harmful methodological decision. Douglas contrasts this restricted, "indirect" role of values in scientific reasoning with what she calls the "direct" role of values in scientific reasoning. When the role of values is direct, their use goes beyond this supplementary function (that is, helping assess the sufficiency of evidence), and instead, trumps evidential ones or replace them. Only the indirect role of non-epistemic values is permissible in scientific reasoning, she claims, as this indirect role does not harm scientific objectivity in an important and meaningful sense of the term objectivity.

To support this argument, Douglas compares the cases of indirect permeation of values with the cases of direct permeation of values during the internal stages of the scientific method, which consists of the methodological steps that concern scientific justification such as selection and interpretation of data, modeling, and confirmation of hypotheses. The latter involves "wishful thinking", "cherry-picking evidence from a wide variety of evidential sources", and "constructing a methodology that will give results serving [one's] own liking" (2009, pp. 150-52). These kinds of epistemic practices would harm scientific objectivity and invite dogmatism, corruption, or radical politicization of science. In contrast, so Douglas argues, the indirect permeation of values in scientific reasoning is different from these cases and does not lead to an erosion of scientific objectivity. She emphasizes that scientists' consideration of non-epistemic values in the indirect manner preserves detachment and non-dogmatism, which she takes to be one of the essential aspects and the core virtues of scientific reasoning (2009, pp. 112-114). Her proposed alternative to the value free ideal preserves these core virtues of scientific reasoning (Douglas, 2004).

Douglas's argument from inductive risk is therefore promising to conceptualize a new normative benchmark for scientific reasoning (although there are important skeptical arguments such as Betz (2013) who questions whether the value free ideal should be abandoned, de Melo-Martín and Intemann (2016) who question whether Douglas's

account *really* rejects the value free ideal). Douglas's account is indeed interpreted in different ways: as part of an alternative to the traditional value-free ideal, as a tool for policing scientists' reasoning, and as a benchmark to distinguish permissible value-ladenness in science from corrupt or politized use of values in science (see, Elliott 2011, for an extensive critical review of these distinct pursuits of Douglas's account). At the same time, various prominent commentators have discussed the scope of the inductive risk arguments such as Douglas', debate to what extent the inductive risk framework can serve as a full-fledged account of values in science, analyze how Douglas' account and the broader inductive risk approach to values in science relate to the other established descriptive and normative arguments about the role and nature of values in science, and demonstrate how the inductive risk framework applies to different cases of scientific practice (Biddle, 2016; Biddle & Kukla, 2017; Brown, 2013; de Melo-Martín & Intemann, 2016; Elliott & Richards, 2017; Hicks, 2014, 2018; Steel & Whyte, 2012).

In this vein, a growing body of philosophical literature contributes to our understanding of the proper place of values in science by applying the inductive risk framework to the assessment of the actual contexts of scientific decision-making. Here, I aim to contribute to these debates by focusing on an episode of scientific practice that instantiates some aspects of scientific reasoning, which are less commonly examined from the perspective of the inductive risk framework. I will analyze the context in which scientists choose between different kinds of evidential sources and evidence-gathering methods and make judgments about the (inductive) risks pertaining to their decisions. I will focus on the case of regulatory toxicologists' gradual adoption of molecular tools. Specifically, I will examine precisely how the inductive risk framework applies to this case, and how Douglas' account can account for some interesting aspects of toxicologists' reasoning such as the graduality of the changes in their evidential judgments and the role of contextual social and institutional processes in the formation of their methodological decisions. Based on my analysis, I will argue that these aspects of the case can be described by Douglas's inductive risk framework when it is considered to be in synchrony with suitable neighboring philosophical approaches, specifically Helen Longino's contextual empiricism. I aim

to demonstrate that these two accounts complement each other in fruitful ways. Let's then focus on the case and discuss how Douglas's inductive risk framework applies to it.

## 2.3. Analyzing the Case of Molecularization from the Perspective of Douglas' Inductive Risk Framework

Applied toxicologists aim to assemble scientific assessments concerning the toxicity of chemical substances to inform regulatory decisions taken by public health institutions such as the National Institute of Environmental Health Sciences (NIEHS), the Environmental Protection Agency (EPA), and the Food and Drug Administration (FDA) in the U.S. context. Because applied toxicology is a "regulatory science" (as initially termed by Jasanoff, 1990), toxicological practice constitutes a perfect case study for Douglas (2000; 2009) and any philosophical discussion regarding the issues arising in the context of policy-relevant sciences and evidence-based policies. In this section, I will investigate the extent to which Douglas' account successfully evaluates various types of decisions that toxicologists make. Specifically, I will focus on the acceptance of molecular evidential sources for conducting policy-relevant toxicological research—the so-called molecularization of regulatory toxicology.

### 2.3.1. Douglas' Norm applied to Toxicologists' Judgments

The practice of regulatory toxicologists is a very suitable case for illuminating Heather Douglas' account of values, as it is a rigorous scientific practice that is not autonomous from society or policy-related processes. In her reflections on regulatory toxicology, Douglas debunks the division of labor that has been widely assumed to hold between "toxicologists as value-free risk assessors" and "regulators as risk managers" (2009, p. 140). Douglas claims that the depiction of toxicologists as *value-free* risk assessors, who operate in isolation from society and who serve the decision-making processes of value-laden policymakers, is descriptively incomplete. She also

argues that trying to achieve such an ideal in practice is normatively undesirable due to the inductive risks involved in toxicological risk assessment. Instead, Douglas proposes that toxicologists should consciously and transparently take responsibility for triggering environmental risk management by adjusting their thresholds for determining the sufficiency of the evidential grounds in response to potential societal consequences of their decisions. In line with the account illustrated in Section 2, Douglas considers this ("indirect") role of values in toxicology-based judgment formation legitimate.

Douglas applies her account to analyze various critical methodological stages of toxicological science (2000):

- Making a judgment about the severity of the changes observed in exposed tissues (Douglas, 2000, sec. 4);
- Choosing an adequate data-analysis method among alternative dose-response curves to fit the available data (Douglas, 2000, sec. 5);
- Interpreting the results in a particular way (e.g., stating whether a study allows scientists to conclude that the substance under consideration is toxic irrespective of its dosage; or that the toxicity of the substance is contingent on the dosage). (Douglas, 200, sec. 6).

How does Douglas' framework represent the structure of scientific reasoning in these contexts? Take the context in which toxicologists interpret the results. Imagine, for instance, two representative toxicologists (say "T1" and "T2") who take 'environmental health' into account in their reasoning when they assess the hazardousness of chemical substances. T1 reports that "the substance is highly toxic" even though she does not unambiguously detect any significant malignant changes in the exposed tissues observed. T1 nevertheless confirms the toxicity of the substance because T1 refrains from making judgments that might lead to undesirable environmental health outcomes. T2 also reports that "the substance is highly toxic". But, different from T1, T2 detects some malignancies in the exposed tissues she observes. Even though available evidence does not fully confirm the hypothesis that "the substance is toxic", T2 lowers the threshold of evidential sufficiency by

considering the undesirability of environmental risks involved and accepts the hypothesis. It is clear that the non-epistemic value (the protection of environmental health) trumps the evidential considerations in T1's judgment. For T2, however, the concern for environmental health only plays a supplementary role in the reasoning. If we follow Douglas' account, it is then pretty straightforward to deem T1 as biased, dogmatic, and unjustifiably subjective. Similarly, it is pretty simple to judge that T2 preserves "objectivity", as some metaphorical distance from evidence is kept in the assessment while allowing non-epistemic considerations to influence the judgment.

The suitability of Douglas' account to evaluate the toxicologists' judgments in cases like the above bears little controversy. There are some pieces of objects, such as slides of the exposed animal tissues, which have an evidential relationship with toxicologists' hypotheses and, more generally, their judgments. These pieces of evidence confirm toxicologists' judgments, such as concerning the toxicity of a substance, to different degrees. Inductive risks are therefore present, and hence Douglas' norms could readily be applied to describe and assess the permissibility of the ways in which non-epistemic values permeate into scientists' judgments in these cases.

The case of molecularization is about a different methodological context than the one I described above. Unlike the sufficiency of evidence for confirmation, the acceptance of molecular data and methods is a methodological issue about the *relevance* of evidence for the aim of regulatory toxicology. This context (i.e., the choice of evidence-types and how scientists make inductive risk judgments about them) is less commonly examined in the literature on inductive risk (see Biddle, 2016; de Melo-Martín & Intemann, 2016; Biddle & Kukla, 2017; Hicks, 2018 for a discussion of the conceptual and theoretical issues regarding how the inductive risk framework applies to this context).

In the following, I will illustrate that Douglas' inductive risk framework properly applies to toxicologists' judgments about the kinds of evidence, too. My aim is not only to demonstrate how Douglas' account applies to this context but also to highlight

some important aspects of the case that the inductive risk framework is not designed or purported to address.

Before I focus on this context in the next subsection, a disclaimer is in order. In my analysis of the case of molecularization, I will specifically refer to some science studies scholars' detailed examination of toxicologists' judgments concerning molecular evidence (such as Frickel, 2004 and Shostak, 2013) with the aim of providing a comprehensive description of the case. Some readers might anticipate that I will pit the science studies scholars' arguments about toxicology against Douglas' arguments. However, that is not my aim. I will not refer to any *philosophical* claims made by these scholars. Rather, I will rely solely on these authors' characterization of the historical facts regarding the case. I will thereafter defend the philosophical significance of the historical details of the case for the inductive risk framework.

I will now start with a brief explanation of the so-called "molecularization of regulatory toxicology". Then, I will offer a description of how regulatory toxicologists choose between molecular evidential sources and conventional pathological ones and examine how the inductive risk framework captures their reasoning. I will then discuss why it is a good idea to integrate the inductive risk framework with a suitable neighboring approach to analyze the specifics of the case.

### 2.3.2. Choosing between the Molecular and Conventional Methods in Regulatory Toxicology

Classical policy-relevant toxicology used to be predominantly based on pathological examination of exposed animals in laboratory conditions. In this method, laboratory observations of exposed tissues, which are then statistically analyzed by toxicologists, are used for making predictions about the hazardousness of toxic substances under changing dosages based on different dose-response models. [1] This primarily pathology-based, observational method was referred to as the gold standard for

[1] Notice that the examples in previous section, which Douglas examines in her work, pertain to classical toxicology, which is based primarily on pathological studies.

performing policy-relevant toxicological analysis until the 2000s (see, for instance, National Toxicology Programs' seminal report (2004), where a vision for a change is introduced).

The received methodology of policy-relevant toxicological science has changed dramatically toward the adoption of molecular methods, especially in the United States. Contemporary toxicological science, at least in the context of the United States, including toxicology practiced for regulatory purposes, has been systematically moving towards becoming a genomic and molecular science (see, a prominent consensus report published by the National Academy of Sciences (2017) that reviews the developments in the last two decades). Though some controversy remains and the widespread application of new methods to the regulatory realm is still a project in progress, the relevance and the usefulness of molecular methods for informing environmental health policies are widely accepted by the major institutions of regulatory toxicology in the U.S. The research outputs of NTP at NIEHS and NCTR at FDA or the activities of ToX21 collaboration attest to this change in major institutions' methodological judgments about molecular methods (see EPA 2019 for a review).[2] Regulatory toxicologists nowadays increasingly rely on data gathered through advanced micro array technologies and advance quantitative tools such as high-throughput screening, which affords a comprehensive and ever more precise measurement of the simultaneous effects of multiple toxic chemicals across genetically different populations and individuals.

Nevertheless, the shift towards the molecular methods in applied toxicology, which is labeled "the molecularization of toxicology" by science studies scholars (Shostak,

[2] The debate over the acceptability of molecular evidence and tools in regulatory toxicology may still persist among different groups of toxicologists (see, for instance, Hicks (2018) analysis about the acceptability of a particular molecular model for endocrinal risk assessment). In the light of my analysis, I will also reflect on how the inductive risk framework may interpret such disagreements among different communities of toxicologists (such as the one described by Hicks) later in Section 4. However, it is beyond dispute that a substantial shift has been taking place toward the adoption of the molecular methods, and that the major institutions of toxicology no longer consider the traditional methods as the gold standard for conducting regulatory risk assessments, and their strategical aim is to move further towards the adoption of molecular tools for conducting regulatory risk assessments.

2005; 2013), has been highly controversial among toxicologists and relevant stakeholders such as regulators, industrial organizations, and environmental advocacy groups. As we will see, toxicologists' reasoning about the adoption of molecular methods was influenced by epistemic considerations such as the high precision of the molecular evidential sources as well as non-epistemic ones such as how regulatory toxicologists should protect public health and reduce ecological and environmental health risks.

Broadly construed, the controversy regarding the choice of the molecular over the pathological method revolves around the following. Many applied toxicologists and relevant stakeholders such as research-based environmentalist advocacy groups initially proposed that molecular evidence-gathering methods and evidential sources were ill-suited for the pragmatic and regulatory roles of toxicology. Among other things, these roles include informing legislators about how the industrial production of chemical substances should be regulated and performing evaluations of ecological and health risks associated with various toxicants so as to conserve environmental health. The traditional pathological evidence-gathering methods and evidential sources were deemed entirely sufficient for serving these purposes, as they were conducive to assessing hazards of one chemical substance in isolation from other causally relevant chemical and genetic factors. Furthermore, to venture into comprehensive analyses of the complex molecular and genetic mechanisms of disease causation was thought to be impractical and inefficient for regulatory toxicology, however interesting it might be for purely scientific purposes. Some toxicologists working at prominent regulatory institutions in the U.S. (e.g., NTP-NIEHS, NCTR-FDA, and EPA) or researchers from environmental advocacy and justice groups (e.g., West Harlem Environmental Action, WEACT) thought that the epistemic benefits of molecular methods, such as greater precision and explanatory power, could only be achieved at the expense of making environmental regulation more complicated and less effective (Shostak 2013, 64-70). Hence, so their view went, these methods would potentially serve the interests of profit-seeking industries, which aim to minimize the financial costs of toxicological regulation. Here are some excerpts from the interviews that a prominent science studies scholar, Sara Shostak, conducted with applied

toxicologists working at NTP, which exemplify these kinds of worries about the increasing use of molecular methods in regulatory toxicology:

> [at NTP] we need some people with practicality. We need some people with skills in toxicology . . . empirical descriptive toxicology. [If] you find out something causes cancer, then let somebody else mess around with the mechanism…. I don't want to know how it does it . . . I want to know, "Is this safe?" (Shostak, 2013, p. 64).
>
> My interest is in what can we change to make people healthier? We can change exposures. . .You can't change your gene pool (Shostak, 2013, p. 65).
>
> Any new technology, it's always a good delaying tactic for environmental health risk assessments (Shostak, 2013, p. 66).

While the molecular kinds of evidence are conducive to more accurate and precise toxicological risk assessment, toxicologists based in prominent regulatory institutions initially did not accept the relevance of molecular kinds of evidence for their inquiry. Their reasoning was that molecular evidence-gathering methods complicated toxicological risk assessment as they are conducive to more detailed results, the significance of which is ever more difficult to translate into judgments that can conveniently be used for regulatory purposes. Working with molecular evidence-gathering methods thereby slows down and renders ambiguous the regulatory process, hence harming their mission of protecting environmental health.

It seems that these toxicologists took the relevant inductive risks into account when making pertinent negative judgments about the acceptability of molecular kinds of evidence. Given the risk of making a wrong methodological decision for the pragmatic purposes of regulatory toxicology, some toxicologists such as the ones quoted believed that they should keep low standards for evidential sufficiency (i.e., the level of evidential precision and certainty they deem sufficient for the regulatory purposes of toxicology). This belief was shaped by taking into account the environmental health risks associated with the purported impracticality of the molecular methods for serving policy-relevant toxicological assessment.

While we can explain toxicologists' reasoning by reference to the language of Douglas' inductive risk framework, the content of toxicologists' inductive risk

judgment in this instance is also puzzling from the perspective of Douglas' account. Note that those toxicologists referred above who were skeptical toward the molecularization *did not* argue that the adoption of molecular kinds of evidence would worsen the precision, accuracy, and the overall quality of evidence used for conducting toxicological risk assessment. On the contrary, it seems that the toxicologists acknowledged that the molecular kinds of evidence would enable them to make more precise and accurate assessments than they could do through the classical kinds of evidence (Shostak, 2013, pp. 48-64). They nevertheless resisted adopting molecular kinds of evidence by considering the magnitude of inductive risks regarding the protection of public health and the ease of regulation. This kind of reasoning is not fully in line with Douglas' account because more precise and accurate evidence is supposed to *decrease* inductive risks, according to her theory. In other words, when the available evidence is more certain, the chance of making wrong decisions becomes lower; hence the magnitude of inductive risks is supposed to be lower, too (Douglas, 2009, p. 96). Yet, puzzlingly, in this case, if the toxicologists' assessment were right, then additional or better evidence would lead to more uncertainty, not less.

The same issue has been recently highlighted by Hicks through a case study about the practice of policy-relevant molecular toxicology (Hicks, 2018). Hicks also observes that additional evidence, of the kind achieved using molecular models in the studies of endocrine disruptors, does not decrease uncertainty. They note that "uncertainty here is not the result of limited evidence but of more accurate or precise evidence" (2018, p. 170). Furthermore, they label this as "an important counterexample or exception" to Douglas' norm (p. 170). For now, I flag this aspect of the case as an important issue to be addressed. I will respond to this problem later, in Section 4.

An equally interesting —in fact the main—aspect of the case of molecularization is that the prominent communities of toxicologists who were initially skeptical about the molecular approaches to toxicological risk assessment later *updated* their judgments and gradually accepted the use of molecular evidential sources and evidence-gathering methods (Shostak, 2013). In so far as this change in toxicologists' judgment

is described in terms of the language of the inductive risk framework, it must be the case that the content of toxicologists' inductive risk judgments changed. The same toxicologists must have concluded that the use of molecular evidential sources did not increase inductive risks, unlike what they had initially thought.

I will now describe how this change in toxicologists' judgment happened, drawing on the relevant historical studies. Thereafter, I will discuss why and how the historical details of the 'molecularization' can suitably be addressed by an analysis of the case from the perspective of the inductive risk framework.

### 2.3.3. The Transition to Molecular Regulatory Toxicology: The Role of Social and Institutional Processes

Let's review the relevant historical facts about how major communities of toxicologists in the U.S. gradually adopted the molecular approaches to toxicological risk assessment, and then ask ourselves whether and how we should fit the details of the case with Douglas' inductive risk framework.

Prominent historical studies of contemporary toxicology (such as Frickel, 2004; Shostak, 2005, 2013) describe the process of molecularization as one that is mediated by a set of sociological and institutional processes. According to these studies, a significant macro-sociological process that contributed to the molecularization of regulatory toxicology was decreasing financial and institutional support for regulatory toxicology. Although not adopting molecular methods would not have practically prevented toxicologists from performing effective policy-relevant analyses, regulatory toxicologists increasingly met criticism for not using what were considered cutting-edge scientific methods (Shostak, 2013, Chapter 3). In particular, these criticisms were raised by commercial scientists working in profit-seeking industries (2013, p. 6). In addition, the popularity of genomic studies had increased in the general public, and the neighboring disciplines had already adopted molecular methods. These macro-scale developments led regulatory toxicologists to worry about losing the financial and social support they received from governmental and non-governmental

organizations. In the 2013 book, Sara Shostak extensively documents how leading communities of applied toxicology in the U.S. interpreted these macro-sociological changes as *threats* to the scientific authority of toxicology for regulatory purposes. Shostak explains that such financial and institutional considerations boosted the toxicologists' gradual adoption of molecular methods.

Shostak also describes that the major communities of regulatory toxicologists initiated a set of strategic micro-institutional processes in response to these macro challenges to their inquiry. Those developments, too, contributed to the gradual adoption of the molecular evidence-gathering methods over pathological ones in regulatory toxicology. Here are some examples of these institutional processes toxicologists deliberately engaged in:

- Initiating new institutional ties with research centers, which were formerly thought to be only remotely related to applied-toxicological practice (such as the National Human Genome Research Institute, NHGRI) (Shostak, 2013, pp. 141-143).
- Forming new research centers that aim to raise funds for promoting policy-relevant molecular research in toxicology, including providing infrastructure for training new toxicologists who are experts in both classical toxicology and molecular methods (Shostak, 2013, pp. 143-150).
- Defining a research agenda for exploring applications of gene expression profiling specific to regulatory toxicology. This includes encouraging the use of adequate tools (such as ToXChips) for translating the new type of data into the old forms of regulatory procedures (Shostak, 2013, p. 147).
- Establishing research initiatives such as the Toxicogenomics Research Consortium to develop new methodological procedures and standards relevant for the use of new molecular risk assessment models, and to advance tools so as to allow the transparent and standardized use of new kinds of evidential sources (Shostak, 2013, pp. 156-157).

- Organizing workshops such as consensus-building forums that aim to convince the relevant stakeholders (such as governmental agencies, funding agencies, private sector participants, environmental justice groups, and practicing scientists), explain the need for molecular methods in performing regulatory toxicological practice (such as the large-scale information forum "the Committee on Emerging Issues and Data on Environmental Contaminants"), and shape users' and producers' expectations of toxicological practice (Shostak, 2013, pp. 160-166).

Now, these are important descriptive details regarding how major communities of regulatory toxicologists gradually changed their judgments in favor of the molecularization of regulatory toxicology. I want to highlight two aspects of the case based on this description. First, according to the historical details of the case, the updating of toxicologists' evidential judgments was mediated by a variety of social and institutional processes working in the background such as disciplinary politics, economic and cultural pressure, and toxicologists' own attempts to build a new consensus through academic and technical innovations. Second, and by implication of the first, the updating of toxicologists' evidential judgments was *gradual* and followed the social and institutional processes I have briefly described.

My contention is that these highlighted observations about the historical details of the case are philosophically significant, and an analysis of the case from the perspective of the inductive risk framework would get conceptually stronger by properly addressing and accounting for them. More generally, I suggest, these aspects of the case are important for our understanding of the nature of toxicologists' inductive risk judgments. In the following, I will first explain why this is the case and then offer a promising way through which Douglas' inductive risk framework can address these details.

### 2.3.4. The Need to Account for the Role of Social and Institutional Processes in the Formation of Toxicologists' Inductive Risk Judgments

As we have seen, toxicologists consider the pertinent inductive risks in judging the acceptability of the molecular evidential sources in regulatory toxicology. They do so by weighing the epistemic quality of molecular evidence (e.g., additional precision and accuracy in toxicological risk assessment) and the perceived non-epistemic consequences of adopting molecular methods (e.g., making toxicology-based regulation more complicated, slowing the pace of risk assessment, and ambiguating the evidence that is supposed to trigger policies which preserve environmental health). Moreover, we have seen that the content of toxicologists' inductive risk judgments pertaining to the adoption of molecular approaches has changed over time. Furthermore, the historical details of the case suggest that social and institutional processes have mediated this change.

The role that sociological and institutional processes may play in the formation of scientists' various decisions are not referenced by the standard presentations of the inductive risk framework such as Douglas'. This is natural because the inductive risk framework is not designed to address the role of social context in the formation of scientists' judgments. Our question is then the following: How should an analysis of the molecularization of toxicology from the perspective of the inductive risk framework account for the relevant historical facts reviewed above? In other words, what would be a plausible interpretation of the relevant historical details of the molecularization from the perspective of Douglas' account?

A plausible response would be to *deny* the philosophical significance of these historical details for an inductive-risk-analysis of the case. It could be that the documented historical facts, which describe how some toxicologists came to accept the relevance of molecular methods and evidential sources, are some nitty-gritty details of precisely how the toxicologists updated their inductive risk judgments. The role of sociological mediating factors, so this response goes, is redundant in an analysis of the case from the perspective of inductive risk. I will ultimately reject this

response. To do so, I will present how the standard presentation of the inductive risk framework can describe the change in toxicologists' judgments, excluding the details about the role of social and institutional context. Then, I will discuss why the standard reconstruction of the case would be incomplete.

Recall the logic of the inductive risk judgments that I summarized in Section 2. Douglas suggests that scientists consider the inductive risks of their methodological choices based on an assessment of the magnitude of the non-epistemic consequences of making a risky decision and an assessment of the evidential uncertainty (that is, an estimation of how accurate and confirmatory the available evidence is). According to this standard presentation of scientists' decision-making, a change in the content of scientists' inductive risk judgments may result from a change in one or two of these variables that enter into the inductive risk calculation. That is, the content of scientists' inductive risk judgments could follow a change in scientists' assessment of the magnitude of the adverse consequences stemming from error, i.e., a change in their evaluation of relevant non-epistemic considerations. The change could also follow an update in the confirmatory power of available evidence (for example, by producing more precise and accurate evidence, possibly through technological advancement).

Now, in describing how toxicologists first rejected the molecularization and then later updated this judgment and adopted the molecular types of evidence, we cannot reasonably attribute this to a change in their evaluation of the importance of public and environmental health. Nor can we cogently claim that the major advisory institutions of regulatory toxicology abandoned their pragmatic mission of producing evidence that can be used for regulatory action. Hence, the acceptance of molecular methods is not reasonably attributable to a shift in toxicologists' assessment of the relevant non-epistemic consequences. Accordingly, if we want to stick to the standard model of the inductive risk framework, we would expect that the shift in toxicologists' judgment should have resulted from some improvement in the perceived confirmatory power (and epistemic quality) of the molecular tools and evidential sources for the purposes of regulatory toxicology. For instance, toxicologists might have reasoned that accepting molecular kinds of evidence would not only improve the accuracy and

the precision of the toxicological risk assessment but also do so without bearing additional costs for the pragmatic, regulatory function of toxicology. Toxicologists might have, therefore, chosen to raise their standards for evidential quality, as that decision no longer posed serious inductive risks – contrary to what they initially thought.

If this rational reconstruction of the case of molecularization is persuasive, then toxicologists' acceptance of molecular kinds of evidence can be explained purely in terms of Douglas' standard presentation of the inductive risk model. The historical details of the case that I highlighted above (that is, how the social and institutional processes mediated the shifts in toxicologists' judgments) would then be philosophically uninteresting details of *precisely* how toxicologists made their judgments about the relevance of molecular kinds of evidence for their inquiry. However, I argue that the reconstruction of the case I have just offered is incomplete in an important respect, and that the role of the social context would better be addressed in an analysis of the case in terms of the inductive risk framework. What is then missing in the presented reconstruction of the case of molecularization?

Admittedly, it is true that the molecular risk assessment models in contemporary regulatory toxicology are highly sophisticated such that they effectively fulfill the regulatory and societal purposes of contemporary toxicology. Nevertheless, this sophistication is not directly followed by some drastic advances in molecular technologies. On the contrary, the molecular methods used in contemporary toxicological research were already being used by scholars in neighboring disciplines such as bio-medicine, biology, and even in less applied parts of toxicology (such as the genetic branch of toxicology that deals with the mechanisms of toxicity irrespective of policy-relevant questions) *decades before* the adoption of these technologies in regulatory toxicology. Applied (regulatory) toxicology appears to be the *last* scientific field among its neighbors that adopted the molecular methods and made use of molecular evidential sources (see Frickel, 2004 for an exhaustive historical account). Indeed, this fact is precisely why the molecularization of toxicology, or researchers' initial resistance to it, is such an interesting case to

investigate in the eyes of science studies scholars in the first place (Frickel, 2004; Shostak, 2013). Hence, it is not as straightforward as to say that the molecular methods in toxicology had advanced so much that adopting these new methods no longer bore the risk of making policy-relevant toxicological analysis less practical.

The reconstruction of toxicologists' reasoning offered in the standard presentation of the inductive risk framework is, thus, not compelling because it does not explain *why toxicologists did not adopt the advanced technologies earlier*. In my reading, the change in toxicologists' judgments and the graduality of the change can be explained by reference to institutional and sociological processes that mediate regulatory toxicologists' acceptance of molecular kinds of evidence. In other words, in the absence of these social and institutional processes, toxicologists' acknowledgment of epistemic virtues of molecular methods over the pathological methods, and only that, would not be enough to change their inductive risk judgments. Only after these contextual changes took place did regulatory toxicologists gradually update the content of their inductive risk judgments and accept the relevance and aptness of molecular approaches for their purposes. The social and institutional processes (including broad sociological processes or the institutional activities of the communities of toxicologists) therefore played a philosophically significant role in mediating the changes in toxicologists' inductive risk judgments.

This argument supports the need to account for the social processes for an analysis of the case like the molecularization through the language of the inductive risk framework. It also motivates my proposal to integrate Douglas' inductive risk analysis with neighboring philosophical accounts that can accommodate the idea that social contextual factors may sometimes scaffold the formation of scientists' evidential judgments. Discussing how this can be done will be my next task.

In order to accomplish this task, I propose that we do not need to radically transform Douglas' inductive risk framework such that it also accounts for the relevant social dimensions of scientific reasoning. Such a modification would be unnecessary as Douglas' account and the broader inductive risk framework are not designed to examine the role of social and institutional factors in scientists' reasoning, and more

importantly, doing so is not their main philosophical function. The main function of the inductive risk framework is that it gives us a plausible rational account of scientific reasoning in the face of factual uncertainty and the pertinent need to consider pragmatic and moral consequences of scientists' decisions. In line with Douglas' compatibilist reading of the literature on values in science (2015), my strategy will be to show that the inductive risk framework is compatible and works in harmony with the relevant philosophical accounts that accommodate the constitutive role the social and institutional processes may play in mediating scientists' research decisions. I will then demonstrate how such an integrated understanding of the inductive risk framework helps us address the novel aspects of the case I have identified so far.

## 2.4. Integrating Douglas' Account with Longino's Contextual Empiricism

I have so far argued that it is desirable for Douglas' account and the inductive risk framework to account for the social and institutional processes that mediate scientists' evidential judgments when examining cases like the molecularization of regulatory toxicology. I have also proposed that a reasonable way to do so is to show how Douglas' account can be complemented with a relevant philosophical framework that gives special attention to the contribution of social and institutional surroundings to individual scientists' research decisions. What immediately comes to mind is Helen Longino's contextual empiricist conception of scientific evidence (1979; 1990; 2008). Longino's framework is suitable to integrate with Douglas' account of values in science in the context of our discussion, not only because the former entails a social perspective on scientists' decisions about evidence (or, more precisely, on the question of why certain objects and states of affairs are considered by scientists as *relevant* evidence), but also because Douglas (2004; 2009, p. 18) considers Longino's broader framework as compatible and complementary to her own project. In this section, I will discuss how Longino's conception of evidence can be integrated with Douglas's inductive risk framework and how this integrated account works in the analysis of the molecularization case.

### 2.4.1. Integrating Longino's and Douglas' Accounts

Helen Longino has an explicit conception of evidence and evidential reasoning that describes how objects and states of affairs acquire evidential status (1979). Longino famously argues that there is no intrinsic feature of any states of affairs or objects that make them evidential for a given hypothesis, and that states of affairs or objects acquire evidential status for hypotheses or theories only in light of some contextual background assumptions (1990). These background assumptions may sometimes be products of the broader social and cultural context within which scientists work, and the micro-sociological institutional context such as issues like disciplinary politics, methodological conventions, financial and technological constraints, institutional interactions, disciplinary preferences and interests (2008).

Unlike the pessimistic philosophical views about the social dimensions of science, Longino famously put forward that the social character of science is key to understanding the rationality and trustworthiness of scientific inquiries. Most prominently, for instance, she argues that social institutions of science that facilitate critical interactions among the members of a scientific community or members of different communities would eliminate the potential dominance of biased viewpoints over others (1990). In the same vein, in her more recent work (2002), Longino has argued that the social and institutional background context of science, which historical studies of science examine thoroughly, can in principle contribute to the rational progress of individual scientists' research decisions rather than deteriorate or bias them.[3]

According to Longino's framework, then, toxicologists' acceptance of new, molecular kinds of data as relevant evidence for their inquiry can plausibly be mediated by the

[3] Longino (2002) contrasts her account with the approaches that consider "the social" harmful to rational decision-making in science. While prominent scholars of sociology of science interpret the documented influence of the social context in scientists' decisions as evidence of the irrationality of science, prominent philosophers of science have developed normative accounts of science which consider the social factors as deteriorating or irrelevant for the rationality in science. Longino (2002) subscribes to *neither* of these traditions and instead seeks a cogent reconciliation between these two canonical perspectives.

sociological and institutional processes such as the ones I have reviewed here. In line with Longino's ideas, these processes can, in principle, play a constitutive role in the sense that they scaffold toxicologists' inductive risk considerations about the kinds of evidence. I suggest that this theoretical possibility provides us with a basis for integrating Longino's and Douglas' accounts.

While Longino's account supplies a prominent theory of how the social processes in science can plausibly change scientists' judgments about evidence and methods, Douglas' account supplies a theory of the principles that make scientists' inductive reasoning reliable and legitimate in the face of factual uncertainty and the pertinent need to consider moral and pragmatic consequences of scientific judgments.

My suggestion is that these two theories complement each other in describing the case of molecularization I have examined here. Specifically, Longino's theory is suitable and helpful in describing the case of molecularization because it accounts for the philosophically significant role that social processes played in changing toxicologists' inductive risk judgments. At the same time, Longino's framework is complemented by the inductive risk framework, as her own account of values in science does not put forward a theory of individual scientific decision-making under inferential risk.

Now, let's put this integrative understanding to use and provide an example of how Douglas' inductive risk framework and Longino's contextual empiricism complement each other in analyzing the case.

### 2.4.2. Back to the Case of Molecularization

Armed with the inductive risk framework integrated with Longino's contextual empiricism, we can address some of the puzzling aspects of the case again. Specifically, we can account for why some toxicologists only *gradually* accepted molecular methods about which they were initially skeptical in a way that is exceptional to Douglas' theory of scientific reasoning. Moreover, we can supply a plausible explanation of some persisting disagreement between different communities of toxicologists about the acceptability of the molecular approaches to regulatory toxicology.

Remember that regulatory toxicologists were initially skeptical toward molecular kinds of evidence and thought that molecular methods increased inductive risks even though these methods were conducive to a more precise and accurate measurement of environmental health risks. This is, however, puzzling from the perspective of the inductive risk framework, which has been noted by Hicks (2018) as an important counterexample to Douglas' norm of reasoning under inductive risk, as I have flagged earlier in Section 3.2. While Hicks gives us a plausible explanation of this apparent exception by suggesting that molecular models increase evidential uncertainty at another level (that is, in relation to the question of what ranges of potency are relevant for toxicological risk assessment (Hicks 2018, p. 170), the integrated account proposed here complements Hicks' interpretation by supplying an explanation of why this kind of apparent exception may have arisen and how toxicologists' reasoning is still in line with the inductive risk framework. It goes as the following.

Individual toxicologists' evidential judgments *can* be intransigent at first but change gradually in response to a set of social and institutional changes, as the case of molecularization manifests. While the observed judgment of toxicologists who were initially skeptical about the adoption of molecular tools seems exceptional to Douglas' theory of scientific reasoning, the toxicologists working at the major regulatory institutions did gradually update the content of their inductive risk judgments in line with Douglas's norms, followed by a set of social changes (i.e., active institutional attempts to re-evaluate the contribution of molecular sources to regulatory toxicology and to establish the relevance of molecular tools for regulatory toxicologists). In other words, toxicologists came to accept that adopting molecular approaches does not necessarily increase inductive risks, but this change in their judgment was mediated at least partly by a set of social and institutional processes that helped major communities of toxicologists establish the new methodological standards. Accordingly, once we acknowledge the mediating role of the social and institutional processes in the formation of toxicologists' judgments, and accept that the updating of toxicologists' judgments can be a gradual process, then the toxicologists' observed reasoning does not conflict with Douglas' inductive risk framework. Douglas' account of scientists' reasoning is still descriptively adequate in this context, but the structure

of reasoning described in Douglas' account seems to be instantiated gradually in practice and scaffolded by a set of institutional and social factors in line with Longino's constitutive understanding of the social dimensions of scientific judgment-formation.

Consequently, by showing how Douglas' account is complemented by a suitable philosophical account such as Longino's contextual empiricism, we can plausibly explain an interesting aspect of the case of the molecularization and do so from the perspective of the inductive risk framework.

The integrated account can also provide us with an explanation of why there are still ongoing disagreements between different communities of toxicologists about the acceptability of molecular approaches to regulatory toxicology. To make this point, I will refer to a concrete example from Hicks' 2018 study. Hicks (2018) focuses on a recent controversy in regulatory toxicology in which toxicologists assess a set of chemicals as potential endocrine disruptors by using a *molecular* data-generating model referred to as the ER model. Hicks observes that different toxicologists' evaluations of inductive risks attached to the use of certain molecular evidence-gathering methods for toxicological risk assessment are inconsistent. Moreover, they report, this is reflected in conflicting estimations of risks calculated by different groups of regulatory toxicologists who made use of the same methods in calculating the risks. Specifically, the NRDC (Natural Resources Defense Council), an environmental advocacy organization, has argued against the use of the molecular model in question because NRDC claims that the method generates more inconclusive data-points than the conventional methods do, however precise and accurate those results might be. Given their purpose of environmental protection, NRDC does not favor the use of the method because, when used for regulatory assessments, it motivates less aggressive regulations of the industrial use of the chemicals in question. Hicks contrasts NRDC's judgment with that of the EPA-NIEHS. The researchers and regulators at the EPA-NIEHS collaboration are in favor of adopting the molecular method (the ER model). Unlike NRDC's judgment, EPA-NIEHS's use of the same model generates less inconclusive evidence and motivates a more aggressive

regulatory action when used for justifying regulatory decisions. Hicks reports that the difference between the two judgments results from "different ways of handling inconclusive chemicals" that score low-range response rates. While EPA-NIEHS's methodological approach allows researchers to interpret such data as evidence of toxicity, NRDC's methodology interprets the same results as "inconclusive" (Hicks, 2018, pp. 168-169).

Hicks hypothesizes that the discrepancy between the views of two groups of toxicologists might be the result of a difference in how much they value the protection of the environment: "it is highly plausible that [NRDC's] calculations are directly motivated by their concern to protect human health and the environment" (p. 170). But, at the same time, they rightly observe that NRDC's resistance to EPA-NIEHS's approach is notable because "EPA-NIEHS's interpretation of inconclusive chemicals would be much more protective of these values" (p.170).
Since Hicks is interested in another aspect of the case he focuses on, they stop there and do not question the persistence of this disagreement. But, the disagreement between these groups of toxicologists about the content of inductive risks is hardly attributable to differences in evaluations of relevant non-epistemic values. Indeed, EPA and NIEHS are also explicitly motivated by the concern to protect environmental health.

The integrated account I have proposed here supplies another plausible hypothesis that explains the disagreement between these two groups of toxicologists, which again complements Hicks' analysis. In my reading of the example, it is plausible that the two groups of toxicologists have gone through different institutional processes. EPA and NIEHS have reached the judgment that the use of molecular methods is effectively conducive to the protection of health, and they have built a new set of shared methodological and technical standards regarding how to use molecular tools for the regulatory purposes of toxicology effectively. (Recall the social and institutional processes that mediated the formation of this consensus, such as those described by the historical studies of toxicology or the activities such as the formation of the ToX21 collaboration). In the context of Hicks' case study, EPA-NIEHS's

judgment is about how to interpret low-range response rates that are generated through the use of molecular methods. While the EPA-NIEHS consensus advises to count low-range potency rates as conclusive evidence of toxicity in the context of endocrinal disruption studies, NRDC does not seem to accept this new methodological standard *even though* doing so would be more protective of the non-epistemic values that they aim to protect.[4]

Accordingly, it is a plausible hypothesis that the researchers at NRDC have not participated in those social and institutional processes that helped the researchers at EPA-NIEHS update their inductive risk judgments about the acceptability of the molecular approaches. Recall, for instance, the consensus-building forums that aimed to introduce and promote the use of new methods in regulatory toxicology, which prominent institutions of regulatory toxicology such as NIEHS initiated (Shostak, 2013, pp. 160-164; see also p. 176 for an example of such forums with the relevant environmental justice groups). From the perspective of the integrated account of the inductive risk framework presented here, such forums are good examples of institutional processes that could help the researchers from environmental advocacy groups such as NRDC share or update their methodological judgments.[5]

[4] In this specific example, it seems rational for NRDC to subscribe to EPA-NIESH's acceptance of the molecular method (the ER model) and the pertinent methodological standards. Note, however, that my account does *not* purport to make a normative assessment of the *success* of regulatory toxicologists' changing judgments or the *quality* of the institutional and social processes they have attended to. Engaging in such an evaluation would require one to qualify the integrated account proposed here in normative terms; specifying the conditions under which the mediating social processes are conducive to reaching the right or desirable content of inductive risk judgments. Obviously, one could ask, for instance, how far Longino's normative criteria for social interaction between scientific communities (such as enhanced inclusiveness) are fulfilled in the case of molecularization (Longino, 1990); or to what extent the financial/industrial interests have permeated into the social and institutional processes similar to the ones described here, as it would rightly be questioned by many philosophers of science (e.g. Elliot, 2014). I thank an anonymous referee of this journal for encouraging me to highlight these important questions that I do not address in this article.

[5] It is also intriguing to analyze how and why the mediatory role of institutional and social processes becomes dysfunctional or fails to generate (desirable) forms of agreement between different communities of scientific researchers in this case or in similar cases. This is another issue I do not attempt to analyze in this article.

This example is yet another demonstration of why the mediating role of social and institutional processes is philosophically significant for our understanding of toxicologists' inductive risk judgments, and why integrating Douglas' account with suitable philosophical frameworks, such as Longino's contextual empiricism, is useful to reflect on the cases such as the molecularization of toxicology.

## 2.5. Conclusion

We have seen that the prominent regulatory toxicologists' choice of molecular kinds of evidence over traditional ones is informed by some assessment of how suitable the different methods are for their mission of protecting public health and the environment. In particular, that assessment involves accounting for non-epistemic considerations such as the seriousness of potentially harmful consequences that could follow from toxicologists' choices over different kinds of evidence (e.g., slowing the pace of producing regulation-related analysis or making toxicological assessments less intelligible for the regulatory action). The case I have presented here, therefore, instantiates a context in which the inductive risk framework applies to scientists' choices across different types of evidence.

In this context, I have offered a description of how and why the content of regulatory toxicologists' inductive risk judgments changed over time, drawing on the relevant historical studies of contemporary regulatory toxicology. In doing so, I have focused on some aspects of toxicologists' judgments that Douglas' inductive risk framework is not purported to address, namely the initial rigidity of their evidential judgments, and the graduality of the updating of these judgments, and the role of institutional and social processes in mediating toxicologists' judgments. I have argued for the philosophical significance of these aspects of the case and suggested that Douglas' account, and the broader inductive risk framework, should be able to suitably address them.

I have proposed an account that integrates Douglas' inductive risk account and Longino's contextual empiricism as a suitable philosophical account of evidence that can accommodate the idea that social contextual factors may sometimes plausibly

mediate scientists' evidential judgments. I have then shown how Longino's contextual empiricism and the inductive risk framework fruitfully complement each other in analyzing the specific questions the toxicology case raises, which are also of interest to the specialized philosophical literature.

The case of the molecularization of regulatory toxicology motivates the need to consider how our sophisticated philosophical and historical accounts of scientific judgment relate to each other. And, the integrated account proposed here instantiates an exploration where such connections between Douglas' account of values in scientific reasoning and Longino's contextual empiricism are drawn and then put into use to describe and understand toxicologists' acceptance of molecular approaches to regulatory toxicology.

# References

Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, *3*(2), 207–220. https://doi.org/10.1007/s13194-012-0062-x

Biddle, J. B. (2016). Inductive risk, epistemic risk, and overdiagnosis of disease. *Perspectives on Science*, *24*(2), 192–205. https://doi.org/10.1162/POSC_a_00200

Biddle, J. B., & Kukla, R. (2017). The geography of epistemic risk. In K. Elliott & T. Richards (Eds.), *Exploring inductive risk: Case studies of values in science* (pp. 215–237). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467715.003.0011

Bright, L. K. (2018). Du Bois' democratic defence of the value free ideal. *Synthese*, *195*(5), 2227–2245. https://doi.org/10.1007/s11229-017-1333-z

Brown, M. J. (2013). Values in science beyond underdetermination and inductive risk. *Philosophy of Science*, *80*(5), 829–839. https://doi.org/10.1086/673720

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, *67*(4), 559–579. https://doi.org/10.1086/392855

Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, *138*(3), 453–473. https://doi.org/10.1023/B:SYNT.0000016451.18182.91

Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press. https://doi.org/10.1177/0094306110367909q

Douglas, H. (2015). Values in science. In P. Humphreys (Eds.), *Oxford handbook of philosophy of science* (pp. 610-630). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199368815.013.28

Elliott, K. C. (2011). Direct and indirect roles for values in science. *Philosophy of Science*, *78*(2), 303–324. https://doi.org/10.1086/659222

Elliot, K. C. (2014). Financial conflicts of interest and criteria for research credibility. *Erkenntnis*, *79*, 917-937. https://doi.org/10.1007/s10670-013-9536-2

Elliott, K. C., & Richards, T. (Eds.). (2017). Exploring inductive risk: Case studies of values in science. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467715.001.0001

Elliott, K. C., & Richards, T. (2017). Exploring inductive risk: Future questions." In K.C. Elliott & T. Richards (Eds.), *Exploring inductive risk: Case studies of values in science*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467715.003.0013.

EPA (Environmental Protection Agency). (2019). *United States federal government tox21 collaboration advancing toxicology to improve environmental health and pharmaceutical safety*. US Environmental Protection Agency.

https://tox21.gov/wp-content/uploads/2019/04/Tox21_FactSheet_Apr2019.pdf

Frickel, S. (2004). *Chemical consequences: Environmental mutagens, scientist activism, and the rise of genetic toxicology*. Rutgers University Press.

Hicks, D. J. (2014). A new direction for science and values. *Synthese*, *191*(14), 3271–3295. https://doi.org/10.1007/s11229-014-0447-9

Hicks, D. J. (2018). Inductive risk and regulatory toxicology: A comment on de Melo-Martín and Intemann. *Philosophy of Science*. *85*, 164-174. https://doi.org/10.1086/694771

Jasanoff, S. (1990). *The fifth branch: Science advisers as policymakers*. Harvard University Press.

Jeffrey, R. C. (1956). Valuation and acceptance of scientific hypotheses. *Philosophy of Science*, *23*(3), 237–246. https://doi.org/10.1086/287489

John, S. (2015). Inductive risk and the contexts of communication. *Synthese*, *192*(1), 79–96. https://doi.org/10.1007/s11229-014-0554-7

Kincaid, H., Dupré, J., & Wylie, A. (Eds.). (2009). Value-free science?: Ideals and illusions. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195308969.001.0001

Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.

Longino, H. E. (1979). Evidence and hypothesis: An analysis of evidential relations. *Philosophy of Science*, *46*(1), 35–56. https://doi.org/10.1086/288849

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Longino, H. E. (2002). *The fate of knowledge*. Princeton University Press.

Longino, H. E. (2008). Values, heuristics, and the politics of knowledge. In M. Carrier, D. Howard, & J. A. Kourany (Eds.), *The challenge of the social and the pressure of the practice: Science and values revisited* (pp. 68–85). University of Pittsburg Press.

de Melo-Martín, I., & Intemann, K. (2016). The risk of using inductive risk to challenge the value-free ideal. *Philosophy of Science*, *83*(4), 500–520. https://doi.org/10.1086/687259

National Academies of Sciences, Engineering, and Medicine. (2017). *Using 21st Century Science to Improve Risk-Related Evaluations*. The National Academies Press. https://doi.org/10.17226/24635.

NTP (National Toxicology Program). (2004) *Toxicology in the 21st Century: The Role of the National Toxicology Program*. https://ntp.niehs.nih.gov/ntp/main_pages/ntpvision.pdf

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science*, *22*, 237-246. https://doi.org/10.1086/287231

Shostak, S. (2005). The emergence of toxicogenomics: A case study of molecularization. *Social Studies of Science*, *35*, 367-403. https://doi.org/10.1177/0306312705049882

Shostak, S. (2013). Exposed science: Genes, the environment, and the politics of population health. University of California Press. https://doi.org/10.1177/009430611456220 1xx

Steel, D., & Whyte, K. P. (2012). Environmental justice, values and scientific expertise. *Kennedy Institute of Ethics Journal*, *22*, 163-182. https://doi.org/10.1353/ken.2012.0010

# Part II

# Philosophy of Evidence Based Policy from the Perspective of Values in Science

# Chapter 3

# Behavioral Policies and Inequities : The Case of Incentivized Smoking Cessation Policies

## 3.1. Introduction

Behavioral policies aim to change people's behavior through interventions to choice contexts and psychological mechanisms of decision-making. Some well-known examples of behavioral policies are nudges (Thaler & Sunstein, 2008), boosts (Hertwig & Grüne-Yanoff, 2017), and behavioral-economics-informed incentivized programs for behavior change (Loewenstein & Chater, 2017). A growing body of research investigates the justifiability of behavioral policies. Behavioral economists, law scholars, public policy specialists, and those invested in evidence-based policy aim to identify the ethical, scientific, and institutional grounds for putting behavioral public policies into practice.

There is also a distinct *philosophical* literature on behavioral policies. Philosophers raise ethical worries about behavioral policies by, for instance, discussing the

desirability of the normative principles these policies aim to instantiate and the attainability of these goals by different types and tokens of behavioral policies (e.g. Bovens, 2010; Hausman and Welch, 2010). Philosophers also raise epistemic worries about behavioral policies, asking "how do we know that a behavioral policy is successful in achieving a desideratum in a certain environment?" (Heilmann, 2014; Barton & Grüne-Yanoff 2015; Grüne-Yanoff, 2016; Grüne-Yanoff et al., 2018). The arguments focusing on this methodological question have been informed by the theories of human action underlying behavioral policies, such as the dual systems approach (Kahneman, 2011), the methodological accounts of experimental social science (e.g. Guala 2005; Steel, 2008), and the philosophical literature investigating the epistemic and methodological requirements for the evaluation of evidence-based bio-medical and social policies (Clarke et al., 2014; Hardie & Cartwright, 2012; Cartwright, 2012).

In this chapter, I contribute to this growing body of specialized methodological literature that assesses how empirical researchers predict and evaluate the success of behavioral policies (see Grüne-Yanoff, 2016 for a paradigmatic account). A common worry raised by the commentators in this literature is that the methodology used for evaluating behavioral policies is not conducive to a comprehensive analysis of these policies, as the evidence typically gathered establishes that these policies work, but without indicating how they do. Here, the focus is on the adequacy of Randomized Controlled Trials (RCTs) in assessing behavioral public policies. The use of RCTs, so the critics argue, does not serve researchers in investigating long-term, unintentional or distributional consequences of behavioral policies without evidence regarding how they work.

I address this literature through an analysis of how Incentivized Smoking Cessation Programs (*henceforth*, ISCP) are evaluated. ISCP are prominent examples of evidence-based behavioral health policies (Bhargava & Loewenstein, 2015, p. 400; Loewenstein & Chater, 2017; Cabinet Office UK, 2011; Volpp et al., 2011). ISCP are thoroughly investigated or implemented in the UK and US in the last ten years. My analysis of ISCP is illuminating for the methodological literature for several reasons.

First of all, ISCP are instances of a prominent type of behavioral public policies, so called "incentivized behavioral policies", that so far have not been investigated by the specialized methodological literature. Secondly, and more importantly, the evidence-based evaluation of ISCP, as it is practiced in the UK, offers us rich resources for investigating specific methodological issues. Specifically, the evaluative perspectives and questions engaged in the evaluations of ISCP tend to be more comprehensive, unlike the evaluations of other commonly known behavioral policies, such as nudges. This includes the alternative methodological approaches to the use of RCTs.

I focus on the evaluation of ISCP's impact on *health inequities*, hence the assessment of ISCP's long-term effectiveness across specific sub-groups in the population. I argue that RCTs, when combined and synchronized with different evidence gathering methods, have distinct advantages in delivering inequity-relevant evidence over primarily RCT-based evaluations. More generally, I contend that this example gives us a reason to believe that a more pluralist evaluative methodology for behavioral public policies rectifies some of the commonly stated methodological limitations of primarily RCT-based extant behavioral policy evaluations.

The chapter proceeds as follows. In section 2, I introduce and explain what Incentivized Smoking Cessation Policies are and argue that assessing how ISCP are evaluated affords promising insights that are relevant for the methodological literature on the evaluations of behavioral public policies. In section 3, I explicate what it means to evaluate a policy's impact on health-inequities, and how ISCP are evaluated in this respect. In section 4, I assess the extant evaluative practices for ISCP's impact on health-inequities and argue for a pluralist methodology. In section 5, I reflect on the implications of this analysis for the philosophical literature on the evaluations of behavioral public policies. In section 6, I conclude by emphasizing the importance of pluralism of evidence gathering methods and the community of researchers' capacity to synchronize diverse methods to achieve more comprehensive evaluations of behavioral public policies.

# 3.2. Incentivized Smoking Cessation Programs (ISCP)

## 3.2.1. What are ISCP?

It is well documented that smoking contributes to the development of serious non-communicable diseases such as type 2 diabetes, respiratory and cardiovascular diseases, and lung cancer (WHO, 2011).[1] Most smokers acknowledge such malignant health consequences of smoking. Yet, they often fail in their attempts to quit. There are various policy instruments to influence people's smoking behavior. These include taxation of tobacco consumption, health-information campaigns, promotion of anti-smoking culture, and more coercive forms of regulation such as limiting the supply of tobacco and mandating smoking-free zones in cities. However, smoking remains a significant public health problem (e.g. NHS, 2017). Governments, therefore, actively seek new approaches to smoking cessation policies (e.g. Department of Health UK 2010, 2011; Commission on Social Determinants of Health [CDSH, WHO], 2008).

Behavioral economics offers a novel approach to smoking cessation (Loewenstein et al., 2007; Dolan et al., 2010; Volpp et al., 2011; Loewenstein et al., 2012; Cabinet Office UK, 2004; 2011). A well-studied example of the behavioral-economics-inspired smoking policies is that of Incentivized Smoking Cessation Policies (*henceforth,* ISCP) (e.g. Volpp et al., 2009; Halpern et al. 2015; Sunstein, 2015). ISCP promote quitting through *monetary or non-monetary rewards*. The supporters of ISCP (such as the Behavioral Insights Team in the UK, and authors such as Kevin Volpp, George Loewenstein, and Cass Sunstein) consider them applicable to small-scale environments such as firms or neighborhoods. They also consider ISCP as decentralized policies: implementers of ISCP could be a regional health service agency, or a non-governmental organization, or a private company. ISCP have been found attractive mostly in the US and in the UK. In the context of the US, large private

[1] The numbers cited by the WHO report about the non-communicable diseases around the globe are striking: "Almost 6 million people die from tobacco use each year, both from direct tobacco use and second-hand smoke […] Smoking is estimated to cause about 71% of lung cancer, 42% of chronic respiratory disease and nearly 10% of cardiovascular disease" (WHO, 2011, p.1).

companies have an interest in smoking cessation among their employees, as it reduces the insurance costs (Strickland, 2014). In the context of the UK, governments find it more desirable to approach smoking cessation at the regional, rather than national scale (Department of Health UK, 2011).

Some unsophisticated variants of ISCP were already implemented in the UK. For instance, the *Quit4U* program in Scotland (Ormston et al., 2014) and the *Give It Up For Baby* program in the UK (Ballard & Radley, 2009; Radley et al., 2013) used financial rewards (such as food and shopping vouchers) to reduce smoking in socio-economically disadvantaged smokers. Based on results from behavioral economics, the proponents of ISCP suggest that the way in which incentives are presented can make a difference in changing smoking behavior (Lowenstein et al., 2012; Adams et al., 2013). It is, therefore, important not to conflate those standard policy-interventions that alter financial incentives with the behavioral incentivized policies that alter incentives in sophisticated ways, based on evidence regarding the cognitive or psychological models of decision-making.

Consider, for instance, the finding that *explicit* financial rewards, such as direct cash payments or holiday vouchers, are more likely to lead to behavior change than 'relatively more invisible incentives' *of the same magnitude*, such as costs tied to insurance premiums (Strickland, 2014; Volpp et al., 2009). The more salient the incentives are, the more likely the behavior change will be, or so the proponents of ISCP argue. Similar to the salience or visibility of incentives, there are other and much more complicated aspects of incentive-provision for behaviour change that the proponents of ISCP deem helpful for designing and implementing ISCP (see Tversky & Kahneman 1974; Congdon et al., 2011; Bhargava & Loewenstein, 2015 for the relevant behavioral economic literature). Importantly, behavioral economic studies suggest that people who engage in non-volitional health behaviors (such as smoking, binge-eating, and excessive gambling) might suffer from the effects of 'cognitive biases' that most humans have,[2] and that policy makers can make use of these biases

[2] For example, the so-called 'peanut effect' consists in underestimation of cumulated long-term effects of minor negative health behavior (Weber and Chapman, 2005). 'Present bias' is the

for incentivized behavior change by harnessing them through offering incentives in varying immediacy, duration, frequency, or timing. Accordingly, the proponents of ISCP contend, for instance, that increasing the immediacy and frequency of incentives may render the incentivized quitting more likely to be effective, as it harnesses certain known cognitive biases of smokers (such as 'choice-bracketing' and 'present bias'; see Loewenstein et al., 2012 for a review).

At first glance, it seems that ISCP are promising evidence-based instruments for smoking policy that should perhaps be used much more widely. Should, for instance, governments encourage private companies and regional health services to adopt ISCP? In the rest of the paper, I will examine the methodological dimension of this question, leaving important ethical and political issues regarding ISCP aside (see Bovens, 2016; Schmitt, 2016; and Kelly, 2016 for a discussion of these matters). Indeed, I am solely interested in the question of whether ISCP should be adopted on *evidential grounds*; more specifically, which evidential sources warrant a positive evidential assessment of ISCP. It is also important to note that by "an evidential assessment of ISCP" or "an evaluation of ISCP' impact", I mainly refer to an ex-ante evaluation of evidence-base for the adoption of ISCP.[3] To this end, I will critically examine the methodologies researchers currently employ to evaluate ISCP. I will then discuss whether the employed evaluative methods are entirely adequate to assess these policies.

## 3.2.2. The importance of ISCP evaluations

While philosophers have analyzed the evaluations of well-studied types of behavioral public policies such as boosts and nudges (Grüne-Yanoff, 2016; Grüne-Yanoff et al., forthcoming), ISCP can neither be defined as boosts nor nudges. Nudges and boosts

label for overvaluation of prospects that occur sooner in time (O'Donoghue and Rabin, 2000). The so-called 'choice-bracketing' is defined as valuing a particular decision in isolation from the larger incentive structure (Read et al., 1999). The so-called 'loss aversion' is the overvaluation of small losses than gains of equal magnitude (Thaler and Kahneman,1997). These are some of the various heuristics and biases that behavioral economists use to explain smoking behavior and its change.

[3]As I will explain in section 3, it is possible that an ex-post evaluation of evidence-base for an already-implemented ISCP is considered as an evidential source in an ex-ante evaluation of ISCP (e.g. the use of systematic reviews in the evaluations of ISCP).

alter people's cognitive biases and heuristics to change behavior *while leaving the incentives intact*. ISCP and other incentivized behavioral policies do not fit this definition because they alter people's cognitive biases and heuristics in order *to increase the effectiveness of an incentive-altering intervention or an incentive-structure existing in the target environment.*

The categorical difference between nudge/boost and incentivized behavioral policies does not render the latter less of an important type of behavioral public policy, for two reasons. Firstly, incentivized behavioral policies are categorically different than the traditional use of incentives in policy making, because the former make use of behavioral economic insights about how people respond to different ways of presenting incentives and information. Secondly, and more importantly, the *proponents* of the nudge agenda regard incentivized behavioral policies as a prominent type of behavioral public policies. Loewenstein and Chater (2017), for example, argue that behavioral public policy making should not be equated with nudging because the majority of available tokens of behavioral-economics-inspired policies are *not* instances of nudge (and nor of boosts for that matter), but rather of incentivized behavioral policies (see also Chetty, 2015; Bhargava and Loewenstein, 2016 for similar arguments).[4]

How incentivized behavioral policies are evaluated has not yet been investigated in the methodological literature focusing on the evaluation of behavioral public policies. ISCP certainly are prominent and well-referenced examples of this latter category of behavioral policies (see, for instance, Sunstein 2015, Bhargava and Loewenstein 2016). An investigation of how ISCP are evaluated is, therefore, an important addition to the methodological literature in its own right. Beyond these, there are two more aspects of ISCP and their evaluation that render them interesting.

[4] Loewenstein and Chater (2017) also argue that incentivized behavioral policies have a wider range of applicability than nudging because the conditions for the proper applications of ideally defined nudges are harder to instantiate in practice than that of incentivized behavioral policies.

Firstly, ISCP are evaluated by a wider range of evaluators than the range of evaluators corresponding to nudges or boosts. Nudges (and boosts, for that matter) have so far been evidentially assessed primarily by behavioral economists or behavioral scientists. As I will demonstrate in the next section, incentivized health policies such as ISCP, on the other hand, are evaluated not only by behavioral economists (e.g. Loewenstein et al. 2007; Giné et al., 2010), but also by public policy scholars specialized in smoking cessation (e.g. Halpern et al., 2015), social epidemiologists (e.g. Adams, 2009), scholars from bio-medicine and preventive medicine (e.g. Bickel et al., 2016; Majestko et al., 2016), as well as prominent systematic reviewers of evidence-based public health interventions such as the Cochrane Collaboration, Campbell Collaboration, King's Fund, and NICE (e.g. Cahill et al., 2015; Jochelson, 2007).[5] Importantly, the plurality of evidential evaluators of ISCP also implies a plurality of policy desiderata with respect to which behavioral public policies can and should be evaluated. For instance, the evidential evaluations of nudges have so far primarily focused on the question of whether nudges are effective, or the extent to which they are. The methodological literature thereafter questioned the extent to which the evaluations of nudging realizes the evaluative goal of short or long term effectiveness (e.g. Grüne-Yanoff, 2016). But the evaluators of ISCP, as we will see, are focusing on further evaluative goals of ISCP, such as the effectiveness of these policies in specific subgroups in a population, or in reducing smoking-related health inequities, or the minimization of unintended consequences (side effects).

Secondly, the evaluations of ISCP, as they are practiced in the UK, are based on a *plurality of methods*, which makes the case of evaluating ISCP different from the evaluation of nudges.[6] The evaluation of ISCP is primarily based on evidence gathered through randomized controlled trials (RCTs) and systematic reviews of

[5] At least, this is the case in the UK where the comprehensive evaluation of ISCP is mostly practiced.
[6] Grüne-Yanoff (2016) argues that the evaluations of nudges are primarily based on lab and field experiments.

RCTs.[7] Systematic reviews of ISCP typically excludes or gives lower grades to non-randomized trials. Yet, importantly, some reviews consist of studies that make use of non-experimental and observational evidence gathered such as through in-depth interviews.[8] Just as the extant methodological studies questioned in how far the evaluation of nudging is practiced on sound methodological grounds, we may also ask in how far the evaluation of ISCP is based on the adequate use of the evidence gathering methods available. As the background research regarding the evaluation of public health policies tends to be comprehensive and well-documented in the UK, the evaluations of ISCP help us to focus on specific methodological problems regarding the use of experiments in the evaluation of behavioral policies discussed in the literature.

I have so far argued that we have good reasons to investigate the evaluation of ISCP from a methodological perspective. In the rest of the article, I will focus on the evaluation of ISCP's *impact on health inequities*. The evaluation of ISCP's impact on health inequities helps to further analyze the interesting and novel characteristics of the evaluation of incentivized behavioral policies that I listed above. More broadly, the health inequity focus will allow me to reflect on the question of whether behavioral public policies, as a new type of public policy, alter existing inequities and how this aspect of behavioral public policies should be evaluated.

In the next section, I will first explain what it means to evaluate a public policy's impact on health-inequities. I will then review what we know about the evaluation of ISCP's impact on health-inequities, introducing diverse perspectives of different kinds of researchers who provide us with relevant information for the assessment of how ISCP are evaluated.

[7] See, for example, Halpern et al. (2015) as a typical example of an evaluation of ISCP based on RCTs and Cahill et al. (2015) for a prominent example of a systematic review which evaluates the Halpern et al.'s study along with other empirical investigations of ISCP.
[8] See Morgan et al. (2015), Thomson et al. (2014), Adams et al. (2013), Graham et al. (2012), for an examples of systematic reviews of ISCP that make use of non-experimental evidence.

## 3.3. The Evaluation of ISCP with respect to Health Inequity

### 3.3.1. What Does It Mean to Evaluate a Policy's Impact on Health Inequity?

What does the term "health-inequity" mean? To start with, inequities in health outcomes should not be conflated with *inequalities* in health outcomes. Health *inequalities* are measurable differences in health outcomes across different populations and individuals. Health *inequities*, on the other hand, are health inequalities that result from people's unequal access to health services and capabilities to sustain healthier lives. For instance, health inequities in health outcomes in a country may result from its citizens' differential access or ownership of health services, nutritional sources, or health-related social capital. Being in worse health condition due to these disadvantaging and contextual factors is different than being worse-off due to unchangeable biological factors or volitional preferences. Hence, health inequities are commonly regarded as unnecessary, avoidable, unfair, and unjustifiable inequalities in health outcomes, which should be addressed by public policy interventions (O'neill et al., 2014; WHO 2012; Tugwell et al., 2006; Whitehead, 1992).

An effective health policy intervention may not be successful in reducing the inequalities in health outcomes between the most and the least disadvantaged individuals. For instance, it might be that an intervention is effective overall but ineffective or less effective for disadvantaged people. Similarly, a specific way to offer a public health intervention may discourage certain group of individuals from taking up the treatment, although that was not intended by the policy designers. The resulting inequities may result directly from the intervention itself or may appear as one of the already existing inequities prior to the intervention but exacerbated by the intervention (Lorenc et al., 2014). Health inequities may, therefore, remain intact or exacerbate due to the mistakes in the implementation and design of public health policies as well as the knowledge gaps in the evaluation of these policies.

Accordingly, when I talk about "evidence-based evaluation of policies with respect to inequities", I refer to the empirical investigation of the potential ways in which a policy may generate inequities once it is implemented or exacerbate a known inequity. These kinds of investigations, then, aim at detecting if there are any inequity-relevant mistakes in the implementation of the policies in question, or if there are knowledge gaps related to that (see O'Neill et al., 2014 for a more detailed description).

Our question is: what sort of methodological practices lend us the evidential basis for making this sort of judgment about ISCP and similar behavioral public policies? I now turn to answering it.

### 3.3.2. A Plurality of Perspectives for the Evaluation of the ISCP's Impact on Health-Inequities

We now have fixed an understanding of what it means to evaluate a policy's impact on health inequity. I will now present an overview of how ISCP are evaluated in this regard. As we will see, it would be misleading to characterize the assessment of ISCP as a scientific activity that is governed by a homogeneous set of methodological principles, advising the use of a single method such as RCTs. As I articulate in this section, it is more appropriate to describe ISCP's assessment as a scientific activity that involves different *Types of Researchers* who have different evaluative goals and who make use of different evidence gathering methods in evaluating ISCP. Each type of researcher delivers *some* evidential output that is relevant for the assessment of ISCP with respect to inequities. It is precisely this plurality of evaluative and methodological perspectives that makes the investigation of ISCP interesting for informing the methodological debates regarding the evaluation of behavioral public policies. In the following, I will first describe the evaluation of ISCP with respect to inequities, demonstrating how the plurality of perspectives plays an important role. I will thereafter focus on the methodological lesson we should draw from this practice.

*Table 3. 1* offers a structured summary of my characterization of the specialized literature evaluating ISCP. I list different *Types of Researchers* who contribute to the evaluation of ISCP. I also give a few *Representative Examples* for the respective type

of researchers. I indicate *Primary Evaluative Goals* that the different types of researchers aim for. I specify *Evidence Gathering Methods* the different types of researchers commonly use to gather the evidence in question. Finally, I mention some *References* from the literature that report or exemplify the kind of research in question.

**Table 3.1:** The characterization of the specialized literature evaluating incentivized smoking cessation policies: a plurality of evaluative and methodological perspectives.

| Type of Researchers | Representative Researcher | Primary Evaluative Goal | Evidence-Gathering Methods | References |
|---|---|---|---|---|
| *Decision-Makers* | National Health Services (NHS) in the UK | Specify evidence required for policy making (e.g. ISCP's effect on health inequities) | | CSDH (2008), Petticrew (2004), Department of Health (2011) |
| *The Proponents of Behavioral Policies* | Behavioral Insights Team in the UK | Defend behavioral policies such as ISCP based on evidence and propose new tokens of behavioral policies | Behavioral economic theory, experimental evidence, RCTs | Haynes et al. (2012), Dolan et al. (2010) |
| *Evidence-based Policy Specialists* | (a) RCT specialist Policy Evaluators (e.g. Kevin Volpp, Scott Halpern) (b) Public Health Evaluators (e.g. Jean Adams, Heather Morgan, Gill Thomson NICE in the UK, Higgins) | Measure the ISCP's (and similar incentivized policies') impacts such as effectiveness, persistence, cost-effectiveness, unintended consequences, effects on specific populations | RCTs, Observational Methods, Systematic reviews, Mixed methods | Halpern et al. (2015), Volpp et al. (2009), Morgan et al. (2015), Higgins and Solomon (2016) |
| *Behavioral Economists* | George Loewenstein, Paul Slovic | Pursuing empirical and theoretical knowledge about the psycho-cognitive factors that determine the differential effects of incentives and information across different types of smokers. | Theory, lab and field experiments (methodologically individualist orientation). | |
| *Social Epidemiologists* | Micheal Kelly, Hilary Graham, Jennie Popay, Stanley Blue | Empirical and theoretical knowledge about the structural factors that determine various aspects of smoking behavior, the health inequalities and differential effects of public health policies. | Theory, observational methods including econometrics descriptive statistics and qualitative methods (with a holistic methodological orientation) | Kelly (2010), Graham (2011), Popay (2008), Blue et al. (2016) |
| *Systematic Reviewers* | Cochrane Collaboration, Campell Collaboration, Kings's Fund | (i) Review, rate, and report available evidence on various impacts of ISCP; (ii) inform the design of new evidence-based policy analysis (e.g. by generating new hypothesis, defining gaps in evidence) | All of above and methods of structured review | Cahill and Perera (2011, 2015), Jochelson (2007) |

*Table 3. 1* characterizes a number of categories for each of the *Type of Researchers* listed on the left. Let's start with *Decision-Makers*. *Decision-Makers* are policy makers or public policy agents who seek information to determine the justifiability of implementing ISCP in a certain context. Although *Decision-Makers* are, strictly speaking, not types of scientific researchers, they search for scientific consultants, commission reports of available evidence, and seek evidence-based arguments for making particular types of policy making. Two good examples of *Decision-Makers* in the context of ISCP are the NHS or the UK Department of Health, who not only consider political, ethical, economic or other, pragmatic, concerns for implementing ISCP, but also seek information about available evidence concerning the impact of ISCP (UK Department of Health, 2011). As policy makers, *Decision-Makers* demand certain types of evidential output, which then indirectly determines the kinds of evaluative goals the other *Types of Researchers* seek to deliver. For instance, evidence regarding health inequities is undoubtedly very important for *Decision-Makers* all around the world (Petticrew et al. 2004, CSDH 2008, UK Department of Health 2011). More specifically, smoking control is one of those areas in public health where the evidence on policies' impact on specific disadvantaged groups and inequities is highly important.[9] Because there are widely accepted smoking-related inequities in health inequities across various socio-economic-demographic strata, health policy's success in reducing health inequities is crucial to investigate.

[9] Statistical data show that although overall smoking has decreased over time in many countries, smoking prevalence is unevenly distributed across different socio-economic classes in all countries (European Commission, 2015; NHS, 2017). For instance, according to the most recent statistics in the UK, where overall adult smoking prevalence is lower than in many other European countries, socioeconomically disadvantaged citizens are more likely to be a smoker than affluent ones (classified as those with lower income, less education, routine and manual jobs; as those with higher income, higher education, managerial and professional jobs; respectively) (NHS, 2017, part 3). In the same way, the influence of being an unemployed person or a member of an ethnic minority in the UK increases the likelihood of being a smoker. Moreover, smoking during pregnancy is highest in the economically poorest regions of the country. Also, smoking during pregnancy is more common among young mothers who are likely to experience financial and social distress. Similar facts about the socio-economic distribution of tobacco use also hold for other West-European countries where overall smoking prevalence is very low in comparison to the rest of the world (Verdurmen et al., 2015; Baha et al., 2016).

Who then are the *Types of Researchers* that are supposed to assemble the evidence relevant for the assessment of ISCP with respect to health inequities? Firstly, there are the *Proponents of Behavioral Public Policie*s who advocate ISCP. Secondly, we have the *Evidence-based Policy Specialists* who assess ISCP and similar policy proposals based on evidence. Thirdly, we have *Systematic Reviewers* who review, rate, and report available evidence-based assessment of ISCP. Finally, there are different kinds of scientists associated with different research disciplines such as *Behavioral Economists and Social Epidemiologists* who provide all the other *Types of Researchers* with the relevant theoretical and empirical input gathered through the practicing so called "primary science" regarding smoking behavior and health inequities. Let's first focus on *Proponents of Behavioral Policies*.

The *Proponents of Behavioral Policies* are policy researchers who advocate a particular policy approach, that is the behavioral approach. They aim to inform or else convince *Decision-Makers* based on scientific evidence regarding the performance of behavioral policies. In the context of behavioral public policies and ISCP, the researchers in the Behavioral Insights Team (BIT) of the UK are a good example of this category of researchers. BIT aims to justify behavioral public policies based on evidence. Hence, its main evaluative goal is to gather or report evidence that supports various tokens of behavioral public policies such as ISCP. Now, it is important for our discussion to note that BIT in the UK, and similar so-called Nudge Units around the world, adopt a particular methodological strategy to evaluate behavioral policies. That strategy prioritizes RCTs as *the* evidence gathering method that should be used for evaluating behavioral policies. This strategy is often explicitly stated. For instance, in one of BIT's methodological reports, the researchers claim that "Randomized Controlled Trials are at the heart of the Behavioral Insights Team's methodology' and that 'RCTs are the best way of determining whether a policy is working'" (Haynes et al., 2012, p. 4). We now have a more detailed characterization of the *Proponents of Behavioral Policies* in the context of UK, and their methodological strategy. Let's focus on the next type of researchers engaged in the evaluation of ISCP: *The Evidence-based Policy Specialists.*

The *Evidence-based Policy Specialists*' primary aim is to measure ISCP's (and similar Incentivized Policies') success with respect to various policy desiderata. Effectiveness is their most commonly presupposed policy desideratum; however, other evaluative goals such as investigating a policy's cost-effectiveness, persistence, unintended consequences, and impacts on specific populations or inequities are also pursued. I consider all advanced researchers who pursue this kind of evaluative goal as evidence-based policy specialists. However, for the purpose of characterizing diverse methodological lines in the ISCP's evaluation with respect to inequities, I focus on two qualitatively different examples of *Evidence-based Policy Specialists*: (i) *RCT-specialists* and (ii) *Public Health Evaluators*.

Similar to BIT's methodological approach, the empirical literature evaluating ISCP is marked by the prominence of randomized trials comparing the effectiveness of various ISCP (e.g. Halpern et al., 2015). *RCT-specialists* evaluating ISCP are primarily interested in determining whether a token of ISCP is effective in an environment for a particular target. As it is well-known, RCTs work perfectly in pursuing that particular evidential output. Hence, it is no surprise that *Evidence-based Policy Specialists* who investigate the effectiveness of ISCP for smoking cessation base their research primarily on RCTs.

*Public Health Evaluators*, on the other hand, are akin to social policy scholars who assess health policies from the perspective of public health. Hence, their primary research interest and evidential output goes beyond the specification of ISCP's effectiveness and includes the specification of ISCP's effectiveness in certain groups in a population; for instance, ISCP's impact on inequities, or ISCP's cost-effectiveness, or other pieces of evidential output relevant for justifying health policies. *Public Health Evaluators* are therefore not a homogenous category of researchers. Rather, they occupy roles in different disciplines such as epidemiology, social epidemiology, social policy, economics, sociology, preventive medicine, etc. This variety is then reflected in the different evidence gathering methods employed by Public Health Evaluators across disciplinary backgrounds. For instance, researchers in NICE tend to use qualitative and observational types of evidence

gathering methods, which are pertinent to disciplines such as epidemiology, social epidemiology, or sociology of health (e.g. NICE, 2007). Preventive medicine scholars, on the other hand, tend to pursue their primary evaluative goal regarding ISCP through laboratory experiments (e.g. Higgins et al., 2012).

I should emphasize that these two categories of example researchers, *RCT-specialists* and *Public Health Evaluators,* are not mutually exclusive. In other words, an RCT-specialist may also be a public health evaluator (e.g. Jean Adams). I distinguish between the two in order to emphasize the following point: *Evidence-based Policy Specialists* use different evidence gathering methods for evaluating various aspects of ISCP. More generally, that is to say, *Evidence-Based Policy Specialists* as a type of researcher does not correspond to a homogenous body of researchers with respect to the primary evaluative goal and evidence gathering methods used.

The next *Type of Researchers* are scientists who deliver primary theoretical input and empirical evidence relevant to the evaluation of ISCP. Since smoking cessation is a complex and multi-faceted scientific subject, there are many such different kinds of "primary scientists" involved. As they are the most relevant ones for the assessment of ISCP's impact on health inequity, I focus here on *Behavioral Economists* and *Social Epidemiologists*. Behavioral economists assemble empirical and theoretical knowledge on the psycho-cognitive factors that determine the differential effects of incentives and information across different types of smokers (Loewenstein et al. 2012; Gine et al., 2010). *Social epidemiologists*, on the other hand, analyze the impacts of social-structural factors on individual, population health states, health-related social practices and health behavior (Honjo, 2004). Social-structural factors are commonly referred as "wider determinants of health" by epidemiologists. Social epidemiologists generally make use of observational evidence gathering methods to investigate health inequities.

Finally, we have the *Systematic Reviewers* who review, rate, report and evaluate available evidence about various impacts of ISCP. The most prominent examples of *Systematic Reviewers*' research in the context of health interventions and ISCP are assembled by major evidence-based policy institutions in public health and

biomedicine such as the Cochrane Collaboration, Campell Collaboration, and Kings' Foundation in the UK. Systematic Reviewers play a major, arguably the most crucial, role in the evaluations of health interventions. The most explicit and the main contribution of *Systematic Reviewers* is the reporting of evidence in a way that is useful for decision-making. Hence, *Systematic Reviewers*' reviews, rating and reporting of evidence directly inform *Decision-Makers* in the case of UK A less salient but a very important contribution of *Systematic Reviewers* is the evaluation of available evidence so as to inform the design and implementation of new evidence-based policy assessment, ex-ante. That is to say, *Systematic Reviewers*' evidential output informs *Evidence-based Policy Specialists'* research. Specifically, *Systematic Reviewers* do so by generating new hypotheses, determining gaps in evidence or theories provided by *Primary Scientists*, and communicating the relevant evidential demands on behalf of *Decision-Makers*.

Scientists delivering primary theoretical and empirical evidence relevant to the evaluation of ISCP (e.g. *Behavioral Economists* and *Social Epidemiologists*) may also benefit from the *Systematic Reviews*' research output in the same way as *Evidence-based Policy Specialists* do. However, the interaction between *Systematic Reviewers* and *Evidence-based Policy Specialists* is much more direct than the one between *Systematic Reviewers* and *Primary Scientists* in practice. This is the case firstly because *Evidence-Based Policy Specialists* draw generally on evidence available from the systematic reviews (e.g. Evidence-based policy specialists' research articles would generally include a 'background' section where the evidential output of relevant Systematic Reviews is reported). Secondly, it is often the case that an Evidence-based policy specialist is also a specialist in systematic reviews (e.g. researchers such as Jean Adams, Gill Thomson, and Heather Morgan).

What do we know about the evidence gathering methods *Systematic Reviewers* use in pursuing their primary evaluative goal? As I described above, *Systematic Reviewers* deliver two types of evidential input for the evaluation of ISCP: one that is relevant for reporting, another one that is relevant for further evidence-based policy assessment. The systematic reviews rely on observational evidence gathering methods

in reviewing available evidence. However, it is also appropriate to speak of heterogeneity of ISCP's systematic reviews in terms of the kinds of evidence selected for the reviews. Depending on the aims and the orientations of *Systematic Reviewers*, systematic reviews of ISCP sometimes draw only on RCT-based evaluations of ISCP (e.g. when assessing overall effectiveness, see for instance, Cahill & Perera, 2011). Yet, they also draw on other evidential output assembled by the use of alternative evidence gathering methods including the *Public Health Evaluators'* assessment based on non-RCT studies, theoretical and empirical evidence delivered by *Social Epidemiologists* and *Behavioral Economists* (e.g. when assessing aspects of interventions other than the effectiveness, see for instance Thomson et al., 2014). Systematic reviews sometimes integrate different evidence gathering methods. Such reviews, therefore, significantly contribute to methodologically more integrative evaluations of ISCP. *Systematic Reviewers*' research output, which may be reinforced by the multiple evidence gathering methods, also serves as observational evidence that informs *Evidence-based Policy Specialists' research* (see, Section 4, for an example).

Here, I have offered a general definition of what it means to evaluate a policy's impact on health-inequities (3.1) and have reviewed the different sources and types of available evidence relevant for judging whether ISCP reduce health inequities in diverse contexts and demonstrated that that there are different methods of evidence gathering involved (3.2).

I will now assess how ISCP's impact on health-inequities is evaluated. Based on this analysis, I will specifically argue that the evaluation of ISCP through the combination of different evidence gathering methods has distinct advantages in delivering inequity-relevant evidence in comparison to primarily RCT-based evaluations [advocated by the proponents of behavioral public policies and some of the RCT-specialized policy evaluators represented in the table (e.g. Haynes et al., 2012)]. More generally, I contend that this example gives us a reason to believe that a more pluralist evaluative methodology for behavioral public policies rectifies some of the commonly

stated methodological limitations of extant behavioral policy evaluations which are pertinent to reliance primarily on RCTs.

## 3.4. RCTs Integrated with Different Evidence Gathering Methods for the Evaluation of ISCP's Impact on Health Inequity

I have offered an overview of different types of evaluators of ISCP and different evidence gathering methods they use for delivering inequity-relevant evidence. I will now examine the adequacy of these evidence gathering methods in investigating ISCP's success with respect to the reduction of health inequities. My purpose is not to propose a single best methodology for the evaluations of ISCP; however, I will argue that RCTs fare better in delivering relevant evidence *when integrated with alternative evidence gathering methods.* This argument, to the extent that it is an argument regarding the use of RCTs in policy evaluation, is concerned with how RCTs can be used in a better way rather than stating how limited RCTs are. In this section, I will first review how well-known limitations of RCTs arise in the context of evaluating ISCP's impact on health inequities. I will then illustrate how some evaluative studies integrate RCTs with different evidence gathering methods and argue for a more integrated evaluative methodology for the evaluation of ISCP.

Let me first review the limitations of RCTs as an evidence-gathering method used for the evaluation of ISCP's impact on health inequities. To do so, I consider the evidential output delivered by the RCT-specialists who are *Evidence-based Policy Evaluators* and the *Proponents of Behavioral Policies* who primarily rely on RCTs.

I will offer two examples of the kind of evidential output that is relevant for evaluating ISCP's impact on inequities, yet not delivered by *Evidence-based Policy Evaluators* and the *Proponents of Behavioral Policies* who only use RCTs.

The first kind of evidence is concerned with sub-groups, specifically the group of socioeconomically disadvantaged smokers. Based on available systematic reviews, what we know is that primarily RCT-based assessments which report that some tokens of ISCP are significantly effective in ceasing smoking (e.g. Volpp et al., 2009;

Halpern et al., 2015) are not informative about effectiveness in *specific subgroups*, such as disadvantaged smokers. The following comment that appeared in a Cochrane Collaboration systematic review concerning these studies is quite telling in this respect:

> "Since both trials enrolled employees of large American companies, who were predominantly white and enjoyed relatively high levels of education and income, their success may not be readily generalizable to other populations of smokers, with different regional, socio-economic and ethnic mixes" (Cahill & Perera, 2011).

The lack of evidence regarding ISCP's impact on specific disadvantaged groups is a major limitation for making evidence-based judgments about ISCP's impact on health inequities. For instance, without comprehensive information about the subgroups of the population under investigation, one cannot judge whether ISCP's effectiveness is modified across smokers with different demographic characteristics, or whether some overall effective ISCP are not successful in ceasing smoking for disadvantaged groups. A possible methodological reply to this challenge might be to carefully stratify the population of the experiment prior to the experiment, thus to define the different subgroups. However, doing so in the right way in fact invites researchers to use alternative evidence gathering methods such as observational studies and descriptive statistics together with RCTs, as I will illustrate further below.

The second example I would like to put forward concerns the lack of evidence regarding the *long-term effectiveness* of ISCP. The long-term effectiveness is crucial for understanding ISCP's impact on health inequities. There is well-known social epidemiological evidence, based on qualitative and quantitative observational studies, reporting on the specific challenges of *long-term* smoking cessation in the context of disadvantaged smokers (e.g. due to stressors associated with social and economic exclusion). Many social epidemiologists would consequently argue that post-ISCP smoking behavior would be different across groups, even if an ISCP is initially successful, anticipating that disadvantaged smokers are more likely to relapse (Popay, 2008; Blue et al., 2016). Although such social epidemiological evidence is theoretically plausible, it gives us at best an *indirect* or *prima facie* reason to believe that disadvantaged quitters are more likely to relapse months after a successful, ISCP-

generated, abstinence. But those who favor ISCP may always demand further and stronger evidence to believe in social epidemiologists' arguments against the effectiveness of ISCP. Although evidence gathered through RCTs are generally considered stronger than observational evidence, RCT-based studies of ISCP fail to provide us with evidence confirming or disconfirming arguments pro or against ISCP's long-term effectiveness in disadvantaged smokers. Specifically, primarily RCT-based studies of ISCP do not deliver information about the distribution of relapse behavior across different strata of smokers in the long term.

Based on the reports of systematic reviews, what we know is that the smoking abstinence generated by effective ISCP usually does not last longer than a couple of months after the incentives are withdrawn (Marteau & Mantzari, 2015; Cahill & Perera, 2011; Jochelson, 2007). Yet, based on the considerations I stated above, it would be crucial to have RCT-based evidence indicating whether the post-ISCP relapse behavior is stratified and modified by the characteristics of disadvantages, such as unemployment or social exclusion, as many social epidemiologists would anticipate. As I will illustrate further below, such information can be more easily delivered when RCTs are integrated with alternative evidence gathering methods that are more suitable for *predicting* the stratification or modification effects in the long-term (e.g. by studying specific mechanisms of behavior change through qualitative studies or further modelling by primary scientists).

Now, it is *not* surprising that the evaluations of ISCP that rely *only* on evidence gathered through RCTs do *not* deliver evidence on these two aspects. RCTs are considered more adequate tools for determining the overall effectiveness of interventions rather than the variation of effectiveness across sub-groups. Similarly, since RCTs are not supposed to give information regarding *how* or through which mechanisms an intervention works, they are similarly not informative about the long-term impacts. These issues have been widely discussed in the relevant philosophical literature on policy evaluation in general (e.g. Cartwright & Hardie, 2012), and the behavioral public policy evaluation in particular (e.g. Grüne-Yanoff, 2016). It is therefore well known that RCTs have limitations in delivering comprehensive

evidence regarding the effects of interventions (such as evidence characterizing the heterogeneity of subgroup in the target population and differential distribution of effects across different subgroups, longer-term effects of the intervention, or how the intervention interacts with the context of the target environment). My aim is not to advance upon on these well-known critiques of RCTs, or to offer a new one. I fully acknowledge these criticisms and point out that the same issues arise in the context of ISCP evaluation as well. I also do not suggest that RCTs are *in principle* unconducive to investigate health inequities. My aim is rather to make a constructive methodological claim regarding how RCTs, as they are currently employed in the case of ISCP, might be designed and harnessed better for the purpose of making judgments about how behavioral policies such as ISCP fare with inequities.

To this end, I suggest that the *Evidence-based Policy Specialists*' assessments of ISCP can and do actually deliver the necessary evidence when they *integrate different evidence gathering methods with RCTs*. To demonstrate this and to exemplify what kinds of evidence gathering methods are needed, let me offer a closer look at those methodologically more integrative evaluations of ISCP and why they perform better.

Consider Morgan et al.'s (2015) investigation of the incentives for smoking cessation during pregnancy conducted for the NHS in the UK. This social scientific research involves multiple methodological steps, but it is possible to represent Morgan and her colleagues' investigation in two parts for the sake of understanding how they integrate multiple evidence gathering methods. The first part of their study involves a systematic review of RCT-based evaluations of various ISCP's effectiveness, a report of qualitative and theoretical literature about the mechanism of incentive-based behavior change, and a collection of socio-epidemiological and behavioral scientific evidence regarding the barriers and facilitators of smoking cessation during pregnancy in the context of socioeconomic disadvantage. Morgan et al.'s aim in the first part is to integrate these different pieces of available evidence so as to inform the *design* and the *scope* of the second part of their study. The second part of the study involves conducting primary qualitative studies (based on structured interviews) to understand how the target audience of ISCP trials, which were carefully pre-selected in the first

part of the study, responds to incentive provision. That is to say, Morgan and her colleagues make use of qualitative studies in order to assemble comprehensive evidence regarding the working and the consequences of RCT-based trials. In a simultaneous study based on the same data, Thomson et al. (2014) gather evidence specifically relevant for making judgments about various *unintended consequences* of ISCP under investigation for disadvantaged smoking pregnant women in the UK. Their results do also inform judgments about ISCP's impact on health inequities based on empirical evidence gathered in these two simultaneous studies.

The integrated methodology, exemplified in Morgan and her colleagues' study, extracts inequity-relevant information from the available RCT-based studies of ISCP which do not necessarily have inequity-relevant content. It does so by making use of different evidential sources such as the evidence delivered by *Primary Scientists* regarding the possible mechanisms, barriers and facilitators of smoking cessation, and evidence delivered by qualitative literature concerning which ISCP trials have failed, for which sub-groups, and how. In doing so, it reveals comprehensive evidence regarding subgroups, which is not readily available from RCTs. Moreover, the integrated methodology also better predicts the potential long-term consequences of selected ISCP by conducting follow up structured interview studies in order to extract information about potential modifications or unintended consequences of the interventions. It thereby succeeds in rectifying the abovementioned limitations of primarily RCT-based evaluations.

Moreover, the studies which integrate multiple methods do also a better job in inequity assessment than those studies which rely *only* on social epidemiological evidence or qualitative methods. While the latter provides prima facie indirect descriptive and theoretical evidence for ISCP's various impacts, the former deliver more direct and comprehensive causal evidence, which also makes use of available social epidemiological evidence.

This kind of more integrated methods are increasingly advocated by the *Evidence-based Policy Specialists* and *Systematic Reviewers* (Petkovic et al., 2017; Welch et al., 2015; O'Neill et al. 2014; Welch et al., 2010), and the number of similar studies

is increasing, as demands for evidence-based evaluations of inequity increases (NICE, 2013). To the best of my knowledge, none of the behavioral public policies or incentivized behavioral health policies has been evaluated in this manner, but there is no principled reason against doing so.

I have suggested that the evaluations of ISCP that integrate different evidence gathering methods with RCTs fare better in delivering pieces of evidence relevant for the inequity assessment in comparison to those which rely only on RCTs, or only on social epidemiological methods. This argument gives us a nuanced view regarding what RCTs can and cannot do in evaluating behavioral public policies. I will now also discuss in what way the argument advanced so far also contributes to the following more general epistemological question: what kind of *evidence* does the justifiability of behavioral public policies require?

## 3.5. Discussing Evidence Gathering Methods in the Philosophy of Behavioral Public Policies

I have illustrated that there are different kinds of evidence gathering methods involved in the evaluations of ISCP. I then have claimed that the evaluation of ISCP through the *integration* of different evidence gathering methods has distinct advantages in delivering inequity-relevant evidence in comparison to studies relying on single methods. I will now specify how my argument relates to more general philosophical debates.

Philosophers contribute to the evidence-based policy making by specifying the kinds of ideal epistemic requirements (pertaining to the nature of causal knowledge necessary for policy-making purposes) the evidence-based policies should meet, and by determining the sources of evidential gaps exists in the practice of evidence-based public policy justification (e.g. Cartwright & Hardie, 2012). Philosophers specialized in this literature often make a distinction between two broad categories of evidence: evidence of difference-making and evidence of mechanisms. There are controversies regarding what mechanisms are (Williamson & Illari, 2012), what counts as

mechanistic or difference-making evidence (Illari, 2011), and how mechanistic evidence relates to the difference-making evidence types of evidence (Clarke et al., 2014).

A general contention of the philosophers regarding the received methodology of evidence-based policy assessment (such as the idea of "evidence hierarchies" in bio-medical and public health interventions) has been that the assessment of these policies is based on a limited array of evidence, although inclusion or prioritization of different categories of evidence is sometimes necessary. Evidence-based public policy assessment is primarily based on randomized controlled trials, which is commonly seen as delivering evidence of difference making. Philosophers often demand evidence of mechanisms for enhancing the evaluation of evidence-based policies (Russo & Williamson, 2007; Clarke et al., 2014). The former denotes the kind of evidence establishing that a policy-intervention or a treatment makes a difference in a target environment. The latter denotes the kind of evidence establishing how (through which mechanisms) a policy intervention makes a difference in a target environment. Grüne-Yanoff (2016) discusses these issues in the context of behavioral public policies such as Nudges. He observes that behavioral economists typically do not gather evidence of mechanisms to assess behavioral policies and considers this an important drawback for the justification of these policies.[10]

The dichotomy between evidence of difference-making and evidence of mechanisms definitely helps us understand some of the limitations of policy evaluations based *primarily on* RCTs (see, for instance, Russo & Williamson (2007) in the context of

[10] He provides examples from the literature so as to demonstrate that a lack of sufficient mechanistic evidence regarding how these policies make a difference in people's behavior avoids us to assess various policy desiderata attached to these policies. For instance, without mechanistic information, Grüne-Yanoff argues, one cannot determine whether a default nudge program will be effective across different target environments, and whether it will lead to robust, persistent, and welfare-improving behavioral changes. This is the evidence that is assembled by behavioral economists corresponding to their explanatory models of choice behaviors. For example, default nudges make a difference in people's behavior because of a mechanistic model A that specifies "inertia" as the main causal entity or process that leads to the conclusion, or model B cites recommendation effect, or C loss-aversion etc.

evidence-based medicine). I am broadly supportive of this conclusion, as I too try to spell out how RCT-based evaluations of behavioral policies can be improved. Yet, the argument I advanced in favor of the pluralism of evidence gathering methods is different from the arguments made in favor of or against mechanistic evidence in at least two ways. Firstly, the former and the latter have different evaluative goals. My analysis aims to answer how behavioral policies' *impact on inequities* should be analyzed; whereas the call for mechanistic evidence is motivated to answer how behavioral policies' *efficacy* should be assessed in different contexts. Secondly, my argument and the argument(s) for mechanistic evidence operate on different levels. While mine is a demand for a pluralism of *methods*, a demand for mechanistic evidence is about the *content of evidence* required for the respective evaluative goals. Of course, an implication of evidential pluralism as a thesis about the content of evidence for policy evaluation might be that we need to integrate different evidence gathering methods to deliver evidence of difference making and mechanisms. To the extent that evidential pluralists are happy to endorse the pluralism of evidence gathering methods for policy evaluation, this article provides further support for evidential pluralism in the context for behavioral public policy evaluation.

I would be broadly sympathetic to draw closer and more precise connections between the demand for pluralism of evidence gathering methods and the demand for evidence of mechanisms. For instance, one might argue be that the integration of evidence gathering for inequity assessment demands only specific kinds of mechanistic evidence and specific kinds of difference-making evidence. Or one might also quite naturally contend that the demands for pluralism of evidence gathering methods and need for mechanistic evidence could imply or complement each other, depending on how these positions are formulated exactly, or to which case they apply. Spelling out an exact relationship between my analysis and the need for mechanistic evidence, however, goes beyond the purpose of the present article. My aim has been to advance a specific perspective on the evaluations of behavioral public policies, one that I hope will be a useful addition to the discussions about the need for evidence of mechanism in this broader philosophical literature as well.

## 3.6. Conclusion

In this chapter, I have offered new insights for evidential evaluations of behavioral policies by focusing on evaluative challenges of a prominent example of behavioral health policies (Incentivized Smoking Cessation Policies). I have contended that evaluators of behavioral policies should go beyond a primarily RCT-based methodology for more comprehensive evaluations. I have also offered a pluralist evaluative methodology in which RCTs are integrated with alternative evidence gathering methods.

I focused on the *evaluation* of incentivized behavioral policies*;* however, I also believe that my comprehensive and practice-oriented analysis of ISCP also expands our understanding of behavioral policies more broadly. My focus on a fundamental public policy problem, the evaluation of health inequity aspects, invites readers to think about my analysis also in a different way: as exploration of a salient way in which behavioral economics, the 'run-away success story of contemporary economics' (Angner, 2015), modifies standard public policies, conventional 'welfare regimes' (Esping-Andersen, 2013[1990]), and claims to rectify the standard economics approach in addressing one of the centuries-old central problems of welfare-provision, how to address inequities. The chapter, therefore, provides a rationale for why philosophers and evaluators of behavioral policies may want to adopt a more interdisciplinary toolbox and mindset in addressing the potential merits and pitfalls of behavioral policies as *evidence based public policies*.

# References

Adams, J. (2009). The role of time preference and perspective in socioeconomic inequalities in health-related behaviours. In S. J. Babones (Ed.), *Social inequality and public health* (pp. 9-25). Bristol: Policy Press.

Adams, J., Giles, E. L., McColl, E., & Sniehotta , F. F. (2013). Carrots, sticks and health behaviours: a framework for documenting the complexity of financial incentive interventions to change health behaviours. *Health Psychol Rev, 8* (3), 286-295.

Angner, E. (2015). To navigate safely in the vast sea of empirical facts. *Synthese, 192*: 3557. doi:10.1007/s11229-014-0552-9

Baha, M., Boussadi, A., & Le Faou, A-L. (2016). French smoking cessation services provide effective support even to the more dependent. *Preventive Medicine*, doi: 10.1016/j.ypmed.2016.06.024

Ballard, P., & Radley, A. (2009). Give It Up For Baby-A Smoking Cessation Intervention for Pregnant Women in Scotland. *Cases in Public Health Communication & Marketing*, *3*, 147-160.

Barton, A., & Grüne-Yanoff, T. (2015). From Libertarian Paternalism to Nudging—and Beyond. *Review of Philosophy and Psychology*, *6*(3), 341-359. doi:10.1007/s13164-015-0268-x.

Bickel, W. K., Moody, L., & Higgins, S. T. (2016). Some current dimensions of the behavioral economics of health-related behavior change. *Preventive Medicine, 92*, 16-23. doi:10.1016/j.ypmed.2016.06.002.

Bhargava, S., & Loewenstein, G. (2015). Behavioral economics and Public policy 102: Beyond nudging. *American Economic Review*, *105*(5), 396-401. doi:10.1257/aer.p20151049

Blue, S., Shove, E., Carmona, C., & Kelly, M. P. (2016). Theories of practice and public health: understanding (un)healthy practices. *Critical Public Health, 26*(1), 36-50. doi:10.1080/09581596.2014.980396

Bovens, L. (2010). Ethics of Nudge. In T. Grüne-Yanoff & S. O. Hansson (Ed.), *Preference change: approaches from philosophy, economics and psychology*. Dordrecht: Springer.

Bovens, L. (2016). Don't Mess with My Smokes: Cigarettes and Freedom. *The American Journal of Bioethics, 16* (7), 15-17. DOI: 10.1080/15265161.2016.1180461.

Cabinet Office UK. (2004). Personal Responsibility and Changing Behaviour: the state of knowledge and its implications for public policy. Retrieved from http://webarchive.nationalarchives.gov.uk/+/http:/www.cabinetoffice.gov.uk/media/cabinetoffice/strategy/assets/pr2.pdf.

Cabinet Office UK, Behavioral Insights Team. (2011). *Applying behavioural insight to health.* Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60524/403936_BehaviouralInsight_acc.pdf.

Cahill, K., & Perera, R. (2011). Competitions and incentives for smoking cessation. Cochrane Database of Systematic Reviews (4), Art. No.: CD004307. DOI: 004310.001002/14651858.CD14004307.pub14651854.

Cahill, K., Hartmann-Boyce, J., & Perera, R. (2015). Incentives for smoking cessation. Cochrane Database of Systematic Reviews (5), Art. No.: CD004307.

Cartwright, N. (2012). Will this policy work for you? Predicting effectiveness better: how philosophy helps. *Philosophy of Science, 79*, 973–989.

Cartwright, N., & Hardie, J. 2012. *Evidence-Based policy: a practical guide to doing it better.* New York, NY: Oxford University Press.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin M. (2003). Regulation for conservatives: behavioural economics and the case for "asymmetric paternalism". *U Penn Law Rev,1151*, 1211-54.

Chetty, R. (2015). Behavioral Economics and Public Policy: A Pragmatic Perspective. *American Economic Review, 105*(5), 1-33. doi:10.1257/aer.p20151108

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi, 33*, 339–360.

Congdon, W. J., Kling, J. R., & Mullainathan, S. (2011). *Policy and choice: Public finance through the lens of behavioral economics.* Washington, DC: Brookings Institution Press.

Commission on Social Determinants of Health (CSDH) World Health Organization (2008). *Closing the gap in a generation: health equity through action on the social determinants of health. Final report.* Retrieved from *: http://apps.who.int/iris/bitstream/handle/10665/43943/9789241563703_eng.pdf;jsessionid=BA3A7DB491AE0306866B416371749C6C?sequence=1*

Department of Health UK. (2010). *Healthy lives, healthy people: Our strategy for*

*public health in England*. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216096/dh_127424.pdf

Department of Health UK. *Healthy Lives Healthy People: Tobacco Control Plan for England.* (2011). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213757/dh_124960.pdf

Dolan, P., Hallsworth, M., Halpern, D., King, D. Vlaev, I. *MINDSPACE: influencing behavior for public policy: Institute of Government*, 2010. http://www.instituteforgovernment.org.uk/sites/default/files/publications/MINDSPACE.pdf

Esping-Andersen, G. (2013) [1990]. *The three worlds of welfare capitalism*. Cambridge: Polity Press.

Executive Office of the President, National Science and Technology Council. 2015. *Social and Behavioral Sciences Team Annual Report.* Retrieved from *https://sbst.gov/download/2016%20SBST%20Annual%20Report.pdf*

Giné, X., Karlan, D., & Zinman, J. (2010). Put your money where your butt is: A commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, *2*(4), 213-235. doi:10.1257/app.2.4.213

Graham, H. (2011). *Understanding health inequalities.* Maidenhead, England: Open University Press.

Graham, H., Sowden, A., Flemming, K., Heirs, M., Fox D (2012). *Public Health Research Consortium: Using qualitative research to inform interventions to reduce smoking in pregnancy in England: a systematic review of qualitative studies.* Retrieved July 02, 2017, from http://phrc.lshtm.ac.uk/project_2005-2011_a810.html

Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, *32*(03), 463-483. doi:10.1017/s0266267115000425.

Grüne-Yanoff, Till. (2017). Reflections on the 2017 Nobel Memorial Prize awarded to Richard Thaler. *Erasmus Journal for Philosophy and Economics*, *10*, (2), p. 61-75, Available at: <https://www.ejpe.org/journal/article/view/307>.

Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). "Toward a Framework for Selecting Behavioural Policies: How to Choose between Boosts and Nudges." *Economics and Philosophy, 34* (2), 43-66.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge: Cambridge University Press.

Halpern, S.D., French, B., Small, D.S., Saulsgiver, K., Harhay, M.O., Audrain-McGovern, J., Loewenstein, G., Brennan, T.A., Asch, D.A., & Volpp, K.G., (2015). Randomized trial of four financial-incentive programs for smoking cessation. *N. Engl. J. Med*, *372*, 2108–2117.

Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy, 18*(1), 123-136. doi:10.1111/j.1467-9760.2009.00351.x

Haynes, L., Goldacre, B., & Torgerson, D. *Test, learn, adapt: developing public policy with randomised controlled trials.* Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/TLA-1906126.pdf.

Heilmann, C. (2014). Success conditions for nudges: A methodological critique of libertarian paternalism. *European Journal for Philosophy of Science, 4*(1), 75-94.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: steering or empowering good decisions. *Perspectives on Psychological Science, 12* (6), 973–986.

Higgins, S. T., Washio, Y., Heil, S. H., Solomon, L. J., Gaalema, D. E., Higgins, T. M., & Bernstein, I. M. (2012). Financial incentives for smoking cessation among pregnant and newly postpartum women. *Preventive Medicine, 55*. doi:10.1016/j.ypmed.2011.12.016

Higgins, S.T., & Solomon, L.J. (2016). Some recent developments on financial incentives for smoking cessation among pregnant and newly postpartum women. *Curr Addict Rep, 3* (1), 9–18.

Honjo, K. (2004). Social epidemiology: Definition, history, and research examples. *Environmental Health and Preventive Medicine*, *9*(5), 193–199. http://doi.org/10.1007/BF02898100

Illari, P. M. (2011). Mechanistic evidence: disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science*, *25*, 139–157.

Illari, P.M. & Williamson (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal of Philosophy of Science, 2*: 119

Jochelson K. (2007). *Paying the patient: improving health using financial*

*incentives. King's Fund*. Retrieved from: www.wpro.who.int/NR/rdonlyres/BCC2F5C3-B685-4FF1 AE7739C52F4ED247/0/payingthepatient.pdf

Kahneman, Daniel. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Kelly, M. P. (2010). The axes of social differentiation and the evidence base on health equity. *J R Soc Med, 103*(7), 266-272. doi:10.1258/jrsm.2010.100005

Kelly, M. P. (2016) The politics of behaviour change. In Spotswood, F. (Ed.), *Beyond behaviour change: Key issues, interdisciplinary approaches and future Directions*. (pp. 11-26). Bristol: Policy Press.

Laibson, D. (2001). A cue-theory of consumption. *Quarterly Journal of Economics, 116*(1), 81–119.

Loewenstein, G. (1999). Visceral account of addiction. In Elster, J. & Skog, J. O. (Ed), *Getting Hooked: Rationality and Addiction*. Cambridge University Press.

Loewenstein, G., Brennan, T., & Volpp, K.G. (2007). Asymmetric paternalism to improve health behaviors. *JAMA, 298*, 2415-2417.

Loewenstein G., Asch, D. A., Friedman, J. Y., Melichar, L. A., & Volpp, K. G. (2012). Can behavioural economics make us healthier? *British Medical Journal, 344*, e3482

Loewenstein, G., & Chater, N. (2017). Putting nudges in perspective. *Behavioural Public Policy, 1* (01), 26-53.

Marteau, T. M. & Mantzari, E. (2015). Public health: The case for pay to quit. *Nature, 523* (7558), 40-41.

Matjasko, J. L., Cawley, J. H., Baker-Goering, M. M., & Yokum, D. V. (2016). Applying behavioral economics to public health policy. *American Journal of Preventive Medicine, 50* (5). doi:10.1016/j.amepre.2016.02.007

Morgan, H., Hoddinott, P., Thomson, G., Crossland, N., Farrar, S., Yi, D., … Hislop, J. (2015). Benefits of incentives for breastfeeding and smoking cessation in pregnancy (BIBS): a mixed-methods study to inform trial design. *Health Technology Assessment, 19* (30): 1–522.

National Health Service (NHS) (2017). *Statistics on Smoking, England 2017*. Retrieved from https://www.gov.uk/government/statistics/statistics-on-smoking-england-2017

NICE (The UK National Institute for Health and Excellence) (2007). *Evidence for the effect on inequalities in health of interventions designed to change behaviour*. Retrieved from https://www.nice.org.uk/guidance/ph6/documents/evidence-for-the-effect-on-inequalitiesdesigned-to-change-behaviour2.

NICE (The UK National Institute for Health and Excellence) (2013). *Stop smoking services.* Retrieved from https://www.nice.org.uk/guidance/ph10/chapter/1-Key-priorities

Ormston, R., van der Pol, M., Ludbrook, A., McConville, S., & Amos, A. (2014). *Quit4u: the effectiveness of combining behavioural support, pharmacotherapy and financial incentives to support smoking cessation*. Health Education Research.

O'neill, J., Tabish, H., Welch, V., Petticrew, M., Pottie, K., Clarke, M., …, & Evans, T. (2014). "Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health." *Journal of Clinical Epidemiology, 67* (1): 56–64. doi:10.1016/j.jclinepi.2013.08.005.

O'donoghue, T., & Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making, 13*(2), 233-250. doi:10.1002/(sici)1099-0771(200004/06)13:2<233::aid-bdm325>3.0.co;2-u

O'donoghue, T., & Rabin, M. 2002. "Addiction and present-biased preferences." Retrieved from http://www.arts.cornell.edu/econ/edo1/addicts.PDF.

Petticrew, M. 2004. "Evidence for public health policy on inequalities: 1: The reality according to policymakers." *Journal of Epidemiology & Community Health, 58* (10), 811–16. doi:10.1136/jech.2003.015289.

Petticrew, M., Tugwell, P., Kristjansson, E., Oliver, S., Ueffing, E., & Welch, V. (2012). Damned if you do, damned if you don't: Subgroup analysis and equity. *Journal of Epidemiology and Community Health, 66*(1), 95-98.

Petkovic, J., Welch, V., Jull, J., Petticrew, M., Kristjansson, E., Rader, T., Yoganathan, M., McGowan, J., Lyddiatt, A., Grimshaw, J.M., Volmink, J., Moher, D., Shea, B., Pottie, K., Pantoja, T., Wells, G.A., & Tugwell, P. (2017). *How health equity is reported and analyzed in randomized trials.* Cochrane Database of Systematic Reviews 2017, Issue 8. Art. No.: MR000046.

Popay, J. (2008). Should disadvantaged people be paid to take care of their health? No. *British Medical Journal, 337,* (Jul08 3). doi:10.1136/bmj.a594

Radley, A., Ballard, P., Eadie, D., MacAskill, S., Donnelly, L., & Tappin, D. (2013). Give It Up For Baby: outcomes and factors influencing uptake of a pilot smoking cessation incentive scheme for pregnant women. *BMC Public Health*, *13*, 343.

Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, *19*, 171–197.

Schmidt, A. T. (2016). Withdrawing versus withholding freedoms: Nudging and the case of tobacco control. *American Journal of Bioethics*, *16*(7), 3–14.

Slovic, P. (2001). Cigarette Smokers: Rational Actors or Rational Fools? In Slovic, P. (Ed.) *Smoking: Risk, perception, and policy*. Thousand Oaks, CA: Sage.

Sun, X., Briel, M., Walter, S. D. & Guyatt, G. H. (2010). Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *British Medical Journal*, *340* (mar30 3), c117–c117. doi:10.1136/bmj.c117.

Sunstein, C. R. (2015). Nudging smokers. *New England Journal of Medicine*, *372*, 22, 2150 -2151.

Steel, D. (2008). *Across the boundaries: extrapolation in biology and social science*. Oxford: Oxford University Press.

Strickland, S. (2014). Does it work to pay people to live healthier lives? *British Medical Journal*, *348*, g2458.

Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, *12*, 183-206.

Thaler R.H., Tversky, A., & Kahneman, D. (1997). The effect of myopia and loss aversion on risk taking: an experimental test. *Quarterly Journal of Economics*, *112*, 647–661.

Thaler, R. H., & Sunstein, C. (2008). *Nudge*. Yale, New Haven, Conn.

Thomson, G., Morgan, H., Crossland, N., Bauld, L., Dykes, F., & Hoddinott, P. (2014). Unintended Consequences of Incentive Provision for Behaviour Change and Maintenance around Childbirth. *PLoS ONE*, 9(10). doi:10.1371/journal.pone.0111322

Tugwell, P., Petticrew, M., Kristjansson, E., Welch, V., Ueffing, E., Waters, E., Bonnefoy, J., Morgan, A., Doohan, E., & Kelly, M. P. (2010). Assessing equity in systematic reviews: realising the recommendations of the Commission on Social Determinants of Health. *British Medical Journal*, *341*, (sep13 1): c4739–c4739.

Tversky, A., & Kahneman. D. (1974). Judgment under uncertainty: Heuristics and biases. *Sciences*. 1124-1131.

Verdurmen, J., Monshouwer, K., & Van Laar, M. (2015). *Factsheet Continu Onderzoek Rookgewoonten 2014.* Utrecht: Trimbos-instituut.

Volpp, K.G., Troxel, A.B., Pauly, M.V., Glick, H.A., Puig, A., Asch, D.A. (2009). A randomized controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*, *360*, 699-709.

Volpp, K. G., Asch, D. A., Galvin, R., & Loewenstein, G. (2011). Redesigning Employee Health Incentives — Lessons from behavioral economics. *New England Journal of Medicine*, *365*(5), 388-390. doi:10.1056/nejmp1105966

Weber B.J., & Chapman, G.B. (2005). Playing for peanuts: why is risk seeking more common for low-stakes gambles? *Organ Behav Hum*, *97*, 31-46.

Welch, V., Tugwell, P., Petticrew, M., de Montigny, J., Ueffing, E., Kristjansson, B., McGowan, J., Benkhalti Jandu, M., Wells, G.A., Brand, K., Smylie, J. How effects on health equity are assessed in systematic reviews of interventions. *Cochrane Database of Systematic Reviews 2010*, Issue 12. Art. No.: MR000028. DOI: 10.1002/14651858.MR000028.pub2.

Welch, V., Petticrew, M., Petkovic, J., Moher, D., Waters, E., White, H., … the PRISMA-Equity Bellagio group. (2015). Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration. *International Journal for Equity in Health*, *14*, 92. http://doi.org/10.1186/s12939-015-0219-2

Whitehead, M. 1992. The concepts and principles of equity and health. *International Journal Health Service, 22* (3), 429e45.

World Health Organization (WHO). (2011). *Global status report on noncommunicable diseases 2010*. Retrieved from http://www.who.int/nmh/publications/ncd_report2010/en/.

World Health Organization (WHO). (2012) *Health impact assessment, glossary of terms.* Retrieved from http://www.who.int/hia/about/glos/en/index1.html.

# Chapter 4

# Reconciling Ernest Nagel's Impartiality with Anna Alexandrova's "Mixed Claims"

## 4.1. Introduction

The boundaries that define where the scientific content begins and ends in social and behavioral sciences are inconspicuous. Social scientific research projects may often be driven by scientists' purpose to change or preserve the existing institutions and policies and defend or criticize a given ideal about how society should be organized (e.g., Stiglitz, Sen, Fitoussi, 2010; Banerjee & Duflo, 2020). The normative aspects of the social scientific inquiries often shape how social scientists conduct their research and may even influence their results. Social and behavioral scientists' assertions thereby commonly express implicit or explicit non-epistemic value presuppositions. Yet, social and behavioral scientists often find themselves in positions to make influential policy recommendations, and they may aim to influence public opinion at large.

All this may sound problematic in the eyes of uninformed observers and the users of social and behavioral scientific knowledge. The value-oriented and value-laden nature of these sciences seems to blur the very reliability and trustworthiness of their input for policy deliberation. One may question the factual character of the social and behavioral experts' policy recommendations, on the grounds that they represent scientists' opinions or policy preferences instead of evidence and a kind of objective warrant for decision-making.

While the nature of objectivity in social scientific inquiries has been debated prominently since Max Weber's time, there is ongoing and indeed renewed interest in recent years as philosophers of science ever more frequently engage with the issues regarding scientific advisory, evidence-based policymaking, and the role of scientific experts in democracy (Kitcher, 2011; Cartwright & Hardie, 2012; Betz, 2013; Alexandrova, 2017, 2018; Rolin, 2020; Brennan, 2017; Reiss, 2017, 2019; Koskinen, 2020).

Anna Alexandrova offers an account that is focused on the social and behavioral sciences (2017, 2018). She contends that the existing accounts of social scientific objectivity cannot accommodate the "mixed" nature of social and behavioral sciences, specifically the mixed nature of the claims asserted by the social and behavioral scientists. "Mixed claims" as a close relative of "thick concepts" include some non-epistemic value presuppositions about them in their definition. For instance, a statement like “reading improves wellbeing” (including its empirical justification), rests on a normative conception of what wellbeing means. Alexandrova's overarching project is to offer a new account of social scientific objectivity that is more suitable for such "mixed sciences." In contrast to the traditional Weberian tenets on the subject, her proposed account of objectivity rejects the idea that social scientists should refrain from making explicit non-epistemic value judgments.

Alexandrova's argument is based on a rejection of Ernest Nagel's treatment of value-inflicted concepts in the social sciences. She focuses on Nagel's famous distinction between characterizing and appraising value judgments, which establishes that social

scientists, in principle, can make impartial descriptive claims conditionalized on value presuppositions. Alexandrova argues that plausible ways to operationalize Nagel's distinction in the context of mixed claims are either impractical or undesirable. She defends that it is desirable for social and behavioral scientists to explicitly attend to making the necessary value judgments, and the Nagelian advice goes against this idea. If Alexandrova is correct in her argument against Nagel, mixed sciences' impartiality should be discussed in radically different ways than traditionally understood.

In this chapter, I claim that the contrast between Nagel's and Alexandrova's approaches to social scientific objectivity is not as strong as it initially looks. I contend their respective contributions address conceptually distinct problems concerning mixed claims in social and behavioral sciences. Specifically, I suggest that Alexandrova's critique and insights should be understood as responding to the problem of determining the sources and the content of value judgments in mixed claims. I call this the "authority problem of values". I distinguish this problem from that of defining the legitimate roles value judgments should play in mixed claims. The latter I call the "legitimacy problem of values". I demonstrate that the Nagelian impartiality approach to mixed claims primarily responds to this latter problem. As such, the Nagelian approach to mixed claims remains intact also when accepting Alexandrova's critique. On this basis, I claim that Alexandrova's insights are reconcilable with the Nagelian framework. Importantly, this does not deflate Alexandrova's critique of Nagel. Rather, it highlights its main import for the philosophical accounts of social scientific objectivity, or so I will argue.

The chapter is structured as follows. In Section 2, I introduce the problem of mixed claims in policy-relevant social and behavioral sciences as a challenge to the traditional conception of social scientific impartiality. In Section 3, I review the Nagelian impartiality approach to address the challenge and Alexandrova's critique of it. In Section 4, I suggest that Alexandrova does not fully reject Nagelian impartiality to the extent that it serves for fulfilling impartiality in mixed claims, motivating a qualification of her critique. In Section 5, I distinguish between the authority and the legitimacy problem of values to qualify Alexandrova's disagreement with the

Nagelian treatment of mixed claims and suggest that the two accounts can indeed be reconcilable, as Alexandrova's disagreement applies to the authority problem of values. In Section 6, I respond to some arguments against the proposed reconciliation, focusing on how the contemporary Nagelian approaches accommodate scientists' use of their own value positions in making mixed claims. In Section 7, I conclude with a brief summary.

## 4.2. Mixed Claims as a Challenge to Social Scientific Impartiality

Social and behavioral scientific claims use concepts such as wellbeing, unemployment, inequality, poverty, nudge. These concepts are, at least in part, defined and measured by relying on some non-epistemic value presupposition about the nature of them. For instance, scientific investigations of poverty and empirical generalizations about poverty rely on some presupposed standard of what it means to be poor. Anna Alexandrova (2018) proposes the notion of "mixed claims" to label empirical generalizations of this kind. Many social and behavioral scientific claims deserve the label mixed because "they mix the normative and the empirical in a way that ordinary scientific claims do not" (2018, p.422). Alexandrova also offers a general definition of mixed claims (2018, p. 424):

> *Mixed Claim*. (1) It is an empirical hypothesis about a putative causal or statistical relation. (2) At least one of the variables in this hypothesis is defined in a way that presupposes a moral, prudential, political, or aesthetic value judgment about the nature of this variable.

Even though mixed claims are part and parcel of social and behavioral scientific practice, their presence challenges the traditional notion of social scientific impartiality which hinges on the separation of facts and values (Weber, 1949; Nagel, 1961; Anderson, 2004; Mongin, 2006; Douglas, 2011; Su & Colander, 2013). Scientific impartiality requires that the justification of the empirical claims or the content of the empirical results does not depend on the non-epistemic values that a

scientific investigation may rely on. On the face of it, the ideal of impartiality does not appear to accommodate mixed claims because a mixed claim, by definition, relies on a wittingly or unwittingly presupposed value position. Consider the following examples.

- *Wellbeing*. All scientific measures of wellbeing rely on *some* philosophical conceptions of what being well means in a given circumstance (see Alexandrova, 2018 for an exhaustive analysis of different conceptions of wellbeing operationalized by social and behavioral scientists). Wellbeing might mean satisfaction of people's subjective preferences; hence, some wellbeing measures have been developed by scientists taking this presupposition into account. Wellbeing may also be measured by reference to people's hedonic experiences; hence, behavioral scientists may sometimes draw on the hedonist theories of wellbeing. It can also be assessed by focusing on a set of non-subjective criteria, such as by measuring how far they are endowed with capabilities that are believed to be conducive to living well in a given context. Settling on a particular conception of wellbeing is primarily a philosophical activity that requires one to address normative questions regarding what conditions are supposed to make a person well in a given context. Accordingly, when the social and behavioral scientist operates on a specific measure of wellbeing or proposes a new one, she also partly engages with primarily non-epistemic content. Her choice of a wellbeing measure reflects a philosophical presupposition regarding the nature of wellbeing, even though this stance is not always made explicitly and endorsed wittingly; or justified by the scientists operating with them; or agreeable by the users of the knowledge about wellbeing.

- *Nudge*. Behavioral policy proposals are ethically sophisticated interventions. They differ from each other regarding how they approach the ethical aspects of behavior change (see Barton & Grüne-Yanoff, 2015 for a review). For instance, nudges are supposed to respect people's freedom of choice and autonomous decision-making processes and motivated as libertarian

> paternalist policies. To assess behavioral policies' effectiveness empirically, the behavioral scientist needs to instantiate these interventions in experimental or quasi-experimental conditions to study whether and how they work. In order to instantiate a nudge intervention, researchers generate a situation that mimics the real-world in which a subtly operating choice architecture respects people's freedom and autonomy in guiding their choices. Doing so requires addressing a set of ethical questions about behavior change. For instance, one has to be clear about how exactly the intervention respects decision-makers' autonomy (e.g., by making the intervention transparent or by providing choosers with a genuine chance to opt-out of the choices they were nudged to). More generally, there are ethical and methodological ambiguities regarding how these policies' ethical requirements should be met in practice (see Heilmann, 2014 for a detailed methodological analysis regarding nudges). Irrespective of how the behavioral policy analyst resolves such methodological ambiguities, the way the intervention is designed and implemented will always reflect *an* understanding of the ethics of the intervention (e.g., regarding how the intervention respects people's freedom or autonomy). Again, this understanding may not always be made explicit or justified, or wittingly attended by the scientists evaluating the intervention, or scrutinized by the users of the scientific assessment such as policy deliberators.

Due to the value-inflicted nature of the concepts “wellbeing” and “nudge”, empirical claims about them such as “4-day-week policy improves employees' *wellbeing*” or “*nudges* work more effectively than information campaigns in dealing with obesity” are mixed claims in Alexandrova’s sense. Unless reformulated and communicated differently, these kinds of claims appear to be partial as described above, violating the received philosophical notion of scientific impartiality.

These examples also highlight that the violation of impartiality matters in a pragmatic sense as well. Non-empirical ambiguities surrounding a mixed claim have potentially decisive consequences in the realm of policymaking and the eyes of policy

deliberating actors. For instance, the wellbeing scholar may measure the effects of a 4-day-week policy based on subjective reports of the employees' satisfaction, and pay less attention to the prudential effects of the leisure activities employees engage in during their extra free time. This may lead to the underrepresentation of some potential benefits of the 4-day-week policy on wellbeing. Similarly, the behavioral policy evaluator may instantiate a less paternalistic nudge for the sake of respecting freedom and autonomy during the experimental implementation, curbing the intervention's effectiveness. Unless the ethical desiderata of the intervention are not explicitly communicated, the results of the nudge experiment may suggest to the policymaker that nudging could be less preferable to an alternative policy intervention that is measured to be more effective even though it may not fulfill the ethical desiderata that the nudge does. The real-world instances of science-policy interactions are, of course, way messier than manifest in these examples, and singular scientific arguments seldomly shape policy decisions so exclusively and directly. The examples simply highlight that our evaluations of mixed claims are important for pragmatic purposes as well as theoretical discussions concerning the social scientific objectivity.

Now, the traditionally accepted response to the challenge from mixed claims is to recommend that mixed claims can be reformulated as impartial claims (Weber, 1949; Anderson, 2004; Douglas, 2011; Mongin, 2006; Nagel, 1961; Su & Colander, 2013). To do so, a conceptual distinction between the descriptive and the normative content in mixed claims should be put forward by explicitly specifying all kinds of non-epistemic presumptions involved in the way they form and communicate their judgments. In this way, social and behavioral scientists present mixed claims as impartial ones for democratic policy deliberation, even though these claims are not strictly free from non-epistemic value presuppositions. Anna Alexandrova's evaluation of mixed claims; however, motivates a rejection of the traditional approach to mixed claims as she argues against reformulating mixed claims as impartial claims, as I explain below. I will first review the disagreement between Alexandrova and Nagel before motivating their reconciliation in sections 4 and 5.

# 4.3. Nagelian Impartiality and Alexandrova's Critique

## 4.3.1. Nagelian Impartiality: Communicating value-inflicted claims through conditionalizing on unwarranted assumptions

The traditional philosophical accounts of social scientific objectivity have understood the issue of value-inflicted language as a practical, not a principled, challenge to social scientific claims' impartiality. The concepts used and formed by social scientists may embed unwarranted value presuppositions in them, but that does not necessarily harm the objectivity (qua impartiality) of the claims made by social scientists, according to many philosophers of science who subscribe to the traditional approach (e.g., Weber, 1949; Nagel, 1961; Anderson, 2004; Mongin, 2006; Douglas, 2011; Su & Colander, 2013). The challenge is surmountable by making the value assumptions explicit and disentangling them from the primarily factual statements. In this way, social scientific claims are properly localized and conditionalized on the value presuppositions. This motivates the idea that the justification of the social scientist's conclusions does not depend on the values inserted into the analysis.

The question, of course, is how exactly value presuppositions can actually be disentangled from social and behavioral scientists' descriptive statements. When a scientific expert claims "X improves wellbeing", doesn't she unavoidably impose an unwarranted subjective standard of wellbeing? One of the most enduring responses to this question has been articulated by Ernest Nagel (1961). Nagel, in the relevant chapters of his book (*Value-oriented bias of social inquiry*, Chapter 36), defends the idea that it is possible to keep the descriptive and the normative content separate from each other even when the language used for reaching and making social scientific claims is deeply entangled with values.[1] He invites the skeptics to appreciate the idea

[1] In the chapter, Nagel's aim is to establish the idea that there is no principled difference between social and natural sciences. He readily accepts that many concepts used by social scientists are indeed laden with normative value judgments. However he rejects that the

that social scientists can, in principle, make "conditional rather than categorical assertions about values" (Nagel, 1961, p.491). In making a value-laden descriptive judgment, social scientists do reason with values but without necessarily expressing appraisals implicated by those values or imposing these values to their addressees. The key, according to Nagel, is to acknowledge a conceptual distinction between two kinds of value judgments. Value presuppositions might express one's approval or disagreement (called *appraising* value judgments). They could also express one's characterization of how far a state of affairs or a normative ideal is instantiated in a given circumstance (called *characterizing* value judgments) (Nagel, 1961, p. 492).

It is useful to revisit Nagel's famous example to illustrate this distinction. Consider the statement "the animal is anemic." Nagel tells us that the "anemic" as a judgment may express an estimation (characterizing how far the animal instantiates the characteristics attributed to the condition *anemia*). It could also express a pejorative judgment about the badness of the animal's state of health. Nagel's example helps us appreciate the idea that a claim about a value-inflicted concept such as wellbeing can be understood and communicated by the social and behavioral scientist as a characterization only. "X improves wellbeing" can be explicitly expressed as "X improves the wellbeing-conducive conditions that the Y-conception of wellbeing describes." Moreover, acknowledging the distinction, the social and behavioral scientist can also make it explicit when she not only characterizes but also appraises. So, when it is the aim of the social scientist to make a normative judgment or support a policy option, "X improves wellbeing" can be explicitly expressed by the social scientist as "it is desirable that X improves the wellbeing-conducive conditions that the Y-conception of wellbeing describes because the Y-conception of wellbeing is desirable."

It may not always be required to reformulate mixed claims in scientific practice (Nagel, 1961, p. 494). Yet it can be desirable to operationalize Nagel's distinction in the context of reaching scientific results and communicating them for policy

problem poses unique challenges for the social sciences, and he argues that the social scientific claims should in principle be treated in the same way as the natural scientific claims.

deliberation. As long as the social and behavioral scientific experts can single out the unwarranted non-epistemic assumptions in their claims, they would be in a position to be transparent about the roles these assumptions play in their analysis and qualify the communication of their results accordingly. This requires separating primarily characterizing claims from the value presumptions made, as suggested by some contemporary philosophers who focus on the communication of scientific results in a democratic context (see, for instance, Betz, 2013; Jeffrey, 2018).

### 4.3.2. Alexandrova's argument "against Nagel" [2]

Anna Alexandrova (2017, 2018) has argued against this particular strategy for dealing with the problem posed by mixed claims in social and behavioral sciences. Her contention is that the objectivity and the reliability of the value-laden claims should not be dependent on whether appraisals are separable from characterizations. She argues for leaving the Nagelian solution illustrated above behind and proposes an alternative account.

In the 2017 book, Alexandrova's project is to understand, evaluate, and defend the sciences of wellbeing. In this project, one of the philosophical questions she tackles is whether the sciences of wellbeing can be objective and, if so, in what sense it is (see also Alexandrova, 2018 for an article-length treatment of the issue). Alexandrova contends that that "there is still no positive story about how projects that rely on mixed claims can be both value-laden and objective" (Alexandrova, 2018, p. 422). She wants to offer one such positive story which accommodates mixed claims as plausibly objective claims and reliable for policy deliberation in a democratic setting.

This project leads her to argue against the received Nagelian approach to deal with mixed claims I described above. Her premise is that when mixed claims are reformulated as characterizing claims conditionalized on a set of value judgments, they would not be mixed anymore (Alexandrova, 2018, p.429). This premise seems

[2] This refers to the title of the section where she presents her arguments against Nagel (2018, 430-432).

to be in tension with the deeply normative character of the scientific claims typically made in the wellbeing sciences. Alexandrova instead defends the idea that mixed claims should remain mixed and value-laden. Specifically, she finds conditionalizing on value presuppositions unsatisfactory.

She argues that the strategy of conditionalization does not eliminate the need to make normative judgments instead "only pushes them to another, less appropriate stage" (Alexandrova, 2018, p.430):

> Suppose we went with Nagel and reformulated mixed claims into estimation claims, then there would still remain a question as to which normative standard scientists should use in their estimation claims.

So, even though the mixed claims can be presented as impartial claims through the use of conditionalization strategy, the scientist who makes these claims still need to settle on some normative presumptions in reaching their judgments, and there seems to be no non-arbitrary way to determine the content of the antecedent of the conditionalized mixed claims.

Alexandrova considers possible strategies through which the content of value judgments is determined while keeping the scientists' impartiality intact and argues that none of them satisfactorily fits the bill. She, for instance, considers the possibility of relying on folk conceptions of value standards to fill in the normative part of the conditional estimation claims. But, she dismisses it as it is often the case that the "folk" disagrees about their presupposed standards of wellbeing (Alexandrova, 2018, p. 430). She also considers whether scientists could avoid taking a stance about value assumptions by running analyses through considerations of alternative normative standards and reporting the resulting judgments for each separately. She also finds this strategy unsatisfactory, as there is no uncontroversial way of defining such a list, so a normative judgment about the relevance of different normative standards must be made in an inquiry (Alexandrova, 2018, p. 430). She then takes it that the most plausible Nagelian response to her challenge would be to rely on a division-of-labor argument that distinguishes between scientific experts as impartial characterizers of

facts and the relevant non-experts as the providers of the necessary normative presuppositions.

Normative theorists or moral philosophers could be considered legitimate suppliers of the mixed sciences' normative input. It is also intuitive to suppose that, in a broadly liberal democratic context, scientists could turn to the democratically legitimized decision-makers or citizens who are at the receiving end of the evidence-based policies as the relevant decision-makers concerning the value presuppositions that would permeate into policy-relevant research projects. Accordingly, the proponents of the conditionalizing strategy may indeed plausibly suggest that, when necessary, the legitimate non-experts should be assumed to make the necessary normative decisions that would then determine scientists' conceptual and discursive choices (Alexandrova, 2018, p. 431). Scientific experts would then make mixed claims without imposing their own normative presuppositions to their addressees. Alexandrova also rejects this promising argument in favor of Nagel's argument's operationalizability.

Specifically, she argues that this strategy faces two challenges. One of these challenges is a practical one. The practical challenge against the operationalizability of the Nagelian division of labor is the abundance of small-scale normative decisions that need to be made in mixed sciences. It is simply impractical that scientists avoid making small normative decisions themselves by deferring to the "outsiders" whenever the normative decision to be made would influence their resulting judgments. Nevertheless, Alexandrova suggests that this challenge is simply a practical one that is in principle surmountable in practice, so it does not give us a cogent reason to reject the advice the conditionalize mixed claims in the context of policy advice.

Her second and main challenge against the Nagelian division-of-labor argument targets the very normative desirability of the conditionalizing strategy. Alexandrova's premise is that social scientists' own value presuppositions should, at least in some contexts, be considered apt for the task. Based on this premise, she argues that the division-of-labor argument "ignores or devalues scientists' knowledge about values,

which they have acquired in virtue of their knowledge of facts. This knowledge enables them to make better normative choices qua scientists" (Alexandrova 2018, p. 431). To support this argument, Alexandrova gives examples of those cases where social scientists make informed and defensible value judgments such as divorce scholars' choice of conceptualizing divorce not as "breaking up" but as a potentially maturing activity, drawing on Anderson 2004:

> In all these cases, value judgments are a result of an epistemic process; they are informed in part by facts and in part by the earlier value judgments made to detect those facts. Because of this process of co-evolution, scientists are in a good position to make some value judgments. Consulting philosophers and the public when making normative choices is important, but that does not mean scientists should refrain from using their own hard earned normative knowledge. So the Nagelian division of labor ignores the methodological realities of mixed sciences and wastes the normative knowledge scientists acquire while in the business of producing mixed claims (Alexandrova 2018, p. 432).

Suppose that Alexandrova is right that the strategy of conditionalization does not really resolve the problem posed by mixed claims, or else does so in undesirable ways. We must then rely on some alternative ways to justify mixed claims in the context of policy advice. The account should be able to protect mixed claims from any arbitrary subjectivism of scientific experts such that the unwarranted or unattended value presuppositions do not permeate into evidence-based policy deliberation processes or imposed on the citizens. Alexandrova highlights problems of "inattention" and "imposition" of values by scientists and suggests that our accounts of social scientific objectivity should be able to address these problems (Alexandrova, 2018, pp. 432-433).

Alexandrova argues that relevant extant accounts of scientific objectivity that accommodate the value-ladenness in scientific judgments (such as Longino's (1990) suggestion to control hidden value assumptions through pluralism and interaction among peer reviews) do not adequately address the task of controlling value presuppositions in the context of mixed claims. This seems to be right, especially if one considers the short-term, decision-oriented nature of evidence-based policy

context. Alexandrova, accordingly, proposes a version of procedural objectivity that is tailored for the specific problem of value-inflicted language in sciences like wellbeing studies.

Her proposal aims to set some ground rules that meet the desideratum that scientists' normative judgments are transparently subjected to the scrutiny of their peers and the relevant stakeholders in public. These ground rules are three-fold (Alexandrova, 2018, pp. 437-440). First, scientists should make the value presuppositions that they insert in their analysis explicit. Second, it must be checked if their results are robust under different or conflicting normative presuppositions. Third, if the second step leads to controversy, scientists should check in with the relevant stakeholders to justify their normative assumptions through, for instance, deliberative democratic avenues. These procedures, she hopes, can "ensure that values undergone an appropriate social control, giving a community reasons to trust this knowledge" (Alexandrova, 2018, p. 436).

If one accepts Alexandrova's critique and proposal, one needs to radically revise the traditionally adopted accounts of social and behavioral scientific objectivity. At the very least, it implies leaving the Nagelian conditionalizing strategy behind. Nevertheless, I will now demonstrate that the contrast between these two approaches is not as stark as it looks, and that there are some important conceptual commonalities between Nagel's and Alexandrova's respective approaches that motivate a reconsideration of Alexandrova's critique and a reconciliation between the two apparently contrasting approaches.

## 4.4. A common ground for Nagel and Alexandrova : “impartiality$_2$”

I will now motivate the idea that Alexandrova's critique of Nagelian conditionalization for mixed claims needs to be qualified and specified further. This will allow me to locate the main disagreement and motivate reconciliation thereafter.

Alexandrova claims that the conditionalizing strategy needs to be rejected because it does not really address how the value presuppositions in mixed claims should be satisfactorily determined. Moreover, she argues, Nagelian ways to address this problem (e.g., deferring to the non-scientists in determining the content of value presuppositions) are unsatisfactory because they deny social scientists to engage with normative questions even though that scientists' attendance to the value presuppositions often seems to be a methodological and epistemic desideratum in social and behavioral sciences.

Alexandrova's critique does not target what seems to be the primary function of the Nagelian conditionalization that their proponents deem important. Recall that the primary goal for conditionalizing mixed claims is to operationalize the impartiality in producing and asserting mixed claims. When the value content in a mixed claim is isolated, and the claim is conditionalized on the value presuppositions made, it becomes possible to demonstrate how the claim is accepted and communicated in an *impartial* manner. Through the separation of the primarily normative content from the primarily descriptive content in a value-inflicted claim, one could demonstrate that the claim's acceptance is not primarily dependent on the value presuppositions made (avoiding the charge of wishful thinking). Similarly, one could communicate a mixed claim as a conditional statement in order to be transparent about the value presuppositions made. This renders it possible to contextualize the claim properly for its users and to accept that the claim may not be accepted when alternative value presuppositions are inserted, which avoids "wishful speaking" (John, 2019)[3].

In more general terms, the goal of reaching impartiality in making mixed claims is independent of the goal of justifying the value presuppositions in mixed claims. Only the latter appears to be the main focus of Alexandrova's critique. It is unclear,

[3] Stephen John offers a helpful definition of wishful speaking: “A scientist engages in wishful speaking when she makes a claim which is not well-established relative to relevant epistemic standards for accepting claims, and where her motivation for making that claim is the predicted non-epistemic benefits that follow from others believing (or believing that she believes) it, regardless of its truth” (John, 2019, p. 66).

however, whether Alexandrova also rejects the former; that is, Nagelian conditionalization *as a strategy for achieving impartiality of mixed claims, too.*

Alexandrova does reject the Nagelian conditionalization when it serves as a tool to *eliminate* the value-inflicted language in social and behavioral sciences for the sake of scientific impartiality. If the elimination of value-ladenness of social sciences is the ultimate aim of Nagel's approach, then the strategy of conditionalization would be meant to be used for making mixed claims *value-free*. Mixed claims that are not reformulated properly as conditional ones cannot then pass Nagel's scientific impartiality test. Thus formulated, Nagel's approach appears to be against mixed claims. Alexandrova seems to presuppose this interpretation of Nagel's approach when she labels it as "the most explicit case against mixed claims" (Alexandrova, 2018, 429):

> But his prising apart of appraisal from estimation has more ambitious intentions. The point of drawing the distinction is to eliminate appraisal from science, leaving only estimation. The ideal science for him is an ethically neutral one (Nagel [1961], p. 495). What I have called mixed claims are plausibly appraising claims on Nagel's picture.

In the same vein, Alexandrova dismisses a version of impartiality that sounds equivalent to value-freedom:

> The problem is that when impartiality is formulated in a standard way as forbidding that values determine our acceptance of hypotheses, mixed claims face a test they could not possibly pass. This is because in a mixed claim, the initial value judgment does preclude certain findings and, to this extent, values do determine what we will find (Alexandrova, 2018, pp. 434-435).

However, the pursuit of value-freedom in science is different from the pursuit of impartiality, as discussed previously, for instance, by Douglas (2004, 2011) and Anderson (2004), among others. Impartiality is an attitude towards values and appraisals in judgment formation and assertion. As such, the pursuit of impartiality in science does *not* imply the elimination of values. Impartiality as a goal of social and behavioral sciences requires that one produces and assesses evidence in a manner that

is detached from the value presuppositions made. This does not imply that values do not influence scientists' conclusions, but it does imply that values themselves cannot be sufficient reasons to determine the conclusions of a scientific study unduly. Impartiality is operationalized when researchers make the value presuppositions transparent and when their reasoning for accepting a claim relies primarily on epistemic considerations, even though value presuppositions do contextualize the claim to be accepted.

Indeed, formulated as an ideal of objectivity that regulates (and *does not forbid*) the use of values in social scientific judgment and assertion, impartiality is not fully rejected by Alexandrova, who offers a less strict version of it when applied to the context of mixed claims:

> *Impartiality*$_2$: A mixed claim is impartial if and only if once all the value decisions about the measures, methods, and required levels of confirmation are made, non-cognitive values to do not play any further role in determining whether the hypothesis is accepted. To be fair, this is probably the version closest to what advocates of impartiality have in mind, and it may well be acceptable to them. All I claim is that Impartiality2 does not help us with imposition and inattention. This rule guards against the imposition of values into claims already stripped of any values but epistemic. Thus it cannot tell us how to deal with claims that are not stripped of them, such as mixed claims (Alexandrova, 2018, p. 435).

Even though Alexandrova does not strictly reject this notion of impartiality, she does not elaborate on how Nagelian conditionalization relates to impartiality$_2$ – most likely because she directly ties Nagel's approach to value-freedom, not impartiality. But, there is nothing in the Nagelian approach to mixed claims that contradicts the principle of *Impartiality*$_2$. In line with *Impartiality*$_2$, Nagelian conditionalization of mixed claims as conditional statements *per se* does not *eliminate* the requirement to attend to the value judgments. Instead, it appears to be a way to operationalize the ideal of impartiality in the context of mixed claims. Conditionalizing on normative assumptions helps achieve impartiality because it compels one to be transparent about the relationship between the resulting claims and the values that went into making that claim. The purpose of conditionalizing mixed claims is to make it possible that the

claim's justification is *not determined* by the initial value presumptions but only contextualized by them. Accordingly, conditionalization per se does not forbid the idea that values may influence what one finds, unlike Alexandrova's description of its main purpose.

I have so far highlighted three premises which, taken together, sounds paradoxical and in need of clarification: (1) Alexandrova rejects Nagelian conditionalization; (2) Nagelian conditionalization serves for the pursuit of *impartiality*$_2$; (3) Alexandrova does not *fully* reject *Impartiality*$_2$. I now will suggest a way out of this paradox and demonstrate that Alexandrova's account does not reject Nagelian conditionalization as a tool for fulfilling impartiality in mixed claims, but as a tool for *determining* the content and source of values embedded in mixed claims. I claim that once we qualify Alexandrova's disagreement with Nagelian conditionalization in this way, it will be apparent that her rejection of the Nagelian approach to mixed claims is not as stark as it initially looks and that the two accounts can indeed coexist and even complement each other.

## 4.5. Distinguishing between the authority and the legitimacy of value presuppositions embedded in mixed claims

I propose to make an explicit distinction between different kinds of challenges regarding the involvement of values in science.

One challenge regarding (non-epistemic) values in science is to define their *legitimate roles* and the limitations of their *use* in scientific judgment or scientific assertion. Call this the *legitimacy* problem about values in science. Another challenge is to define the proper *source*s and the *content* of the value judgments made in scientific reasoning and to define what makes value judgments more *authoritative.* Call this the *authority*

problem about values in science.[4] On the basis of this distinction, I will now argue that it is, in principle, possible to be Nagelian *and* a proponent of Alexandrova's suggestions. More specifically, one could favor Alexandrova's procedures to determine the value judgments' content (with respect to the *authority* problem) and be a Nagelian in suggesting how mixed claims should be justified and communicated by scientists (with respect to the *legitimacy* problem).

The legitimacy and the authority problems are conceptually separable challenges to respond to by different extant accounts of values and objectivity in social sciences. As such, the way they are addressed in practice need not rely upon the same ground principles on the values and objectivity in sciences. One could, for instance, in principle, follow Weber's insights in response to the legitimacy problem of values and subscribe to Kitcher's insights in response to the authority problem.

Let's first revisit Nagel's treatment of mixed claims in the light of this distinction. Recall that the primary purpose of the distinction between characterizing and appraising value judgments is to conceptualize and operationalize impartiality: Nagel's distinction helps social scientists clarify and distinguish the descriptive and the normative elements in their study as transparently as possible. In doing so, it helps scientists stay detached from the values presupposed during the process of reaching their claims and regulate how they assert these claims, serving to realize the impartial sense of objectivity. As such, Nagelian conditionalization of the mixed claims is meant to be a conceptual tool to address the *legitimacy* problem. A Nagelian about values in social sciences needs to commit to the idea that scientists (can) in principle and (should) in practice strive to keep the descriptive and the normative content conceptually distinct from each other to avoid any forms of deceptive assertion or unduly permeation of values in judgment formation.

Apart from this primary function of it, the practical application of Nagel's distinction is also helpful to address the authority problem, as illustrated by Alexandrova in detail: Reformulating mixed claims into conditional claims allows scientists to isolate the

[4] (Rolin, 2020) introduces a similar distinction.

value presuppositions in claims and *not commit to* them, *if necessary*. In other words, operationalizing Nagel's distinction helps social scientists avoid *endorsing* any value positions themselves, and the *authority* of these judgments can simply be sought elsewhere. However, I submit, a Nagelian does not have to address the authority problem in this way.

An important confusion arises because, as a matter of historical fact, some prominent commentators in the recent history of philosophy of science have employed the Nagelian conditionalizing on value assumptions for addressing the authority problem in this specific way. For instance, Betz 2013 defends the value free ideal based on arguments from individual autonomy and democratic legitimacy, and he finds the Nagel-type conditionalization of value-laden claims particularly useful for fulfilling these meta-values in scientific judgment because the conditionalization strategy allows scientists to not inflict their own value judgments and thereby respects the broader principles of democracy and autonomy in decision-making. Nevertheless, there is no principled requirement for Nagelians to address the legitimacy and authority problems on the same conceptual grounds, as Betz does.

Indeed, Nagel himself implicitly adopts the distinction I operate on here and does not take an explicit position regarding the normative *authority* of value judgments, and he finds it helpful to add the following footnote regarding the matter:

> It is irrelevant to the present discussion what view is adopted concerning the ground upon which such judgments supposedly rest—whether those grounds are simply arbitrary preferences, alleged intuitions of "objective" values, categorical moral imperatives, or anything else that has been proposed in the history of value theory. For the distinction made in the text is independent of any particular assumption about the foundations of appraising value judgments, "ultimate" or otherwise (Nagel, 1961, p. 493).

Accordingly, one could put forward Nagel's distinction to address the legitimacy problem of values in social sciences; *and* she would *not* contradict herself when she does *not* employ the same distinction to address the authority problem, too. By implication, a Nagelian *can* agree with Alexandrova and in principle support the idea

that social scientists sometimes should *not* refrain from employing their own value judgments and that scientists' value judgments can sometimes be considered normatively authoritative, or that the authority of the value content may sometimes need to be substantiated through procedures of social control. It is then, in principle, possible to be Nagelian *and* a proponent of Alexandrova's suggestions. More specifically, one could favor Alexandrova's procedures to determine the value judgments' content (with respect to the authority problem) and be a Nagelian in suggesting how mixed claims should be justified and communicated by scientists (with respect to the legitimacy problem). A clear distinction between the authority and the legitimacy problem of values in mixed claims, therefore, illuminates that the contrast between Alexandrova's and Nagel's approaches is not as stark as it initially appears.

Moreover, the distinction also puts us in a better place to qualify Alexandrova's disagreement with the Nagelian framework. Alexandrova's criticisms regarding reformulating mixed claims as conditional statements are only concerned with the limitations of the former qua a strategy for validating the authority of value presumptions inserted in mixed claims. She does not explicitly criticize the Nagelian conditionalization qua a tool for achieving impartiality in making mixed claims (and so, as a tool for responding to the problem of legitimacy). Indeed, as I illustrated, she does not entirely reject impartiality$_2$ and only finds it limited in addressing the compelling problems she identifies regarding mixed claims, such as the imposition and inattention of values by scientists. In the light of the distinction proposed here, we can understand these problems as challenges pertaining to the authority of problem of values.

Let’s elaborate on the problems of inattention and imposition to clarify why they, strictly speaking, concern the authority problem of values. According to Alexandrova, the problem of inattention means that scientists may fail to identify the value presuppositions that a mixed claim rests upon. Relatedly, the problem of imposition means that scientists may import some value judgments into mixed claims that the outsiders (such as the subjects or the users of their investigation) may have good

reasons to reject (Alexandrova, 2018, pp. 432-433). Both of these problems suggest that the value judgments embedded in mixed claims need to be scrutinized thoroughly, especially when mixed claims are used for policy evaluation purposes. For instance, Alexandrova suggests that committing inattention "amounts to misusing the authority of science" and committing imposition amounts to underappreciating citizens' value preferences regarding matters that interest them. These problems indeed motivate Alexandrova's suggestions as to how values appearing in mixed claims should be determined and justified (e.g. through means of procedural checks and, when necessary, through participation of citizens). Accordingly, addressing the problems of inattention and impositions are ways to address the authority problem of values in mixed sciences. More generally, any accounts that aim at resolving these problems would consist of a set of prescriptions regarding how the content of value presuppositions in mixed claims should be determined such that the value judgments scientists rely upon can be said to be more authoritative. However, this exercise does not arm one up to properly address the legitimacy problem of values in mixed claims. Avoiding inattention and imposition in the context of determination of values does not address the issue of how a supposedly well-scrutinized (authoritative) set of value judgments should inform the way scientists make and assert mixed claims. The latter concerns the legitimacy problem of values in mixed claims. The Nagelian conditionalization of mixed claims primarily addresses the latter (though it can also be used for addressing the former, as illustrated above).

In sum, the legitimacy-authority distinction allows us to demonstrate the reconcilability of Alexandrova's account with the Nagelian impartiality: the conditionalization strategy, or impartiality$_2$ for that matter, is not coupled with a principled suggestion as to how the content of value judgments should be determined. Indeed, as I will illustrate further below, conditionalizing on value assumptions is also compatible with Alexandrova's specific desiderata for addressing the authority problem of values in mixed sciences.

## 4.6. Is Nagel's Approach Congruent with Social Scientists' Ownership of Value Presuppositions?

I have illustrated a conceptual understanding in which Alexandrova's rejection of the Nagelian approach to mixed claims is not as strong as it initially looks. I have also suggested that the two accounts can indeed coexist and even be reconcilable with each other.

Now, one might disagree with this conclusion by arguing that the Nagelian account cannot really accommodate the cases in which social and behavioral scientists could plausibly determine value presuppositions themselves and insert their own value judgments into their analysis. Indeed, one of the main reasons for Alexandrova to reject the Nagelian treatment of mixed claims is that it forbids social scientists to utilize their "perfectly valuable" and "hardly earned normative knowledge." Can the Nagelian conditionalizing strategy *really* incorporate Alexandrova's more radical proposals regarding how the authority problem of values in mixed sciences should be addressed by our philosophical accounts of social scientific objectivity?

It is straightforward for the following kinds of cases: where the assumed authority of the normative assumptions is not scientists themselves or where scientists do not want to or need to employ their own normative knowledge. There are a number of different contexts where scientists' own normative knowledge may be irrelevant or unnecessary to put in use. In contexts where scientific experts serve as consultants, for instance, the necessary normative judgments are supposed to be made by policymakers. There are also common cases in social and behavioral scientific practice where scientists want to explore the implications, applications, limitations of certain value stances for the sake of normative theorizing. In such contexts, scientists may employ specific values without necessarily endorsing them.

However, it is not straightforward how the Nagelian account works in cases where scientists themselves are the sources of necessary normative input in their mixed

claims. The Nagelian impartiality principle seems to be in conflict with social scientists' endorsement of any value positions. If scientists are supposed to be impartial about the values in their analysis and assertion, how is it plausible that scientists employ the values presuppositions they endorse? In terms of the distinction I introduced in the section above, the question is whether a Nagelian in response to the legitimacy problem of values could also accept the idea that scientists' ownership of value judgments can sometimes be considered authoritative in response to the authority problem of values.

The answer to this question is positive. Indeed, such a seemingly hybrid position is not a minority taste in the contemporary philosophy of science. Some prominent commentators endorse impartiality as an indispensable conception of objectivity, but they also believe that scientists can plausibly insert their *own* non-epistemic value judgments into their analysis. Contemporary feminist philosophers of science in the analytical tradition (e.g., Longino, 1990; Anderson, 2004; Douglas, 2003; Douglas, 2009) are perhaps best representatives of this position. Elizabeth Anderson, for instance, emphasizes avoidance of dogmatism, which is a desideratum concerning the legitimacy problem of values, while promoting feminist social scientists' employment of feminist values in their inquiry (Anderson, 2004), which is a recommendation in response to the authority problem of values. Douglas offers the most detailed account. She formalizes and defends an ideal of impartiality ("objectivity as detachment") in replacement of objectivity as value-freedom (Douglas 2004, 2009). Douglas also explicitly subscribes to Nagel's advice in her approach to social and behavioral scientific objectivity (Douglas, 2011, pp. 23-24). More precisely, Douglas is a Nagelian with respect to the legitimacy problem of values in science. [5] Simultaneously, she disagrees with the value-free ideal of science; and concerning the

[5] It must be noted that while Douglas keeps the Nagelian impartiality qua a response to the legitimacy problem of values, she transcends it as her account permits a more extensive set of legitimate roles for values to play than Nagel would permit. Moreover, different from Nagel, she proposes to keep an eye on epistemic and non-epistemic consequences of using appraising/thick language in assessing the legitimacy of value-inflicted language in sciences (Douglas, 2015, p. 624).

authority problem, she supports the idea that scientists should not refrain from making their own value judgments when their research informs policy decisions (Douglas, 2009, Chapter 4, Chapter 7).

Nagel also holds that social scientists' importation of their own values is a manageable practical difficulty and lists methodological steps that can be taken to fulfill the objectivity of the factual content in social scientific claims; however they might be laden with scientists' own value presuppositions (Nagel, 1961, p.489).[6] Neo-Nagelian philosophers such as Douglas and Anderson would respond to the authority problem in ways that Alexandrova would also endorse: by not devaluing scientists' own normative knowledge and by allowing scientists to make appraising value judgments. When scientists' claims are used for legitimizing controversial policy decisions, the methodological steps recommended by the Nagelian accounts can be extended by the procedures of social control recommended by Alexandrova. Nagel himself does not seem to be conceptually far away from this hybrid position. In his chapter, he submits that "it is surely beyond serious dispute that social scientists do in fact often import their own values into their analyses of social phenomena" (Nagel, 1961, p. 488) and that social scientists may often aim at changing the society towards normative ideals they subscribe.[7]

Finally, as a second-order objection, a skeptic might worry that in contexts where scientists employ their own value presuppositions, the need for reformulating mixed claims into conditional statements becomes groundless because conditionalizing a claim on value presuppositions is meant to disown the values presupposed. Betz (2013) clearly articulates this worry in his assessment of Douglas's account:

[6] Nagel recommends social scientists to be explicit about their value assumptions. As a corrective force against the potential problem of undetected hidden value assumptions, Nagel highlights the self-corrective mechanisms of science as a social inquiry, preceding Helen Longino (1990).

[7] Though different from Alexandrova or the Neo-Nagelians mentioned above, he takes it that this is descriptively true for social scientific inquiries and is neither a methodological required nor epistemically desirable.

> "Curiously, conditionalizing on normative assumptions is exactly the strategy favored by Douglas (2009, p. 153), herself. It should be noted that such conditional scientific results comply with the value free ideal, because once uncertain or (non-epistemic) normative assumptions are placed in the antecedent of a hypothesis, they are clearly not maintained by the scientist anymore" (Betz, 2013, p. 213).

Betz seems to presuppose that conditionalizing on normative assumptions clearly eliminates the possibility of scientists' ownership of values. Conditionalizing mainly helps the expert to single out the normative judgment and separate it from the descriptive judgment being made. However, doing so does not eliminate the possibility that the expert may indeed maintain or endorse the normative assumption in question. More generally, disowning a value-judgment is different from distancing oneself from it. Of course, the conditionalization strategy is especially convenient *if* the scientist does not endorse the normative assumption placed in a hypothesis's antecedent. If she does endorse the assumption, she may employ the conditionalization strategy to operationalize the impartiality of her judgment to avoid any possible biases in the ways she reaches her judgments or communicates them.

## 4.7. Conclusion

The problem posed by mixed claims is that when social and behavioral scientific claims are inflicted with potentially unwarranted value presuppositions (as they often are), their objectivity may be compromised. Moreover, the communication of mixed claims for deliberative policy purposes is challenged in a broadly democratic paradigm. Policymakers and the people at the receiving end of policymaking may simply disagree with social and behavioral scientists' values on which mixed scientific claims rest. Unless properly communicated or controlled, it seems as if scientists unduly impose unexamined value presuppositions to the policy deliberation.

The received Nagelian strategy to address these problems recommends that the value presuppositions be disentangled from the mixed claims, and that social and behavioral scientists should conditionalize their claims on explicitly stated value presuppositions.

The contrasting strategy, recently articulated by Alexandrova, denies the tenability and desirability of the Nagelian approach and suggests that social and behavioral scientists should instead legitimize their value presuppositions through proper means of social control similar to the deliberative democratic procedures in the political realm.

I have argued that the contrast between these two approaches is not as strong as it initially looks. I have demonstrated that Alexandrova's critique of the Nagelian way to deal with mixed claims can be qualified as a critique of Nagel's approach as an account that is used for defining the cogent sources and the content of value presuppositions inserted into mixed claims. Thus understood, the Nagelian treatment of mixed claims as conditional statements, as a recommendation for fulfilling mixed claims' impartiality, remains intact. Moreover, I have argued that the Nagelian approaches are compatible with Alexandrova's methodological responses to the problem of defining the content of value assumptions in mixed claims. I have also illustrated that neo-Nagelian accounts of social scientific objectivity (such as Douglas's) supports the reconciliation I have offered.

# References

Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford University Press https://doi.org/10.1093/oso/9780199300518.001.0001

Alexandrova, A. (2018). Can the science of well-being be objective? *British Journal for the Philosophy of Science*, *69*(2), 421–445. https://doi.org/10.1093/bjps/axw027

Anderson, E. (2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia*, *19*(1), 1–24. https://doi.org/10.1111/j.1527-2001.2004.tb01266.x

Banerjee, A. V., & Duflo, E. (2020). *Good economics for hard times: better answers to our biggest problems*. Penguin Books.

Barton, A., & Grüne-Yanoff, T. (2015). From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, *6*(3), 341–359. https://doi.org/10.1007/s13164-015-0268-x

Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, *3*(2), 207–220. https://doi.org/10.1007/s13194-012-0062-x

Brennan, J. (2017). *Against democracy*. Princeton University Press.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.

Douglas, H. (2003). Hempelian insights for feminism. In S. Clough (Eds.), *Siblings under the skin: Feminism, social justice, and analytic philosophy*. The Davies Group Publishers.

Douglas, H. (2004). The Irreducible Complexity of Objectivity. *Synthese*, *138*(3), 453–473. https://doi.org/10.1023/B:SYNT.0000016451.18182.91

Douglas, H. (2009). Science, policy, and the value-free ideal. University of Pittsburgh Press. https://doi.org/10.1177/0094306110367909q

Douglas, H. (2011). Facts, values, and objectivity. In I. Jarvie & J. Zamora-Bonilla (Eds.), *The SAGE handbook of the philosophy of social sciences* (pp. 513–529). SAGE Publications Ltd. https://doi.org/10.4135/9781473913868.n28

Douglas, H. (2015). Values in science. In P. Humphreys (Eds.), *Oxford handbook of philosophy of science* (pp. 610-630). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199368815.013.28

Heilmann, C. (2014). Success conditions for nudges: A methodological critique of libertarian paternalism. *European Journal for Philosophy of Science*, *4*(1) 75–94. https://doi.org/10.1007/s13194-013-0076-z

Jeffrey, A. (2018). Limited epistocracy and political inclusion. *Episteme*, *15*(4), 412-432. doi:10.1017/epi.2017.8

John, S. (2019). Science, truth and dictatorship: Wishful thinking or wishful speaking? *Studies in History and Philosophy of Science Part A*, *78*, 64–72. https://doi.org/10.1016/j.shpsa.2018.12.003

Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.

Koskinen, I. (2020). Defending a risk account of scientific objectivity. *The British Journal for the Philosophy of Science*, *71*(4), 1187–1207. https://doi.org/10.1093/bjps/axy053

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Mongin, P. (2006). Value judgments and value neutrality in economics. *Economica*, *73*(290), 257–286.

Nagel, E. (1961). *The structure of science: problems in the logic of scientific explanation*. Harcourt, Brace & World, Inc.

Reiss, J. (2017). Fact-value entanglement in positive economics. *Journal of Economic Methodology*, *24*(2), 134–149. https://doi.org/10.1080/1350178X.2017.1309749

Reiss, J. (2019). Expertise, agreement, and the nature of social scientific facts or: Against epistocracy. *Social Epistemology*, *33*(2), 183–192. https://doi.org/10.1080/02691728.2019.1577513

Rolin, K. H. (2020). Objectivity, trust and social responsibility. *Synthese*. https://doi.org/10.1007/s11229-020-02669-1

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (Eds.). (2010). *Mismeasuring our lives: Why GDP doesn't add up*. New Press.

Su, H., & Colander, D. (2013). A failure to communicate: The fact-value divide and the Putnam-Dasgupta debate. *Erasmus Journal for Philosophy and Economics*, *6*(2), 1–23. https://doi.org/10.23941/ejpe.v6i2.131

Weber, M. (1949). *The methodology of the social sciences*. Free Press.

# Chapter 5

# Nudges, Well Ordered

## 5.1. Introduction

More than ten years have passed since Sunstein and Thaler have published their popular book *Nudge* (2008). The book famously proposed and defended a specific policy intervention to change people's behavior called "nudge". A nudge is a sophisticated intervention on choice-environments that aims at enabling people to make better choices. Nudges make simple and small changes in the way choices are presented to choosers. Therefore, they are often considered "softer" than conventional paternalistic policies that work by shifting people's incentives or limiting the number of choices available to decision-makers. Instead, typical nudges, such as alternating or setting defaults or rearranging the ordering of the foods in a cafeteria, work by harnessing scientifically studied cognitive heuristics and biases people employ when making decisions.

Since the beginning of the nudge agenda, many governments have gradually adopted the nudges in a wide range of public policy areas, including health, finance, education, and environmental protection (OECD, 2017). Many "nudge units" or "behavioral insights teams" have been founded across the globe hosting evidence based behavioral policy specialists dedicated to advising governments in evaluating and implementing nudge-like policies (Sanders et al., 2018). In the meantime, philosophers, scientists, public policy scholars, and proponents of nudge have been discussing the justifiability of nudges (see Barton & Grüne-Yanoff, 2015 for a review). The philosophical debates are centered around two main themes: the ethical aspects of nudges and the epistemic conditions for nudge success. The main question regarding the ethics of nudge is whether these policies can be motivated as "libertarian paternalist" interventions instead of classical paternalist ones (e.g. Thaler & Sunstein, 2008; Sunstein, 2015; Dworkin, 2020). In relation to that, philosophers have questioned whether nudges circumvent citizens' freedom of choice and autonomy or even count as unjustifiably manipulative (e.g. Bovens, 2009; Hausman & Welch, 2010, Wilkinson, 2013; Waldron, 2014; Sunstein, 2015). On the epistemic side, philosophers have investigated the evidential underpinning of nudges and methodological requirements for implementing and evaluating them as evidence-based policies, asking how one knows that a behavioral policy is effective or how these policies should be evaluated concerning different policy desiderata and in the face of the ethical challenges (Heilmann, 2014; Barton & Grüne-Yanoff, 2015, Grüne-Yanoff, 2016; Grüne-Yanoff et al., 2018; Dede, 2019; Malecka, 2021).

In this chapter, I focus on a broader normative question regarding the justifiability of nudges as evidence based policies, asking in how far nudges can be argued to be consistent with the values of autonomous democratic self-government. Given that nudges are evidence-based policies and hence legitimized through behavioral policy experts' evaluations, it is an open question whether the normative ends of these policies can be sufficiently scrutinized in the light of the citizens' values and interests. I propose a perspective to address this question by offering a conception of nudging called *Well Ordered Nudge*, drawing on the philosophical literature on the democratization of science-based policies (e.g. Maasen & Weingart, 2005; Kitcher,

2011; Alexandrova, 2018; Rolin, 2020; Douglas, 2005, 2009, 2021; Kurtulmuş 2021; Irzık & Kurtulmuş, 2021). Well Ordered Nudge specifies how nudges, qua evidence-based policies, can respect citizens' autonomy in guiding policy decisions regarding their behavior and conceptualizes how nudges can be made congruent with democratic self-government. I show that Well Ordered Nudge is different from the extant philosophical accounts of nudges, and it advances them with respect to the problem of reconciling nudges with democratic governance.

The chapter is structured as follows. In Section 2, I discuss why typical nudges are considered to be ethically controversial in relation to autonomy and democratic self-government and motivate my proposal to "well order" nudges to address these controversies. In Section 3, I propose the Well Ordered Nudge approach, discuss its relevant characteristics and argue that it puts a solution to the table concerning nudge's stance in relation to democratic governance by rooting public's democratic control of nudges in scientifically informed deliberative democratic practices. In Section 4, I compare the Well Ordered Nudge with some of the prominent extant conceptions of nudging that offer a spectrum of democratic perspectives over nudges and highlight the advantages of following the idea of Well Ordered Nudge. In Section 5, I conclude with a brief summary of the article.

## 5.2. Invisible nudges and their acceptability

Nudges are purposeful policy interventions on choice environments that aim to change people's behavior in ways that benefit them (e.g., adopting healthier diets, choosing optimum saving plans). Unlike traditional paternalist policies, nudges do not eliminate any options and leave monetary incentives intact. Nudges instead trigger scientifically studied decision-making mechanisms such as cognitive biases and heuristics to steer people's choices in directions that they would find desirable upon reflection.

To ease ourselves into the relevant ethical aspects of nudging, let us focus on a paradigmatic example. Consider setting a default option or alternating existing ones.

Administering smart default rules is one of the most famous examples of nudging, and it has been implemented in a wide range of policy contexts, including environmental protection, organ donation, personal financial decision-making. A policy designer sets a default option that she believes to the best option for the choosers. Defaults *work* when the choosers do not opt-out of the choice they are defaulted into or select another option among the menu of alternative options if the default is not the best option for them. Defaults are known to be highly effective in changing people's decisions (see Johnson & Goldstein, 2003 for a seminal review). This way of changing people's choices counts as sufficiently "libertarian" because it does not mandate a choice, nor does it (dis-)incentivizes it. It also does not eliminate any alternative options; hence the freedom of choice seems to be intact. Defaults also nicely instantiate the paternalistic aspect of the nudge policies as they do not leave the choosers fully reflective in making decisions and operate through less-than-fully-rational decision-making mechanisms. Unlike, say, an active choice context, where the choosers have to deliberate over which options are preferable to them and why, defaults usually work when people do not deliberate and let some cognitive heuristics and biases lead their choices. For instance, choosers may take defaults to be an expert's recommendation without question; they may stick with the options they are defaulted into due to inertia or procrastination; they may simply avoid the cognitive effort of decision-making; they make take the defaults as a reference point leading them to biasedly evaluate alternative options as loss or gain (and they tend to stick with the defaults when choosing an alternative is perceived as a loss); they might feel guilt when the alternatives to the default option are perceived as the desirable social norm. (see Grüne-Yanoff, 2016 for a review of distinctive psychological mechanisms that make defaults work).

This particular kind of paternalism nudges promote, which relies on unreflective psychological mechanisms of decision-making (i.e. the cognitive biases and heuristics people employ when making choices), has been the focus of ethical controversies over nudging. The main concern has to do the with the fact that these mechanisms often work "in the dark" (Bovens, 2009). That is, they are not detected or acknowledged by the people who are being nudged. Indeed, subtle nudges such as defaults are often not

even perceived by ordinary people as purposeful policy interventions that aim to change their behavior, and defaults are one of such invisible forms of government policy. The invisibility and undetectability of the nudges raise a set of controversial questions about the nudge's relation to agency, autonomy, and democratic self-government (see Saghai, 2013 for an analysis). Consider the following pressing questions: When a decision one makes is not sufficiently reflected over, could one say that people remain the authority behind their choices? (e.g. Bovens, 2009) Are these kinds of interventions to people's behavior unacceptably manipulative? (e.g. Wilkinson, 2013) If choosers do not sufficiently scrutinize the intervention and its aims, how can we know that the results promote people's well-being from a subjectivist point of view? Are these kinds of interventions and their mechanisms of operation transparent and visible enough to legitimize through democratic means?[1]

A pressing problem that amplifies the worries articulated by these questions is the asymmetry of information between those who nudge and those who are being nudged (e.g. Reijula & Hertwig, 2020 for an introduction to this problem). While the mechanisms behind nudges are thoroughly investigated by the behavioral scientists and utilized by the policy designers (also known as the "choice architects" in the nudge literature), they are not accessible to or acknowledged by the uninformed citizens who are being nudged. This asymmetry of information, when left unaddressed, makes it harder to defend and scrutinize the claims in defense of the ethical acceptability of nudging. Moreover, when nudges are invisible, undetectable, and unsuitable for deliberative scrutiny, they are more likely to be questioned and

[1] It must be noted that not all nudges and nudge-like policies operate in the same ways, and some nudges embody different ethical characteristics concerning the way through which they attempt to change people's behaviors. For instance, some behavioral interventions target at enhancing people's decision-making repertoire by teaching them new heuristics (called as "boosts"); and some do not override people's active decision-making facilities and target at steering people towards further deliberation about their decisions ("sometimes called as "system 2 nudges" or educative nudges). The ethical concerns about the visibility, transparency, and democratic accountability of nudging specifically apply to those nudges that are less detectable and that target at harnessing people's cognitive biases and heuristics (sometimes called as "system-1 nudges"). For a detailed typology of nudging with respect to the dual process framework see Heilmann (2014).

reacted on the grounds that they may be serving for illicit policy goals or normative ends that may not be unambiguously aligned with some citizens' real interests, needs, and preferences.

A promising way to address this problem would be to promote the public's democratic control over nudge policies by providing the public with more transparent and detailed information about nudges so as to receive their democratic endorsement of these kinds of policies. In the following, I will propose a conception of nudges, called "Well Ordered Nudge", that promotes the public's democratic control of nudges from a perspective of philosophy of science that focuses on evidence based policies' congruence with democratic governance (Kitcher, 2011; Douglas, 2021).

## 5.3. Well Ordered Nudge

In this section, I will first introduce the philosophical motivation behind the idea of well ordered nudge: seeking a proper balance between evidence based justification of nudges with the democratic control of these policies. I will then describe the main features of the proposal.

### 5.3.1. Evidence based policy evaluation and the democratic control of policies

Any broadly liberal democratic framework for policy evaluation requires that every individual who is influenced by a policy intervention should be able to scrutinize and exercise control over the ends of the intervention in the light of her interests, needs, and values. In this way, the members of the public (the receivers of the policies) are supposed to be the authority behind the interventions that affect their lives and behaviors. Call this *the requirement of democratic self-government*.

In the context of evidence based policies such as nudges, we could think that the democratic requirement for policymaking is fulfilled indirectly: the governmental units (who are supposed to be democratically elected representatives of the public)

communicate the public's policy-preferences with the scientific experts, and the experts take these preferences into account in evaluating policy proposals based on evidence. However, in practice, evidence based policies are not always scrutinized in this way. Paternalistic policies such as nudges, for instance, are often designed and evaluated based on scientific experts' own normative presuppositions and knowledge about what the needs, interests, and values of people in a given community are and what policy interventions would best serve these needs, interests, and values. Moreover, evidence based policy analysis alone does not fully justify the policies in the face of ethical and political uncertainties and disagreements (see Ongaro & Andreoletti, *forthcoming* for a recent overview). When scientific experts may not explicitly attend to making such judgments themselves, resolving some of the methodological questions they face (such as how to design an experiment, which methods to use for policy evaluation, what evidence counts as relevant, how much evidence is required to confirm a hypothesis) can put them in a position to inform the making of such normative judgments wittingly or unwittingly (see Douglas, 2009; Alexandrova, 2018; Reiss, 2019 among others for philosophical overviews on the normative aspects of evidence based policy evaluation).

In the light of this understanding of how scientists contribute to normative matters regarding informed policy decisions, evidence based policy's relation to democracy is also analyzed by philosophers of science. The value-ladenness of evidence based policy evaluations is argued to stand as a challenge to democratic self-governance (e.g. Kitcher, 2011; Alexandrova, 2018; Douglas, 2021). When scientists' judgments of policies are not free from value considerations, and the policymakers primarily rely on scientists' conclusions to justify policy proposals, the *requirement of democratic self-government* is not fully met. Scientific experts guide and influence the normative considerations regarding the values, interests, needs of the public. This condition challenges the requirement of democratic self-governance, as the people who are at the receiving end of the policies are not in a position to be the primary normative authority behind the interventions that affect their lives. This challenge is more pressing in the context of nudges because typical nudges are prone to being

undetectable by their receivers, thus less controllable, and they are suspected of undermining people's agency and autonomy.

In the face of this challenge, one of the favorite remedies is to call for democratizing the justification of evidence based policies: scientific experts should inform the public more, and the public should have more control over the normative choices regarding evidence based policy justification (Douglas, 2005, 2009; Kitcher, 2011; Alexandrova, 2018; Irzık & Kurtulmuş, 2018; Rolin, 2020; Kurtulmuş, 2021). This requires facilitating appropriate avenues and procedures for citizens to participate in the processes of informed policy justification. In the following, I will propose a conception of nudging that is justifiable through such a democratized approach to scientifically informed policy evaluation.

To develop this conception, I will mainly draw on Philip Kitcher's approach as the latter defines a suitable account of the public's democratic participation in science and describes the main desirable characteristics of the mechanisms that enable citizens to join scientists to scrutinize policies (Kitcher, 2011). Kitcher's approach to science in a democratic society is known as the "well ordered science," and it also gives rise to the label of the conception of nudge I propose: Well Ordered Nudge.

### 5.3.2. Well Ordered Nudge

Well Ordered Nudge is a conception of how nudges can be justified as evidence based policies that promote the requirement of democratic self-government. Consequently, it aims to offer a perspective that reconciles two desiderata for successful nudges that are apparently in tension: their reliance on scientific experts who design and evaluate nudges and their congruence with democratic control in ways that promote autonomous behavioral changes. Following Kitcher's 2011 "well ordered science" framework, Well Ordered Nudge seeks to achieve this through enlisting institutionalized mechanisms (such as citizen juries, deliberative mini-publics) in which the relevant scientific experts and the representatives of policy-receiving public deliberate over the evaluation of nudges.

These deliberative mechanisms' primary role is to facilitate an exchange of information between scientists and the members of the public. On the one hand, scientists are supposed to collect citizens' normative perspectives about behavioral policies (their needs, values, preferences, and interests; and how acceptable they find different kinds of nudges) to inform the design and evaluation of nudges. On the other hand, the members of the public are supposed to be tutored by the scientific experts about the relevant evidence, the information about how nudges operate, deliberate with citizens about their policy preferences and value perspectives. The result of such a deliberative process is the public's informed democratic endorsement of particular tokens of nudges.

Note that *all* types of behavioral policies can be scrutinized through such informed deliberative democratic procedures. This implies that Well Ordered Nudge is not a particular kind of behavioral policy, and so the conception does not constrain the set of justifiable behavioral policies. That is to say, the idea is not that only those kinds of nudges that are sufficiently transparent and conducive to deliberative scrutiny (e.g. "boosts", "self-nudging", "active choice", "educative nudges", "system 2 nudges") are suitable for public's scrutiny. The typical, less detectable nudges such as setting default options that operate through triggering people's less-than-rational, non-deliberative decision facilities can also be scrutinized through deliberative democratic procedures, as these procedures' aim is to inform the public's opinion about also these kinds of interventions. Indeed, Well Ordered Nudge's motivation is to make these kinds of otherwise invisible nudges sufficiently controllable by an informed public so as to promote policy-recipients' autonomy in guiding the courses of behavioral changes.

Let's now focus on the relevant characteristics and the principles of the two-sided deliberation among experts and the public to describe the principles of Well Ordered Nudge.

#### 5.3.2.1. Normative authority behind well ordered nudges

The Kitcherian approach to normative authority (that is, the authority to decide what ends are right, what policy interventions are desirable, or more broadly, conceptions of the good and the right) presupposes that there is no normative expertise (Kitcher, 2011, p.112). According to Kitcher, the conceptions of the right and the good should ideally be determined through an ideal conversation where the representatives of all kinds of value perspectives and interests participate in a mutually engaging and scientifically informed deliberation. The outcome of such a conversation would not be known by any agents prior to deliberation, as such way of determining normative goods is describable as a "multidimensional balancing act" on which there is no expert (because "no single person, not an enlightened philosopher or a religious teacher has the answer" (Kitcher 2011, p. 112). This means that "normative authority" or "normative superiority" is not a characteristic that is plausibly ascribable to the members of the public, or scientists, or philosophers. Rather, normative authority comes from the *quality of the conversation*. For Kitcher, the closer the deliberations about normative issues gets to his ideal, the more authoritative they are. This broad principle of normative expertise informs Kitcher's "well ordered science": He suggests that the conceptions of what normative ends the pragmatic use of science should serve should be decided in deliberative democratic settings that approximate his "ideal conversation" as much as possible.

Kitcher's well ordered science approach thereby provides us with the key insight to conceptualize Well Ordered Nudges: none of the relevant agents (i.e. scientists who are evaluating or designing nudges, governmental units, or the members of the public) should be considered as the normatively authoritative agent who determines the ends of the behavioral interventions. Rather, what comes as the outcome of a properly structured deliberation among these agents should count as normatively authoritative. This premise has an important implication for the *requirement of democratic self-government*. For the public's control of behavioral policies to count normatively authoritative, it is not sufficient for scientists to simply collect citizens' normative

perspectives to inform the design and evaluation of nudges. Well Ordered Nudge calls for a more refined and demanding conception of democratic self-government that is articulated by Kitcher's ideal: public's normative input for the evaluation of nudges should be the product of an informed deliberation among a heterogeneous group of people that represent values, needs, and interests of diverse groups in public. Moreover, the scientists should tutor (or perhaps also moderate) this deliberation so that individuals' normative perspectives are sufficiently informed and challenged based on relevant pieces of evidence and the considerations of the needs, values, and interests of other members of the public. How could this idea be instantiated in practice?

#### 5.3.2.2. Mechanisms of well ordered nudges

The traditional democratic mechanisms such as voting or public surveys are not able to deliver the kind of normative input that Well Ordered Nudge requires because they yield results that reflect the public's insufficiently informed and non-deliberate opinions. Indeed, Sunstein (2016, pp. 116-158) extensively reviews the surveys that reflect people's opinions about nudges, and the results are not conducive to a cogent defense of citizens' democratic control over nudges: citizens tend to assess policies in politically biased ways without sufficient acknowledgment of the relevant facts. Their preferences might be too self-directed (e.g. formed without considering the relevant other people's preferences or relevant externalities or minority group's needs or interests). The public opinion surveys are also very much vulnerable to framing effects or unduly influence of more powerful agents (e.g. commercial interest groups).

Alternatively, as Kitcher proposes, the contemporary deliberative democratic mechanisms such as citizen juries, deliberative polling, mini-publics, consensus conferences are apt for the task precisely because the opinion surveys and majority voting have the kind of limitations that Sunstein reviews (see Brown, 2006; Mansbridge et al., 2010; Fishkin, 2018 for relevant background literature on deliberative democratic practices; and see Douglas, 2005; Kitcher, 2011; Alexandrova, 2018; Steel et al., 2018; Rolin, 2020 among others for applications in

the philosophy of science). A deliberative democratic mechanism for evidence based policy analysis aims to elicit citizens' *tutored* normative opinions and encourage deliberators to reflect on non-self-serving (other-regarding) considerations. Consider, for instance, the practice of mini-publics for evidence based policy evaluation.

Mini-publics are participated by (typically randomly selected) members of the public who are representative of the different socio-economic groups and stakeholders relevant to a policy question. The participants of the mini-publics are tutored by scientists about all the relevant evidence, open normative questions, points of disagreements and consensus among experts (e.g. why a given behavioral policy is needed, what alternative policy options are available, how it works, what are people's relevant psychological propensities, what the ethical and practical challenges are, and the costs and benefits for the community or for different groups). Citizens are encouraged to share their perspectives and concerns about the scientists' brief, and they deliberate together with experts about the relevant open questions regarding the policy evaluation or design. The deliberation is supposed to be properly moderated so as to reach some consensual decisions at the end. For instance, irrelevant or scientifically ungrounded ideas can be dismissed by the moderator. Similarly, participants can be encouraged to reflect on the alternative points of view or arguments that are not obvious to them prior to deliberation. The discussion can also be moderated in ways that avoid more powerful agents (such as interest groups) to hijack or frame the debate in ways that serve them.

#### 5.3.2.3. The outcome of well ordered nudges

Well Ordered Nudge (and its conception of the public's democratic control of nudges) rests on the normative input resulted from deliberative democratic mechanisms such as mini-publics. When this kind of input informs scientific experts' evaluation and design of nudges, nudges deserve the label "well ordered". Nudges that are scrutinized through deliberative democratic mechanisms envisioned by the Well Ordered Nudge conception promote democratic self-government of nudges as evidence based policies. In this way, the people at the receiving end of nudges can be said to retain their autonomy as they sufficiently control the agenda of the behavioral interventions

that influence their lives. They also do so by reflecting on their values, needs, and interests in the light of the essential epistemic insights provided by scientists that they otherwise may not have access to.

I will now revisit the extant alternative philosophical conceptions of how nudges conform to the public's autonomy and democratic self-government from the viewpoint of Well Ordered Nudge. In doing so, my aim is to make the Well Ordered Nudge's stance more precise and highlight some of the advantages of following the well ordered nudge idea for the ongoing discussions over the justifiability of nudges.

## 5.4. Alternative Conceptions of Nudges in comparison to Well Ordered Nudge

In this section, I will review the current philosophical discussions of nudges that rely on alternative conceptions of nudges' congruence with autonomous self-government. I will discuss how Well Ordered Nudge differs from these alternatives, which offer different conceptions of how nudges can be justified. I will highlight in what ways Well Ordered Nudge differs from them and discuss some of the advantages of following the Well Ordered Nudge idea. Though I do not rely on any empirical claims over the potential advantages and disadvantages of my proposal, the proceeding analysis can also be useful for comparative empirical analyses of Well Ordered Nudges in comparison to its alternatives (e.g. analysis of whether people find well ordered nudge more acceptable, or how effective they are in comparison to alternative nudges).

### 5.4.1. "Technocratic" Nudge

The founding fathers of nudge, Sunstein and Thaler, propose that nudges make people better off, *as judged by themselves* (Thaler & Sunstein, 2003, Thaler & Sunstein, 2008). The phrase "as judged by themselves" captures the idea that nudges are supposed to lead behavior changes that the people who are nudged would have

endorsed. Indeed, the nudges' reliance on the subjective conceptions of well-being enables Thaler and Sunstein to motivate nudges as *libertarian* paternalist interventions in contrast to the traditional paternalist ones that rely on the welfare analysts' conceptions of the good and desirable behaviors. The nudge agenda's aspiration to be the subjectivist about well-being is, however, in tension with the empirical findings that the nudges are based on.

The scientific evidence on people's judgment formation regarding the matters of their own well-being demonstrates that people are commonly misinformed and cognitively biased evaluators of their well-being (see Heilmann 2014; Barton & Grüne-Yanoff 2015; Grüne-Yanoff et al., 2018 for detailed accounts of the theories of human action that the nudge agenda draws on; the so-called dual systems theories of decision making and the "heuristics and biases" program of behavioral economics). In response to this challenge, the subjectivist proponents of nudges rely on reconstructive, laundered notions of people's subjective preferences; that is, those preferences which are, at least hypothetically, rational and free from cognitive biases and based on full relevant information. Accordingly, Sunstein and Thaler's version of "as judged themselves" is famously followed by the qualification "*had they been free from all cognitive biases, had access to all relevant information*" (Thaler & Sunstein, 2008; see Sugden, 2018 for a critical review).

This way of conceptualizing nudge's normative success is essentially concerned with the *authenticity* of the preferences derived in this way. By authenticity, I mean a measure of closeness between the laundered preferences identified by scientific experts and the *real* preferences nudgees *themselves* would form had they been fully informed. The closer these two are to each other, the more authentic the laundered preferences are. If nudges are to be motivated with subjectivism about well-being, then the laundered preferences experts reconstruct must not be *alien* to the nudgees as autonomous decision-makers.

Now, when the nudges are supposed to change people's behaviors towards directions that their actual self may not endorse but an informed and hypothetically unbiased version of them do, a great deal of normative justification depends on the welfare

analyst's success who determines people's laundered, ideal, unbiased subjective preferences. When the welfare experts fail to elicit the correct laundered preferences that correspond to the targeted individuals' actual informed and debiased preferences, the subjectivist aspiration of the nudge agenda loses its ground. The welfare expert faces the risk of substituting her own normative conceptions of the good (which may be also informed and shaped by the theoretical presuppositions of behavioral economics) with the people's conceptions' the good, and the nudges fall back into being instances of traditional paternalist policies justified on non-subjectivist grounds.[2]

Conceptualized in this way, the normative success of nudges (that is, whether they effectively serve targeted individuals' needs, interests, and values) becomes a primarily *technocratic* issue as opposed to a democratically determined one. The crucial normative questions about nudges' justifiability are resolved by the welfare experts, and this activity is successful when experts are able to measure people's preferences in the right way.

Compare this with the conceptualization of the nudge's normative success that Well Order Nudge articulates. Remember Well Ordered Nudge's normative authority rests on the people's tutored preferences which are products of democratic argumentation

[2] Indeed, the debates about the justifiability of nudging on libertarian or subjectivist grounds have naturally questioned Sunstein-Thaler's argument's normative and epistemic cogency (e.g. Rizzo & Whitman, 2009; Grüne-Yanoff, 2012; Qizilbash, 2012; Guala & Mittoene, 2016; Sugden, 2018b, 2021; Hausman, 2019). Setting aside the benign cases where the relevant welfare analysis is straightforward or plausibly unobjectionable on subjectivist grounds (Sunstein 2016, p. 115), the skeptics argue that the welfare analysts are not in a good position to determine people's laundered preferences in such a way that these preferences are indeed welfare-inducing, and the welfare analyst does not reconstruct people's preferences without replacing her conceptions of the good with the people's (see Rizzo & Whitman, 2009; Infante, Lecouteaux, & Sugden, 2016 for earlier formulations of this argument, see Sugden, 2021 for a recent version of this argument which is now called as the "knowledge problem" in behavioral welfare economics). This argument does not only consider the epistemic limitations of the nudge agenda but also highlights that nudges cannot be justified as genuinely libertarian policies. When the laundered preferences do not reflect nudgee's authentic preferences, successful nudges that rely on them will not preserve people's autonomy (see, for instance, Grüne-Yanoff, 2009; Qizilbash, 2012; Guala & Mittione, 2016, Sugden, 2018; Hausman, 2019). Moreover, there are deeper conceptual and methodological problems regarding what counts as successful and justifiable laundering of preferences.

and deliberation, not a state to be discovered or reconstructed by a welfare expert. (Recall Kitcher's conception of normative authority!) Well Ordered Nudge does not rely on the welfare analysts' hypothesized projections of people's informed conceptions of the good. It relies on real people's informed conceptions of the good *when* they have access to relevant information and are presented with evidence regarding their biases in evaluating their preferences (*not* had they been free from all cognitive biases, had access to all relevant information). For this reason, the criticisms of the Technocratic Nudge (as listed in footnote 2), which are based on the normative and epistemic success of experts' reconstruction of preferences, are not applicable to Well Ordered Nudge. Well Ordered Nudge is not directly targeted by these critiques as the former relies on a different *domain of normative justification* than the latter.

Note that the criticisms of the technocratic conception of nudges may still be relevant to evaluate the success of Well Ordered Nudge in eliciting people's authentic preferences as the Technocratic and Well Ordered Nudge conceptions may be both needed or potentially complement each other in practice to elicit people's preferences to design nudges (though I do not make any claims regarding this). However, my point here is to make a comparison between the Technocratic and Well Ordered conceptions of nudge. Their respective domain of normative justification is the most salient and important difference-maker between the two conceptions, and the criticisms of the Technocratic Nudge are useful to highlight this difference.

### 5.4.2. Autonomy-Preserving Nudge

Another prominent argument in favor of the justifiability of nudging is that nudges retain people's power to control the nudge by virtue of nudge's minimally interventionist nature (see Sunstein, 2015 for a review). The idea is that even though nudges may sometimes steer people towards behaviors that do not align with their preferences, these policies leave alternative options intact and that people could opt-out or reverse their nudged position when they do not endorse the choices they are nudged into. So, the proponents argue that nudges do not limit alternative choices and are easily reversible, and so people remain to be the authority behind their choices when they are nudged (e.g. Thaler & Sunstein, 2008).

This particular defense of nudging poses a set of desiderata for the successful implementation of these policies in practice. For instance, a nudge does not restrict or ban the set of choices available to the nudgees. Similarly, a nudge preserves the incentive structure in the choice environment. For instance, no negative taxing is involved in nudges that discourage unhealthy food. Finally, a nudge is supposed to give the nudgees a chance to opt-out easily. The commentators in the literature have discussed whether and how nudges can fulfill these success criteria in practice (e.g. Bovens, 2009; Wilkinson, 2013; Saghai, 2013; Heilmann, 2014). For this to work, nudges should be transparent and detectable (Bovens, 2009; Saghai, 2013), and the individuals who are being nudged should have a genuine chance to reverse and control them (e.g. the option of opting out should be "genuine") (Heilmann, 2014).

This cautious attitude regarding the proper implementation and design of a nudge motivates a slightly different conception of nudging than the technocratic one. By emphasizing the importance of designing and implementing nudges in autonomy-preserving ways, this conception of nudges acknowledges the limitations of a primarily technocratic justification of nudges. Call this conception of nudge the "Autonomy-Preserving Nudge".

According to the Autonomy-Preserving Nudge, the normative success of nudges (that is whether they effectively serve for targeted individuals' needs, interests, and values) derives not only from welfare experts' success in determining the right normative criteria but also from designing and implementing nudges in ways that retain people's capacity to scrutinize and control policy interventions: When a nudge fails to serve for one's interests or needs, people should be able to resist them and opt out of their nudged positions.

Autonomy-Preserving Nudge articulates a less technocratic understanding of the normative authority behind nudges as it rests on a number of operations and methodological procedures that limit the normative power of experts over the people who are being nudged in guiding nudges. Policy-recipient individuals are supposed to determine whether nudges serve them well in the light of their interests, values, and

needs and should be able to do so by virtue of the soft, autonomy-preserving character of nudge interventions. Conceptualized in this way, a lot of normative justification for nudges hinges *on the successful implementation of nudges* in ways that meet the requirements of transparency, detectability, salient/genuine presentability of chances to opt-out of the intervention such that people can *actually* steer nudges and sufficiently scrutinize and control their ends.

The Autonomy-Preserving Nudge approximates Well Ordered Nudge conception to the extent that the conditions for the successful implementation of nudges are understood to require extensive, institutionalized interactions between the nudge-designing experts and the people who are being nudged. Consider the transparency condition. The transparent implementation of nudges requires the policy designer to disclose information about a nudge so that the targeted individuals can perceive it as a purposeful policy intervention and consider the option of resisting it in the light of their policy preferences and interests. Notice however that depending on how democratically a nudge will be controlled, the information disclosure procedures would change.

A procedure of information disclosure that is almost as democratized as the information-sharing practices described in Well Ordered Nudge would look like the following: the nudge-designing experts could extensively inform the targeted individuals about the nudge (explaining them the aims of the nudge, the relevant scientific evidence, the psychological mechanisms through which the intervention works, and highlighting them that they could opt out). Now, compare this with the examples from the extant empirical literature that investigates the transparency condition. In these examples, scientists typically disclose little information about nudges and information disclosure does not presuppose any deliberative exchanges between the policy designing experts and people being nudged (see Loewenstein et al., 2015; Arad & Rubeinstein, 2018 as the most cited examples, and see Sunstein, 2016, pp. 153-158 for a review). Rather, the information disclosed about nudges appears as a warning such as "you are being nudged" without any reflection-triggering explanations about the normative aspects of the nudge in question.

Accordingly, while the domain of normative justification overlaps much more than with Technocratic Nudge, Well Ordered Nudge differs from the Autonomy-Preserving Nudge with respect to the kind of mechanisms they rely on to preserve people's autonomy.

### 5.4.3. Democratic Nudge

Some political philosophers (Nys & Engelen, 2017; Schmidt, 2017) propose an alternative conception of nudge, the justification of which is based on democratic practices, including the deliberative democratic practices favored by Well Ordered Nudge. Nys & Engelen (2017) argue that people at the receiving end of the nudges are the ultimate holders of the decisions concerning the normative ends of nudges under a well-functioning democratic system. As nudges are policies implemented by democratically authorized institutions, people who are being nudged can always punish governments when their interests and preferences are not served, and so the policymakers have an incentive to implement nudges that work best for their recipients (see also Schubert, 2017 for discussions about the political implications of this incentive structure for policy makers).

Accordingly, these commentators believe that the normative ends of nudges are determined through people's democratic consent and control: "What 'they' do, 'they' do in 'our' name and because 'we' enable 'them' to do it." (Nys & Engelen, 2017, p. 208). In the same vein, Schmidt (2017) argues that the republican and relational conceptions of autonomy in political philosophy motivate this Democratic Nudge conception: nudged people's autonomy can be empowered under well-functioning democratic institutions as people can enlist nudges to expand the realm of public control over behavior change which is otherwise controlled more by private companies or arbitrary power of unchecked governments.

The Democratic Nudge conception is in line with the Well Ordered Nudge approach in two key respects: firstly, the domain of normative justification is similar, and secondly, both approaches aim at a democratic justification. The main difference

between them is a significant one, however. These philosophers' conceptualization of nudge underestimates or else disregards the epistemic requirements of democratic control emphasized by the Well Ordered Nudge. An important premise of the Well Ordered Nudge is that people's democratic control of policies is shaped by the kind of knowledge they have about the policies (see Kurtulmuş, 2021 for a brief review of this argument). The scientific input informs (and is supposed to change) people's preferences over policies. A "well-functioning democracy" is possible only if the epistemic division of labor in society is properly ordered, as Kitcher teaches us: Absent from well ordered expertise, democracies could work against the interests and preferences of their citizens. Hence, according to the Well Ordered Nudge approach, the autonomy-empowering control of nudges sought by these political philosophers would better be promoted when scientists, who know more about the behavioral interventions, join the public in controlling behavioral interventions.

### 5.4.4. Citizen Nudge

A recent conception of nudge that promotes more interaction between scientific experts and the public is the so-called "citizen nudge" approach developed by Reijula & Hertwig (2020). This conception is similar to Well Ordered Nudge in that it motivates the idea that behavioral policy experts should share extensive information about nudges with the public so that individuals can design nudges in an autonomous manner.

The main difference between the Citizen Nudge approach and Well Ordered Nudge is that the former promotes a one-way information flow from scientists to the public while the latter promotes a two-way information flow where public is informed by scientists and informs scientists about their normative perspectives. As such, the Citizen Nudge concept promotes *self-nudging* (people's own initiatives to nudge themselves) and does not have of an account of how better-informed citizens can contribute to the designing of nudges as evidence based *public policy* instruments.

## 5.5. Conclusion

I have proposed a conception of nudges that is rooted in the deliberative democratic accounts of informed policymaking. This conception is called Well Ordered Nudge, and it is informed by Kitcher's seminal work on science in a democratic context. I have elaborated on the main characteristics of this conception and explained how it puts a solution to the table concerning nudge's stance towards the public's autonomous self-government. I have also discussed how Well Ordered Nudge differs from alternative conceptions of nudging and highlight some of its advantages.

Further research is needed to operationalize Well Ordered Nudge in practice and concretize its demands. As far as this chapter goes, my aim has been to articulate the distinct and useful perspective the Kitcher-type "well ordered science" approach provides for the ongoing debate about the democratic justification of nudges, and similar types of behavioral policies.

# References

Alexandrova, A. (2018). Can the science of well-being be objective? *The British Journal for the Philosophy of Science*, *69*(2), 421–445. https://doi.org/10.1093/bjps/axw027

Arad, A., & Rubinstein, A. (2018). The people's perspective on libertarian-paternalistic policies. *The Journal of Law and Economics*, 61(2), 311–333. https://doi.org/10.1086/698608

Barton, A., & Grüne-Yanoff, T. (2015). From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, *6*(3), 341–359. https://doi.org/10.1007/s13164-015-0268-x

Bovens, L. (2009). The ethics of nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference change* (pp. 207–219). Springer Netherlands. https://doi.org/10.1007/978-90-481-2593-7_10

Brown, M. B. (2006). Survey article: Citizen panels and the concept of representation. *Journal of Political Philosophy*, *14*(2), 203–225. https://doi.org/10.1111/j.1467-9760.2006.00245.x

Dede, O. Ç. (2019). Behavioral policies and inequities: The case of incentivized smoking cessation policies. *Journal of Economic Methodology*, *26*(3), 272–289. https://doi.org/10.1080/1350178X.2019.1625223

Douglas, H. (2005). Inserting the public into science. In S. Maasen & P. Weingart (Eds.), *Democratization of expertise*? (pp. 153–169). Springer-Verlag. https://doi.org/10.1007/1

Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh University Press. https://doi.org/10.1177/0094306110367909q4020-3754-6_9

Douglas, H. (2021). The role of scientific expertise in democracy. In M. Hannon & de Ridder, Jeroen (Eds.), *The Routledge handbook of political epistemology*. Routledge.

Dworkin, G. (2020). Paternalism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 Edition). <https://plato.stanford.edu/archives/fall2020/entries/paternalism/>

Fishkin, J. S. (2018). *Democracy when the people are thinking* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198820291.001.0001

Grüne-Yanoff, T. (2009). Welfare notions for soft paternalism. Papers on Economics and Evolution, *Max Planck Institute of Economics*, Jena, 2009-17

Grüne-Yanoff, T. (2012). Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, *38*(4), 635–645. https://doi.org/10.1007/s00355-011-0636-0

Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, *32*(3), 463–483. https://doi.org/10.1017/S0266267115000425

Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). Toward a framework for selecting behavioural policies: How to choose between boosts and nudges. *Economics and Philosophy*, *34*(2), 243–266. https://doi.org/10.1017/S0266267118000032

Guala, F., & Mittone, L. (2015). A political justification of nudging. *Review of Philosophy and Psychology*, *6*(3), 385–395. https://doi.org/10.1007/s13164-015-0241-8

Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, *18*(1), 123–136. https://doi.org/10.1111/j.1467-9760.2009.00351.x

Hausman, D. M. (2019). Enhancing welfare without a theory of welfare. *Behavioural Public Policy*, 1–16. https://doi.org/10.1017/bpp.2019.34

Heilmann, C. (2014). Success conditions for nudges: A methodological critique of libertarian paternalism. *European Journal for Philosophy of Science*, *4*(1), 75–94. https://doi.org/10.1007/s13194-013-0076-z

Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, *23*(1), 1–25. https://doi.org/10.1080/1350178X.2015.1070527

Irzik, G., & Kurtulmus, F. (2018). Well-ordered science and public trust in science. *Synthese*. https://doi.org/10.1007/s11229-018-02022-7

Irzik, G., & Kurtulmus, F. (2021). Distributive epistemic justice in science. *The British Journal for the Philosophy of Science*, 715351. https://doi.org/10.1086/715351

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, *302*(5649), 1338–1339. https://doi.org/10.1126/science.1091721

Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.

Kurtulmus, F. (2021). The democratization of science. In D. Ludwig, I. Koskinen, Z. Mncube, L. Poliseli, & L. Reyes-Galindo (Eds.), *Global epistemologies and philosophies of science* (pp. 145–154). Routledge.

Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science & Policy*, *1*(1), 35–42. https://doi.org/10.1353/bsp.2015.0000

Maasen, S., & Weingart, P. (2005). *Democratization of expertise?: Exploring novel forms of scientific advice in political decision-making*. Springer. http://site.ebrary.com/id/10134079

Małecka, M. (2021). Knowledge, behaviour, and policy: Questioning the epistemic presuppositions of applying behavioural science in public policymaking. *Synthese*. https://doi.org/10.1007/s11229-021-03026-6

Mansbridge, J., Bohman, J., Chambers, S., Estlund, D., FÃ¸llesdal, A., Fung, A., Lafont, C., Manin, B., & MartÃ-, J. luis. (2010). The place of self-interest and the role of power in deliberative democracy. *Journal of Political Philosophy*, *18*(1), 64–100. https://doi.org/10.1111/j.1467-9760.2009.00344.x

Nys, T. R., & Engelen, B. (2017). Judging nudging: Answering the manipulation objection. *Political Studies*, *65*(1), 199–214. https://doi.org/10.1177/0032321716629487

OECD. (2017). *Behavioural insights and public policy: Lessons from around the world*. OECD. https://doi.org/10.1787/9789264270480-en

Ongaro, M., & Andreoletti, M. (forthcoming). Non-empirical uncertainties in evidence-based decision making. *Perspectives on Science*. http://philsci-archive.pitt.edu/19353/

Qizilbash, M. (2012). Informed desire and the ambitions of libertarian paternalism. *Social Choice and Welfare*, 38(4), 647–658. https://doi.org/10.1007/s00355-011-0620-8

Reijula, S., & Hertwig, R. (2020). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, 1–31. https://doi.org/10.1017/bpp.2020.5

Reiss, J. (2019). Expertise, agreement, and the nature of social scientific facts or: Against epistocracy. *Social Epistemology*, *33*(2), 183–192. https://doi.org/10.1080/02691728.2019.1577513

Rizzo, M., & Whitman, D. G. (2009). Little brother is watching you: New paternalism on the slippery slopes. *Arizona Law Review*, *51*, 685–739.

Rolin, K. H. (2020). Objectivity, trust and social responsibility. *Synthese*. https://doi.org/10.1007/s11229-020-02669-1

Saghai, Y. (2013). Salvaging the concept of nudge. *Journal of Medical Ethics*, *39*(8), 487–493. https://doi.org/10.1136/medethics-2012-100727

Schmidt, A. T. (2017). The power to nudge. *American Political Science Review*, *111*(2), 404–417. https://doi.org/10.1017/S0003055417000028

Schubert, C. (2017). Exploring the (behavioural) political economy of nudging. *Journal of Institutional Economics*, *13*(3), 499–522. https://doi.org/10.1017/S1744137416000448

Steel, D., Fazelpour, S., Gillette, K., Crewe, B., & Burgess, M. (2018). Multiple diversity concepts and their ethical-epistemic implications. *European Journal for Philosophy of Science*, *8*(3), 761–780. https://doi.org/10.1007/s13194-018-0209-5

Sugden, R. (2018). 'Better off, as judged by themselves': A reply to Cass Sunstein. *International Review of Economics*, *65*(1), 9–13. https://doi.org/10.1007/s12232-017-0281-8

Sugden, R. (2018b). *The community of advantage: A behavioural economist's defence of the market* (First edition). Oxford University Press.

Sugden, R. (2021). How Hayekian is Sunstein's behavioral economics? *Behavioural Public Policy*, 1–10. https://doi.org/10.1017/bpp.2021.11

Sunstein, C. R. (2015). The ethics of nudging. *Yale Journal on Regulation*, *32*(2), 413–450.

Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press.

Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, *93*(2), 175–179. https://doi.org/10.1257/000282803321947001

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Waldron, J. (2014, October 9). It's all for your own good. *The New York Review of Books*.

Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies*, *61*(2), 341–355. https://doi.org/10.1111/j.1467-9248.2012.00974.x

# Conclusions

The thesis has offered a perspective from the philosophy of science to understand and approach the justifiability of Evidence Based Policymaking (EBP). This perspective focuses on identifying the proper sources and the content of (non-epistemic) values that inform and guide scientific research and the legitimate roles these values play (conceptualized in the thesis as the "authority" and the "legitimacy" problem of values in science, respectively). The thesis has consequently analyzed the justifiability of EBP from the philosophical perspective of "values in science." Approaching the subject from this perspective, the chapters in the thesis have conceptualized EBP as a practice of scientifically informed decision-making that can be considered complete and justifiable only when properly identified non-epistemic value considerations properly inform it. Specifically, the chapters contribute to understand, highlight, clarify, conceptualize, and resolve some of the major challenges of EBP that come to the fore when the non-epistemic dimensions of EBP are explored.

We have seen that the challenges of EBP pertaining to its value-ladenness can be analyzed from the perspective of values in science, specifically, by addressing the authority and the legitimacy problems of values in EBP. Addressing the authority problem of values in science in the context of EBP helps us discuss which values, projects, interests, and needs should guide the policy-advising scientific research and how these non-epistemic desiderata can and should be identified (e.g., should "lay citizens" or "scientific experts" determine these values? Through which procedures or avenues may these values be justifiably determined?) Addressing the legitimacy problem of values in the context of EBP helps us to discuss under which conditions allowing non-epistemic values to guide scientific research does not compromise the reliability of science for making informed policy decisions. Although society may ask scientists to serve for the general public's valuable projects, the instrumental value of science decreases when non-epistemic pursuits of science override its epistemic quality (e.g., unduly harming science's objectivity, reliability, and trustworthiness). In short, the practices of EBP should then ideally strike the right balance between these two goods: the pragmatic and the epistemic value of science.

The thesis has provided philosophical insights on the authority and the legitimacy problems by engaging with the values-in-science literature in philosophy of science which is devoted to responding to these compelling problems. The philosophical conclusions on these problems are meant as contributions to the conceptual progress in the ongoing theoretical discussions regarding values in science, social dimensions of science, and science in a democratic society. In order to make meaningful contributions to these discussions, each chapter of the thesis has provided insights on different aspects of the relevant topical issues and done so by analyzing some of the paradigmatic and societally important examples of EBP such as toxicological risk evaluation, incentivized behavioral interventions, and nudges. In the following, the thesis will conclude with a summary of some of the general insights drawn from the preceding chapters.

The first part of the thesis has introduced the values in science perspective to approach and understand the justifiability of the EBP in a democratic context. We have reviewed the challenges and potentials of the contemporary philosophical approaches to the values in science in the light of the debates on the value free ideal of science. *Chapter 1* has highlighted the normative appeal of the "value free ideal" of science and invited the scholars in the sciences to pay special attention to the consequences of transcending the value free ideal of science. It also elaborated on the need for formulating new ideals for science in society that also convincingly address the broader worries of the proponents of the value free ideal (e.g., concerning the epistemic authority of science or its place in a democratic society). The chapter has suggested that the extant accounts that transcend the value free ideal of science (such as Douglas's, or Longino's) indeed do, in different ways, address these worries, even though we may still lack a full-fledged theory of values in science that replaces the traditionally accepted value free ideal of science in a comprehensive manner.

*Chapter 2* has motivated the idea that our philosophical theories of value-laden decision-making in science can benefit from resting on more detailed descriptions of scientific practice (such as those offered by empirical studies of science). It has done so in the context of the most prominent contemporary account of values in science

(i.e., the inductive risk framework). This philosophical exercise is not only important for the sake of the descriptive adequacy of our philosophical theories of values in science but also fruitful in exploring comprehensive and normatively appealing philosophical accounts of science. Consider, for instance, the analysis in *Chapter 2* regarding how social and institutional context may shape or change regulatory toxicologists' evaluations of the relevant inductive risks. It is in and of itself essential that our philosophical theories of values in science are able to adequately describe and assess these processes as defended in the thesis (especially concerning the societally interesting cases of EBP such as regulatory toxicology). Moreover, it would also be desirable that our philosophical approaches can advance prescriptions regarding how value judgments are determined and how they change in response to the social environment of scientists (pertaining to the authority problem of values in science). The latter may be a subject for future work in the relevant field, and the kind of extended/integrated account of inductive risks proposed in the chapter is suitable for critically analyzing the conditions under which the social interactions between different groups of scientists (or between scientists and the public, or between scientists and institutions surrounding them) lead to epistemic or non-epistemic progress.

The second part of the thesis has focused on understanding and addressing some of the philosophical issues that come to the fore when we ask how the public's needs, values, and interests properly inform and guide scientific research for policy making purposes. The focus on the non-epistemic dimensions of EBP leads us to examine scientists' own value presuppositions, the proper use and the communication of these values in policy-relevant contexts. The resulting philosophical analysis lends itself to a pluralistic view of science practiced by groups of researchers pursuing analytically distinct research questions, epistemic and non-epistemic projects. This view of science is consistent with the thesis's conceptualization of EBP as a case of informed decision-making where the issues such as the fit-for-purpose of scientific evidence and the balancing of the evidence and values come to the fore. Contrast this with the philosophical conceptualization of EBP as a case of causal inference where the primary focus of analysis is the extrapolation of scientific evidence irrespective of

their fit-for-purpose or pragmatic value. An advantage of the former conceptualization of EBP is that it helps us to critically evaluate the use of scientific findings in policymaking without making universalized or strict methodological claims. This has been nicely illustrated in *Chapter 3.* The chapter examines the limitations of randomized controlled trials as an evidence-gathering method (in informing behavioral health policies targeting at reducing health-related inequities) *not* by reference to its inherent methodological problems but its fit-for-purpose: to what extent it is conducive to achieve policy desiderata in question.

The last two chapters have offered more direct insights into the values-in-science perspective on the justifiability of EBP in a broadly liberal and democratic context. *Chapter 4* has instantiated the advantages of approaching the issues in EBP from the perspective of values in science by clearly introducing the distinction between the authority and the legitimacy question of values in science. The distinction helps us appreciate the idea that the claims regarding the authoritativeness of value judgments in science (e.g. those presupposed by scientists) do not necessarily put one in a position to endorse a claim about the legitimacy of value judgments in science. As illustrated, this idea could help us advance philosophical views on how to address and resolve the legitimacy and authority problems. *Chapter 4*, for instance, specifies the view that we could integrate the most robust responses to these problems even though they may seem conflicting. It supports the arguments of those philosophers (such as Heather Douglas and Elizabeth Anderson) who combine the novel and promising responses to the authority problem of values (e.g. feminist and democratic responses such as Alexandrova's) with more traditional responses to the legitimacy problem of values (e.g. impartiality-centered views such as Nagel's).

While *Chapter 4* has advanced the distinction between the authority and legitimacy problems, *Chapter 5* explores a specific response to the authority problem of values in the context of EBP. Following the overarching aim of the thesis (that is, analyzing the justifiability of EBP with respect to meta-values such as democratic self-government), the chapter sides with Philip Kitcher's view on normative expertise, which rests on the procedural and interactionist accounts of normative justification

where neither of the relevant actors (e.g. the members of the public, scientists, or moral philosophers) should count as normatively authoritative. *Chapter 5* has elaborated how this promising take on the authority problem of values can fruitfully contribute to the ongoing discussions on the democratic justifiability of nudges, which is a controversial contemporary case of EBP.

I hope to have demonstrated that the preceding chapters' socially relevant and practice-oriented philosophical insights on the use and formation of values in science can improve our understanding of how evidence based policymaking practices can be made more accountable and justifiable in a democratic context.

# Appendices

# Appendix A

## Samenvatting (Dutch Summary)

Wetenschap wordt over het algemeen gezien als de meest betrouwbare bron van kennis.. Beleidsmakers hebben als doel om beleid te ontwikkelen met effect op de maatschappij. Op wetenschap gebaseerd overheidsbeleid wordt ook wel 'evidence based policymaking' genoemd. Dit proefschrift biedt een wetenschapsfilosofisch perspectief op de rechtvaardiging van evidence based policy (EBP) met betrekking tot algemene democratische en liberale waarden.

Het rechtvaardigen van EBP op basis van democratische en liberale waarden houdt in dat EBP waarden zoals vrijheid en autonomie bevorderen en versterken, of in ieder geval in overeenstemming met die waarden zijn. Het identificeren van de voorwaarden waaronder EBP in overeenstemming met deze waarden is vergt begrip van de rol van wetenschap en wetenschappers bij het realiseren van publieke waarden, menselijke behoeften, interesses en het streven naar vrijheid. Dit proefschrift biedt een filosofisch perspectief met als doel om te conceptualiseren hoe wetenschap in overeenstemming met maatschappelijke waarden, behoeften en interesses kan zijn, of deze zelfs kan promoten. Meer specifiek neemt behandelt het proefschrift een wetenschapsfilosofisch perspectief dat focust op de relatie tussen wetenschap en (maatschappelijke waarden). Dit type wetenschapsfilosofische perspectief op 'waarden in wetenschap' heeft twee overkoepelende doelen. Ten eerste heeft het als

doel om bij te dragen aan debat over welke waarden en niet-epistemische overwegingen wetenschappelijk onderzoek zouden moeten dienen. Het gaat bijvoorbeeld in op de vraag welke sociale mechanismen de niet-epistemische waarden die een rol spelen in wetenschap zouden moeten blootleggen en wat adequate bronnen voor deze niet-epistemische waarden zijn (bijv. Kourany, 2010; Kitcher, 2011). Ten tweede biedt het proefschrift theorieën over de rol van niet-epistemische waarden in wetenschappelijke redeneringen en onderzoek. Deze theorieën laten zien dat niet-epistemische waarden de epistemische kwaliteiten en doelen van wetenschap niet ondermijnen. Het waarden-in-wetenschap perspectief balanceert daarbij tussen de instrumentele waarde van wetenschap (i.e., het gebruik van wetenschap om bepaalde maatschappelijke doelen en waarden na te streven) met haar epistemische autoriteit (i.e., de objectiviteit, non-dogmatisme en betrouwbaarheid). Het proefschrift bevordert begrip van EBP vanuit het perspectief van waarden in wetenschap door het adresseren van problemen die naar voren treden wanneer erkend wordt dat EBP een waardegedreven vorm van geïnformeerde besluitvorming is.

Het proefschrift bestaat uit twee delen die ieder op zichzelf staande wetenschappelijke artikelen bevatten. In deel 1 (getiteld *Values in Science: Beyond the "Value Free Ideal" of Science),* gaat het proefschrift in op de wetenschapsfilosofische literatuur over waarden in wetenschap. Het laat zien hoe deze literatuur zich verhoudt tot EBP en gaat in op hoe hedendaagse filosofische benaderingen de rol van waarden in wetenschap zien en hoe die waarden kunnen veranderen in relatie tot de context waarin ze zich voordoen. In deel 2 (getiteld *Philosophy of Evidence Based Policy from the Perspective of Values in Science*) behandelt het proefschrift problemen met betrekking tot de rechtvaardiging van EBP vanuit het waarden-in-wetenschap perspectief.

In het eerste deel van het proefschrift (getiteld *Values in Science: Beyond the "Value Free Ideal of Science*), worden de consequenties van het verlaten van het waardenvrije ideaal van wetenschap onderzocht. *Hoofdstuk 1* evalueert de argumenten voor en tegen het waardenvrije ideaal van wetenschap en illustreert dit door gebruik te maken van de gereguleerde toxicologie. Het hoofdstuk beschrijft de risico's van het verlaten

van het waardenvrije ideaalbeeld. Het beargumenteert dat de prominente filosofische perspectieven die het waardenvrije ideaal van wetenschap overstijgen het waardenvrije ideaal niet afwijzen, maar de meta-waarden van wetenschap die de voorstanders van het waardenvrije ideaal belangrijk achten juist behouden. Onder deze waarden vallen de epistemische autoriteit van wetenschap, wetenschappelijke onpartijdigheid en het vermijden van ondemocratisch (technocratisch) beleid.

*Hoofdstuk 2* reflecteert op de vraag hoe hedendaagse filosofische frameworks die de meest prominente waarde-geladen concepties van wetenschap ontwikkelen en evalueert voorbeelden uit de wetenschappelijke praktijk. In het bijzonder wordt er aan aandacht besteed aan het inductief risico framework van Heather Douglas en bediscussieert hoe het framework van toepassing is op een maatschappelijk relevant en filosofisch significante historische episode uit de regulerende toxicologie. Het hoofdstuk biedt een empirisch geïnformeerde beschrijving van toxicologen die methodologische beslissingen maken met betrekking tot het relevante bewijs voor hun doelen en die kiezen tussen verschillende methoden van bewijsvergaring. Het hoofdstuk gaat specifiek in op interessante theoretische aspecten van deze casus (d.w.z., de rol van sociale en institutionele processen bij veranderende perspectieven van toxicologen op inductieve risico's) die nieuw zijn voor het inductieve risico framework. Het laat zien hoe het inductieve risico framework kan worden uitgebreid om ook op deze casus van toepassing te zijn.

In het tweede deel (getiteld *Philosophy of Evidence Based Policy from the Perspective of Values in Science*), reflecteert het proefschrift op de filosofie van evidence based policy vanuit het waarden-in-wetenschap perspectief.

*Hoofdstuk 3* onderzoekt hoe een specifieke vorm van gedragsinterventies, namelijk anti-rook programma's, wordt geëvalueerd door wetenschappers. Het laat zien dat de evaluaties van deze programma's in de praktijk worden gekenmerkt door een hoeveelheid aan wetenschappers die gebruik maken van verschillende bewijsvergaringsmethoden, die verschillende wetenschappelijke disciplines representeren en verschillende beleidsperspectieven op roken als een publiek gezondheidsprobleem hanteren. Het hoofdstuk beargumenteert dat een empirische

evaluatie van de vraag of anti-rookbeleid effectief is in het verminderen van rook-gerelateerde gezondheidsverschillen (wat over het algemeen gezien wordt als het meest belangrijke doel voor veel anti-rookbeleid in Europa) integratie van verschillende methoden (zoals gerandomiseerde controlestudies en kwalitatieve methoden uit de sociale epidemiologie) vergt. Het hoofdstuk bespreekt de implicaties van dit pluralistische perspectief voor de filosofische discussies over de voorwaarden voor bewijs voor gedragsinterventies.

*Hoofdstuk 4* richt zich op speciale cases waarin de claims die gemaakt worden door sociale wetenschappers concepten includeren die gebaseerd zijn op ongewenste waarde-vooronderstellingen. Anna Alexandrova definieert zulke claims als "gemixte claims", omdat ze normatieve en empirische inhoud hebben. Gemixte claims zijn bijzonder controversieel in het geval dat ze gebruikt worden voor beleid. Tenzij ze heel goed onderzocht zijn, beïnvloeden deze mixed claims impliciet het beleidsdebat met waardegedreven perspectieven. Het hoofdstuk analyseert het contrast tussen Nagel en Alexandrova en biedt een perspectief dat beide kanten van de controverse verzoent. In het bijzonder laat het hoofdstuk zien dat, gebaseerd het verschil tussen autoriteitsprobleem van waarden en de rechtvaardiging van waarden, Alexandrova's kritiek uit 2017 van Nagel (1961) is beperkt tot het autoriteitsprobleem van waarden en laat 'Nageliaanse' conditionering van gemixte claims als strategie dat legitimering of rechtvaardiging van waarden intact. Het laat tevens zien dat het hedendaagse Neo-Nageliaanse concepten (zoals E. Andersons en Douglas' concepten van objectiviteit) de verzoening met betrekking tot gemixte claims tussen Alexandrova en Nagel ondersteunen.

*Hoofdstuk 5* richt zich op een belangrijk voorbeeld van evidence based policy, namelijk Nudges. Nudge beleid is een prominent en vaak toegepast type evidence based policy. Nudges roepen belangrijke ethische en politieke vragen op die relevant zijn in het kader van de in dit proefschrift besproken filosofische vraagstukken. Gebaseerd op de literatuur over de democratisering van wetenschap (specifiek op Philip Kitcher's benadering), bevat dit hoofdstuk een voorstel voor een Nudge concept, oftewel een 'well-ordered nudge', die rechtvaardiging van dit soort

gedragsbeleid zoekt via deliberatieve democratische besluitvorming. Het hoofdstuk biedt daarmee een specifieke praktische benadering en laat de conceptuele implicaties zien voor een belangrijk type evidence based policy. In het hoofdstuk worden de well-ordered Nudges vergeleken met de reeds bestaande Nudge concepten en gaat in op de vraag hoe het filosofische perspectief dat dit proefschrift biedt de huidige discussies over de democratische rechtvaardiging van Nudges en nudge-achtige evidence based beleidsvormen kunnen verreiken.

# Appendix B

## Curriculum Vitae

**O. Çağlar Dede**
Email: o.c.dede@vu.nl

## EDUCATION

2014 - 2021 **PhD in Philosophy**
Erasmus Institute for Philosophy and Economics (EIPE), Erasmus School of Philosophy, Erasmus University Rotterdam, Netherlands

*Dissertation title*: "Science and Values: A Philosophical Perspective on the Justifiability of Evidence Based Policymaking"
*Supervisors*: Jack Vromen and Conrad Heilmann

2012 - 2014 **MA (research) in Philosophy and Economics**
Erasmus University Rotterdam, Erasmus Institute for Philosophy and Economics (EIPE), Erasmus School of Philosophy, Netherlands

*Thesis Title*: "Preference Elicitation and the Normative Basis of Nudge Policies"
*Supervisor*: Conrad Heilmann

2007 - 2012 **BA in Economics** (*with honors degree*)
Boğaziçi University Istanbul, Department of Economics, Turkey

## AREAS OF SPECIALIZATION / INTEREST

**AOS** Philosophy of Economics, Ethics and Sciences of Behavior Change, Social Dimensions of Science, Philosophy of Science: Values, Objectivity, and Evidence Based Policy

**AOI** Social Epistemology of Science, Science and Democracy, Interaction and Argumentation in Science, Deliberative Democracy

## CURRENT APPOINTMENT

2020 - 2023 **Postdoctoral Researcher**
**"**Social Epistemology of Argumentation", (ERC SEA, *PI: Catarina Dutilh Novaes*)
Department of Philosophy, Faculty of Humanities
Vrije University Amsterdam (VU), Amsterdam, Netherlands

## PUBLICATIONS

### Peer Reviewed Research Articles/Book Chapters

Dede, O. Ç. (2021). The value free ideal of science and non-epistemic values in regulatory toxicology. In A.S. Arcuri & F. Coman-Kund, (Eds.) *Technocracy and the law: Accountability, governance and expertise*. **Routledge**. https://doi.org/10.4324/9781003174769

Dede, O. Ç. (2020). Integrating Heather Douglas' inductive risk framework with an account of scientific evidence: Why and how?. ***Perspectives on science***, *28*(6), 737–63. https://doi.org/10.1162/posc_a_00358

Dede, O. Ç. (2019). Behavioral policies and inequities: The case of incentivized smoking cessation policies. ***Journal of economic methodology***, *26* (3), 272–89. https://doi.org/10.1080/1350178X.2019.1625223

### Other Publications

The Community of Advantage: An Interview With Robert Sugden. ***Erasmus journal for philosophy and economics***, 13(1), 61–78. doi.org/10.23941/ejpe.v13i1.483 (with Erwin Dekker)

## MANUSCRIPTS IN PREPARATION

*Reconciling Ernest Nagel's Impartiality with Anna Alexandrova's "Mixed Claims"* (manuscript conditionally accepted for publication in ***The British Journal for Philosophy of Science***)

*Nudges, Well Ordered* (manuscript in preparation for submission).

## TALKS & PRESENTATIONS

### Talks at Peer-Reviewed International Conferences and Workshops

**"**Nudges, Well Ordered: A positive story the scientist-citizen interaction**"** **British Society for Philosophy of Science**, June, 2021

"Epistemic Conditions for Libertarian Paternalism", **The Amsterdam Graduate Conference in Political Theory**, May, 2019

"The Values in Science Approach to Behavioral Policies", **European Network for Philosophy of Social Sciences**, Roundtable Meeting. (ENPOSS), Hannover, August 2018.

**"**Well Ordered Nudge: A New Epistemic Conception of Libertarian Paternalism", **4th International Conference "Economic Philosophy" Network**, Lyon, June 2018.

"Well Ordered Nudge", **Democratisation of science: epistemological issues and new perspectives Workshop**, **The ANR funded Democrasci Project Workshop Series**, Lyon, May 2018.

"Heather Douglas's Account of Values and Evidence in Science", **European Society for Philosophy of Sceince (EPSA)** 2017, Exeter, September 2017.

"Values in the Evidential Evaluations of Behavioral Policies", **International Network for Economic Methodology**, INEM 2017, San Sebastian, August 2017.

"How to Evaluate Behavioral Public Policies", **Economics and/or Psychology Workshop**, TINT, Helsinki, May 2017.

"A Fresh Look at the Evidential Evaluations of Behavioral Public Policies" at the **Nordic Network for Philosophy of Science Conference**, Copenhagen, April 2017.

Comment on Sarah Wieten and Donal Khorowski's "Putting Rigor Back into Pragmatic Trials", **Nordic Network for Philosophy of Science Conference**, Copenhagen, April 2017.

"The Interplay between Ethical and Evidential Evaluations of Behavioral Policies", at the EIPE 20TH Anniversary Conference, March 2017.

**"**Can the Evaluations of Behavioral Policies be Objective?**" Dutch Research Scbool for Philosophy (OZSW)**, Utrecht, 2017.

"Evidential Justifiability of Behavioral Public Policies: Is Mechanistic Evidence Enough?" (*In Turkish*) at **the First Graduate Student Conference of the Middle East Technical University**, Ankara, Turkey, June 2016.

"Assessing the Evidence-Base of Behavioral Public Policies" at the **Dutch Research Scbool for Philosophy OZSW Graduate Conference**: How Philosophy Meets the World, Twente University, Netherlands, April 2016.

"Assessing the Evidence-Base of Behavioral Policies: The Case of Smoking Cessation

Programs", at **Explanation, Normativity, and Uncertainty in Economic Modelling Workshop, London School of Economics (LSE)**, March 2016.

"Evidence in Heather Douglas's Account of Values in Science" at the **Annual OZSW Conference**, Vrije University, Amsterdam, December 2015.

"Evidential Justifiability of Behavioral Public Policies: Why Traditional Evidence is not enough" at the 12th **International Network for Economic Methodology Conference**, University of Cape Town, South Africa, November 2015.

**Invited Talks at Seminars and Workshops**

"On the Politics of Behavioral Science" **Perspectives on Science Seminar Series**, University of Helsinki. November, 2020 (cancelled).

Invited talk on the behavioral approach to public health at the multidisciplinary workshop "***Healthy People: How the medical humanities and the health sciences study the complexity of health***", **Lorentz Center** in Leiden, organized by Federica Russo, Phylis Illari, Brendan Clarke, Virginia Gharia, October 2018 • The "brain writing sessions" during the workshop resulted in the following communication piece published the British Medical Journal (BMC).

"How to Deal With Values in the Evidence-based Evaluations of Nudges", **EIPE-VU Joint Research Day**, **Vrije University, John Stuart Mill College**, April 2018.

Invited talk and panel discussion on "Values in Regulatory Toxicology: A Philosophy of Science Perspective", Multidisciplinary workshop on ***Unpacking the 'Accountability Paradox' in Expert-based Decision-making, Erasmus School of Law,*** November 2017.

"Evidence in Heather Douglas's Account of Values in Science" at the Value-Free Ideal in Economics' workshop, **Centre for Humanities Engaging Science and Society (CHESS)**, Durham, December 2015.

"Assessing the Behavioral Public Policy Alternative: A Methodological Perspective" at the **BETAM Economics Seminar**, Bahcesehir University, Istanbul, December 2015.

Comment on Erik Weber *et al.*'s "Woodward, Experiments, and Evidence", **Ghent-Tilburg-Rotterdam Graduate Workshop in Philosophy**, February 2015.

**Talks at Departmental Seminars and Workshops**

Comment on Vincent S.' "Mini-public's epistemic performance in deliberative democracy", EIPE PhD Seminar, May 2019.

Comment on Merve Burnazoglu's "EU Skills Profiling Tool for Refugees' Integration to Labor Markets", EIPE PhD Seminar, April 2018.

Comment on Lieke Martens' "Economic Forecasts and the Principle of Reflexivity", EIPE PhD Seminar, June 2018.

"Can the Evaluations of Behavioral Policies be Objective?" EIPE PhD Seminar, November 2017.

Comment on Mario Castellano's "On Corporate Moral Responsibility", EIPE PhD Seminar, April 2017.

"Heather Douglas's Account of Values and Evidential Judgments in Toxicological Practice", EIPE PhD seminar, Erasmus University Rotterdam, December 2016.

Comment on Nicolas Bernaman's "Consensus in Economics", EIPE PhD Seminar, November 2016.

Comment on Virginia Ghiara's "Big Data: Collection and Selection", EIPE PhD Seminar, March 2016.

"Evidential Justifiability of Behavioral Public Policies: Why Traditional Evidence is not enough", EIPE PhD seminar, November 2015.

"Evidence and Values in Heather Douglas's Framework" Cursos de Verano: the INEM/CHESS Summer School in Philosophy & Economics, San Sebastian, Spain , July 2015.

Comment on Elisabeth Does, "Experimenting with Experiments: a systematic approach to mixed methods design in research on economic decision making", EIPE PhD Seminar, April, 2015.

Joost Hengstmengel, "Philosophy and Economics: Aristotelianism in Early Mercantilism", EIPE PhD Seminar, November, 2014.

"Nudge as Performativity", EIPE PhD Seminar, March 2014.

Comment on Guus Dix, "Governmentality to performativity. The 'incentive' as a locus of power/knowledge", EIPE PhD Seminar, March 2013.

## TEACHING

### Courses designed and taught

**Lecturer** • *Ethical Aspects of Economics* • Erasmus University Rotterdam • 2018 – 2019

Fully responsible for course design, teaching and examination of the course "*The Ethical Aspects of Economics*" for 3rd year undergraduate economics and business students following a double degree program in philosophy

### Guest lectures for postgraduate teaching

**Guest Lecturer** • *Women in Philosophy: Helen Longino* • Erasmus University Rotterdam • February 2020

Co-teaching a lecture on Helen Longino's philosophy of science covering themes such as values, objectivity, and pluralism in science

**Guest Lecturer** • Erasmus University Rotterdam • January 2018

A master-level seminar lecturing on "Nudge" in course "*Welfare and Cooperation*"

**Workshop Leader** • Erasmus University Rotterdam • December 2014

Lead a course on nudge policy, preference formation and behavior change. Part of the OZSW winter school on "Philosophy, Policy, and Social Science" for research master students.

### Undergraduate teaching

**Tutorial Instructor** • Erasmus University Rotterdam • June 2017
Three weeks of lecturing on "*Business Ethics*"

**Tutorial Instructor** • Erasmus University Rotterdam • May/June 2016
Four weeks of instruction in the course "Philosophy of Economics"**,** Erasmus School of Economics (ESE), including lecturing, exam preparation, and examination

**Tutorial Instructor** • Erasmus University Rotterdam • May/June 2015
Four weeks of instruction in the course "Philosophy of Economics"**,** Erasmus School of Economics (ESE), including lecturing, exam preparation, and examination

**Teaching Assistant** • Boğaziçi University • Spring semester, 2012
Teaching assistant for the courses "Introduction to Macroeconomics" and "Economic Policy"

**Teaching Assistant** • Boğaziçi University • Autumn semester, 2011
Teaching assistant for the course "Introduction to Microeconomics"

## SERVICE TO THE PROFESSION

### Editor

*Erasmus Journal for Philosophy and Economics* (http://ejpe.org/) • April 2015 – August 2018

Managing article submissions and book reviews, organizing special issues and interviews, preparing annual reports. Editor responsible for a book review symposium on D. McCloskey's *Bourgeois Virtues* (link), two commentaries on the 2017 and 2016 Nobel Prize Winners in economics (link, link).

### Refereeing

*Perspectives on Science*,
*Social Theory and Practice*,
*Minds and Machines*,
*Journal of Economic Methodology*,
*Erasmus Journal for Philosophy and Economics*,
*Erasmus Student Journal of Philosophy*

### Selected Workshop/Discussion Group Organization

2021 Departmental meeting: Philosophy of Science and Technology Discussion Group, Vrije University Amsterdam

2021 Symposium: Epistemic Diversity and Argumentative Practices in Science, British Society for Philosophy of Science

2018 Workshop: Evidence-based Policies and Values in Science, Erasmus School of Philosophy

### Student Assistant

*EIPE Research Assistant* • Erasmus University Rotterdam • 2013 – 2014

Assistant for Dr. Conrad Heilmann, co-director of Erasmus Institute for Philosophy and Economics (EIPE).

*Support for Conference Organization*:

Formal Ethics Conference 2014 • EIPE, Erasmus University Rotterdam • May 2014
Annual Conference of the Dutch Research School in Philosophy (OZSW) • Erasmus University Rotterdam • November 2013
11th International Network for Economic Methodology (INEM) Conference • Erasmus University Rotterdam • June 2013
9th Conference of the European Society Ecological Economics (ESEE) • Boğaziçi University • June 2011.
15th Conference of the European Society for History of Economic Thought (ESHET) • Boğaziçi University • May 2011.

## AWARDS AND SCHOLARSHIPS

**The Dean's Award for Multidisciplinary Excellence** • Erasmus Graduate School of Social Sciences and the Humanities, Erasmus University Rotterdam • November 2016 • Awarded for the best multidisciplinary research presentation on the theme: aging.

**PhD Researcher Scholarship** • EIPE, Erasmus University Rotterdam • 2014 – 2018

**Non-EU student fee waiver** • Erasmus University Rotterdam • 2012 – 2014, Scholarship awarded to non-resident students without external funding.