Formal models of the scientific community and the value-ladenness of science

Word count: 9,561

Dr. Vincenzo Politi

Faculty of Medicine

Department of Biostatistics

IKlaus TorgŒrds vei 3

0372 Oslo (Norway)

vin.politi@gmail.com

vincenzo.politi@medisin.uio.no

Abstract. In the past few years, several formal models of the social organisation of science have been developed. While their robustness and representational adequacy has been analysed at length, the function of these models has begun to be discussed in more general terms only recently. This paper is a contribution to the general philosophical debate on the formal models of the social organisation of science. Its aim is to understand which view of science these models end up supporting as a consequence of their philosophical presuppositions, which are also in need to be explicated. It will be argued that, because of some of their philosophical underpinnings, current formal models of the scientific community do not explain the internal role that moral and societal values play in scientific research. It will also be discussed whether formal models of the social organization of science can actually capture the value-ladenness of science. At the same time, it will be shown that the discussion on the formal models of the scientific community may contribute in fruitful ways to the ongoing discussions about value judgements in science.

Keywords: Social epistemology of science; Agent-Based Models; Non-epistemic values in science; value-free ideal

1. Introduction

In the past few years, social epistemologists have developed a series of *formal models of the social organisation of scientific research* - or, to borrow from Martini and Pinto (2017), 'SOSR models' - which provide highly idealised representations of the scientific community. The pioneering work on SOSR modelling was made by Kitcher (1990), who developed a 'marginal contribution/reward' (MCR) analytical framework. More recent approaches in SOSR modelling make use of Agent-Based Models (ABM), computer programs in which social groups are represented through a number of individual agents, the rules for their behaviour, and their environment. The simulations run on ABMs explain how, through the interactions of the individual agents, macro-phenomena emerge at the group level (Axelrod 1997; Gilbert and Terna 2000). ABMs have been used by social scientists for a long time. Social epistemologists have applied this tool to study the social aspects of science, such as the impact of knowledge acquisition on subjective beliefs (Zollman 2007, 2010), the optimal division of cognitive labor (Weisberg and Muldoon 2009), the creation of new theories within a scientific community (De Langhe 2014), and various other topics. Overviews on the most discussed SOSR models in philosophy of science have been provided by, among the others, Weisberg (2010) and Reijula and Kuorikoski (2019).

The conspicuous literature generated by SOSR modelling often focuses on some technical aspects of specific models. Recently, however, the debate has begun to shift towards more general issues. Aside from the technical issues, in fact, there are many questions about the function and use of SOSR models in general which deserve to be considered. For example, what is the philosophical underpinning of SOSR models, and what is the philosophical argument that they are trying to make? Which image of science and of the scientific community do they end up supporting? What can philosophers, scientists, and policy makers learn from SOSR modelling? Are computer based methods the best ways to argue about the social organisation of science? This paper is a contribution to the general philosophical debate on SOSR models. Its aim is to understand which view of science these models end up supporting as a consequence of some of their philosophical presuppositions, which are also in need to be explicated.

In the next section, I reconstruct debate around SOSR models. In section 3, I explicate the philosophical presuppositions of these models. The problem will be then to understand to what extent, because of such philosophical presuppositions, SOSR models end up supporting a value-free image of the scientific community. In section 4, I provide a short overview of the current debates on the role of non-epistemic values science. In section 5, I explain in which sense current SOSR models fail to capture the value-ladenness of science. In section 6, I question whether such models can actually capture the value-ladenness of science after all. Finally, in section 7, I consider how reflecting some limitations of SOSR models may actually enrich the current debate on values in science.

2. The current debate on SOSR models

A model is, very generally speaking, an abstraction representing a portion of reality, its entities and their relations. Models can be used to explain some aspects of the target they represent, and may have different forms: mathematical abstractions, as in the case of differential equations, physical objects, as in the case of the double-helix model of the DNA, and so on. SOSR models represent the scientific community, its 'agents' (that is, scientists), and their relations. Like all the other models, they can be used to potentially explain some features of their target, namely the some aspect of the dynamics of scientific communities. At least in the intentions of their developers, the explanations obtained from the SOSR models can be used for making policy suggestions.

SOSR models have had such a 'normative ambition' since their inception. Kitcher (1990) regarded his own work as providing the basis for policy suggestions: "How do we best design social institutions for the advancement of learning? The philosophers have ignored the social structure of science. The point, however, is to change it" (Kitcher 1990: 22). Similarly, Zollman (2007) uses the results of his simulations to suggest which communicative structures are preferable than others in some circumstances. Weisberg and Muldoon (2009) also use the results of the simulations run on their model to make suggestions about which populations of agents best distribute cognitive labor and is therefore preferable to other scenarios.

The proliferation of SOSR formal models has led to a growing philosophical literature which often focuses on some technical issues, such as their robustness, that is their capability of producing outputs which remain consistent even when some of their input parameters are modified. Many contributions to the debate on SOSR models analyse the technical problems and limitations inherent to particular models. Issues with Zollman's epistemic network have been discussed by Alexander (2013), Frey and Šešelja (2019), Rosenstock *et al.* (2017), while the Weisberg-Muldoon landscape model has been criticised, among the others, by Alexander *et al.* (2015), Thoma (2015), Pöyhönen (2017), Avin (2019), Bedessen (2019), Hessen

(2019). By questioning and modifying some of the problematic assumptions of a particular model, some philosophers even end up developing their own SOSR model, or at least their own ‘version’ of an SOSR model. For example, Holman and Bruner (2017) have developed an epistemic network in which the number of agents does not remain fixed over time and in which some external factors may play a role in the agents’ decision. The already mentioned Thoma (2015), Alexander *et al.* (2015), and Pöyhönen (2017) have designed more robust versions of the Weisberg-Muldoon epistemic landscape. Grim *et al.* (2013) have combined an epistemic landscape with a social element of communication: in their model, agents do not just explore the landscape, but they also communicate their results with one another, thus forming an epistemic network.

In the past lustrum or so, the focus of the debate on SOSR models has been gradually shifting. Instead of focussing on the technical aspects of particular models, some philosophers have begun to discuss the functions of SOSR formal models *in general*. One of the earliest contributions in this sense is provided by Martini and Pinto (2017) who argue that, contrary to the ABMs employed in the social sciences, SOSR formal models fail to connect with empirical data. As Martini and Pinto rightly remind us, modelling involves (at least) three steps: first, the identification of a problem and the selection of a theoretical framework to tackle it; second, the construction of a model in the light of some specific assumptions; third, the establishment of a connection between the model and the empirical data. In the social sciences, the last step is taken by following a number of strategies, such as parametrization and model validation; in SOSR modelling, however, no step is taken to bridge the gap between models and empirical data. As Martini and Pinto explain, the failure of a model to connect with the empirical data is the failure of that model to adequately represent its target system. In their view, therefore, it is crucial that SOSR modelling will take up their ‘empirical challenge’.

The observations of Martini and Pinto (2017) open the way to questioning the normative aims of SOSR models. Thicke (2019), in fact, argues that if these models are descriptively inadequate, not only because of some of their problematic assumptions (second step of the modelling process), but mainly because they do not connect in any way with the empirical data (third step of the modelling process), then they should not be used to support normative claims about the social organisation of science. Bedessen (2019) also points out that, since they do not capture some important aspects of scientific research, such as pluralism and the co-existence of interlocking objectives and sub-objectives, SOSR models ends up misrepresenting scientific communities and may lead to problematic policy advices.

While some philosophers have questioned the descriptive adequacy and the normative function of SOSR models, others have questioned *whether* their actual function is to provide adequate descriptions and to make normative claims after all. Frey and Šešelja (2018) suggest that these models should not be viewed as representations in need of empirical validation, but as formal tools which supplement the empirical basis used within some debates in the history and philosophy of science. Šešelja (2020), furthermore, claims that SOSR formal models are 'abstractions', the function of which is neither representational nor explanatory but, rather, 'exploratory'. She argues that the epistemic function of these models is not to provide plausible explanations of the dynamics of actual epistemic communities, but to explore philosophical questions such as those concerning the theories of scientific rationality.

The philosophical and theoretical function of SOSR models is also advocated by Aydinonat *et al.* (2020), who feel that focussing only on the model-target dyadic relation is limiting. In their view, SOSR models are some of many possible argumentative devices which can be used to defend and reinforce philosophical positions. Following this reasoning, an SOSR model is to be considered successful if it generates further arguments. For example, the Weisberg-Muldoon epistemic landscape *together* with its subsequent versions, which were developed by modifying some of its original assumptions, are all part of the same 'argumentative landscape', each new version of the model corresponding to an argumentative move. As such, and despite the limitations of the initial version, the Weisberg-Muldoon epistemic landscape is a successful SOSR model.

Considering SOSR models as possessing a philosophical and argumentative function, on the one hand, would rescue them from the criticisms about some of their limitations: if their function is *not* to provide an adequate representation of actual scientific communities, then the fact that they are not empirically validated is not a problem anymore. By doing so, at the same time, the idea that these models can be used to justify science-policy recommendations will have to be given up, or at least weakened, despite the fact that such an idea is explicitly held by some of their developers. But even accepting the idea that SOSR models are tools to be used in philosophical arguments, however, there are still a lot of questions to be answered. What is, indeed, the philosophical argument these models are making? Which image of science do they support? To answer some of these questions, I will uncover the philosophical underpinnings shared by many of the current SOSR formal models.

3. The ‘Kuhnian project’

I regard SOSR models as articulations of what I call the ‘Kuhnian project in philosophy of science’, with which I mean something different from the famous model of scientific change developed in *The Structure of Scientific Revolutions* (Kuhn 1962) and in other writings. Regardless of the tenability of that model, it is possible to appreciate Kuhn for establishing a new way of practicing philosophy in which normative claims about science are grounded on a descriptive account of the dynamics of the scientific community. Even though he has been often misinterpreted as claiming that science is irrational, what Kuhn was attempting to do is to develop a new ‘social’ theory of scientific rationality. This is why his philosophy should be regarded as one of the first attempts at a properly social epistemology of science (Wray 2011). As De Langhe (2013) suggests, furthermore, the more recent approaches to the study of the scientific community, such computer simulations, ABMs, network systems, and big data analysis, can all be regarded as an expansion and articulation of Kuhn’s original philosophical method. As I will show, current SOSR models are continuous with the ‘Kuhnian project’ not only because they aim at describing the scientific community, but also because they are developed in an attempt to solve some of the problems Kuhn introduced in philosophy of science (see Table 1).

Kuhn’s interest in the dynamics of the scientific community predates the publication of *Structure*. In his (1959), he described the scientific community as being pervaded by an *essential tension*: on the one hand, there is a majority of scientists who have a rather dogmatic attitude towards the dominant theory, methodology or research programme, and who are engaged in routine work; on the other hand, within the scientific community there is also a minority of ‘divergent’ and iconoclastic scientists, ready to challenge the dominant tradition of research and to develop innovative approaches. Both parties perform vital tasks for science: the dogmatic scientists develop, articulate, and apply the established knowledge, while the fewer divergent thinkers, by working on alternative approaches, guarantee a ‘reserve tank’ of alternative theories and methods to be used in case of crisis. Kuhn’s description of the essential tension supports a normative argument about risk-spreading (D'Agostino 2010). If every member of the scientific community bets on the same theory or on the same problem-solving strategy, then the group as a whole would risk too much: the theory everybody has endorsed may lose the bet after all, while potentially successful and yet ‘non-mainstream’ research strategies would have remained underdeveloped or even completely abandoned. By allowing a minority of divergent scientists to work on alternative theories and methods, the community as a whole hedges its bets and spreads the risk of error. Not only the scientific community is pervaded by an essential tension, but it also ought to be so.

While Kuhn showed why it is rational for the community as a whole to maintain an essential tension between dogmatic and divergent tendencies, he did not explain why it would be rational for individual scientists to adopt the latter. Kitcher (1990) re-elaborates and complements Kuhn's risk-spreading argument in order to solve this problem. His MCR analytical model explains why it is rational for self-interested individual scientists *not* to choose the project with the highest probability of success. The probability of making relevant contributions by working on a research line which does not have the highest probability of success, in fact, can be higher than the probability of making an equally relevant contribution in an overcrowded research line. By choosing a scientific project with a lower probability of success instead, some scientists may actually increase their 'expected utility' in terms of career advancement, visibility, and prestige. The self-interested choices of rational individuals allows the community as a whole to hedge its overall bets. In this way, the cognitive labor necessary to reach the epistemic ends of the community as a whole is divided across different individuals, so that the essential tension is preserved.

While Kitcher's MCR framework complements Kuhn's risk-spreading argument, the more recent computer-based SOSR models were developed by modifying some of its problematic assumptions. One of them is that individual agents possess a complete knowledge of the current state of research, of its internal division in discrete projects, and of the probability of their success. Zollman (2007, 2010) overcomes this difficulty: instead of being all-knowing credit-seekers, the agents within an 'epistemic network' can and do learn from one another, and have their preferences modified accordingly. In this way, Zollman's model provides a refined account of how scientists choose their projects.

In both Kitcher's and Zollman's models, however, agents do not show important individual differences: they are either rational agents calculating marginal contributions, in one case, or Bayesian agents learning from one another, in the other. These models, in short, do not consider that agents possess different attitudes and may be driven by different motivations. The 'epistemic landscape', first developed by Weisberg and Muldoon (2009), overcomes this issue. Their model represents a field of research as a landscape containing two 'peaks', which agents aim at discovering. Agents move from one patch of the landscape to the other in an attempt to find the best path to get at the top of the peaks. In the model, however, different kinds of agents adopt different exploratory strategies: while ‘mavericks’ are adventurous and always choose the paths less travelled, ‘followers’ take less risk and prefer to walk through the paths which have already been opened by others. In this way, the Weisberg-Muldoon model provides a more sophisticated and refined view of the distribution of cognitive labor.

Finally, all the models discussed so far take for granted the availability of scientific theories but cannot explain where new theories come from. De Langhe (2014), who explicitly manifests his debts towards Kuhn, develops a ‘unified’ ABM which shows how the distribution of cognitive labor leads to both the exploitation of existing alternatives and to the endogenous generation of new theories within the scientific community.

In short, the SOSR models discussed in this paper can be regarded as laying on an ideal continuum originated by Kuhn, with each new model attempting to overcome some of the limitations of the previous models by modifying some of their assumptions. To use the terminology of Aydinonat *et al.* (2020), all these models belong to the same ‘argumentative landscape’. It remains to be seen whether, in virtue of their Kuhnian roots, these models also end up supporting a value-free image of the scientific community.

Table 1: the development of the Kuhnian project in philosophy of science

	Problem	Name of the Model	Type of Model
Kuhn	Showing why it is rational for the scientific community as a whole to spread the risk by maintaining internal diversity	Essential Tension	Generalisation from personal experience
Kitcher/Strevens	Resolving the tension between collective and individual rationality	Marginal Contribution Model	Analytical framework
Zollman	Explaining the mechanisms of information transmission within the scientific community	Epistemic Network	Computer simulation
Winsberg & Muldoon	Showing the optimal division of cognitive labor for epistemic progress	Epistemic Landscape	ABM
DeLanghe	Showing the optimal division of cognitive labor for balancing the exploitation of existing theories and the exploration of new theories	Unified model	ABM

4. Value-laden science

Even though he is widely known for acknowledging the social dimension of scientific research, Kuhn regarded himself as an internalist philosopher of science. In his view, social factors may occasionally influence the direction a research may undertake, or accelerate the rate of scientific change within a particular field, but they do not play a relevant internal role in science. For Kuhn, scientific research activity is ultimately driven by epistemic factors alone (Bird 2000, 2013; Wray 2010, 2011). That Kuhn developed a 'social epistemology of science', in other words, does not make him a 'sociologist of science'. In fact, he promoted an insular model of the scientific community mainly devoid of immediate social concerns. For Kuhn, science makes progress by solving the problems dictated by a dominant paradigm, which sets the research agenda and contributes to isolate the scientific community from the rest of society: "A paradigm [...] can even insulate the [scientific] community from those socially important problems that are not reducible to the puzzle form" (Kuhn 1993: 37). In some passages, Kuhn even seems to suggest that scientific progress *depends* on the segregation of the scientific community from the rest of society, especially when he applauds "the unparalleled insulation of mature scientific communities from the demands of the laity and of everyday life" (Kuhn 1993: 164).

For Douglas (2009: ch. 3), Kuhn's insular conception of the scientific community led to the establishment of the so-called *value-free ideal*. While they admit, like Kuhn, that non-epistemic values (i.e., social, political, or moral values) can be important 'external factors' for science, for instance in the agenda-setting stage, supporters of the value-free ideal claim that such values ought not to play any 'internal role'; for example, they are not admitted in the evaluation of empirical evidence or in the justification of scientific theories (Sober 2007; Betz 2013; Hudson 2015). The main motivation behind the value-free ideal is the defense of the epistemic authority of science: should it be driven by social, political, or ethical values, then science would no longer be regarded as providing impartial and objective knowledge.

Several philosophers have questioned the feasibility of the value-free ideal and have argued that non-epistemic values have a legitimate internal role in science: in their view, science is *value-laden* (Longino 1990, 1995; Kitcher 2001, 2011; Kourany 2010; Solomon 2001). Some argue that non-epistemic values are necessary when dealing with the so-called *inductive risk* (that is, the risk of making errors which may lead to undesirable consequences for society) inasmuch as they do not replace evidence: their role would be, therefore, internal but 'indirect' (Douglas, 2009; Biddle, 2013). The very direct/indirect distinction has been questioned (Elliott, 2011, 2013, 2017), while some philosophers have argued that sometimes non-epistemic values may even have a priority with respect to evidence (Brown, 2013). Other philosophers have put into question the validity of the very distinction between epistemic and non-epistemic values, enriched the taxonomy of values, and studied their interplay and their functions in questioning problematic background assumptions (Anderson, 1995, 2004; Longino, 1990, 1995, 2002; Solomon, 2001). Since SOSR formal models articulate the Kuhnian project, and since Kuhn appears to hold the value-free ideal, it is necessary to assess whether SOSR formal models also end up holding the value-free ideal of science.

A specification is here in order: describing the scientific community as isolated from society does not necessarily imply the value-free ideal. In principle, it is possible to conceive a scientific community which is separated from the demands of the outside society, but which is nevertheless internally driven by moral and social values. Kuhn characterises the scientific community as being insulated from society *and* also as being internally driven only by epistemic values (with some social factors playing an occasional external role): in this case, it is plausible to consider his view as at least reinforcing the value-free ideal.

Many SOSR formal models also provide a rather insular view of the scientific community, which they represent as a complex but ultimately 'closed' system. The evolution of such a closed system is not influenced by the external environment but, rather, by an internal 'invisible hand' mechanism (Wray 2000). Such a closedness is a rather strong idealisation, since actual social groups, including the scientific community, are influenced by a larger social context and, often, social groups and their wider context co-evolve together. This issue has been noticed, among the others, by Viola (2015) who urges the inclusion of external factors such as the funding mechanisms of science in SOSR modelling. SOSR models also rely on other dubious assumptions: for instance, the number of agents within a scientific community is usually represented as remaining constant. In this way, SOSR models do not realistically represent the turn over of qualified scientists working in scientific research. Moreover, they do not seem to capture those instances in which an agent belonging to an epistemic community may decide to contribute to the achievement of the epistemic goals of *another* community, not to seek credit or recognition but out of her sense of 'epistemic responsibility' (Koskinen 2016).

On the one hand, however, many of the simplifying assumptions of the SOSR models are justified: such models are idealisation with a limited scope, their aim is to account only for the *internal* characteristics of the scientific community. It could be said, therefore, that the existence of external influences on the scientific community is not denied, it is just not represented. On the other hand it is also true that there have been some attempts to relax at least some of these simplifying assumptions. For instance, the epistemic network model developed by Holman and Bruner (2017) is characterised by an 'exit-entry' dynamics, which captures the turn over characteristic of real scientific communities, and represents the external influence of industrial funding mechanisms in agent's decisions. Alvin (2018), instead, develops his own version of the epistemic landscape model which includes an explicit representation of the impact of centralised funding schemes in the distribution of cognitive labor.

In short: although SOSR models do often represent scientific communities as isolated and closed systems, some of their most recent versions show that it is possible to include a representation of the impact of some external factors on the dynamics of such communities. However, in the same way in which an insular representation of the scientific community does not necessarily imply a value-free image of science, so opening such a representation to some external factors does not necessarily make the models able to capture the value-ladenness of science. It is still to be seen, therefore, whether SOSR models can also represent the internal role played by social, moral, political values in science.

5. Do SOSR models represent the value-ladenness of science?

Some SOSR models represent the scientific community in a way which looks very similar to the Kuhnian insular and value-free image. In De Langhe's unified model, for example, both the exploitation of old knowledge and the creation of new theories are produced within the scientific community, which appears to be driven by epistemic factors alone. Other SOSR formal models seem to indirectly address some ethical issues arising from scientific research. Epistemic network models, for example, can be used to study how false or unreliable information spreads throughout the scientific community. Clearly, accepting false theories or bad evidence may have very important social and ethical consequences. Modelling how error or false information can spread through the network, however, is far from capturing the internal role of non-epistemic values in scientific research. Current epistemic networks, in fact, do not show how such moral, social, or political values influence the acceptance of some particular piece of information, or how critical decisions are taken in the face of inductive risk. It is not to be excluded that further developments will lead to an enrichment of epistemic networks in this sense.

There are some SOSR models which at least *prima facie* seem to account for the value-ladenness of science more directly. SOSR models presuppose the existence of an objective (O) and of various methods (M_i) for reaching it. Both Kitcher's MCR framework and the Weisberg-Muldoon epistemic landscape make use of the notion of *epistemic significance* to define O. In both models, the agents are represented as actively engaged in the pursuit of 'significant truths', which are truths that are relevant and useful to particular groups of people with practical interests. To look for significant truths means, in a sense, to work for the 'common good'. The social, moral and political values that scientists use to make judgements are, in a sense, exogenously given in the models through the notion of epistemic significance. By motivating and determining the behaviour of the agents, the externally given values captured by epistemic significance clearly play a role in the internal working of epistemic communities: some would argue, therefore, that this is all SOSR models would need to provide a value-laden image of science. There are at least three reasons for why this line of argument does not work.

First, it is not entirely clear how using the concept of epistemic significance to define the objective of scientific research is sufficient to provide a value-laden image of science. Even the supporters of the value-free ideal of science do not deny that non-epistemic values can play an important role in the research agenda setting. What they deny is that *after* such a research agenda has been fixed, and *after* scientists have begun to work to reach the objective (which in fact may be the discovery of a significant truth), non-epistemic values play no further role in scientific research. In the way in which they use the notion of epistemic significance to define their objective, SOSR models do not seem to propose anything relevantly different from what is endorsed by the supporters of the value-free ideal.

Second, as already explained, SOSR models pre-define *O* as a significant truth and assume that all the agents agree on how to recognise and assess significance. However, *O* could be defined in many other ways. For instance, it could be defined as 'the solution of a theoretical puzzle internal to the scientific community which has little or no relevance for the rest of society'. By modifying the definition of *O* nothing in the internal dynamics of the SOSR models would change: the cognitive labor necessary to reach *O* would still be divided across the agents in the same way. It is important to stress that this issue has nothing to do with robustness. A model is robust when its outputs are not sensible to drastic changes in the value or quantities of the input variables and parameters. For example, to assess the robustness of an epistemic landscape model it is possible to modify the shape and size of its peaks and of the landscape itself, its homogeneity, and so on. If, however, we assume that the peaks of an epistemic landscape do not represent 'significant truths' but, instead, 'solution of a theoretical puzzle internal to the scientific community', then we are not modifying the value of one of the initial parameters, we are just interpreting that parameter differently. The fact that the dynamics of the model does not change by changing how we interpret *O* does not indicate that the model is indeed robust, but only that with or without the notion of epistemic significance the dynamics of the model remains the same. Which in turn means that the concept of epistemic significance does no real job in the model.

Third, for the supporters of the value-laden ideal, non-epistemic values play a role in many aspects of science: not only to fix the objectives of research, as even the proponents of the value-free ideal would agree, but also in the choice of the methods for reaching them, in the assessment of the available evidence, in the way in which scientific results are communicated, and so on. Both Kitcher's framework and the Weisberg-Muldoon model assume that O is a significant truth, but they explain how agents choose how to reach it in different ways. In the first case, by assuming that all the agents use Rational Choice Theory to maximize their expected utility. In the second, even though 'mavericks' and 'followers' have different exploratory attitudes, all the agents follow the rule of moving from lower to higher patches of the landscape. Neither of these models represent how non-epistemic values determine the choice of the methods for reaching O. The problem is that methods of scientific research do not serve only the instrumental function of helping scientists to reach their goals: in many cases, research methods and approaches are value-laden and the choice between them requires critical decisions and value judgements.

Following an example provided by Elliott (2017, ch. 4) and based on an actual case study, let's assume that the O of a research project in the field of agricultural science is to find a way to produce more food in poor countries suffering from hunger issues. Some scientists may decide to pursue O by following M_1, which consists in the use of biotechnology for the production of genetically modified food. Other scientists prefer to adopt M_2, which consists in studying the characteristics of the local land in order to develop and implement better agricultural strategies. The probability of success of M_1 is higher than the probability of success of M_2. It must also be said that working on M_1 may be more profitable in financial terms, since it may lead to the development of new techniques which could be patented and sold to biotech companies. On the one hand, the scientists who choose M_2, from their part, are not motivated by the probability of future rewards, but by the need to avoid methods which they perceive as leading to environmental issues, such as the impoverishment of the terroir. For them, in short, M_1 is simply too risky, if not plain dangerous for the local communities. On the other hand, however, some of the scientists choosing M_1 are not motivated by the higher probability of success of the method, by the mere possibility of a financial return, or by a mixture of the two. If not a definitive solution, M_1 can still offer a very quick fix to the problem of hunger. In the view of some of the scientists choosing this method, therefore, what is really too risky is waiting too long in a context in which people's actual life is at perils.

From this example, it is clear that scientists make value judgments not only in deciding what they ought to pursue, but also *how* they ought to pursue it. Furthermore, not only scientists *qua* scientists make value judgments, but they can also endorse different values and, consequently, make different value judgments. In short, like cognitive labor, 'ethical labor' too might be divided across the agents of an epistemic community (Politi and Grinbaum 2020). Current SOSR models do not account for the agents' value-judgments in choosing the methods for reaching the community objectives, even though such decisions may influence the configuration of the distribution of cognitive labor within the whole community.

It must be said that, when he first developed his MCR model, Kitcher (1990) openly recognized its limitations: "appealing to human ambition is only the beginning of the story. Other psychological mechanisms might bring scientists closer to the [optimal distribution of cognitive labor] than they would otherwise have been. Not only many vices from greed to fraud play a constructive role, but community ends may be furthered by more salubrious traits. Perseverance, personal investment, personal and national loyalties, and devotion to political causes may, on occasion, help to close a [community-individual] discrepancy" (Kitcher 1990: 18). However, he has never provided an indication on how to incorporate these other traits into his model. As it stands now, that model depicts scientific research as being driven by a mixture of epistemic reasons (the desire of finding out a piece of truth) and self-regarding interests (the desire of recognition). This means that Kitcher's MCR model does not account for the crucial role that social, moral and political values may play in the distribution of cognitive labor and, therefore, it ends up providing a rather value-free image of the scientific community.

The epistemic landscape model also remains silent about how different agents choose different research strategies as a result of their own value-judgments. In an epistemic landscape, agents are instructed to move from lower to higher patches, because higher patches are epistemically more significant. Nothing, in the model, indicates how they deal with the potential impact of choosing a path of the landscape rather than another (i.e., risk of errors, risk of catastrophic consequences, and so on). It is possible to argue that the epistemic landscape represents moral decisions and not epistemic ones, or even a combination of moral and epistemic decisions, so that each patch of the landscape constitutes one of such decisions. The agents could move through the landscape in different ways, and perhaps they will reach a peak when they find one of a number of morally adequate decisions. This suggestion, however, seems a modified version of the idea that the objective of the community represented by the model has to be defined *a priori* as a significant truth. Such a suggestion, in other words, tells us to interpret the patches of the epistemic landscape in a certain way, without changing any parameters or introducing any new 'moral rule' for the agents. With or without such an interpretation, however, the dynamics of the model does not change.

In short: while some SOSR models explicitly represent the epistemic community as isolated and value-free, others define such a community as aiming at the discovery of epistemically significant truths. As I showed in this section, however, the assumption that the objective of scientific research is epistemic significance is not necessary, it has no real impact on the dynamics of these models, and it does not capture the role of value-judgments in the choice of methods and, consequently, in the distribution of cognitive labor.

6. Can SOSR models represent the value-ladenness of science?

At the end of the previous section I argued that interpreting parts of an epistemic landscape as representing moral decisions would not really capture the value-ladenness of science. What is really needed is not a different way of interpreting some of the assumptions of the SOSR models but, rather, providing the agents of the models with some 'moral makeup' in order to represent how scientists use non-epistemic values to make value judgments. Such a task, however, poses some problems.

As already mentioned, some of the most recent SOSR models are computer based Agent-Based Models (ABMs) like those which are widely employed in the social sciences. ABMs are useful tools for the analysis of problematic social phenomena which, for obvious reasons, cannot be explored empirically. One of the first and most known ABM, for example, provides a formal representation of the group dynamics leading to the phenomenon of ethnic segregation (Schelling 1971). More recent models, just to mention a few, represent ideal scenarios in which criminality can be successfully inhibited (Birks *et al.* 2012), or the dynamics leading to the escalation of radicalization (Neumann 2014). In short, ABMs are often employed by social scientists to represent and frame several social phenomena with clear moral overtones: these models are, in a sense, 'morally charged'.

It is important to stress that what the ABMs used in the social sciences do is to model individual preferences and actions leading to large-scale group phenomena which the individual agents did not necessarily intend. To make an example: the famous model of the production of social segregation developed by Schelling (1971) represents how, as a result of individual preferences on the choice of the neighborhood, a social group ends up segregating sub-groups belonging to ethnic minorities even though none of the agents is programmed to have racist preferences. In this model, residential segregation of ethnic minorities emerges from the complex and re-iterated interactions of virtually non-racist agents: the macro-phenomenon is an unintended consequence of individual actions. Although used as a tool to understand morally charged social phenomena, Schelling's model is not a model of the moral makeup of its individual agents. The claim that science is not value-laden, by contrast, amounts to the claim that scientists consciously apply non-epistemic values to make critical decisions and to act responsibly. The social responsibility of science is not the spontaneous or even accidental property emerging from the interaction of a-moral individuals. Rather, science is socially responsible inasmuch as scientists act ethically. This means that the value-ladenness can be properly integrated in SOSR formal models if such models could represent scientists as moral agents.

The possibility of designing moral agents in computational ABMs is the object of an on-going debate within the AI community. As suggested by Ruvinsky (2008), 'rights', 'liberties', 'duties', and other elements constituting a moral framework can be regarded as parameters which can be implemented in what she calls 'computational ethics'. Such parameters can be used to model agents holding different ethical standpoints. In this way, it is even possible to represent an idealised society in which agents hold different ethical standpoints. Ruvinsky's view is purely theoretical, since she does not clarify how the parameters making up 'moral frameworks' can be implemented in practice into a computational ABM.

An ABM computer simulation which actually implements an ethical theory for the design of its agents has been developed by Mascaro et al (2010). Their ABM represents an evolving world, in which the interacting agents can reproduce and pass some of their traits and behaviors to the next generation. With this model, they study how phenomena such as altruism or suicide emerge and spread across the idealised evolving world. The programmers have modelled the moral agents by relying on utilitarian theories. In fact, they explicitly claim that utilitarianism is the *only* ethical theory which can be successfully implemented in a computer simulation: "In order for computer simulation studies to be informative about ethics, we must adopt a point of view which allows us to measure the outcomes. Utilities are the natural currency for measuring ethical outcomes. Utilities also support a very natural ethical system, namely utilitarianism, the thesis that what action is best collectively is what action is best. Utilitarianism is, in fact, the only ethical system which allows us to measure the outcomes of computer simulations and judge them as better or worse" (Mascaro et al.

2010: 5). By contrast, other ethical theories, such as deontological ethics or virtue ethics, "depend upon the exact semantics of the deontic principles or the virtues, respectively, and incorporating semantic understanding into artificial life simulation in any kind of sophisticated way requires a prior solution to the problem of natural language understanding" (Mascaro et al. 2010, p. 32).

In order to represent scientists as moral agents in an ABM, it is first necessary to specify which ethical theory can be successfully implemented in an ABM but also which ethical theory best describes socially responsible scientists. If, on the one hand, utilitarianism seems a good candidate for modelling moral agents in an ABM, on the other hand it is questionable whether responsible scientists are best described by such an ethical theory. For Jonas (1984), ethically responsible scientists ought to reflect upon the potential impacts of their research and anticipate its consequences: this requires a completely new and never before seen future-oriented ethical theory. Other philosophers have argued that what responsible scientists need is the far more familiar virtue ethics (Grinbaum and Groves 2013). As argued by Mascaro et al. (2010), however, virtue ethics cannot be implemented in computer simulations unless one first develops an AI capable of understanding natural language: this is one of the most debated issues in AI, going under the name of 'symbol grounding problem' (Harnad 1990).

Constructing ABM models with moral agents is not impossible in principle. It is still not clear, however, whether an SOSR formal model would be able to *adequately* represent moral scientists. It is also unclear whether such a model would be able to represent the interplay between the ethical and the epistemic dimension: to describe, that is, how and to what extent scientists are driven by *both* epistemic *and* moral or social values.

7. Value-free models in social epistemology, social-free debates about the value-ladenness of science?

It looks like SOSR formal models do not represent the role and function of moral and social values in science and it is also questionable whether they could provide such a representation. But *should* they? Some would argue that it is not fair to demand that these models must explicitly and necessarily represent the value-ladenness of science. After all, formal models in the social epistemology of science, like those in the social sciences, are not expected to provide a *complete* picture of their target. The value-ladenness of science may just be something that SOSR formal models do not aim to account for. When reflecting upon the general function of SOSR models in general, however, it is important to ask *why* such value-ladenness is something that such models do not aim to account for.

One possible answer is merely pragmatic. Constructing artificial models of the scientific community is hard, it involves a lot of preliminary assumptions, methodological choices, and so on; so far, it has been easier to construct value-free toy-models of science. This, however, does not exclude that in the future some new SOSR model will be able to account for the role of non-epistemic values in science. If this is the answer to why current SOSR models do not include the value-ladenness of science, then the argument of this article amounts to something similar to the 'empirical challenge' launched by Martini and Pinto (2017): it would be, in fact, a sort of 'ethical challenge' or, in less grandiose terms, an invitation to consider the role of non-epistemic values in the agents' decisions.

There is, however, another possible answer. SOSR modellers could say that they are interested in modelling the scientific community as being driven only by purely epistemic values, or by a mix of epistemic values and personal interests, because such factors are the most important internal drivers of the scientific community. This would amount to an explicit endorsement of the value-free ideal, in which case a number of questions may arise: Is the representation of science provided by SOSR models realistic? Is it even desirable? Could these value-free representations of the scientific community be used to advance science policy suggestions? Reflecting on the value-free characters of SOSR models shed a new light on some of the general questions surrounding their use and function.

Finally, it is worth stressing how the reflection upon some of the limits in SOSR models also contributes to highlight some important limits in the current debates on the value-ladenness ideal of science. Despite their problems and limitations, SOSR models have done a great deal to advance our theoretical understanding on a number of issues, such as: the difference between individual and collective rationality in science, the optimization of the distribution of cognitive labor in the scientific community, how communication may influence scientists' choices, the variety of epistemic agents within the scientific community, and so on. Even though they can be challenged under several respects - for failing to connect with empirical data, for not providing an adequate representation of the scientific community or, indeed, for not taking into consideration the value-ladenness of science - these models do provide invaluable philosophical insights about the social organisation of science. If, on the one hand, SOSR models miss the value-laden dimension of science, on the other, the debates about the value-ladenness of science miss the social dimension of actual scientific research. What is usually argued, within such debates, is that, to say it like Rudner (1957), "the scientist *qua* scientist makes value judgments". The problem is: *which* scientist makes value judgments? The value judgment maker invoked in this kind of debate is *the* scientist: an ideally moral and responsible individual driven by both epistemic and non-epistemic values and who takes the right decisions in order not to hurt society. This ideal moral scientist is as much an inadequate representation of actual scientists as the ideal rational agents of the SOSR models. But while SOSR models were developed to illustrate, and possibly resolve, the tension between individual and collective aims, the debate on the role of non-epistemic values in science has not started yet to frame the problem of the tension between individual moral character and collective responsibility. As already mentioned in section 5, and as argued by Politi and Grinbaum (2020), different members of the scientific community may endorse different sets of values and make different value judgements. Not only is it questionable whether all individual scientists endorse exactly the same non-epistemic values all the time, but it is also not obvious that they always *apply* such values in the same way. What is needed, therefore, is not only a claim about the role of non-epistemic values in science, but also a description and an explanation of how critical decisions at the group level are taken in the face of individual variations in the application and ranking of non-epistemic values.

The 'ethical challenge' to SOSR models and the 'social challenge' to the debate on non-epistemic values in science are, in a sense, complementary. Attempting to bridge these two separate branches of the philosophy of science might lead to more complete representations of the scientific community and, consequently, to an increased understanding of its dynamics.

8. Conclusions

Like Kuhn's original framework, SOSR formal models represent the scientific community as a value-free system. While it is possible to justify such a representation by specifying that it is an idealization or an abstraction, it remains to be seen whether such an idealization or abstraction is either useful or even desirable. It could be argued that, since non-epistemic values do play a crucial internal role in scientific research, the failure to capture the value-ladenness of science makes SOSR models descriptively inadequate, if not philosophically problematic. However, integrating the non-epistemic values in the makeup of the agents of the current computer-based SOSR models may be a difficult task. Nevertheless, reflecting upon the limitations of SOSR models to capture the value-ladenness of science also illuminates some limitations in the current discussions about the role of non-epistemic values in science.

References

Aydinonat, N. E., Reijula, S., and Ylikoski, P. (2020). Argumentative landscapes: the functions of models in social epistemology. *Synthese*. In press: https://doi.org/10.1007/s11229-020-02661-9

Alexander, J. M. (2013). Preferential Attachment and the Search for Successful Theories. *Philosophy of Science* 80: 769–782.

Alexander, J. M., Himmelreich, J., and Thompson, C. (2015). Epistemic Landscapes, Optimal Search, and the Division of Cognitive Labor. *Philosophy of Science* 82: 424–453.

Anderson, E. (1995). Knowledge, human interests, and objectivity in feminist epistemology. *Philosophical Topics* 23: 7-58.

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: 1-24.

Anderson, M., and Anderson, S. (2011). *Machine Ethics*. Cambridge: Cambridge University Press.

Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration.* Princeton: Princeton University Press.

Bedessem, B. (2019). The division of cognitive labor: two missing dimensions of the debate. *European Journal for Philosophy of Science*. In press: https://doi.org/10.1007/s13194-018-0230-8

Biddle, J. (2013). State of the field: Transient underdetermination and values in science. *Studies in History and Philosophy of Science* 44: 124-133.

Bird, A. (2000). *Thomas Kuhn*. Chesham: Acumen and Princeton: Princeton University Pres

Bird, A. (2010). Social knowing: the social sense of 'scientific knowledge'. *Philosophical Perspectives* 24: 23-56.

Bird, Alexander. 2013. Kuhn, Naturalism and the social study of science. In Kindi, V. and Arabatzis, T. (Eds.). *Kuhn's Structure of Scientific Revolutions Revisited.* New York: Routledge, pp. 205-23

Birks, D., Townsley, M., and Stewart, A. (2012). Generative explanations of crime: using simulation to test criminological theory. *Criminology* 50: 221-254.

Brown, M. (2013). Values in science beyond underdetermination and inductive risk. *Philosophy of Science* 80: 829-839

Churchman, C. W. (1948). Statistics, pragmatics, induction. *Philosophy of Science*: 249–268.

D'Agostino, Fred. 2010. *Naturalizing Epistemology. Thomas Kuhn and the 'Essential Tension'.* New York: Palgrave Macmillan.

De Langhe, R. (2013). The Kuhnian paradigm. *Topoi* 32: 65-73.

De Langhe, R. (2014). A unified model of the division of cognitive labor. *Philosophy of Science* 81: 444-459.

Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.

Douglas, H. (2014). Pure science and the problem of progress. *Studies in History and Philosophy of Science* 46: 55-63.

Elliott, K. (2011). Direct and indirect roles for values in science. *Philosophy of Science* 78: 303–324.

Elliot, K. (2013). Douglas on Virtues: from indirect roles to multiple goals. *Studies in History and Philosophy of Science* 44: 375-383.

Elliott, K. (2017). *A Tapestry of Values: An Introduction to Values in Science.* Oxford: Oxford University Press.

Frey, D., and Šešelja, D. (2018) What is the epistemic function of highly idealised Agent-Based Models of scientific inquiry?, *Philosophy of the Social Sciences.* In press: https://doi.org/10.1177%2F0048393118767085

Frey, D., and Šešelja, D. (2019). Robustness and idealisations in Agent-Based Models of scientific interactions. *The British Journal for the Philosophy of Science*. In Press: https://doi.org/10.1093/bjps/axy039

Gilbert, N., and Terna, P. (2000). How to build and use agent-based models: alternative and prospects. *Mind and Society* 1: 57-72.

Grim, P., Singer, D., Fisher, S., Bramson, Berger, W., Reade, C., Flocken, C., and Sales, A. (2013) Scientific networks on data landscapes: question difficulty, epistemic success, and convergence. *Episteme* 4: 441-464

Grinbaum, A., and C. Groves. 2013. What is 'responsible' about Responsible Innovation? Understanding the ethical issues. In Owen, R. and Bessant, J. (Eds.) *Responsible Innovation*. Chichester: John Wiley & Sons, Ltd, pp. 119-142.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42: 335-346

Hicks. D. (2014). A new direction for science and values. *Synthese* 191:3271–3295.

Holman, B. and Bruner, J. (2017). Experimentation by industrial selection. *Philosophy of Science* 84: 1008-1019.

Hudson, R. (2016). Why we should not reject the value-free ideal in science. *Perspectives on Science* 24:167-191.

Jonas, H. (1984). *The Imperative of Responsibility. In search of an ethics for the technological age.* Chicago: University of Chicago Press.

Kincaid, H., Dupré, J. and Wylie, A., (Eds.)(2007). *Value-Free Science: ideals and illusions?*, Oxford: Oxford University Press.

Kitcher, P. (1990). The Division of Cognitive Labor. *Journal of Philosophy* 87: 5–22.

Kitcher, P. (2001). *Science, truth, and democracy*. New York and Oxford: Oxford University Press.

Kitcher, P. (2011). Science in a democratic society. New York: Prometheus Books.

Kourany, J. (2010). *Philosophy of science after feminism*. Oxford and New York: Oxford University Press.

Koskinen, I. (2016). Where is the epistemic community? On democratisation of science and social accounts of objectivity. *Synthese*: 4671-4686.

Kuhn, T. S. (1959). The Essential Tension: Tradition and Innovation in Scientific Research. Reprinted in Kuhn, T. (Ed.) *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press., 1977, 225-239.

Kuhn, T. S. (1962 [1996]). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 3rd edition.

Kuhn, T. (1977). Objectivity, Value Judgement and Theory Choice. Reprinted in Kuhn, T. (Ed.). *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press., 1977, 320-339.

Hands, D. (1995). Social epistemology meets the invisible hand: Kitcher on the advancement of science. *Dialogue* 34: 605–621.

Hands, D. (1997) Caveat Emptor: economics and contemporary philosophy of science. *Philosophy of Science* 64: ss107–ss116.

Lewens, T. (2019). The division of advisory labour: the case of 'mitochondrial donation'. *European Journal for Philosophy of Science*. In press: https://doi.org/10.1007/s13194-018-0235-3

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

Longino, H. (2002). *The Fate of Knowledge.* Princeton: Princeton University Press.

Martini, C., and Pinto, M. F. (2017). Modeling the social organization of science: Chasing complexity through simulations. *European Journal for Philosophy of Science* 7: 221–238.

Mascaro, S., Korb, K., Nicholson, A., and Woodberry, O. (2010). *Evolving Ethics: The New Science of Good and Evil.* Exeter: Imprint Academic.

Mayo-Wilson, C., Zollman, K., and Danks, D. (2011) The independence thesis: when individual and social epistemology diverge. *Philosophy of Science* 78: 653-677.

Mirowski, P. (2011). *Science-mart*. Boston: Harvard University Press.

Neumann, M. (2014). The escalation of ethnonationalist radicalization. *Social Science Computer Review* 32: 312-333.

Okasha, Samir. 2011. Theory choice and social choice: Kuhn versus Arrow. *Mind* 477: 83-115.

Pinto, M. F. (2016) Economic imperialism in social epistemology. *Philosophy of the Social Sciences* 46: 443-472.

Poincaré, H. 1920[1958]. *The Value of Science*. English translation by Halsted, G., New York: Dover Publications.

Politi, V., and Grinbaum, A. (2020) The distribution of ethical labor in the scientific community. *Journal of Responsible Innovation.* In press: https://doi.org/10.1080/23299460.2020.1724357

Pöyhönen, S. (2017). Value of cognitive diversity in science. *Synthese* 194: 4519–4540.

Rosenstock, S., Bruner, J., and C. O'Connor (2017). In epistemic networks, is less really more? *Philosophy of Science*: 234–252.

Rudner, R. (1953). The scientist *qua* scientist makes value judgments. *Philosophy of Science* 20: 1–6.

Ruvinsky, A. (2008). Computational Ethics. In M. Quigley (Ed.), *Encyclopedia of Information Ethics and Security*. Hershey, PA: Information Science Reference, pp. 76-82

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology* 1: 143-186.

Sober, E. (2007). Evidence and value-freedom. In H. Kincaid, J. Dupré, & A. Wylie (Eds.), Value-free science (pp. 109–119). Oxford: Oxford University Press.

Solomon, M. (2001). *Social Empiricism*. Cambridge: MIT Press.

Šešelja, D. (2020). Exploring scientific inquiry via agent-based modeling. *Perspectives on Science* (forthcoming): http://philsci-archive.pitt.edu/17266/ Retrieved: 20th January 2021.

Thicke, M. (2020). Evaluating formal models of science. *Journal for General Philosophy of Science* 51: 315-335

Thoma, J. (2015). The epistemic division of labor revisited. *Philosophy of Science* 82: 454–472.

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research* 3: 429-457.

Weisberg, M. (2010). New approaches to the division of cognitive labor. In Magnus, P.D. and Busch, J. (Eds.) *New Waves in Philosophy of Science*. Palgrave Macmillan, pp. 250-269

Weisberg, M., and Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science* 76: 225–252.

Wray, B. (2000). Invisible Hands and the Success of Science. *Philosophy of Science* 67: 163-175.

Wray, B. (2010). Kuhn's constructionism. *Perspectives on Science* 18: 311-327.

Wray, B. (2011). *Kuhn's Evolutionary Social Epistemology*. Cambridge.: Cambridge University Press.

Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science* 74: 574-587.

Zollman, K. (2010). The epistemic benefit of transient diversity. *Erkenntnis*: 17–35.