European Journal for Philosophy of Science (2024) 14:8
https://doi.org/10.1007/s13194-024-00571-y

PAPER IN THE PHILOSOPHY OF THE SOCIAL SCIENCES AND HUMANITIES

Reactivity in the human sciences

Caterina Marchionni[1] · **Julie Zahle**[2] · **Marion Godman**[3]

Received: 1 June 2023 / Accepted: 15 January 2024 / Published online: 8 February 2024

Abstract

The reactions that science triggers on the people it studies, describes, or theorises about, can affect the science itself and its claims to knowledge. This phenomenon, which we call *reactivity*, has been discussed in many different areas of the social sciences and the philosophy of science, falling under different rubrics such as the Hawthorne effect, self-fulfilling prophecies, the looping effects of human kinds, the performativity of models, observer effects, experimenter effects and experimenter demand effects. In this paper we review state-of-the-art research that falls under the remit of the philosophy of reactivity by considering ontological, epistemic and moral issues that reactivity raises. Along the way, we devote special attention to articles belonging to this journal's Topical Collection entitled "Reactivity in the Human Sciences".

Keywords Reactivity · Interactive kinds · Looping effects · Self-fulfilling prophecies · Performativity · Reflexivity · Human sciences

✉ Caterina Marchionni
caterina.marchionni@helsinki.fi

Julie Zahle
julie.zahle@uib.no

Marion Godman
m.godman@ps.au.dk

[1] Practical Philosophy, University of Helsinki, P.O. Box 24, Unioninkatu 40a, SF- 00014 Helsinki, Finland

[2] Department of Philosophy, University of Bergen, Sydnesplassen 12-13, 5007 Bergen, Norway

[3] Department of Political Science, Aarhus University, Bartholins Allé 7, Bygning 1340, 248 Aarhus C, 8000 Aarhus, Denmark

1 Introduction

The reactions that science triggers on the people it studies, describes, or theorises about, can affect the science itself and its claims to knowledge. This phenomenon, which we label *reactivity*, has been discussed in the philosophy of science as well as in many different areas of the human sciences under different rubrics, such as the *Hawthorne effect, self-fulfilling prophecies, the looping effects of human kinds, the performativity of models, observer effects, experimenter effects* and *experimenter demand effects.* In this paper, we adopt a comparatively broad conception of reactivity for the purpose of bringing together the disparate discussions on different forms of reactivity.

We identify two principal junctures at which science and people come into contact and may trigger reactivity. The first is at the stage of *data collection*, when the ongoing research may influence the research participants' behaviour and attitudes in ways that may affect the quality of the data or, relatedly, the quality of the inferences made from it. The second is in connection with the *uptake of scientific results*, when a scientific category, a prediction, or a model is acted upon and may change the behaviour of the relevant agents in ways that either undermine or increase the accuracy of the original claim or representation.

These kinds of interaction between science and people, which we take reactivity to encompass, have been singled out as raising special epistemic worries. Not every reaction of people to science counts as reactivity, however. During data collection, for example, research participants may comply with a researcher's request that they carry out various tasks. Likewise, learning about a scientific result may make people understand the social world better or put them in a better position to control and predict it. Neither of these reactions qualifies as reactivity, as they do not, even in principle, question the epistemic capacities of science. In addition, our focus here is on the human sciences, in other words the social, psychological, and medical sciences that study people. This is not to claim that phenomena analogous to reactivity cannot occur in the study of other subject matters.

Thus understood, reactivity raises ontological, epistemic, and ethical issues. First, insofar as the human sciences change people's behaviour and attitudes more than just fleetingly, ontological questions arise concerning whether these sciences somehow also partly constitute (rather than merely represent and describe) social reality. Second, insofar as human agents are at the heart of the phenomenon (it is about how *people* react to scientists and science), reactivity has often been taken to mark a significant ontological and epistemological difference between the human and the natural sciences. Third, there are worries about what inferences may be drawn from data and about the capacity of scientific representations to capture "a moving target" accurately, to paraphrase Hacking's well-known terminology. Finally, reactivity raises ethical issues, especially concerning the objectives and responsibilities of scientists. Are scientists morally responsible for reactive responses to their science? If so, should the science rather also be evaluated from an ethical as well as an epistemic perspective? Our aim in the remainder of this paper is to review the literature

on reactivity and examine these ontological, epistemic, and ethical concerns.[1] Along the way, we devote special attention to contributions to the Topical Collection *Reactivity in the Human Sciences*, recently published in this journal. As such, our paper also functions as an introduction to the said collection.

2 Reactivity in data collection

Reactivity in data collection is exemplified by an interviewee saying what she thinks the researcher wants to hear rather than what her view on the matter is, and by participants in an experiment adapting their behaviour to what they think is likely to impress the researcher. Although these are uncontroversial examples, there is no agreement among discussants on why exactly these scenarios illustrate the occurrence of reactivity. In other words, there are diverse views on how the phenomenon of reactivity in data collection should be characterised.

Reactivity in data collection has attracted the attention of researchers and philosophers, because its occurrence may affect the quality of the data collected or, relatedly, the quality of the inferences made from data. It is useful to distinguish between two views on the implications of reactivity. One is that its presence in data collection *always* undermines the quality of a study. Consequently, the researcher should at all times try to prevent reactivity from occurring or, if it transpires, attempt to eliminate it. The other is that reactivity in data collection *need not* undercut the quality of a study. Accordingly, a researcher should not necessarily aim to prevent it from occurring and, if it ensues, she should seek to determine whether it detracts from the quality of the study or take steps to ensure that it does not. While the former position – that reactivity in data collection is always problematic – has traditionally been dominant, the latter more tolerant approach, viz. that reactivity is only a problem sometimes, is currently prevalent, at least within the philosophy of science.

In the following discussion of reactivity in data collection we first outline various conceptions of reactivity put forward in the human sciences and philosophy. Next, we zoom in on recent philosophical views on how its occurrence affects the quality of a study. As noted above, these views contend that reactivity need not pose a threat to the quality of data and inferences from them.

Before proceeding, however, we should make one more point. Within the human sciences, the extent to which reactivity has been and remains an issue of concern varies across disciplines, study designs, and methods of data collection. Still, it is fair to say that, overall, reactivity in data collection has long been a widely recognized phenomenon in the human sciences. The situation is different in the philosophy of science. One reason for this is that philosophers have only recently started to concern themselves with the philosophical issues raised by data collection and

[1] The debate on the ethical issues specifically raised by reactivity has only emerged recently, and hence literature on it is still scarcer compared to the other two sets of issues. This explains why we devote less space to such concerns in this paper. We do hope that our discussion will encourage further attention to this very important theme.

scientific practice more generally – issues that were disregarded in traditional philosophy of science. Consequently, the current philosophical preoccupation with reactivity in data collection does not reflect a concern with the demarcation of the human sciences from the natural sciences. Rather, it should be seen as part of, and as contributing to, the reorientation of the philosophy of science towards the entire process of doing science in practice.

2.1 Conceptions of reactivity in data collection

There are at least two standard conceptions of reactivity in data collection articulated in the human sciences and contemporary philosophical debates. They have in common that reactivity is taken to occur when the ongoing research somehow affects the research participants whilst data are being collected about them. They differ in how they further spell out the phenomenon.

According to the *awareness conception*, as it may be labelled, there is reactivity when the research participants' awareness of being studied influences their behaviour or attitudes while data about them are being collected. Proponents of this understanding of reactivity in philosophical discussions include Jimenez-Buedo and Guala (2016) and, in this Topical Collection, Runhardt (2021), Jimenez-Buedo (2021) and Feest (2022).

The conception is sometimes used in the human sciences to articulate or highlight a difference between the study of humans and of the inanimate world, namely that differently from the inanimate world, humans may respond to their awareness of being studied (see e.g., Orne, 1962), thereby coming closer to the view that reactivity is a special challenge for the human sciences. This emphasis is not evident in philosophical analyses of reactivity in data collection, however. It is also worth pointing out that research participants' awareness of being studied should be taken to refer not only to their awareness that research is taking place, but also to their beliefs and desires relating to the researcher, the research setting, and the like.

The other conception, call it the *unintended effect* conception, states that reactivity occurs when the ongoing research has unintended effects on the research participants' behaviour or attitudes while the data are being collected. Jimenez-Buedo (2021) and Zahle (2023), both included in this Topical Collection, are among those adopting this view in philosophical discussions. One thing it brings to light is that research may have both intended and unintended effects on research participants. Research has intended effects when the researcher aims to influence the research participants by, say, prompting them to perform a task as part of an experiment or asking them questions during interviews. Reactivity is then identified with all the effects that the researcher did not intend the research to have on the research participants' behaviour and attitudes.

The awareness and unintended effect conceptions are best described as overlapping. It is accepted in both that reactivity occurs in data collection when the research participants' awareness of being studied has unintended effects on their behaviour and attitudes. Where they disagree, however, is on whether or not it occurs when research participants' awareness of being studied has *intended* effects on their

behaviour and attitudes. For instance, a researcher may, on purpose, appear very enthusiastic about a task so that the research participants will do their best when they perform it. And picking up on this, the research participants make an effort (see Jimenez-Buedo & Guala, 2016:13). Proponents of the awareness conception maintain that reactivity ensues in such situations (because the effects are brought about by the research participants' awareness of being studied), whereas this is at odds with the unintended effect conception (because the researcher intended the effects). Likewise, the conceptions diverge if the research has unintended effects on the research participants' behaviour and attitudes that do *not* go via their awareness of being studied. For example, a researcher may unintentionally influence the research participants' behaviour without their being aware that she is conducting research. In this case, the unintended effect conception holds that reactivity occurs (because the effects are unintended), whereas the awareness conception holds that it does not (because the effects are not a function of the research participants' awareness of being studied).

Both conceptions may be further elaborated, and their specifications regarded as sub-versions. Thus, one may spell out the causes of the reactive behaviours or attitudes, which are often called the mechanisms of reactivity. The causes may be either aspects of the ongoing research (the researcher's characteristics and activities, the setting, and the like) and/or the research participants' beliefs and desires relating to the ongoing research. One may also expand the conceptions of reactivity by detailing different forms of reactive behaviours or attitudes.

It is acknowledged in both the standard conceptions that reactivity occurs when the ongoing research somehow affects the research participants while data about them are being collected. However, broader, divergent conceptions have also been propounded. In this Topical Collection, for example, Uljana Feest characterises reactivity as "a *disposition to react*" (Feest, 2022:3 – italics in the original). Accordingly, in research contexts *any* reaction (behaviour or attitude) displayed by the research participants exemplifies reactivity. It does not matter whether the reaction is a response to an awareness of being studied or whether it is an unintended effect of the ongoing research. Barbara Paterson provides another example of a broader conception of reactivity, which she considers suitable in the context of qualitative data collection, namely "the response of the researcher and the research participants *to each other* during the research process" (Paterson, 1994:301 – our italics). From this perspective reactivity refers not only to the reactions of the research participants to the researcher, but also to their effects on the researcher.

Conceptions of reactivity in data collection are also put forward in the literature under various other headings. These include "the Hawthorne effect" (see e.g., Payne & Payne, 2004, Adair, 1984), "observer effects" (see e.g., Monahan & Fisher, 2010, Risinger et al., 2002), "experimenter demand effects" (Zizzo, 2010), "researcher effects" (see e.g., Monahan & Fisher, 2010), "experimenter effects" (Rosenthal, 1963) and "demand effects of experimentation" (Orne, 1962). Depending on how they are specified, such conceptions are identical to standard conceptions of reactivity, or to broader or narrower versions of them. For instance, in light of John Adair's definition, the Hawthorne effect is the same as reactivity in the standard awareness sense (Adair, 1984:334). Or consider Torin Monahan and Jill Fisher's

characterization of observer effects as occurring when the researcher's presence influences the research participants (Monahan & Fisher, 2010:357). It implies that observer effects may be either intended or unintended and hence it is a broader version of the unintended effect conception of reactivity. Or, to mention one last example, Daniel John Zizzo characterises experimenter demand effects as "changes in behaviour by experimental subjects due to cues about what constitutes appropriate behaviour (behaviour 'demanded' by them)" (Zizzo, 2010:75). This may be regarded as a sub-version of the awareness conception of reactivity in that it refers to behavioural effects of the experimental subjects' awareness of what is demanded of them.

In the above discussion we have not mentioned one element that is sometimes included in characterisations of reactivity, namely that it has a negative impact on the quality of a study. Geoff Payne and Judy Payne, for example, define the Hawthorne effect as people's tendency "to modify their behaviour because they know they are being studied, and *so to distort (usually unwittingly) the research findings*" (Payne & Payne, 2004:108 – our italics). In that such conceptions assume that the occurrence of reactivity in data collection is always detrimental to the quality of a study, they are not useful in terms of establishing that the presence of reactivity need not be problematic. Thus, unsurprisingly, recent philosophical discussions exploring this approach do not rely on them.

2.2 Is reactivity in data collection a problem?

A common thread in philosophical discussions about reactivity in data collection is their rejection of the view that reactivity *always* has a negative effect on the quality of the data collected or, relatedly, on the quality of the inferences made from data. The argument is rather that reactivity is *sometimes* unproblematic, a claim that is supported via the mapping out of circumstances under which it does not adversely affect the quality of a study.

Thus far, most philosophical analyses have focused on reactivity in experiments conducted within the social sciences. The debate here concerns the validity of inferences from experimental data to causal hypotheses (see Jimenez-Buedo, 2015, 2021; Jimenez-Buedo & Guala, 2016, Teira, 2013). Two types of validity are at issue: the first is the internal validity of inferences, that is, the validity of inferences from experimental data to hypotheses about causal relations *within the experimental setting*; the second is the external validity of inferences, that is, the validity of inferences from experimental data to hypotheses about causal relations *in the world outside of the experimental setting*.

At first sight at least, the occurrence of reactivity may seem to pose a threat to both forms of validity. For instance, assume that the reaction of research participants to the experimental treatment is also influenced by their desire to please the experimenter (reactivity). In this case no valid inference from data about their behaviour to a causal hypothesis about the *exact* effect of the treatment on their behaviour can be drawn because the experiment, in itself, does not allow the researcher to the determine that effect. In short, an inference of this sort would lack internal validity. Likewise, an inference from the data to the world outside the laboratory would be invalid

(lack external validity). This is because, given that the behaviour of the research participants was affected by their desire to please the researcher, something similar to the experimental treatment will not give rise to the same kind of behaviour in the outside world where the researcher's influence on it would be non-existent. Considerations along these lines underlie the view that reactivity in experiments is always problematic. Recent philosophical discussions challenge this, however, by providing a more nuanced picture of the practice of experimentation in the social sciences.

In an early contribution to these debates, Teira (2013) maintains that reactivity occurs when the behaviour of research participants is influenced by cues as to what is demanded of them. He argues that such experimenter demand effects are harmless in field experiments if the research participants' guesses about the goal of the experiment do not correlate systematically with the experimenter's true goal (viz. to test a certain causal hypothesis). Accordingly, research participants should be partly blinded, but not deceived, as to the real goal of the experiment. As Teira points out, these claims resemble Zizzo's assertions about laboratory experiments (see Zizzo, 2010). Teira adds that whether researchers are conducting field or laboratory experiments they should always test to see if the partial blinding was successful – simply assuming it was will not do.

In the view of Jimenez-Buedo and Guala (2016), reactivity transpires when awareness of being studied during the data collection process affects the behaviour and attitudes of research participants (the awareness conception). On that basis, they argue that reactivity does not undermine the external (and internal) validity of inferences from experimental data as long as the researcher is cognizant of it. More precisely, they describe the following situation: participants in an experiment pick up on cues intended by the researcher to indicate that she expects them to follow certain norms and, in consequence, the research participants do indeed act in accordance with the norms. This scenario exemplifies the occurrence of reactivity: the research participants are aware of the experimental cues that, in turn, influence their behaviour. Nevertheless, as Jimenez and Guala argue, the external validity of inferences from the resulting data is not undermined insofar as the research participants may encounter similar norms in the outside world: in this case, the experimental data reveal the behaviour that the norms in question prompt in the outside world, too. Thus, for them reactivity is only a problem if it goes undetected, in which case it is uncontrolled and unintended (2016:12–13).

Jimenez-Buedo (2021) elaborates on the above analysis in this Topical Collection by taking up the issue of whether uncontrolled and unintended reactivity does indeed undermine the validity of inferences from experimental data. To this end, she combines the awareness conception of reactivity with the unintended effect conception. Accordingly, she takes it that reactivity occurs when research participants' awareness of being studied affects their behaviour *and* this effect is an uncontrolled and unintended by-product of the experimental intervention (ibid.12). Thus specified, she contends, reactivity is sometimes unproblematic. To begin with, the introduction of an experimental treatment (an intervention) may cause reactive (i.e., unintended and uncontrolled) behaviour that is independent of the experimental effect (the variable to be measured). In this case, reactivity does not pose a threat to the validity of causal inferences from data, and is thus benign. By comparison, reactivity is

malignant when experimental manipulation affects the experimental effect (the variable to be measured) in an uncontrolled or unintended manner. However, malignant reactivity does not pose a problem insofar as the experiment involves both a treatment and a control group, and the malignant reactive effects are identical for the two groups. In this case, the reactive effect may be subtracted via the control group. It is only when the malignant reactive effects in the treatment and control groups are not the same that it undermines the validity of causal inferences. Jimenez-Buedo refers to this as idiosyncratic malignant reactivity.

Feest (2022), in this Topical Collection, also discusses malignant reactivity in the above sense. As noted earlier, she identifies reactivity broadly with a disposition to react. Accordingly, she regards reactivity in the awareness sense as just one form of reactivity in her broad sense. Awareness-reactivity is only a problem insofar as it is malignant, that is, the research participants' awareness of being studied has an impact on the experimental effect (the variable to be measured). In this case, inferences from data to a hypothesis about the outside world lacks external validity. Accordingly, Feest maintains, experimental data should be viewed as reliable or good only if supplemented with several true assumptions, including one to the effect that apart from the treatment, no causal factors such as research participants' awareness of being studied influenced the experimental effect. When these assumptions are confirmed, the researcher has successfully handled the research participants' reaction - reactivity in Feest's broad sense - and may draw externally valid inferences from her experimental data.

As noted above, most philosophical analyses of reactivity in data collection focus on social scientific experiments. More recently, however, the claim that reactivity is in many cases unproblematic has also been extended to data collected through surveys/questionnaires and by way of qualitative methods.

To introduce the issue of reactivity in survey research, consider a research participant who fills out a questionnaire in which she rates her well-being or responds to questions meant to determine whether or not she is depressed. Further, imagine that her answers are affected by the wording of the questions or by an earlier survey she has taken as part of the same study. In short, reactivity ensues, and this is reflected in the research participant's answers; in other words in the data or measurement results.

One possible response to scenarios like these is to maintain that the resulting data do not reflect the research participant's well-being, whether she is depressed or not, and the like and that no inferences from data about a research participant's reported mental states to her actual mental states are possible.

Runhardt (2021), in this Topical Collection, dismisses this view. She takes the above cases to involve reactivity in the awareness sense, insisting that its occurrence is sometimes unproblematic. First, she considers cases in which, say, the wording of a survey measuring a research participant's well-being (or some other psychological phenomenon) affects her answers. Here, she maintains, the data are nonetheless accurate: they correctly reflect the research participant's well-being. To think otherwise would be to disrespect the research participant's authority on how she is faring. Second, Runhardt points to situations in which a survey makes a research participant

redefine the phenomenon being measured such that when she takes the survey the second time, her answers reflect her new conception of the phenomenon. Runhardt argues that with respect to psychological phenomena such as depression, these are only partly constrained by their biological aspects, and as a result there is room for variation in their characterisation. Consequently, it would be to disrespect the research participant's right to have a say on how, say, depression should be defined if the measurement results based on her revised understanding were considered inaccurate. Thus, Runhardt concludes, reactivity in survey research is legitimate when it doesn't undercut the accuracy of the measurement results.

Let us now consider qualitative methods of data collection, such as qualitative interviewing, participant observation and focus-group interviewing. Assume that during an interview, an interviewee says what she thinks the interviewer wants to hear rather than expressing her view on the matter, or that research participants change their behaviour whenever the researcher is around carrying out participant observation. In these scenarios, reactivity transpires in situations in which the researcher is collecting her data.

Again, it might be held, this means that the researcher cannot use her data to make inferences about the research participants' social life independently of their being studied. Zahle (2023), in this Topical Collection, opposes this view, relying on a version of the unintended effect conception of reactivity. In her opinion, good data are not reactivity free (collected in situations devoid of reactivity), but *reactivity transparent*. Data have this feature in combination with true assumptions about whether reactivity occurred in the data collection situation and, if it did, about how the research participants' doings, sayings, and so on, were reactively affected, and/or what caused the reactivity. In light of these assumptions, the researcher may determine in what ways her data are informative about social life independently of its being studied. For instance, if she supplements her data with the true assumption that a research participant's reactive behaviour was caused by her being a woman, then she may see her data as informative about behaviour towards her in her capacity of being a woman (rather than a researcher). As Zahle also notes, there is of course no guarantee that a researcher will always be able to confirm her reactivity assumptions.

As this examination has shown, current philosophical discussions investigate reactivity relative to specific research designs (e.g., experiments) and forms of data collection (e.g., surveys and qualitative methods). Accordingly, the claim that reactivity in data collection does not need to cause trouble is defended only relative to a specific research context.

3 Reactivity in the uptake of scientific results

In the context of scientific results, reactivity occurs when a scientific claim or representation (whether it be a category, a prediction, or a model) is acted upon, thereby changing the behaviour of the relevant people or agents in ways that either

undermine or confirm its epistemic status. The most frequently discussed cases of such reactivity include Hacking's *looping effects of human kinds* (Section 3.1), *reflexive predictions*, which could be both self-fulfilling and self-defeating (Section 3.2), and *the performativity effects of economics* (Section 3.3).

The differences between these three forms of uptake reactivity are not clear cut, but as a first approximation they could be distinguished based on what the scientific claim concerns.

Scientific *classifications* of people facilitate the accumulation of inductive knowledge to make further predictions and generalisations based on an individual's (supposed) membership of a particular kind. Such categories have an essential role in diagnostics aimed at identifying treatments and other interventions. The worry is that people's reactions to the classifications (i.e. reactivity) will undermine the reliability of the diagnosis and the accumulation of knowledge through the destabilisation or inconsistent use of the category (Laimann, 2020).

In the case of a self-fulfilling prophecy, a false scientific *prediction* is made, but in response agents start acting in ways that make it true. As Robert Merton (1948, p.) writes, "the self-fulfilling prophecy is, in the beginning, a false definition of the situation evoking a new behaviour which makes the originally false conception come true". A self-defeating prophecy instead is one that is initially correct but instigates behaviours that end up defeating it. An example of a self-fulfilling prophecy is that of a bank run, which starts with a false rumour of insolvency but then causes the bank to go bankrupt (Merton, 1948). An example of a self-defeating prophecy is when the prediction of an epidemic changes people's social behaviours so that the epidemic is contained as a consequences of the changed behaviours (cf. van Basshuysen et al., 2021).

In terms of performativity, it is the adoption of a *model or theory* by agents that contributes to aligning the phenomenon more closely with the way in which it was originally depicted. Donald McKenzie (2006) describes how the Black-Scholes model for pricing options came to be adopted by traders on the stock market, thereby bringing the price of options observed there in line with the model's predictions.[2]

Some philosophers have developed frameworks that effectively subsume all cases of reactivity caused by the uptake of scientific results (see e.g., Guala, 2016a, who considers them all correlation devices, and Lowe, 2021 who sees them all as cases of increased conformity due to dissemination). The main reason for treating them separately here is that they have provoked philosophical debates that have proceeded largely independently.

[2] Institutional design is often also discussed as a case of performativity. It is unclear whether it should be considered an instance of reactivity as we define it, however, in that it does not chiefly aim at representation. There are also two other debates on performativity whose connection to reactivity we regrettably do not have space to discuss here because they are not chiefly instigated by science or scientists. The first is speech act theory, and in particular John Langshaw Austin's famous claim that performative speech acts are not genuine assertions or truth evaluable statements (1962). The other is Judith Butler's influential account of gender performativity (1990).

3.1 The looping effects of human kinds

Whether we like it or not, people are constantly categorised into kinds: those with different kinds of disease, those who belong to different religions, and those without religious beliefs. People's interactions with classifications and categorisations of people originate and are also pervasive in everyday life without any input of science whatsoever. But the sciences are expected to maintain particularly high standards in this regard. Given the special epistemic authority science has over these categories, they also affect the identity, self-perception, or "ways of being a person" in distinctive ways (Hacking, 2007, p. 285). Scientific researchers ask new questions about a certain category; new hypotheses offer tempting suggestions to people belonging to the category such that they adopt the suggested behaviour; and some uncomfortable scientific truths might also be resisted as a result. In such cases science and science-based policy have not just described reality, but, also, to some extent, changed it. Finally, this change to human kinds also affects science such that the existing category and taxonomy should typically be modified (or at least the generalisations based on the category)[3]. This is reactivity in the case of human kinds and categories.

The fact that certain categories, labels and other linguistic devices can change the way people think of themselves and others is not new. The issue is central in the sociology of mental illness, for example, wherein so-called "labelling effects" are scrutinised and raise some doubts about whether the reality of illnesses is independent of their labels (Scheff, 1974). However, it was Ian Hacking who without doubt set the *philosophy* of reactivity, or what he calls the *looping effects* of human kinds, in motion (1995). Hacking principally sparked a controversy about the ontology of human kinds because he thought the looping effects, or reactivity, could be used to demarcate the human from the natural sciences (whose kinds he did not think undergo such looping). Accordingly, this phenomenon – which Hacking also calls the "making of people" (1986) or "kinds of people as moving targets" (2007) – marks off human kinds from natural kinds.

Although few would deny that there are looping effects in the human sciences, whether this amounts to a mark of the human sciences is contested. According to Cooper (2004) there is also plenty of reactivity in many parts of the biological sciences such as with the speed of selection processes that occur (artificially) in animal breeding. Cooper argues that this continuous interaction between human choices and the selective environment has ongoing effects on practices of classification in much the same way as the cases described by Hacking. If so, the natural or at least the biological world produces cases of looping as well. From the opposite direction, Tsou (2007) argues that despite reactivity in cases such as psychiatry (the chief targets of categorisation discussed by Hacking), there remain many identifiable biological

[3] The case of taxonomic or epistemic modification is, of course, more complicated in cases in which reactivity is self-fulfilling (i.e. should the category really be modified?), but at the very least such cases have a bearing on the *moral responsibilities of* scientists, as discussed in Section 4. We thank the reviewer who made this clear.

regularities also for psychiatric kinds. Thus, in his view the distinction between natural and human kinds is obsolete.

Another writer who has had lots of influence on philosophers' thinking about natural and human kinds in recent decades is Richard Boyd. He argues that human kinds, just like natural kinds, lie on a *continuum* in terms of the projectibility of their properties. He therefore insists that there are very good reasons for considering human kinds on the model of natural kinds (1991: 129). Thus, Boyd's work also implies that there is *no* sharp demarcation between natural and human kinds, or at least not in terms of the epistemic grounds of projectibility. Indeed, one can detect an emerging consensus that, even if human kinds are reactive, they are still projectible, or even *multiply* projectible (capable of supporting multiple inductive generalisations) (see also Godman, 2020).

Although these arguments may cast doubt on the demarcation of the human and the natural sciences, they arguably miss another important point in Hacking's work, namely that there is at least an interesting *human-specific form* of reactivity connected to how a certain (scientific) classification changes the kinds through *awareness of the person classified* or of the surrounding community (2007). This is where the scientific introduction of a new classification simply generates new ways of describing not only existing behaviour, but also novel intentional actions. In her recent defence of this point, Allen (2021) argues that attention to such reactivity may well be necessary to conceive of people as *mistaking* certain behaviours (for something which it is not) and, also, for *faking* certain behaviours to *convince* others that they either are, or are doing something, that they are in fact not (we give an example of this below).

The outcome of some of these debates probably hangs on what the predominant *mechanisms* responsible for human kinds and their reactivity are. According to Mallon (2016), humans tend to conform to what is expected of them and so human kinds tend to self-stabilise. He suggests, for example, that it is often strategic, or at least rational, for members of vulnerable groups or minorities to modify their actions in accordance with those of the majority since they are the ones who determine the reward structure of conforming to a particular behaviour (see also Guala, 2016a). Khalidi (2010) agrees that many human kinds are interactive in that they are susceptible to self-stabilising reactive mechanisms whereby the projectible properties stabilise over time. However, he urges that attention be paid also to important cases of *self-defeating* mechanisms whereby people resist the classifications made about them.

Recently, however, Laimann (2020) takes issue with the assumption that human kinds could be thought of as *either* self-stabilising or self-defeating. She argues, instead, that looping or reactivity renders them fundamentally *capricious*. Consequently, they lack a meaningful sense of projectibility because their members behave in wayward and unexpected ways that defy theoretical understanding. Laimann thereby resurrects Hacking's original demarcation precisely on the grounds that human kinds simply lack projectibility.

Indeed, Ian Hacking was notoriously weary of general claims about looping and human kinds, preferring instead to focus on the particular histories of each case and category: "I do not believe there is a general story to be told about making up people.

Each category has its own history" (1986: 168). Nevertheless, taking a cue from one of his cases might still be helpful in developing a systematic picture of the reactivity of human kinds in its different guises. The case of the "apathetic children", referring to a group of refugee children in Sweden in a particular period between 2003 and 2005, is a relatively recent case discussed in Hacking (2010). These children all developed what appeared to be a rare childhood disorder known as pervasive refusal syndrome (PRS). In fact, it was so rare that it was unheard of among Swedish health care professionals and child psychiatrists. Many of the children gradually withdrew from life: they stopped eating, communicating and moving about, and some even ended up in a comatose state (Ahmadi, 2005).

Hacking's account of this case builds on the fact that the media gave it a lot of attention in the form of images and descriptions of a couple of bedridden children who had suffered from a form of post-traumatic stress, or PRS, since infancy. He then argues that these images and accounts were disseminated within certain refugee families and communities (chiefly from post-Soviet states) whereby senior family members compelled their typically oldest children to simulate this modelled behaviour to secure better chances of asylum. This represents the "ecological niche" that Hacking typically also focuses in discussing "transient mental illnesses" such as mad travelling (1998) and multiple personality disorder (1992). In this case, however, he also focuses on the psychological mechanisms of reactivity that led to this new classification of "Apathetic children": the so-called "imitation & internalization model" (2010). Hacking suggests that many of these children imitated others who really had PRS, but progressively ended up also internalising its typical behaviour patterns. Some children who were diagnosed as "apathetic" at the time have since come forward in the Swedish media confirming that they were indeed subjected to pressure from their parents to simulate the condition in much the same way as Hacking described (Sandstig, 2019; cf. Tamas, 2009).

How should one understand this new kind, namely apathetic children? Is it really different from the original PRS, considered by many to be a psychosomatic disorder (see e.g., Godman, 2013)? Harriet Fagerberg, in this Topical Collection (2022), develops a taxonomy to address the different ways in which kinds of people can undergo reactive effects. First, she explains the difference between kinds that are natural and those that are nominal: it lies in so-called *super-explanatory properties* – a term used to distinguish the privileged "underlying" properties that explain why many other properties correlate or co-occur within a kind (Godman et al., 2020). The correlations of properties or symptoms in the case of diseases constitute the explanandum, and the super-explanatory properties are the explanans. In the case of natural kinds, the super-explanations are natural (chemical, biological, genetic, physiological or neurological), whereas the correlated properties of nominal kinds are super-explained by the introduction and dissemination of a (novel) category in *itself*.

Some human kinds, such as in Hacking's example of mad travelling (1998), might therefore be nominal kinds where any shared properties of the kind are due purely to the existence of the classification, but what about the Apathetic children? Let us remember Hacking's suggestion that the first genuine cases of PRS – those that served as models for imitation and internalisation – should be thought of as

natural kinds because the children really *did* suffer from the somatic condition. It is only in particular ecological niches that the kinds become reactive. This, Fagerberg suggests, is quite typical of disease kinds: they are neither entirely indifferent, screened off from any reactive effects; nor entirely nominal, mere products of classification. Instead, they are natural kinds that undergo reactive effects. She goes on to argue that, for some, the classification and knowledge of the kind only affect the secondary properties and not the basic super-explanatory properties. As an example, she considers breast cancer, super-explained by uncontrolled cell division in breast tissue and hence a natural kind, but where further correlated secondary properties nevertheless emerge due to classification. For example, whether breast cancer is known and stigmatised in society also affects people's disposition to seek out help, and the typical timing of screening and diagnosis. Hence, the mortality rate depends on the different settings in which women develop breast cancer – a highly relevant projectible feature at that. If so, there are biological regularities of breast cancer but also potentially new projectible features of the kind (cf. Laimann, 2020). It is thus both a human and a natural kind, which undermines a strong sense of demarcation between the natural and the human sciences.

Fagerberg (2022) also describes the possibility that super-explanatory properties are reactively affected (whereby reactivity goes all the way down, as it were). She cites Covid-19 as an example, the classification and knowledge of the virus interacting with the genetic, super-explanatory structure. In such cases knowledge of the virus's behaviour prompts restrictions, which prompts a selective environment, which prompts adaptation in the genetic properties of the virus. Perhaps, then, something like this essential dependence also occurred in the case of the apathetic children. In other words, knowledge of the condition generated the classification of "an apathetic child" – and a new model of behaving – which in turn affected the super-explanatory properties of the kind. In this case, however, it seems that it was not a *change* to natural super-explanatory properties; it was simply that the categorisation generated *new* super-explanations.

This change reveals itself in the increase in cases – indeed, in the epidemic characterisation – of the apathetic children. Some cases of the condition continue to be super-explained by means of organic causes (genuine cases of PRS), but among a new group the apparent condition is now super-explained at least in part by the classification (i.e., a nominal kind). Another implication of this analysis is that there are two kinds of apathetic children (PRS and the nominal kind), which should be distinguishable at the level of two different sets of symptoms and hence have different profiles in terms of projectibility. If so, this raises a new set of epistemic challenges for medical professionals and policymakers: what are the differences between the two different kinds, and to which of them does a particular individual belong?

As the case of the Apathetic children illustrates, what often makes reactivity striking is not necessarily the change of properties of a kind, but the prevalence or epidemic character of the phenomenon that undergoes reactive effects. This character of reactivity is also the focus of Riin Kõiv's article in this Topical Collection (2023). Kõiv argues the recent increase and prevalence of obesity is caused by certain scientific claims themselves; namely claims about obesity being genetic. Upon learning about such claims, those that are classified as obese feel like their weight is less

within their control, so they end up shifting their behaviour, expending less effort to control their nutritional intake, thereby favouring those very genes that lead to obesity. Kõiv points out that the same moral holds for much of what science deems to be "caused by genes". As a result, there is a shift in the selective environment in which genes targeted by the scientific claims are favoured. Classifying or describing something as caused by genes then leads to many traits actually be caused by genes.

Kõiv's case is a good illustration of the potentially profound impact reactivity has, but which often risks going unnoticed. It also highlights the importance of *who* delivers the claims and classifications for them to undergo reactivity. Kõiv cites the evidence in Dar-Nimrod et al.s' (2014) study of how people respond to precise scientific claims about the existence of genes associated with obesity: they act as if the trait (obesity) is outside of one's control and is somehow essential to oneself. However, even if we all tend to essentialise (Gelman, 2004), at least most adults do not do so indiscriminately. It rather seems to matter that those who deliver claims about certain traits have the authority to do so. The beliefs that are trusted the most originate in science. But with this authority arguably comes responsibility. This is an issue to which we return in Section 4.

3.2 Reflexive predictions

Let us recall that a self-fulfilling prediction is a false prediction that becomes true as agents get to know about it and act on the basis of that knowledge (and vice versa when the prediction is self-defeating). Early debates about self-fulling prophecies covered the following two questions: Do they pose a problem for scientists? Is their occurrence unique to the social sciences?

According to Merton (1948), the only way of breaking the vicious circle of self-fulfilling prophecies is to abandon the false definition of the situation that set off the vicious circle to begin with. As it might not be easy to convince people that their definition of the situation is indeed incorrect, Merton thought that the solution would often need to rely in deliberate institutional control. For example, in the case of the bank insolvency, legislation may be deployed to avoid spreading panic about insolvency.

Strictly speaking, self-fulfilling prophecies need not have much to do with science. When he delineated the phenomenon Merton was concerned mainly about its social consequences, although he did mention that it marked a difference between the social and the natural sciences. In a classic exchange that occurred a few years later, Grünbaum (1956) and Buck (1963) explicitly debated whether the logic of self-fulfilling prophecy, or *reflexive predictions* in their terminology, did mark a significant difference between the social and the natural sciences.

This is how Buck (1963, 359) defines a *reflexive prediction*.

> A prediction comes true because it comes to the attention of actors on the social scene whose actions will determine its truth-value. Or a prediction turns out false because those same actors become aware of the prediction, and its falsity issues from the actions they are thus led to initiate.

Grünbaum (1956) and Buck (1963) agreed the reflexive predictions did not pose any special problems for the social scientist because the question of whether a prediction operates reflexively can be investigated and possibly addressed. If it turns out that a prediction is indeed reflexive, then it can be corrected by considering how it will affect people's behaviour, or its dissemination can be restricted to safeguard its validity. Where the two authors disagree is on the question of whether the existence of reflexive predictions marks a philosophically interesting difference between the social and the natural worlds. Grünbaum points out that the same dynamics are to be found in the natural world. His example is that of a computer predicting that on its current trajectory a missile will miss its target, communicating this information to the missile in the form of new instructions, and thereby causing it to change its course and hit its target. Buck disagrees, claiming that there is a major difference between the missile following instructions and a person acting on their beliefs. Unfortunately, he does not explain what exactly this difference entails.

Romanos (1973) took this up, arguing that even if Buck (1963) was right and the concept of reflexive predictions should be limited to predictions involving beliefs and actions based on them, this was a difference without consequences, one among many possible mechanisms of reflexivity. The apparent problem of reflexive prediction in science is that it makes genuine testing of a theory impossible. More recently, Kopec (2011) showed that reflexivity makes the genuine testing of a theory difficult within a Bayesian and likelihood-confirmation framework. In the absence of knowledge concerning whether the event predicted by a theory came about as postulated or whether it was the product of the theory's dissemination, observation e cannot be said to provide evidence for the theory. Although Kopec (2011) is right in that this might be a problem, an obvious rejoinder would be to say that scientists could anticipate reflexivity and either devise alternative ways of testing the hypothesis (against alternatives including the reflexivity hypothesis) or revise the theory accordingly. This is the case with responses to the well-known Lucas critique in economics, for example. The idea is that macroeconomic theory should render agents' reactions to economic policies endogenous, meaning that the reactions of (rational) agents to the policy should be explicitly modelled and the prediction corrected accordingly. The question then becomes whether it is, in fact, possible to anticipate reflexivity (namely, whether and how people will react to the prediction or the model), and what is the proper way of addressing it. Grünbaum and Buck would call this a technical rather than a methodological challenge: as it will become clear later on, we believe that it is a technical challenge with significant methodological consequences.

As Northcott (2022) argues in this Topical Collection, the challenge of reactivity is not so much that predictions are reflexive, but rather that it is often hard to predict whether reflexivity will occur in the first place. He describes predictability as a function of both what is known about the causal relations involved and of their features. Some causal relations are very fragile - they hold under very specific conditions. If correct, however, this explains why endogenizing reactions will not always work: there may be too much contingency in people's reactions and their consequences to allow the effective refinement of predictions to take reactivity into account. If Northcott is right, then there is a difference between the social and the natural world: causal relations in the former are generally more likely to be fragile, therefore

reflexivity will be more difficult to predict for the social than for the natural sciences. This comes back to Buck's original suggestion that "beliefs and acting on beliefs" matter. It may well be that it is the fragility of individual-level mechanisms that lies behind the intuition that reactivity is a more pressing epistemic problem for the human and social sciences than it is for the natural sciences.[4]

3.3 Performativity

The idea of economic theory as performative was put forward by sociologist of science Michell Callon, who claimed that "economics, in the broad sense of the term, performs, shapes and formats the economy, rather than observes how it functions" (Callon, 1998: 2).[5] The STS literature on performativity documents several ways in which economics affects the economic world; the epistemic import of performativity remains unclear, however.

In its most comprehensive meaning, performativity is taken to speak in favour of some form of social constructionism: the models and theories of economics do not represent an independent social world, but rather contribute to its creation. Callon's thesis is rather broad, but later contributors sought to narrow it down. MacKenzie (2006), for example, distinguishes between three different forms of performativity in his discussion of the Black and Scholes' formula for option pricing. *Barnesian performativity* is the most interesting of these from a philosophical perspective, and the one that falls more squarely under our conception of reactivity. It is based on the premise that the practical use of an aspect of economics makes economic processes more like their depiction in an economic theory or model. In the case of the Black-Scholes formula, Barnesian performativity holds because the formula did not initially constitute an accurate representation of how options were priced in financial markets but turned into a good empirical description because of its repeated use by financial agents. In Callon's words, the *theory was made true by its application*. This implies that there is no point in talking about true or false theories or models: there is no independent target out there that our theories or models can describe accurately; rather the target is "constituted" by them.

Mäki (2013) argues that performativity is nonetheless compatible with scientific realism about the social sciences (or parts of it): there is still a phenomenon out there independent of scientific efforts at theorising it, even if representations, by being disseminated, may cause changes to it. Accordingly, performativity might be an interesting sociological phenomenon but that does not imply the rejection of a realist outlook on economics or social science more generally. In keeping with Mäki's suggestion, Guala (2016b) reconstructs performativity as a case in which a scientific theory or model functions as a coordinating device; in other words, by being disseminated or used the theory or model creates a set of mutual expectations

[4] On the fragility of individual-level mechanisms, see Steel (2007), for example.

[5] The literature on performativity has tended to focus on economics, but other social theories could also be performative in principle. See, for example, Healy (2015) on the performativity of social network analysis, and van Basshuysen et al. (2021) on epidemiological models.

that solve a given coordination problem. As such, performativity is no different from any other social convention, such as driving on the left side of the road, and a theory or model is one among many possible ways of converging towards an equilibrium in a coordination game. From a metaphysical perspective, Guala agrees with Mäki that there is nothing particularly suspicious in performativity: in itself its existence does not threaten realism about the social world or the capacity of theories and models to represent it.

Taking an epistemological standpoint, Bergenholtz and Busch (2016) argue that as long as self-fulfilling changes can be predicted via a *meta-theory* that explains them, then there is no real threat to realism understood as a claim about the success of social scientific theories in describing the social world. Such a meta-theory is one that "(a) predicts if there is a self-fulfilling impact of the adoptions of (first order) theories, but also (b) identifies the specific theoretical mechanisms that constitute this impact." (Ibid., 36).[6] Clearly, this kind of meta-theory is hard to come by, and Bergenholtz and Bush are aware of this. Theirs is an in-principle argument aimed at showing that "there is nothing in the nature of the phenomena to be theorized about that prohibits such an investigation" (Ibid., 37). However, the problem remains for any theory to determine the extent to which its empirical performance is due to its causal contribution to bringing about the event rather than the event itself.

Hence, even if performative models do not "constitute" their targets in any substantive sense, it remains unclear in what way their performance should be evaluated. First, in some cases it may be impossible to pull apart the evidence that is not affected by the model from the evidence that is. This makes it hard to decide whether the model is descriptively good because it captures its target well, or because it *shaped* the target well. This is again Kopec's problem discussed in Section 3.2 above. Secondly, could the "shaping" of a model in itself be considered a sign of its success? This apparently counterintuitive idea has been explored in some recent contributions to the literature on the philosophy of scientific models.

To address the latter question, let us distinguish two separate cases: when the shaping is the main purpose of a model as is the case in institutional design, and when the shaping is an unintended consequence of the use of a model built with an entirely different epistemic or practical purpose in mind. As van Basshuysen (2022) argues for example, if one purpose of models of market design is that of "performing markets", it should count in favour of a model that the implemented market functions as intended: the model has been practically successful. This raises the question of how to balance practical success with epistemic success when these are in tension. The only suggestion van Basshuysen gives is that scientists should not intentionally *deceive* policy makers and the public. Within this space, however, there is plenty of room for resolving the trade-off between practical and epistemic purposes (and values) in different ways, a topic on which contributions to performativity have only now begun to reflect (see Khosrowi, 2023; Godman & Marchionni, 2022).

[6] Bergenholtz and Busch's proposal comes close to the endogeneizing solution mentioned in the previous section – although their in principle argument in favour of realism does not imply it is possible in practice explicitly to represent the reactions in one's model.

Concerning unintended effects, in a recent paper van Basshuysen et al. (2021) discuss the case of epidemiological models, and in particular policy-influential models built and disseminated during the COVID19 pandemic. Some of these have been criticised for delivering overly pessimistic predictions on the course of the epidemic. van Basshuysen et al. (2021) suggest that, in this and similar cases, predictive failures were not attributable to representational deficiencies, but rather to the fact the models shaped the behaviour of agents in ways that undermined their initial predictions. Insofar as these performative effects contributed to modifying the course of the epidemic for the "better", the argument continues, they should be considered practical successes and hence taken into account in evaluations of the model's performance. Practical success however should not be a relevant criterion at the stage of model construction, let alone justify deceiving the public in order to achieve the desired outcome (van Basshuysen et al., 2021).

In their contribution to this Topical Collection, Vergara-Fernández et al. (2023) argue that model evaluation should factor in performative effects alongside standard epistemic criteria. Having revisited the context of use of the Capital Asset Pricing Model (CAPM), they argue that comprehensively evaluating a model is a matter of identifying two kinds of contribution, namely the epistemic and the practical, and their interaction, and most importantly to embed such contributions in the historical and social environment of its development and use. This amounts to endorsing what they call a *contextual approach to model evaluation*, namely a general framework that not only acknowledges the relevance of reactive effects for the evaluation of scientific models, but also pushes the philosopher to engage with the historical and sociological setting in which exist.

This suggests that, in addition to purely epistemic concerns, pragmatic ones as well as other types of values are relevant to the evaluation of scientific claims and representations.

4 Reactivity and the moral responsibilities of scientists

Ethical concerns arise at both junctures of data collection and uptake of scientific results. Issues such as making sure the research does not harm participants, are well known and are already encapsulated in ethical guidelines covering research in the human sciences. Ethical issues that arise from reactivity following the uptake of scientific results however have yet not been systematically discussed. This is why, in the following, we bring together the sparse literature concerned with the latter.

As Hacking (1995) recognised, the human kinds more likely to be susceptible to looping effects are those that people care about, that they want to be or not to be. Thus, reactive changes are likely to be imbued with values. Along similar lines with regard to performativity, MacKenzie (2006: 275) suggests that: "performativity prompts the most important question of all: what sort of a world do we want to see performed?" In other words, changing the world (alongside representing it) raises questions concerning which changes are desirable and which not, namely questions of value.

Addressing normative questions concerning which changes are desirable is clearly not the sole responsibility of science. There are nevertheless reasons to think that scientists ought to be concerned. In a way, of course, we humans react to claims, reproduce behaviour and reject statements that are made about us in many contexts, such as within social movements, education and personal relationships. Nevertheless, there is a good reason why scientific and related institutions such as the media and, in particular, medical professions and bureaucrats have been the focus of Hacking's work all along (1995). It is because scientific research and its institutions, particularly the medical sciences, have special power and authority when it comes to knowledge about humans and kinds of people (see also Douglas, 2003).

What kind of responsibility and what kind of effects are at issue here? Let us go back to the distinction between the intended and unintended effects of performativity discussed earlier. In the case of intended effects, let us again take the case of market design. It could be argued that if a design is implemented such that the set goals are achieved, it is simply a case of a model doing its job, just like a blueprint succeeds in delivering a working machine (van Basshuysen, 2022). Nevertheless, two kinds of ethical concerns are relevant here from the perspective of scientists. The first, which is discussed as standard practice in fields such as engineering and technology, has to do with the legitimacy of the practical purpose for which something is built. The second is whether it is acceptable for scientists to use the model to bring about outcomes *they* (rather than the policy makers) deem desirable, without being transparent about it. Both kinds of concern fall within the standard ethical guidelines for conducting research.[7]

Cases of unintended effects instead raise dual-use kind of dilemmas concerning whether the epistemic benefits of a new piece of research should trump its possible harmful effects. The question is whether this is a dilemma for scientists. In other words, should scientists be responsible for effects they did not intend to bring about? According to most accounts the answer is yes, but only if they were able to foresee the unintended effects.

However, it may be that reactive effects are so unpredictable that scientists cannot be held responsible for them (Bergenholtz & Busch, 2016; cf. Laimann, 2020). The argument is then that (1) the individual researcher is generally unable to predict the performative effects, and (2) the materialisation of performative effects requires much more than the dissemination or application of the theory or model (Bergenholtz & Busch, 2016).

Godman and Marchionni (2022) reach a different conclusion in their contribution to this Topical Collection, arguing that the extent to which performative effects cannot be predicted has been overstated (see also Northcott, 2022). Moreover, scientists' responsibility need not be tied to the causal contribution of their research to a given harm, but may stem from the epistemic position they occupy. One responsibility for scientific researchers *qua* scientists to assume is to acquire context-specific

[7] The case becomes more complicated, however, if one goes beyond the neat case of 'deception' and enters the realm of choices, such as about the risks that are acceptable (Douglas, 2009).

knowledge of any likely harmful reactive effects. Another is to adopt strategies for mitigating the harm, such as rethinking the way in which scientific claims are communicated outside science.

Koskinen (2022) in this Topical Collection develops an account of how a participatory model might work and generate reactivity that is good for the groups to which the claims pertain. Traditional science of indigenous life and communities has typically had an external role and is also typically highly tainted by ideology and colonial thinking. In contrast, indigenous activist research is based explicitly on the idea of mental decolonisation, such as by gearing research and developing innovations in education to secure the transmission of Sami languages and culture. As Koskinen (2022) points out, when such research succeeds it effectively replaces earlier harmful looping effects with new emancipatory effects. The extent to which it is possible to compromise representational adequacy in favour of beneficial or emancipatory reactive effects remains an open question, however (see van Basshuysen et al., 2021; Khosrowi, 2023).

5 Concluding remarks

We have examined ontological, epistemic, and ethical issues raised in connection with reactivity in data collection and the uptake of scientific results. Whether the potential for reactivity implies that the human sciences constitute or bring about social reality, rather than merely represent and describe it, has been the subject of debates on looping effects and the performativity of models. With regard to human kinds, there have been attempts in the literature to be more careful in distinguishing between cases in which reactivity leads to the fundamental alteration of the kind in question, and when science "merely" changes some of its properties or cultural presentation. Not all reactive changes are alike in ontological terms. Similarly, the sweeping antirealist thesis that all social phenomena are performed rather than discovered does not follow from the fact that models contribute to changing the reality they are about. The interesting issue, then, relates to what a given instance of performativity *implies* for a model and for the phenomenon it both describes and changes.

Although reactivity has traditionally been regarded as epistemically problematic, current contributions maintain that this is not always the case. It has become clear that the extent to which reactivity does constitute an epistemic problem, and how, depends very much on the forms and circumstances of data collection, as well as on the context and consequences of scientific uptake. In some cases, the reactions that data collection may provoke in research participants may indeed adversely affect the quality of the data. In other cases scientists are able to control for, or otherwise take account of, the reactivity to allow reliable inferences from the data. Analogously, when the dissemination of a result is potentially hugely self-fulfilling and its consequences very hard to predict, scientists should monitor and attempt to mitigate reactivity, and not only for epistemic reasons; recent debates in fact have highlighted the possibility that reactivity may raise moral concerns. Yet an analogous conclusion

holds for moral as for epistemic assessments, namely that reactivity should not always be seen as a problem.

All this indicates that the demarcation between the human and the natural sciences is far from the only way of thinking about the philosophical significance of reactivity; reactivity raises plenty of other issues for philosophers to tackle. The variation in its contexts and consequences also casts doubt on the idea that reactivity could constitute a demarcation criterion between the human and the natural sciences: it is often untroubling (epistemically, ontologically and morally), but when it is troubling, it is not for reasons that are unique to the human sciences. Some moral reasons connected to harms might generalise to the study of other non-human mammals, for example, just as some epistemic and ontological concerns about self-fulfilling prophecies might generalise to virology for example. This is not to deny that more care should be applied to sciences that deal with people because of reactivity concerns, but it does not amount to a clear demarcation between the human and the non-human sciences on either moral, epistemic or ontological grounds.

Acknowledgements We would like to thank both the participants of the four reactivity workshops for engaging discussions and the authors to this Topical Collection for the insightful contributions. We are also grateful for the comments of two anonymous reviewers.

Author contributions The authors all contribute equally to this paper.

Funding Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). Independent Research Fund Denmark 9062-00049B (Marion Godman), Academy of Finland (Caterina Marchionni), Joint Nordic research Councils (NO-HS) (all).

Data availability N/A.

Declarations

Ethical approval N/A.

Informed consent N/A.

Conflict of interest N/A.

References

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 69*, 334–345.

Ahmadi, N. (2005). *Asylsökande barn med uppgivenhetssymtom: Kunskapsöversigt och kartläggning [Asylum-seeking children with withdrawal symptoms: An overview]*. Statens Offentliga Utredningar.
Allen, S. R. (2021). Kinds behaving badly: Intentional action and interactive kinds. *Synthese, 198*(Suppl 12), 2927–2956. https://doi.org/10.1007/s11229-018-1870-0
Austin, J. L. (1975). *How to do things with word*s (Vol. 88). Oxford University Press.
Bergenholtz, C., & Busch, J. (2016). Self-fulfillment of social science theories: Cooling the fire. *Philosophy of the Social Sciences, 46*(1), 24–43.
Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies, 61*(1/2), 127–148.
Buck, R. (1963). Reflexive predictions. *Philosophy of Science, 30*, 359–369.
Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.
Callon, M. (1998). The embeddedness of economic markets in economics. In M. Callon (Ed.), *The laws of the markets* (pp. 1–57). Oxford Blackwell.
Cooper, R. (2004). Why hacking is wrong about human kinds. *British Journal for the Philosophy of Science*, *55*(1), 73–85
Dar-Nimrod, I., Cheung, B. Y., Ruby, M. B., & Heine, S. J. (2014). Can merely learning about obesity genes affect eating behavior? *Appetite, 81*, 269–276.
Douglas, H. E. (2003). The moral responsibilities of scientists (tensions between autonomy and responsibility). *American Philosophical Quarterly, 40*(1), 59–68.
Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press.
Fagerberg, H. (2022). Reactive natural kinds and varieties of dependence. *European Journal for Philosophy of Science, 12*(4), 72.
Feest, U. (2022). Data quality, experimental artifacts, and the reactivity of the psychological subject matter. *European Journal for Philosophy of Science, 12*(1), 13.
Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8*(9), 404–409.
Godman, M. (2013). Psychiatric disorders qua natural kinds: The case of the "apathetic children." *Biological Theory, 7*, 144–152.
Godman, M. (2020). *The epistemology and morality of human kinds*. Routledge.
Godman, M., & Marchionni, C. (2022). What should scientists do about (harmful) interactive effects? *European Journal for Philosophy of Science, 12*(4), 63.
Godman, M., Mallozzi, A., & Papineau, D. (2020). Essential properties are super-explanatory: Taming metaphysical modality. *Journal of the American Philosophical Association, 6*(3), 316–334.
Grünbaum, A. (1956). Historical determinism, social activism, and predictions in the social sciences. *British Journal for the Philosophy of Science, 7*, 236–240.
Guala, F. (2016a). *Understanding institutions: The Science and Philosophy of living together*. Princeton University Press. https://doi.org/10.2307/j.ctv7h0sjc
Guala, F. (2016b). Performativity rationalised. In I. Boldyrev & E. Svetlova (Eds.), *Enacting dismal science: New perspectives on the performativity of economics* (pp. 29–52). Springer.
Hacking, I. (1986). Making up people. In T. C. Heller, M. Sosna, & D. E. Wellbery (Eds.), *Reconstructing individualism: Autonomy, individuality, and the self in western thought.* Stanford University Press.
Hacking, I. (1992). Multiple personality disorder and its hosts. *History of the Human Sciences, 5*(2), 3–31.
Hacking, I. (1995). The looping effects of human kinds. In D. Sperber & A. Premack (Eds.), *Causal cognition* (pp. 351–394). Clarendon Press.
Hacking, I. (1998). *Mad travelers: Reflections on the reality of transient mental illnesses*. University Press of Virginia.
Hacking, I. (2007). Kinds of people: Moving targets. *Proceedings of the British Academy, 151*, 285–318.
Hacking, I. (2010). Pathological withdrawl of refugee children seeking asylum in Sweden. Studies in History and Philosophy of Science. Part C, Studies in History and Philosophy of Biological and Biomedical Sciences 41.4: 309–317. Web.
Healy, K. (2015). The performativity of networks. *European Journal of Sociology/Archives Européennes De Sociologie, 56*(2), 175–205.
Jimenez-Buedo, M. (2015). *The last dictator game? Dominance, reactivity, and the Methodological Artefact in Experimental Economics*. International Studies in the Philosophy of Science.
Jimenez-Buedo, M. (2021). Reactivity in social scientific experiments: What is it and how is it different (and worse) than a placebo effect? *European Journal of Philosophy of Science, 11*, 42.

Jimenez-Buedo, M., & Guala, F. (2016). Artificiality, reactivity, and demand effects in experimental economics. *Philosophy of the Social Sciences, 46*(1), 3–23.
Khalidi, M. (2010). Interactive kinds. *British Journal for Philosophy of Science, 61*(2), 335–360.
Khosrowi, D. (2023). Managing performative models. *Philosophy of the Social Sciences*, *53*(1), 371–395
Kõiv, R. (2023). *Genetically caused trait* is an interactive kind. *European Journal for Philosophy of Science, 13*, 31. https://doi.org/10.1007/s13194-023-00527-8
Kopec, M. (2011). A more fulfilling (and frustrating) take on reflexive predictions. *Philosophy of Science, 78*, 1249–1259.
Koskinen, I. (2022). Reactivity as a tool in emancipatory activist research. *European Journal for Philosophy of Science, 12*(4), 65.
Laimann, J. (2020). Capricious kinds. *The British Journal for the Philosophy of Science, 71*(3), 1043–1068. Web.
Lowe, C. (2021). *Self-fulfilling Science*. De Gruyter.
MacKenzie, D. (2006). *An engine, not a camera: How financial models shape markets*. MIT Press.
Mäki, U. (2013). Performativity: Saving Austin from MacKenzie. In B. Karakostas and D. Dieks (Eds.), *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*, 443–53. Springer.
Mallon, R. (2016). *The construction of human kinds*. Oxford University Press.
Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review, 8*, 193–210.
Monahan, T., & Fisher, J. A. (2010). Benefits of 'Observer effects': Lessons from the field. *Qualitative Research, 10*(3), 357–376.
Northcott, R. (2022). Reflexivity and fragility. *European Journal for Philosophy of Science, 12*(3), 43.
Orne, M. Y. (1962). On the social psychology of the psychological experiment: With Particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.
Paterson, B. L. (1994). A framework to identify reactivity in qualitative research. *Western Journal of Nursing Research, 16*(3), 301–316.
Payne, G., & Payne, J. (2004). The Hawthorne effect. In *Key concepts in Social Research* (pp. 108–111). Sage Publications.
Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden problems of expectation and suggestion. *California Law Review, 90*(1), 1–56.
Rosenthal, R. (1963). On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Scientist, 51*(2), 268–283.
Romanos, G. (1973). Reflexive predictions. *Philosophy of Science, 40*, 97–109.
Runhardt, R. W. (2021). Reactivity in measuring depression. *European Journal for Philosophy of Science, 11*(3), 77.
Sandstig, O. (2019). Ohörda rop. Fokus, 70. https://magasinetfilter.se/granskning/apatiska-barn-ohorda-rop/. Accessed 23 Sept 2023.
Scheff, T. J. (1974). The labelling theory of mental illness. *American Sociological Review, 39*, 444–452.
Steel, D. (2007). *Across the boundaries: Extrapolation in biology and social science*. Oxford University Press.
Tamas, G. (2009). *De Apatiska*. [The apathetic] Natur och Kultur.
Teira, D. (2013). Blinding and the non-interference assumption in medical and social trials. *Philosophy of the Social Sciences, 43*(3), 358–372.
Tsou, J. Y. (2007). Hacking on the looping effects of psychiatric classifications: What is an interactive and indifferent kind? *International Studies in the Philosophy of Science, 21*(3), 329–344.
van Basshuysen, P. (2022). *Austinian model performativity*. Forthcoming in Philosophy of Science.
van Basshuysen, P., White, L., Khosrowi, D., & Frisch, M. (2021). Three ways in which pandemic models may perform a pandemic. *Erasmus Journal for Philosophy and Economics, 14*(1), 110–127.
Vergara-Fernández, M., Heilmann, C., & Szymanowska, M. (2023). Contextualist model evaluation: Models in financial economics and index funds. *European Journal for Philosophy of Science, 13*(1), 6.
Zahle, J. (2023). Reactivity and good data in qualitative data collection. *European Journal for Philosophy of Science, 13*(1), 10.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*, 75–98.