

SCIENCE AND VALUES: A TWO-WAY DIRECTION

Emanuele Ratti¹, University of Bristol
Federica Russo, Utrecht University and University College London

Authors have contributed equally and are listed in alphabetical order

Abstract. In the science and values literature, scholars have shown how science is influenced and shaped by values, often in opposition to the ‘value free’ ideal of science. In this paper, we aim to contribute to the science and values literature by showing that the relation between science and values flows not only from values into scientific practice, but also from (allegedly neutral) science to values themselves. The extant literature in the ‘science and values’ field focuses by and large on reconstructing, *post hoc*, how values have influenced science; our reconstruction of the case studies, instead, aims to show that scientific concepts and methods, *because of specific identifiable characteristics*, can promote some values rather than (or at the expense of) others. We explain this bidirectional relation in analogy to debates on the normativity of technical artifacts and on feminist approaches in science, and we illustrate our claims with cases from the health sciences and machine learning. While our arguments in this paper also draw on *post hoc* reconstructions, we intend to show where we should engage not only with the question of whether a practice is value-laden, but also how specific conceptual and methodological choices can influence values down the road. All in all, these considerations expand the ways in which philosophers can contribute to more value-aware scientific practices.

Keywords: science and values; health sciences; machine learning; feminist philosophy of science; philosophy of technology

Conflict of interest: The authors declare no conflict of interests

1. Introduction

The ‘value-free’ ideal of science has sparked vivid debates in the past few decades. While various versions of the value-free ideal have been defended until recently, the thesis that science cannot be value-free (be it a descriptive or prescriptive statement) has been favored in some philosophical circles. ‘Science and values’ is now the name for a rich field, which includes debates (just to name a few) on inductive risk, scientific communication, feminist philosophy of science, etc. This literature has discussed in great detail how scientific practice can be indeed influenced by adopting certain values rather than others. Moreover, it has shown that value-laden choices in scientific practice are inevitable, as well as that values can even

¹ mnl.ratti@gmail.com

have causal efficacy in bringing about certain outcomes in science. As Ward (2021) puts it, values filter “the research process, making downstream decisions like hypothesis appraisal value-laden” (p 56). But values can also be *promoted* as a result of scientific practice (Ward 2021; Elliott 2017; Lacey 2005). As we will discuss throughout the paper, the idea of ‘value promotion’ is equally important, as one can make scientific choices because of value-laden motivations, but then those choices end up having other effects at the level of values. While this dimension has been recognized as part of the value-science entanglement, it has not been systematically explored and discussed. In particular, the ‘science and values’ literature does not offer systematic discussions about dynamics of ‘values (or promotion of values) as effects’: is it just the uncontroversial thesis that choices can have moral consequences? What is exactly the relation between the technical and conceptual apparatus of science and values when values (or their promotion) are ‘effects’? That science influences values and norms down the road is also a point made outside the ‘science and values’ circle, for instance in the pragmatist philosophy of John Dewey (1993) or in the sociology of science of Robert Merton (1973). Although their arguments differ from ours in motivation and scope, both pleaded for the idea that science has an important role for fostering democratic values. This is certainly one way in which, to anticipate the terminology of this paper, scientific concepts and methods may promote values. We are however interested in a finer grained type of argument, one that goes to the details of techno-scientific practices, and shows how scientific concepts and methods promote values – at times even unintended ones.

In this article, we take up this insight coming specifically from the ‘science and values’ literature, with the intention of contributing to more ‘value-aware’ techno-scientific practices. To this end, we disentangle the ways in which values can be promoted as a result of methodological and conceptual choices. The ways in which scientific concepts and methods *promote* values are not in contrast with the framing of the ‘science and values’ scholarship, according to which values influence science: in fact, they can be at work independently, simultaneously, and at times even clashing with one another. Our primary aim is to complement the existing debate by showing how specific aspects of science (notably, concepts, methodologies, etc) *promote* certain values (whether scientists want it or not), and do that on top of the values motivating the choice of those concepts and methodologies. We are interested in this direction of influence, from science to values, because we hope to ultimately contribute to science in the making. Through interdisciplinary collaborations between philosophers and

scientists, we intend to promote ‘value-aware’ scientific practices, in which we anticipate, as much as possible, the potential impact of our concepts and methods at the level of values.

Thus, the central thesis of our paper is that the relation between science and values can have ‘directions’. Specifically, there are two main directions which, as we will see, are rather complementary as they sometimes work in iteration, and sometimes independently. A first direction (discussed in Section 2.1) is *from values to science*, in the sense that values shape the concepts, the methodologies, and the choices made by scientists throughout the entire scientific process. This is the classic framing in the field of ‘science and values’, and the literature on inductive risk well exemplifies the first direction; it is shown that science is not and cannot (and sometimes even should not) be value-free, and that choices made on the basis of (non-epistemic) values are not only ubiquitous, but they cannot even be avoided. A second direction (discussed in Section 2.2 and 2.3) is *from science to values*, in the sense that methodologies or scientific concepts create the conditions for some values, rather than others, to be promoted. Specifically, our analysis intends to show how certain conceptual or methodological choices have implications for – or promote – values. The second direction, we submit, has not received systematic attention (especially if compared to the first); here we further develop the argument formulated in (Russo 2021) for the ‘from-science-to-values’ direction in the field of medicine and public health (Section 3), and extend the argument to the use of algorithmic methods in the criminal legal system (Section 4).

2. Two directions in the relationship between science and values

2.1 From values to science

The first direction going from values to science has been explored significantly in philosophy of science. Values influencing science have been seen especially controversial when non-epistemic values are concerned, even though choosing between epistemic/constitutive/cognitive values can require substantial value-judgment (Kuhn 1977; McMullin 1983)². The literature on inductive risk is representative of this first direction. The debate on the so-called inductive risk and the relation between science and values dates back

² There have been attempts to reject the distinction epistemic/non-epistemic altogether. For instance, Longino seems to imply that the distinction constitutive/contextual (mirroring epistemic/non-epistemic) cannot be retained (1990), and Rooney (1992) points to some challenging issues in understanding the distinction and sees it more as a continuum. We focus especially on moral and political values, and the problem of whether the distinction between epistemic and non-epistemic is tenable is not a concern of this article

at least to the end of the 1940s (Churchman 1948; Rudner 1953; Levi 1960; Hempel 1965). The core of the problem stems from the epistemological consideration that scientific hypotheses are never completely verified, and they are accepted/rejected in this situation of uncertainty. Famously, Rudner discussed the problem of accepting/rejecting a hypothesis in connection to its ethical consequences, over and above the strength of the evidence supporting said hypothesis. In determining if evidence is strong enough, we resort to the importance of making a mistake, and we inevitably make use of value judgements. Hempel (1965) made similar considerations and he claimed that accepting a hypothesis in a situation of uncertainty carries the so-called *inductive risk*, namely “that the presumptive law may not hold in full generality, and that future evidence may lead to scientists to modify or abandon it” (p 92). In certain cases, hypothesis acceptance/rejection requires value judgment (p 92) because of the severity of the consequences of either wrongly accepting or wrongly rejecting hypotheses.

Douglas’ work (2000; 2009) gave new life to this problem, by showing how considerations of inductive risk play a significant role not just in hypothesis acceptance, but also in the choice of methodology, in the characterization of data, and in the interpretation of data. In other words, values can shape virtually every phase of scientific practice. She illustrates her claims with several examples. For instance, she considers the case of setting the statistical significance levels for a study aimed at establishing the effects of an air pollutant. Once a scientist has obtained the results, she has to establish if the evidence is enough to support the hypothesis that the air pollutant is toxic. Considering the pollutant as toxic would lead to alarm and regulatory costs, which would be unnecessary if the result were a false positive. A false negative, on the other hand, could harm people and have bad effects on the local community. Weighing, which is influenced by ethical and social values, must occur because of “[t]he social and ethical costs of the alarm and regulation on the one hand, and the human health damage and resulting effects on society on the other” (Douglas 2009, p 105). The values of protecting public health and costs of regulation are then *indirectly adjusting the amount of evidence required to accept or reject hypotheses*. Since Douglas’ contribution, a literature showing the ubiquity of inductive risks in scientific processes has grown fast and big. But it is also important to point out that the value-ladenness of science has not been widely accepted. As Douglas points out (2009), there have been two main lines of arguments against inductive risk. The first argues that it is not scientists’ business to accept or reject hypotheses; what they should do instead is just assign probabilities and then turn them to the public. This position was introduced by Jeffrey (1956) and recently defended by Betz (2013) when he argues that value-laden decisions can be avoided by making “uncertainty explicit and articulating findings

carefully" (p 209). A second line of argument admits that science is value-laden, but qualifies this statement by saying that only epistemic values should play a role in science. This position of epistemic purity (Brown 2020) was defended at first by Levi (1960).

Recently, the argument from inductive risk has been expanded. Biddle and Kukla (2017) have formulated a more general framework to make sense of the connection between scientific uncertainty, risks, and values. They argue that several types of risk in scientific processes are not inductive: traditional inductive risk is a specific instance of a broader category of risks, which they call epistemic risk. This is understood broadly as "any risk of epistemic error that arises anywhere during knowledge practices" (p 218). They distinguish several types of epistemic risks, and they focus specifically on what they call 'phronetic risks', defined as those risks - which have to be managed in light of values/interests - arising during activities that are part of empirical reasoning (p 220). They claim that inductive risk is just a type of phronetic risk, and that some of the cases Douglas focuses on are in fact phronetic risks.

Let us make clear why in the case of inductive/epistemic risk the direction of influence is from values to science. The important aspect to emphasize is that in this context value-laden choices are catalyzed by uncertainty. Because science always operates in a situation of uncertainty, risks of epistemic errors arise "anywhere during knowledge practices" (Biddle and Kukla 2017, p 218). The consequence is that these risks "need to be managed and balanced in light of values and interests" (p 220). For this reason, value-laden choices are inevitable in science: scientists proceed by balancing and managing risks and uncertainty via values – they go from values to scientific choices and, in a sense, this is inescapable.

Other examples of the 'values-to-science' direction abound, especially in feminist philosophy of science. In particular, one can consider the so-called 'gap feminist empiricism' (Solomon 2012), which starts considerations about the underdetermination of theories by the available evidence, where the "gap between the constraints of empiricism and theory choice is bridged with values" (Solomon 2012, p 439). A classic example of this type of scholarly work is Longino (1990). The direction is from values to science because the impasse of theory choice is resolved by adopting certain values.

There are nuances in the way values have an effect on science, as shown by Ward (2021). First, she points out that the view that (non-epistemic) values *must* enter in the scientific decision process assumes a conception of values as 'justifying reasons' for scientific choices. This influence is quite active rather than implicit: in referring to Winsberg's analysis of climate modeling, Ward says that "a thorough justification of the choices made by modelers would require frequent appeal to values" (p 2021, p 56). Appealing to values to justify scientific

choices comes, of course, with responsibility for endorsing those values in the first place. Second, Ward also talks about values as ‘motivating reasons’, where “motivating reasons need not be conscious or arrived at through deliberation” (p 55). Finally, the direction from values to science implies that there can be a ‘causal’ relationship with respect to choices, in the sense that “values can bring about certain outcomes” (p 56).

2.2 From science to values

The core idea of the second direction (i.e. from science to values) is that scientific choices concerning concepts and methods used have effects on which values are effectively promoted or not. This direction, notice, does not necessarily exclude the other one. In fact, one may motivate or justify a methodological choice by appealing to values, but then the concept and methods used – independently of the values motivating them – can promote other values down the road. We prefer to say ‘promote’ rather than ‘cause’, because whether a value will be endorsed as a result of certain scientific methodologies or concepts will depend also on other factors - therefore, the effect is the ‘promotion’, rather than the value itself³. This direction has not received systematic attention in the literature, with only a few exceptions standing out (and two others that will be discussed in Section 4).

First, there is Elliott’s work (2017). He makes the interesting point that scientists’ choices “support some social values while weakening others” (p 13). He also adds that “it is virtually inevitable that (...) standards of evidence (...) will serve some social values rather than others” (p 99). Details of such claims are offered in Chapter 6 of his book (2017). In particular, Elliott mentions several examples of how a specific way of framing a scientific study, or terminological choices like metaphors, or even the use of certain categories, may result in supporting or reinforcing some values rather than others. Scientists certainly are in the driver’s seat (i.e. by modulating frames and terms we can expect to promote some values rather than others), but there is surely a direction of influence going from certain scientific concepts to the promotion of values that is different from the mere inductive/phronetic risk.

Second, Ward also discusses the direction from science to values. In her words, “values can bring about certain outcomes”, or put it differently, values are (also) effects rather than causes. This is, at a very general level, certainly not surprising, because given “the social authority of science, scientific choices influence a wide range of values, including public health, environmental preservation, and individual and corporate wealth” (Ward 2021, pp 56-

³ We thank a reviewer for noticing this distinction.

57). But the fact that the thesis is not surprising has misled philosophers of science into thinking, as Ward points out, that the thesis is ‘trivial’. However, she explicitly recognizes that how “specific scientific choices advance specific values can be surprising” (p 57).

Third, Lacey, in his book (2005), reconstructs previous work of his own and says “I was interested [...] not only in the impacts of values on scientific methodology, but also [...] in how scientific practices and results may have impact in the realm of values, in how, for example, science may have implications for and contribute to the quest for social justice and human well-being” (p 1). Lacey lays down an argument that is strikingly similar to our own: the second direction of influence, from science to values (to use our terminology) is analysed by showing that certain methodological strategies can indeed reinforce values, which he illustrates through a well-documented case study in biotechnology in agriculture. Even though these authors clearly identify this second direction, their discussions leave several issues open. For instance, the examples provided by Elliott are well developed, but we think that they do not depict a comprehensive picture of *how* science promotes values. In fact, the examples are mostly limited to the evocative nature of certain terms and metaphors and the way things are communicated to the public. Ward’s article seems to suggest that scientific methodologies and seemingly neutral concepts can play such causal roles, but to our knowledge these are not thoroughly discussed. We also rejoin Lacey in hoping that science and technology can promote values such as social justice or human well-being, but our analysis is also to send a warning: even concepts and methods that seem minimally laden with value and maximally neutral, in fact do promote values, and sometimes not those we would like to.

A final and partial overlap has been explored in sub-fields in philosophy of science other than the ‘science and values’ debate, albeit with quite different framing and objectives than our own. In a recent Topical Collection on ‘Reactivity’ published in the European Journal for the Philosophy of Science⁴, various scholars investigate how individuals (or groups) react to social studies and modify their behavior as a consequence of that; in one of these contributions, Godman and Marchionni (2022) explore how individual behavior changes, also in terms of how values change.

⁴ https://link.springer.com/journal/13194/topicalCollection/AC_6c944455d353f163d21e50b3bb4f6a47
accessed 26/11/2023

In sum, Elliott's, Ward's, and Lacey's analyses pave the way to some important questions that nonetheless remain quite open: What is the relation between allegedly neutral foundational concepts and values? Should we characterize this direction as a relation of cause-effect between scientific methods and values, or are there alternative - and weaker - frames? Are there mediators? To what extent can certain methodologies and allegedly neutral concepts be modified and/or modulated to promote the values that we want? These are the questions we aim to answer in the remainder of the paper, and that we think have not been explored systematically enough when the direction 'science-to-values' has been identified.

2.3 The science-to-values direction: insights from feminist economics and philosophy of technology

In this section we develop a general argument for showing how science can indeed promote values. But before formulating the argument, let us specify what the 'science-to-values' direction is not. Our work is not contributing to 'values as evidence' works (Goldenberg 2015), namely that trajectory in feminist philosophy of science looking at how we can investigate values and value-judgements empirically (Anderson 2004; Yap 2016) and, in some cases, how values can be "themselves bearers of empirical content" (Clough 2013, p 1). These works seek to evaluate values not via deliberation of communities, but rather by "using the same empirical models of inquiry used to scrutinize empirical claims" (Goldenberg 2015, p 15). While this strand of research might incorporate a direction going from science (i.e. a certain empirical investigation) to values, we understand it as being connected to the direction of 'values-to-science', given that it is a strategy to investigate criteria "to distinguish legitimate from illegitimate ways of deploying values in science" (Anderson 2004, p 2). Take Anderson (2004) for instance. She explicitly develops the idea of a bidirectional influence of empirical inquiry and values, but in our understanding this is not the two-way relation of science and values we talk about here. 'Bidirectional influence' in Anderson's view denotes the fact that value judgements can guide scientific inquiry, and also that scientific inquiry can tell us a lot about our value judgments, and whether we should change them. We are not concerned with this trajectory; rather, we are interested in how scientific methodologies and concepts can promote certain values, while the empirical nature of value judgements is, in a way, orthogonal to our argument⁵.

⁵ As rightly pointed out by reviewers, this literature occasionally entertains the idea that science can indeed promote values in the sense of our view of the 'two-way relation'. For instance, the already cited Anderson's article on divorce (2004) makes the point that certain conceptions of the object of

In order to understand the ‘science-to-values’ direction consider a well-known discussion in feminist economics and epistemology. Mainstream economics has for a long time not considered housewives' contribution as proper labor, and also did not study in detail how gender permeates several aspects of the labor market (Figart 2005; Jacobsen 2017; Barker 2005). This is typically considered as an example of how science is value-laden; in this case, considerations about what ‘work’ is and is not is laden with bias against women’s labor or deliberately ignoring gender gaps, etc. But one can turn the argument upside down and see this *also* as a case of influence from science to values. The ‘old’ conceptualization of ‘labor’, ‘work’, ‘individual socio-economic roles’, and the methodologies for data analysis that follow, are compatible with, and even promote, values that perpetuate gender discrimination, block any policy to reduce gender pay gaps etc., independently of whether those who promote such methodologies support a value rather than another. Different ways of conceptualizing and studying labor and work should go hand in hand with *other* methodologies for data collections and data analysis, and *can* promote more inclusion and diversity, not just in studying a given phenomenon, but also in the way of intervening on it. Outside the mainstream of economics, reflective and reflexive economics tries to liberate economic thinking from the ‘diktats’ of classic economic theory, and to make it a more diverse and inclusive field (for an overview and discussion, see van Stigt 2017).

Seeing the argument upside down reveals that scientific concepts (in this case, ‘labor’ or ‘work’) are, in fact, promoting some values rather than others: this is the direction from science (in this case, a concept) to values (in this case, discrimination or equality with respect to gender). Because of the nature of these concepts and some of their characteristics, they will promote some values, and this may happen independently of, or in combination with, the biases and the values motivating or justifying the concepts themselves. The reason why we think this direction of influence is important is that scientists (and philosophers) hold a responsibility (moral and epistemic) for the concepts and methods they develop. In this case, given that the results of economists’ work routinely inform socio-economic policies, it is fair to say that we

inquiry (in this case, ‘divorce’) may effectively lead to certain empirical results reinforcing those same conceptions. The value-promotion lies in the fact that the reinforced conception of divorce is ‘thick’, and it has a factual as well as an evaluative negative component (e.g. divorce as a ‘trauma’). There is certainly value promotion here, but we think it is not exactly the same idea that we propose here: in our understanding, there is a promotion of a certain value because the value itself is presupposed in the inquiry, and the value reinforces itself through science - our case is slightly different in saying that there can be promotion of values via specific features of scientific practice, but these features are not obviously related to the (non-cognitive) value dimension that is promoted, as in the case of divorce.

hold a chance to influence, one way or another, what these policies will tackle, depending on how we conceptualize, operationalize, and measure ‘labor’ and ‘work’. Because we may really obtain different results by using different concepts and methods, the discussion of this direction of influence lends support to arguments for methodological pluralism and for diversifying methods and approaches (Russo 2022, ch.5).

To further explain the science-to-values direction, consider an analogy to a widely discussed thesis in philosophy of technology according to which artifacts are not normatively neutral (Winner 1980). Intuitively, one may think that artifacts are, for instance, morally neutral at the design stage, and that it is their specific *use* in given circumstances that makes them subject to moral evaluation. Against this idea, some philosophers argue that any piece of technology (by its very nature *and* by design) affords certain uses while impeding others (Winner 1980; Akrich 1992; Radder 2019). At a general level, artifacts are compatible with certain uses because of their design, where such uses can be bad or good (Vallor 2016); they may also be designed with specific uses in mind, and in virtue of this they may require the realization of certain conditions to work properly, including social and political conditions (Winner 1980). Therefore, it is embedded in the artifacts themselves that they facilitate praiseworthy or blameworthy uses, or that they promote certain desirable or controversial social and political conditions.

To be sure, there is more than an analogy here at stake, because scholars in Science and Technology Studies have long studied and documented the ways in which technologies promote values⁶. However, we illustrate our point through the work of Hans Radder (2019), who has formulated in a precise way the idea of the non-neutrality of technology. He says that a technology can be *inherently normative*, in the sense that “its realization implies one or more norms or normative claims about what to say or do” (pp 56-57). This happens because the stable realization of a technology requires that people should behave in a way that enables - and not disturb - the intended functioning of the technology. In other words, if one wants a given piece of technology to work properly in a given context, then individuals are required to follow certain norms about what one ought to do or not do. Radder formalizes his argument as a conditional: “if we want to realize a working technology, [then] the normative conditions for its successful realization ought to be satisfied” (p 59). Referring to norms can be problematic in this context, given that norms are not values. In order to get rid of the confusion right away

⁶ The contributions of Sheila Jasanoff (e.g. 2016; Jasanoff and Kim 2015) are in this respect pioneering and fundamental.

and go back to ‘value’ and ‘value-promotion’ immediately, we can say that norms (be they epistemic, moral, political, etc) presuppose values, as a norm is supposed to contribute to achieve or realize a *desideratum* (in this case, the well-functioning of a technology). This implies that a technology is not neutral in the sense that its successful realization requires the promotion of certain values, which are presupposed in the normativity described by Radder. The effect of using a certain technology is the promotion of values, rather than the values themselves because, as Radder rightly points out by discussing Winner’s famous argument, the normativity of a given technology should not be understood as a form of technological determinism: the magnitude of the impact of normativity will depend on whether people involved will in fact follow the norms implied by the technology. Moreover, Radder (and Winner) distinguishes between technologies that are contingently normative and inherently normative. In the case of inherent normativity a given technology, to function properly, requires the realization of a certain social organization, which is enacted in the form of norms, which presuppose values. In the case of contingent normativity, a given technology promotes a value, but the promotion depends on something already pre-existing within a certain context.

We bring in these discussions from Philosophy of Technology, because we think that concepts and methods in scientific contexts can promote values in a way that is similar to the way technology can promote values in Radder’s view. At a general level, scientific practice is value-promoting when its correct realization implies one or more normative claims about what to say, do, or think, which in turn presuppose some values. The direction is from science to values: it is science that, by being practiced in a certain way or in virtue of having specific characteristics, promotes a given value through the (metaphorical) enforcement of a norm. The previous discussion of how feminist economics has changed the conceptualization and measurement of labor is a case in point. We give further elements of this direction of influence in the rest of the paper. In Section 3 we show how a scientific concept such as ‘health’ may in fact create the condition for certain value-laden interventions that are more than laden with values: they themselves *promote* some values rather than others, by defining which interventions ought to follow to take place, and that is because it defines the target of that intervention. In Section 4, we show that there can be more subtle ways of promoting values. In particular, a range of methodological choices may promote some values rather than others because there is a compatibility between the methodology itself and certain values that just happen to characterize a certain context. In this case, scientific methodologies and values reinforce each other. We show these dynamics in action within the debate on the use of algorithmic procedures in criminal justice.

In both cases, the ‘value-free’ ideal falls short also in a way other than the one usually acknowledged in the science and values literature: scientific concepts and methods are not just laden with values, but they also promote them. We submit that *value-promotion* arises especially when we consider science for its *action-oriented* aspect, for instance, in cases in which the results of scientific research are meant to inform and shape policies, or other types of decisions. It is fair to say that, often, the consequences of scientific concepts and methods at the level of values fall beyond the direct control of scientists, but we think that this is no reason for scientists to shake responsibility off their shoulders. In fact, our interest in delineating this direction of influence from science-to-values is ultimately to positively contribute to more ‘value-aware’ scientific practices. If our argument stands, concepts and methods should be chosen and used for their inherent scientific aptness, but also depending on which values promote.

3. Value-promoting in the health sciences: scientific concepts

In this section we illustrate how scientific concepts promote value, looking at the health sciences as a paradigmatic example of ‘value-promotion’. We present a case study and the main line of argumentation originally outlined in Russo (2021) and complement it with further considerations that stem from the previous discussion.

‘Health sciences’ refers to a very large class of fields and disciplines that study health and disease. The health sciences include many approaches, from experimental to observational, from clinical to preventive. A core philosophical question is how to conceptualize health and disease. We find numerous approaches here too, from the classic distinction between ‘normal’ and ‘pathological’ of Canguilhem (1991) to the ‘statistics-based’ approach of Boorse (1977). Bioethics and public health ethics have also pointed out that ‘health’ is far from being a neutral concept. However, the normativity of ‘health and disease’ is typically *not* considered at the level of scientific investigation, but in normative contexts such as ethical assessments, health policy, and evaluation of health interventions. In this section, instead, we want to make a stronger claim for the normativity of concepts such as ‘health and disease’, i.e. in the context of generation of scientific evidence and results.

Following Clarke and Russo (2017), we can understand ‘medicine’ as an umbrella term under which to gather very different practices and approaches for the study of health and disease. One way to study health and disease is to study their causes and effects. Here again, causal approaches are fundamentally plural (Clarke and Russo 2017). Thus, for instance, in

studying a phenomenon like obesity, we may be interested in what causes obesity, and also in the effects of obesity on other clinical conditions and/or at the level of public health. It is far from obvious to establish which factors are causes and/or effects, for any given health condition. Some difficulties come from availability of technical and/or experimental tool – for instance, while correlations between asbestos exposure and mesothelioma of the lungs has been long established, *how* exactly asbestos particles lead to the onset of disease and to clinical conditions also requires understanding of the phenomenon at the molecular level (Russo 2019). Other difficulties, however, concern what counts as a proper cause, or proper effect, and socio-economic-political-demographic factors are the objects of a long controversy: are these merely determinants, or proper causes? Obesity well illustrates that the way in which we understand and conceptualize causes and causation in the biomedical contexts is far from being value-free. But while the literature on science and values has shown on multiple occasions how biomedicine is value-laden, here we aim to show how it can also be *value-promoting*.

If our conceptualization of health and disease is very reductionistic in character, allowing *only* for biochemical factors to properly cause health and disease, *then* the consequences at the level of diagnosis, individual treatment, and public health intervention are relevant: we will (and ought to) intervene on these factors only, and exclude *other* factors, such as socio-cultural-economic ones. Conversely, if we hold a conceptualization of health and disease in which *both* biochemical and socio-economic factors can be proper causes, then the situation will be very different in terms of what actions should follow from the knowledge that we establish in the health sciences. Scholars in the 'values in science' literature may reconstruct the case of obesity as yet another example of value-ladenness, in the sense of how certain values would motivate the choice of certain concepts and methods. The point we are making here is orthogonal: independently of what values motivate certain choices, those choices end up promoting some values in turn. Sometimes these values can be the same as the one motivating the choices, other times they are different. But we also need a discussion about value-promotion, because more often than not the scientific literature on obesity, or on any other disease, does not discuss the value-promoting dimension

Causes and effects of obesity are studied in literatures that are often disconnected (Russo 2012), and so it is reasonable to infer that interventions such as food labeling rely on an evidence base that includes the biology, but not socio-economic, sphere of obesity. *If* socio-economic factors are proper causes (and not merely classificatory factors to stratify the studied population) *then* interventions should also tackle these factors in a (more) direct way. At the level of individual treatment, this may mean, *ceteris paribus*, considering interventions that

educate individuals to healthier dietary habits or stress management before jumping to surgical operations to reduce stomach size. At the level of public health intervention, this may mean tackling regulation of the food industry, besides the initiatives about food labeling that are already in place, for instance in the European contexts. Regulating the food industry may mean giving strict indication of what food can and cannot be processed, advertised, or sold, which clearly presupposes a paternalistic stance, as opposed to more libertarian ones.

The problem easily gets outside the scope of science and policy design, and becomes *socio-ethico-political*, in two ways. On the one hand, studying the causes and effects of obesity, and designing interventions, is already an ethico-political exercise, precisely because identifying and acting on some causes rather than others promotes some values rather than others. Specifically, interventions on individual causes (e.g. hormone regulation) or on structural causes (e.g. food industry) are very different and would promote different values, or so our argument. Conceptualizing obesity in terms of psycho-social factors (rather than genetic or bio-chemical factors only) can influence treatment and therapy, for instance promoting peer support, rather than stigma, in the patient's close environment. On the other hand, those who are supposed to identify causes and design interventions may belong to different organizations, institutions, departments, or directorates that may have conflicting interests. Thus, for instance, in the EU context, regulation of the food industry is not part of the directorate of health, but of economic affairs, and the goal of this directorate is to promote the economy, not to protect health. Clearly, none of the actions issued by either the health or economics affairs directorates of the EU will be value-free, and the way in which we conceptualize health and disease carries with it *implicit ethico-political considerations*. Notably, conceiving of disease causation as bio-social promotes views of health as a public good, and therefore indicates that governments do have responsibility in protecting it. Instead, conceiving of disease causation as merely biological in character promotes individual freedom and responsibility over and above collective responsibility. Such considerations are not made explicit, and yet they are at work and they do impact on policies, and ultimately on our lives. But analyses like ours can be instrumental in making more explicit the connection between science and values in a way that promotes value-aware scientific practices.

Our intention here is *not* to argue in favor or against a given conceptualization of health and disease, and of the ethico-political implications that follows, but to signal that, beyond being value-laden, science is *also* promoting values, by defining scientific concepts in specific ways. The health sciences are an easy target to make this point, as the vast majority of the concepts used in the field are subject to this type of evaluation. Another one is 'evidence'.

Evidential pluralists (Parkinnen et al 2018) have long been pleading for enlarging the ‘basket’ of what evidence is in order to include evidence of mechanisms as an important component to establish causal claims. Kelly and Russo (2017) have followed up evidential pluralism to show that conceiving of these mechanisms as *mixed* bio-social mechanisms have very important implications for the design of public health interventions. Andreoletti and Teira (2019), instead, have pointed out that putting the bar for the ‘right amount’ of evidence in drug regulation higher or lower carries with it important implications at the normative level, for instance endowing evidence with a paternalistic stance.

The ways in which science (and the health sciences, for the matter) promote values are multiple. In some cases there may be pretty clear, recognizable intentions, while in others the promotion of some values rather than others may be a ‘side effect’ that escapes any intentional design. Molecular epidemiology, and its focus on biomarkers of disease, well illustrates this point. Since the late 1970s, epidemiologists have been exploring new ways to study exposure and disease. The ‘molecular turn’ in epidemiology has been revolutionary *and* controversial at the same time (Vineis and Russo 2018; Vineis et al 2017; Russo and Vineis 2017). Revolutionary, because it shifted the unit of epidemiological analysis from the ‘individual’ (and aggregation of individuals) to ‘molecules’ (and aggregation of molecules, at the population level). This move has been controversial, because some objected that this approach did not belong to epidemiological thinking altogether. At the same time, the development of molecular approaches in epidemiology has been promoted to redress the hype of genetics and genomics. In order to understand the biochemistry of exposure we need to consider the *whole* class of -omics *as well*: genomics, metabolomics, proteomics, etc. Only in Europe, there has been massive investment in this research programme, and with pretty visible and tangible results. Thanks to projects such as Envirogenomarkers⁷ or EXPOsOMICS⁸, we gained invaluable insights and knowledge into the biochemistry of a number of diseases. However, the promise of informing policies has been far less successful than the basic understanding we achieved, and the reason is very simple: biomarkers are not actionable, at the level of public health. One could say that, unintentionally, fundamental research in molecular epidemiology *has* reinforced a reductionist approach in the health sciences, one in which proper causes are biochemical factors, and that has been redressed in more recent projects such as Lifepath⁹, in

⁷ <http://www.envirogenomarkers.net>

⁸ <https://exposomics-project.eu/live-exposome.pantheonsite.io/index.html>

⁹ <https://www.lifepathproject.eu>

which the biochemical sphere is studied *in the context of* life-course approaches, where the social sphere also has its proper causal role. Our discussion of molecular epidemiology is to show that we should consider the question of how scientific concepts and values may promote values continuously, and also make the effort to anticipate unintended effects. And well-beyond medicine and epidemiology, value-promotion characterizes other disciplines in the life sciences. For instance, Fine (2012) makes a similar point for neuroscience, examining how assumptions based to study possible differences between male or female brains ends up sustaining gender differences that are *de facto* perpetuated in the investigations of neuroscientists. Fine concludes that these ethical implications are important and should be addressed

To sum up, in the health sciences scientific concepts *promote* some values rather than others because the knowledge produced in this field is meant to feed and shape decisions, whether at the public health level or at the level of individual diagnosis and treatment. This case study pointed specifically to ways in which a concept such as ‘health’ promotes value, but analogue arguments can be made about about methods used in this field¹⁰. A case in point here are methods such as potential outcome models, that are based on strong assumptions about manipulability of the causes. These methods have been considered inadequate in certain contexts, for instance when studying health outcomes due to factors such as gender, race, or ethnicity, that are clearly not manipulable. We lack space to develop the argument in full, but the idea is that a methodological approach that does not allow to consider race or ethnicity as proper cause may, in turn, lead to (public health) interventions that do not tackle structural problems, for instance fostering inclusiveness and counteracting racist behaviour.

4. Value-promoting in data science: methods

In the previous section we have shown a case of value promotion in the context of health sciences. In this section, we describe a case study from data science (in particular risk assessment tools), which points to a different facet of ‘value-promoting’ science. In particular, this case shows how scientific methods can promote values by restricting the goals that one can achieve in a specific context, and in doing so it limits users to the values that are reinforced by achieving those goals. Analogue arguments have been made more broadly, specifically about how AI may ‘reinforce and naturalise’ bias, see e.g. (Keyes et al 2021) . In this section,

¹⁰ For a detailed argument, see Chapter 5 of (Russo 2022).

we first show how predicting recidivism can be analysed through the classic lenses of inductive risk and the ‘values to science’ direction, and then gradually show that, in fact, the ‘science to values’ direction is also at work, and we explain which values are promoted through these ML practices.

4.1 The value-ladenness of predicting recidivism

In the past decade, machine learning (ML) algorithms have been increasingly used in the American legal system. Such algorithms are generally used as risk assessment tools for recidivism, in particular to “inform decisions about who can be set free at every stage of the criminal justice system” (Angwin et al 2016). Risk scores are assigned to individuals on the basis of individual features that, by means of algorithmic procedures, are found to be correlated with a certain outcome in a particular (population) sample. The risk score *per se* is just a number, but depending on certain numerical thresholds one can be ‘high risk’ or ‘low risk’. As Pruss notes (2021), in deciding prison sentences, parole, bail amount, etc. these numerical thresholds play an important role. Algorithms for predicting recidivism have been promoted within the context of ‘evidence-based sentencing’, as it is often argued that “using risk assessment algorithms to inform sentences could reduce judge bias” (Pruss 2021, p 1). Because of their ‘algorithmic’, ‘formalized’, and ‘quantitative’ nature, these tools are considered ‘value-free’ and, as long as they are value-free, the argument continues, they should be employed (Pruss 2021). However, algorithmic tools for predicting recidivism are far from being ‘value-free’.

As in the case of the health sciences, the value-ladenness of data science tools can be analyzed also from the standpoint of the traditional direction ‘from values to science’. In fact, the framework of epistemic/inductive risk is a fruitful lens to illustrate how ML tools in the penal system are value-laden. For instance, Johnson (2020) shows how misleading the value-free ideal for algorithmic tools is, and also the central role that inductive risk plays. In another comprehensive article, Biddle (2021) has identified how value-laden decisions are pervasive and take place at almost every single step of an idealized ML pipeline. These value-laden decisions are invoked because of epistemic risk or uncertainty: at each step of that pipeline, there is more than one legitimate technical choice that can be taken; choosing to go in a direction or another comes with risks of failure, and it involves value-judgement of which risks or failures are acceptable and which are not. Biddle talks about ‘tradeoffs’ (e.g. between explainability and accuracy), in the sense that in making a technical choice, a practitioner will endorse a particular value at the expense of another. For instance, in creating a data set that will

be used to train algorithms, ML designers must make judgements calls about the data they intend to collect and in general how the data set must be constructed. Worries about biases are here well-motivated. But eliminating biases is not a value-neutral endeavor, as it reflects judgements about “how much diversity should be reflected in the data, and what sort of diversity is important (...) [t]hese decisions (...) reflect values” (2021, p 6). This problem emerged explicitly in the back-and-forth between ProPublica and COMPAS’s company (Northpointe/Equivant), where one side accused the other of using the wrong notion of ‘fairness’ (Pruss 2021). In terms of tradeoffs, it is the one “between rates of accuracy and error among different groups” (Biddle 2021, p 7), and different conceptions of fairness embedded in predictive tools will reflect different tradeoffs – Verna and Rubin (2018) identify at least twenty distinct conceptions, and it is impossible to satisfy them all at once.

While the analysis in terms of inductive/phronetic risk is rich, here we show that if we see algorithmic tools from the standpoint of value-promoting science (i.e. the direction ‘values to science’), it is possible to identify other important facets of how data science is not value-free, and can promote more value-aware practices.

4.2 From inductive risk to value promotion

Pruss (2021) provides an interesting analysis of the value-ladenness of ML tools in the legal system, which in part departs from the values-to-science direction. First, she makes the convincing case that the use of ML tools presupposes “a formalist interpretation of legal principles – namely, that laws have one correct, mechanically discoverable meaning” (p 3). Second, she shows that risk assessment algorithms blur the line between liability assessment and sentencing. In the American context, liability assessment refers to the choice of a verdict (i.e. guilty or not-guilty), and it is usually done by juries, while sentencing is usually the domain of judges. In fleshing out this second thesis, she notices that the use of risk assessment algorithms like COMPAS, which concentrates on predictive features, “presupposes that the purpose of punishment is consequentialist (crime control) rather than deontological (retributive)” (p 15). Let us call the former CVP (consequentialist view of punishment), and the second DVP (deontological view of punishment). Deontological views of punishment see offenders’ blameworthiness and culpability (such as seriousness of harm and intents) as central to determine punishment, while consequentialist views emphasize the importance of consequences (good or bad) of punishment (Monahan and Skeem 2016) Pruss’ arguments - at first glance - illustrates the idea of a value-promoting science: the use of certain tools and methodologies, at a minimum, presupposes certain morally-charged views such as CVP (which

would be a typical statement of value-ladenness), but in fact it even *promotes* such views. It is also worth noting that Biddle himself seems to entertain similar positions at times. For instance, he mentions that “instruments that use socioeconomic status to predict recidivism might *facilitate* the goal of protecting the public, but they do little to determine what an offender deserves” (p 17, our emphasis), thereby implying that the use of certain tools is indeed compatible with certain (value-laden) conception of justice.

However, both Pruss and Biddle do not take this argument as far as we want to do in this article. For instance, Pruss adds that the CVP turn happens when algorithmic tools take into account ‘morally insignificant’ variables (such as demographic information, familial relationships, etc). This does not make explicit whether value-promotion depends either on the use of the morally insignificant variables, or in the algorithmic tools themselves. Similarly, Biddle emphasizes the importance of variables indicating socioeconomic status and not of the characteristics of algorithmic tools *per se*.

Given the premises of their works, we think that both might agree with our amendments. What we add is a more explicit claim: using ML tools in this context *reinforces* CVP, and it does so *independently of the variables that are taken into account*: it is the very idea of using such tools that comes with a commitment to promote - whether a data scientist wants it or not - CVP. Let us show in detail what we mean.

4.3 How ML tools promote some values and not others

In (2016), Starr argues that in deciding punishment, conceptions of what the purpose of sentencing is plays an important role. The goal of sentencing, Starr continues, is either

- (1) to assign a morally just punishment (which corresponds roughly to DVP), or
- (2) to defend the public from the defendant's future crimes (which corresponds to CVP).

As we noticed in Section 2, in philosophy of technology it has been forcefully argued that any piece of technology is not morally neutral. What scholars in this field mean by this is that certain technologies are compatible, *per se*, with certain moral and political consequences, and hence they promote certain values and impede others *independently* of the intentions behind their use. The same holds for scientific methodologies used in this context: what we claim is that algorithmic tools such as ML are in principle conducive of CVP because they are compatible with CVP; if ML tools are used in the legal context, then they will tend to reinforce CVP. Let us see how, first by understanding the nature of this compatibility, and then why

compatibility means that one reinforces the other. It is important to point out that, as in the case of health and disease, we are not arguing for or against a certain conception of punishment. The compatibility between ML and CVP should be spelled out from the point of view of goals that ML and supporters of CVP are taken to achieve. What we claim is that *the goals* of algorithmic tools such as ML are in principle compatible with *the goals* sought in the penal context by those who support CVP and think that CVP can be reinforced by achieving those goals; therefore, if ML tools are used in the penal context, then they will achieve *goals* that are compatible with goals reinforcing CVP.

Consider ML tools first. The purpose of (most supervised) ML tools is to predict or classify on the basis of past occurrences. The more these predictions or classifications in training are accurate, the more the tools are taken to potentially fulfill their purpose: a virtuous ML tool is one that predicts well, *independently* of the variables it takes into account to predict. In this particular context, the goal is restricted to predicting recidivism.

What about CVP? In CVP, punishment is justified by appealing to its ability to decrease future criminal acts perpetrated by the offender. In other words, CVP are interested in limiting the occurrences in which the public will be exposed to criminal activities. Without a well-vetted ability to assess the profile of individuals when it comes to risk of committing crimes, the crime control effects are difficult to achieve. One way to promote CVP is to have tools that will predict whether a person is likely to be a danger to the public. In other words, those who endorse CVP are interested in limiting the occurrences in which the public will be exposed to criminal activities (i.e. their goal), and one way to achieve this goal is to have tools that predict the extent to which an individual will be a danger to the public (i.e. the predictive tool is the means to achieve CVP's goals).

But ML tools, in principle, will make exactly that kind of predictions if they are used in the criminal justice context with the goal of predicting recidivism. Therefore, what promotes CVP and the goals of ML tools are compatible: one way to reinforce CVP is to *predict* who is going to commit crimes, and the goal of ML is in principle to *predict* well. In fact, the goal sought by those supporting CVP is a typical ML goal (i.e. predicting), instantiated in a specific context (i.e. the legal system), and targeting specific types of prediction (i.e. future crimes). In other words, by using ML in this context, it is implicitly chosen to frame the problem to solve in a way that is amenable to CVP.

The consequences of this compatibility are noteworthy. First, using ML tools in the justice system means implicitly endorsing the tenets and best practices of those who support

extreme forms of consequentialism. This happens even at a basic level, e.g. given the characteristics of ML tools, when it is asked “are ML tools better than judges?”, the question is not “are ML tools better at establishing what people deserve?”, but rather “are ML tools better at predicting who is going to commit a crime than judge?”. Moreover, pursuing the goals of ML alter the perception of the goals that can be possibly achieved, and restrict them to the ones sought by supporters of CVP, thereby obscuring other goals that are more readily connected to other views of punishment. In other words, when we decide to employ ML tools in the legal system (independently of the explicit motivations), we use tools that are designed to achieve certain goals, and these goals will reinforce ideas compatible with CVP (given its focus on predictive tasks because of crime control) independently of whether the dataset we use is biased (if, as it often happens, the data set is biased, we clearly have additional issues and concerns to deal with). ML as a method then *promotes* a certain conception of justice in virtue of the fact that its goals and the goals underpinning that conception of justice *are the same* . To put it more precisely: *given* the specific characteristics of ML tools, *if* such tools are used in the justice system, *then* the notion of punishment and justice that will be promoted is CVP. Please note that the effect is the promotion of CVP, and not CVP itself - in fact, judges can decide to disregard the ML outputs, the tools might not be used in the first place, etc. But this does not change the fact that using ML tools in this context promotes CVP because of their compatibility, which is based on identifiable characteristics of both CVP and ML tools. Identifying more clearly how ML - as a methodology - can reinforce CVP may promote more value-aware ML practices.

Let us close with a further point. The importance of considering this ‘science-to-values’ angle in this context lies in what this angle reveals. ‘Values-to-science’ arguments such as the one developed by Biddle are precise in detailing the values at stake in making certain choices when designing ML tools. But when we consider ML tools in the criminal justice system, then the ‘values-to-science’ angle is already constrained within a CVP perspective. The advantage of complementing analyses like Biddle’s with the ‘science-to-values’ angle is to show that the moment you decide to use ML to address the problem of recidivism, you are already framing the problem in a way that resonates with CVP. Especially for those who explicitly endorse DVP theories (or CVP-DVP hybrid), it can be a powerful argument against the use of ML tools in the first place.

5. Conclusion

In this article, we have advanced the view that science and values stand in a two-way direction. The first direction emphasizes the importance of values in taking methodological and/or conceptual decisions in the practice of science, while the second reveals that concepts or methods can *promote* certain values rather than others. As we have noticed, these two perspectives show that science can be value-laden in two ways. In a first sense related to the literature on ‘inductive risk’, scientists inevitably have to make value-laden decisions in order to fill the uncertainty gap. In the second sense, the value-ladenness is not created by uncertainty; rather, it is a feature of the method or the conceptual apparatus that promotes a range of value-laden consequences. Within the direction from science to values, we have shown two ways in which science can promote values. In section 3, we have discussed a case from the health sciences on how a conceptual apparatus of science can promote some values rather than others, because concepts contribute to the creation and promotion of specific interventions, notably in socio-economic policy or in public health, that in turn are value-laden. This may happen, for instance, because the scientific concept ends up defining the target of the policy itself, or important aspects of such a target. To put this idea in a conditional form, *if* the science that is supposed to inform policies has some characteristics, *then* we ought to shape policies accordingly. This is important especially because the effects of using a concept may just escape the intentions of those who employ the concept in the first place: it just so happens that certain concepts, because of their characteristics, lead to the adoption of certain policies. In section 4, we have discussed how scientific methodologies can also be promoters of values. We have illustrated this point through the use of AI in criminal justice systems. In this case, a scientific method is normative when its implementation implies the realization of an epistemic goal, which is compatible with and supports another goal that is itself value-laden. Briefly put, when one uses a scientific method with specific characteristics, he/she necessarily limits the goals that one can possibly achieve. What the method can in principle achieve (i.e. its goal) directly contributes to the achievement of another goal, where this other goal is usually motivated or supported by specific cultural, moral, and/or social values. To put this idea in a conditional form as the previous one, *if* we use a scientific method in a certain context, *then* we promote certain values by supporting certain goals at the expense of others. We hope that, by drawing attention to this direction of influence from science to values, we can contribute to a more effective discussion of the implications of certain conceptual and methodological choices, as part of usual scientific protocols.

To elaborate on this last point, our interest in the directions of influence from values to science and vice-versa lies in the possibility of contributing to more ‘value-aware’ scientific practices. The philosophical methodology typical of the ‘science and values’ literature proceeds by careful historical, socio-political, and philosophical examination of case studies. We are often able to show that scientists took rather ‘active’ choices, in the sense that they made decisions which, albeit implicitly, were laden with values. Our concern is, instead, with science in the making, where scientists usually do not pay sufficient attention to the ways in which the concept and methods they use can be promoters of values. Ethics guidelines and ethics screenings protocols remain vital, but with this discussion we intend to start a discussion about values in science *from within*. The fact that concepts or methodologies promote values greatly increases their responsibility. Rather than passively accepting concepts and methodologies or - even worse - incorporating them instrumentally, scientists must include in their assessment evaluations of how science can possibly promote some values or others. This requires rerouting responsible conduct of research training in a way that facilitates for scientists the cultivation of abilities such as moral attention (Ratti and Graves 2021) or moral imagination (Brown 2020), which can be fundamental to anticipate how scientific concepts or methodologies promote values. We lack space here to develop the argument in full, but we think that, if our point about ‘value promotion’ stands, then the next steps is to explore synergies with the literature on value-sensitive design (Friedman and Hendry 2019) which, in the field of technology and design, has long provided ways to embed values in the design of technical artifacts. But we miss a ‘scientific’ counterpart, which is in part what our paper contributes to. Some authors have explored the potential of virtue-based ethics and epistemology in the context of teaching and research in science (Bezuidenhout and Ratti 2020; Caniglia et al 2023; Melville and Kerr 2021). The two ways in which our article shows how science can promote values can be used as an inspiration to create ethics exploratory exercise in the context of, for instance, responsible conduct of research training. These exercises can make scientists more familiar with the unintended ethical consequences of their technical choices, and gradually embed this thinking into their own practice. This in turn promotes habituation, which is the backbone of virtue-based approaches (Annas 2011).

Acknowledgment. We would like to thank the audience at a seminar in the Department of Philosophy at the University of Bristol, and at the conference *Philosophy of Science in the Forest* for invaluable insights and comments. Finally, reviewers of EJPS have provided important suggestions on how to improve the claims made in this article.

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>
- Andreoletti, M., & Teira, D. (2019). Rules versus standards: What are the costs of epistemic norms in drug regulation? *Science, Technology, and Human Values*.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias - There's a Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. *ProPublica*
- Annas, J. (2011). *Intelligent Virtue*. Oxford University Press.
- Barker, D. (2005). Beyond Women and Economics: Rereading “women’s work.” *Signs: Journal of Women in Culture and Society*, 30(4).
- 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. (2020). On Predicting Recidivism: Epistemic Risk, Tradeoffs, and Values in Machine Learning. *Canadian Journal of Philosophy*, 1–21. <https://doi.org/10.1017/can.2020.27>
- Biddle, J., & Kukla, R. (2017). The Geography of Epistemic Risk. In K. Elliott & T. Richards (Eds.), *Exploring Inductive Risk - Case Studies of Values in Science*. Oxford University Press.
- Biddle, J., & Winsberg, E. (2010). Value judgements and the estimation of uncertainty in climate modeling. In P. D. Magnus & J. Busch (Eds.), *New Waves in Philosophy of Science* (pp. 172–197). Palgrave-McMillan.
- Boorse, C. (1977). Health as a Theoretical Concept. *Philosophy of Science*, 44(4), 542–573.
- Canguilhem, G. (1991). *The Normal and the Pathological*. Princeton University Press.
- Caniglia, G., Freeth, R., Luederitz, C., Leventon, J., West, S. P., John, B., Peukert, D., Lang, D. J., von Wehrden, H., Martín-López, B., Fazey, I., Russo, F., von Wirth, T., Schlüter, M., & Vogel, C. (2023). Practical wisdom and virtue ethics for knowledge co-production in sustainability science. In *Nature Sustainability*. Nature Research. <https://doi.org/10.1038/s41893-022-01040-1>
- Clarke B. and Russo F. (2017) Causation in medicine. In *Companion to Contemporary Philosophy of Medicine*. Edited by J. Marcum. Bloomsbury, 297-322.
- Clough, S. (2013). Feminist Theories of Evidence and Biomedical Research Communities: A Reply to Goldenberg. *Social Epistemology Review and Reply Collective*, 2(12), 72–76
- Dewey, J. 1993. *The Political Writings*. Edited by Morris, D., and Shapiro, I. Hackett Publishing Company

- Douglas, H. (2000). Inductive Risk and Values in Science. *Philosophy of Science*, 67(4), 559–579.
- Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.
- Elliott, K. (2017). *A Tapestry of Values - An Introduction to Values in Science*. Oxford University Press.
- Elliott, K., & Richards, T. (Eds.). (2017). *Exploring Inductive Risk*. Oxford University Press.
- Figart, D. M. (2005). Gender as more than a dummy variable: Feminist approaches to discrimination. In *Review of Social Economy* (Vol. 63, Issue 3, pp. 509–536). <https://doi.org/10.1080/00346760500255692>
- Fine, C. (2012). Explaining, or sustaining, the status quo? The potentially self-fulfilling effects of “hardwired” accounts of sex differences. In *Neuroethics* (Vol. 5, Issue 3, pp. 285–294). Kluwer Academic Publishers. <https://doi.org/10.1007/s12152-011-9118-4>
- Friedman, B., & Hendry, D. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press.
- Godman, M., & Marchionni, C. (2022). What should scientists do about (harmful) interactive effects? *European Journal for Philosophy of Science*, 12(4). <https://doi.org/10.1007/s13194-022-00493-7>
- Goldenberg, M. J. (2015). How can Feminist Theories of Evidence Assist Clinical Reasoning and Decision-making? *Social Epistemology*, 29(1), 3–30. <https://doi.org/10.1080/02691728.2013.794871>
- Hempel, C. (1965). Science and Human Values. In *Aspects of Scientific Explanation and other Essays in the Philosophy of Science* (pp. 81–96). The Free Press.
- Jacobsen, J. P. (2017). Women and the labor market: A feminist perspective. In *The Oxford Handbook of Women and the Economy* (pp. 623–641). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190628963.013.26>
- Jasanoff, S. 2016. *The Ethics of Invention - Technology and the Human Future*. W. W. Norton
- Jasanoff, S. and Kim, S. (Eds.). (2015). *Dreamscapes of Modernity - Sociotechnical Imaginaries and the Fabrication of Power*. University of Chicago Press
- Jeffrey, R. C. (1956). Valuation and Acceptance of Scientific Hypotheses. *Philosophy of Science*, 23(3), 237–246. <https://doi.org/10.1086/287489>
- Kelly M. and Russo F. (2017) Causal narratives in public health: the difference between mechanisms of aetiology and mechanisms of prevention in non-communicable diseases. *Sociology of Health & Illness*. DOI: 10.1111/1467-9566.12621
- Keyes, O., Hitzig, Z., & Blell, M. (2021). Truth from the machine: artificial intelligence and the materialization of identity. *Interdisciplinary Science Reviews*, 46(1–2), 158–175. <https://doi.org/10.1080/03080188.2020.1840224>
- Kuhn, T. (1977). Rationality, Value Judgment, and Theory Choice. In *The Essential Tension* (pp. 320–339). Chicago University Press.
- Levi, I. (1960). Must the scientist make value judgment? *The Journal of Philosophy*.

- Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- McMullin, E. (1983). Values in Science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 686–709
- Melville, W., & Kerr, D. (Eds.). (2021). *Virtues as Integral to Science Education*. Routledge.
- Merton, R. (1973). The normative structure of science. In *The Sociology of Science - Theoretical and Empirical Investigations*. The University of Chicago Press.
- Monahan, J., & Skeem, J. L. (2016). Risk Assessment in Criminal Sentencing. *Annual Review of Clinical Psychology* 12(1).
- Pruss, D. (2021). Mechanical Jurisprudence and Domain Distortion: How Predictive Algorithms Warp the Law. *Philosophy of Science*, 88(5), 1101–1112.
- Rudner, R. (1953). The scientist qua scientist makes value judgement. *Philosophy of Science*, 20(1), 1–6.
- Ratti, E., & Graves, M. (2021). Cultivating Moral Attention: A Virtue-Oriented Approach to Responsible Data Science in Healthcare. *Philosophy and Technology*, 34(4), 1819–1846. <https://doi.org/10.1007/s13347-021-00490-3>
- Russo F. and Vineis P. (2017) Opportunities and challenges of molecular epidemiology. In G. Boniolo and M.J. Nathan (eds) *Philosophy of Molecular Medicine: Foundational Issues in Research and Practice*. Routledge, ch.12.
- Russo F. (2019) L'esposizione all'amianto causa il mesotelioma? Domande scientifiche e analisi filosofiche. *Diritto e Questioni Pubbliche*, XIX (1), 211-239.
- Russo F. (2021). Public Health Policy, Evidence, and Causation: Lessons from the Studies on Obesity, *Medicine, Health Care, and Philosophy*. 15, 141-151
- Russo, Federica. 2022. *Techno-Scientific Practices: An Informational Approach* Rowman & Littlefield, 2022, 9781786612328
- Solomon, M. (2012). The web of valief: an assessment of feminist radical empiricism. In S. Crasnow & A. Superson (Eds.), *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy* (pp. 435–450). Oxford University Press
- Starr, S. (2016). The Odds of Justice: Actuarial Risk Prediction in the Criminal Justice System. *Chance*, 29(1), 49–51. <https://doi.org/10.1080/09332480.2016.1156368>
- Vallor, S. (2016). *Technology and the Virtues - A Philosophical Guide to a Future Worth Wanting*. Oxford University Press.
- Van Stigt, Arwen. 2017. *A Philosophical Exploration into Qualitative Characteristics of Quantitative and Theoretical Economics*. Universiteit van Amsterdam.

- Veli-Pekka Parkkinen, et al. (2018) *Evaluating evidence of mechanisms in medicine. Principles and procedures*. Springer.
- Vineis P., Illari P., Russo F. (2017). Causality in cancer research: a journey through models in molecular epidemiology and their philosophical interpretation. *Emerging Themes in Epidemiology*, 14:7. DOI 10.1186/s12982-017-0061-7
- Vineis P. and Russo F. (2018) Epigenetics and the Exposome: environmental exposure in disease etiology. In *Oxford Research Encyclopaedia of Environmental Science*. Oxford University Press. DOI: 10.1093/acrefore/9780199389414.013.325
- Ward, Z. B. (2021). On value-laden science. *Studies in History and Philosophy of Science*, 85, 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>
- Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109(1).
- Yap, A. (2016). Feminist Radical Empiricism, Values, and Evidence. *Hypatia*, 31(1).