Penultimate draft (March 2013).

Forthcoming in *Studies in History and Philosophy of Science.*

Making the Abstract Concrete: The Role Of Norms And Values In Experimental Modeling[1]

Isabelle F. Peschard and Bas C. van Fraassen

Department of Philosophy, San Francisco State University.

ABSTRACT

Experimental modeling is the construction of theoretical models hand in hand with experimental activity. As explained in section I, experimental modeling starts with claims about phenomena that use abstract concepts, concepts whose conditions of realization are not yet specified; and it ends with a concrete model of the phenomenon, a model that can be tested against data. This paper argues that this process from abstract concepts to concrete models involves judgments of relevance, which are irreducibly normative. In section II, we show, on the basis of several case studies, how these judgments contribute to the determination of the conditions of realization of the abstract concepts and, at the same time, of the quantities that characterize the phenomenon under study. Then, in section III, we compare this view on modeling with other approaches that also have acknowledged the role of relevance judgments in science. To conclude, in section IV, we discuss the possibility of a plurality of relevance judgments and introduce a distinction between locally and generally relevant factors.

Keywords

Relevance judgments, relevant factors, model, theory, experiment, phenomenon, norm, value, experimental modeling, concrete models, abstract concepts, data generating procedure, plurality.

Can focusing on *modeling* in scientific practice make a difference in the appreciation of the role of norms and values in science? We shall concentrate on modeling activity that is conducted hand in hand with experimentation and takes place between the initial selection of problems and the final confrontation of a hypothesis or fully constructed model with the data or evidence to be accounted for.

I. Models and data gathering procedures

Some roles that values play in science have become uncontroversial in philosophy of science, even if not counted among the so-called epistemic values. They are uncontroversial in being not only generally acknowledged but also not regarded as a threat to scientific objectivity. But these uncontroversial roles for values are located within specific moments of scientific inquiry: *before* any process of modeling and experimenting starts, in the choice of problems and methods, and *after* that process is already complete, in the interpretation and application of scientific results (Dorato 2004).

Before the empirical inquiry, the dual enterprise of modeling and experimentation, begins, values enter into the selection of research programs, of problems to be solved, phenomena to be investigated, or even the means to be used for the investigation. For example, in his writings on science in a democratic society Philip Kitcher (2000, 2011) has explored the proper role values play in the selection of, in his

[1] The authors wish to acknowledge support for this research by NSF grant SES-1026183.

terms, the significant truths that are worth seeking; even his critics generally agree on this (cf. Dupré 2004).

At the end stage, when both the data and the hypothesis or theory to be assessed are fully definite, the hypothesis will be assessed, perhaps accepted or rejected, in light of the evidence. Concerning this assessment, discussion on the role of values in science are generally structured around a distinction between epistemic and non-epistemic values. Epistemic values are supposed to be possible characteristics, such as explanatory power, predictive power, or simplicity, of the relation between the hypothesis and the data that are to be accounted for. By contrast, non-epistemic values, typically, bear on the semantic content of the hypothesis, the terms of the explanation it provides, e.g. in terms that can be read as sexist, racist, or feminist. What is hardly controversial is that certain values, not among those traditionally counted as epistemic, might enter in settling on a measure of fit or on how good the fit should be, hence how much evidence suffices for acceptance, depending on a preference for avoiding type I or type II errors and on what is at stake in each case.

This was first argued by Richard Rudner (1953) and more recently thoroughly explored by Heather Douglas (2000, 2006).

But the actual activity of modeling takes place between the initial selection of problems and the final confrontation between a hypothesis or model and the data or evidence to be accounted for. How we view modeling today gives reason to expect some novel insight in the role of values in science there. As Cartwright et al. (1995: 139) put it, theories are not bellies full of models, models are not derived from theories. Modeling is more akin, as Boumans (1999) writes, to 'baking with no recipe'. Even for modeling that is governed by theoretical principles, there is some leeway between the constraints from these principles and what a model can be (Morrison 2007). Within this leeway, there may be room for value judgments.

I.1 Concrete models and DGP statements

To focus the discussion, assume just for now that, given a phenomenon under investigation, the evaluation of the model for the phenomenon is merely based on its empirical accuracy. Setting aside the determination of criteria of fit, it seems then possible to see the selection of a model as a strictly empirical matter and, in that sense, free of value judgment. But this assumption can hold only under certain conditions, where the model can be 'directly' tested.

To see what these conditions are, suppose that the model is described by an equation such as PV = rT. Then claiming that the model is empirically accurate is tantamount to claiming that the outcome of measurements of P, V, T *while properly shielded from interference*, will display that relationship. Such claims about a model can be read as claims about a kind of *data generating procedure*; briefly, they are *DGP statements*. A DGP statement is a statement about the procedures that generate certain data; in this sort of case, the data are taken to be values for the quantities that figure in the model. The DGP statement says about these procedures that the generated values will satisfy the relation displayed in the model. If that is an empirical matter, the testing of the model can be regarded as an empirical matter as well.

Let us call a model a "concrete model" exactly if claims about the model can be read as claims about data generating procedures, that is, exactly if the model consists in relations between quantities that can be ascribed a value by data generating procedures.

I.2 Abstract concepts, concrete models

To only consider concrete models leaves out a crucial part of the process of modeling.

1) Before being tested models need to be constructed. The construction of models not only draws on a diversity of theoretical models and theoretical principles but also incorporates empirical data; that is, they are constructed so as to account for them. For this reason, Boumans (1999, 2005) speaks of models, specifically in economics, as having some 'built-in justification'. Similarly, Dorling's (1971) analysis of Einstein's modeling of the photo-electric effect shows how Einstein built in various results of Boltzmann and Planck, with earlier empirical results.

When a concrete model is tested, the model prescribes which data are relevant to the test and what type of DGP needs to be constructed. But in the earlier phase of model construction, the direction of constraint is not from the model to the data (concrete model → DGP), but from the data to the model (DGP → concrete model). How do we, then, specify the kind of procedure that generates data that have to be accounted for by the model? It is tempting to think that the phenomenon under investigation will somehow dictate the answer to that question. But:

2) Even when a phenomenon is ostensibly specified for investigation, that phenomenon is not, in general, already well-defined in terms of data generating procedures. Modeling is not just concerned with producing models that integrate and account for some given data. The task is also, in some cases, to determine the kinds of data generating procedures that produce the appropriate ones: data that need to be accounted for. Call them 'the *relevant* data'. Then modeling is associated or, to speak more accurately, *articulated* with experimental activity. We call this form of modeling 'experimental modeling.' The experimental activity articulated with modeling consists, to use Joseph Rouse's words, in creating "novel re-arrangements of the world that allow some aspects that are not ordinarily manifest and intelligible to show themselves clearly and evidently" (Rouse 2008). The phenomenon, target of the modeling, will show itself through the data generating procedures that produce values for the quantities that characterize these 'novel re-arrangements of the world'. To create such arrangements is to give form to the phenomenon under study.

This 'giving form' is at the heart of model construction. As we will show with examples, it can proceed in different directions. In the course of the inquiry, a series of *judgments* leads up to the final characterization of the targeted phenomenon.

3) Some philosophers (Longino 1990, Rolin 2011) have argued that values may indeed play a role in the specification of the relevant data: "in order to be able to assess the evidential warrant of a hypothesis, a scientist will have to decide what kind of evidence is relevant for the hypothesis. In some cases, a decision concerning relevance is value-laden". (Rolin 2011)

This issue has remained marginal in studies of modeling, because these studies have largely been concerned with the testing of concrete models. In that case, the kind of data generating procedure that needs to be realized to produce the testing data is prescribed by the model itself. If we want to test $PV = rT$, we simply need a DGP that measures P, V and T.

The problem of evidential relevance arises when the formulation of the problem, making reference to the phenomenon under investigation, involves *abstract* concepts, in the sense of Nancy Cartwright's distinction between concrete and abstract concepts. Cartwright illustrates the idea of abstract concept with the concept of force (1999: 44). Two features characterize its being an abstract concept. One is that it is abstract relative to a set of more concrete descriptions such that the abstract concept does not apply unless some of these descriptions also apply. There is a force only where there is some specific kind of force, electric, magnetic, gravitational, etc. and a specific kind of physical arrangement obtained. This physical arrangement can be characterized in terms of quantities that can be ascribed a value. For instance, there is a gravitational force between two masses where there are two bodies of respective mass m and M, located at a distance r from one another: "in this arrangement the small mass m is subject to the force GmM/r^2" (loc. cit.).

In the present context, we take the concrete descriptions associated with abstract concepts to characterize the physical conditions, the arrangements in which the abstract concept is instantiated. A concrete description we construe as a description of a kind of data generating procedure: it describes an arrangement in terms of the data that can be generated within the arrangement, such as values for the masses, m and M, and for the distance r[2].

The other feature is that the abstract concept applies just in the situations that satisfy the associated concrete descriptions. There is nothing more to a force, no 'new, separate property', than the satisfaction of at least one of these descriptions. There is no general property in addition to being a specific kind of force. In addition to these two features, another important aspect of the distinction abstract/concrete, qualifying the previous statement, is that the abstract concept is not reducible to the associated concrete descriptions. This is illustrated by another of Cartwright's examples of abstract concept, the concept of work. Washing dishes and writing a grant are both cases of work. But the concept of work has connections with other abstract concepts: washing dishes is an instance of the concept of work only insofar as it is not leisure. The connection of *work* with *leisure* puts a constraint on the sort of situations in which *work* can be realized.

We then use the distinction abstract/concrete in the following way: a claim about a phenomenon that is not yet well-defined refers to it using an *abstract* concept; by contrast, a model of the phenomenon that can be directly tested for its empirical accuracy is a *concrete* model. The modeling process goes from the abstract concept to the concrete model via the specification of the conditions of realization of the abstract concept, i.e., of the kind of data generating procedures producing data to be accounted for by a concrete model of the phenomenon. Suppose that one is interested in making a concrete model of the dependence of work productivity on some physiological or social factors. Most often the construction will start with some data that need to be accounted for and as such guide, constrain, inform the construction of the model. But for these data to play this role they must be gathered in conditions that are recognized as *work conditions*: conditions in which the concept of work is realized.

We will now proceed as follows. In section II several case studies in social and in natural science will show how the specification of the conditions of realization of an abstract concept involves making judgments about what kind of data is relevant. These judgments are what we call 'judgments of relevance' or 'relevance judgments'. Differences in relevance judgments will lead to differences in the concrete models of the phenomenon. But these judgments are not empirical; they are normative judgments and through them norms and values are incorporated to the modeling process. In section III,

---

[2] We speak of a 'kind' of data generating procedure because the description does not specify how exactly the quantities should be measured.

we compare this view, which locates normativity within the modeling stage, to other accounts of modeling and to a pluralist thesis on scientific knowledge. Section IV will make the notion of relevance judgment more precise and discuss the compatibility, complementarity, or exclusivity of different relevance judgments.

II. What guides the process of concretization

II.1 The role of relevance judgments in social science

Suppose we want to study the evolution or incidence of rape or violence, maybe as a function of, say, education, and income level. *Rape*, *violence* are abstract concepts. In order to study the evolution of rape we need to specify the sort of conditions in which data about rape are to be gathered. That amounts to specifying a kind of DGP that can produce data that can be used to test claims using this concept. The study of rape is an example offered by John Dupré to illustrate the role of values. That values may have guided the *selection* of this problem for study is uncontroversial. But there is more to it. Not long ago (and perhaps not in all cultures), Dupré explains, a wife could not be raped by her husband; it was conceptually impossible. Rape involves violation of human rights and, until not long ago, wives did not have the sort of rights that could have been violated in that way. So anything husbands did within marriage would not have been regarded as data to be gathered for the investigation of a claim about rape. There may be some empirical considerations involved in the judgment that rape does, or does not, include sexual violence in marriage. But it is not a merely empirical judgment. It is a *relevance judgment*. It is, in this specific case, a judgment that involves moral values, and Dupré focuses on *rape* being an evaluative and therefore normative concept.

The study of social mobility analyzed by Kareem Khalifa (2004) offers a different illustration of the need for relevance judgments. Khalifa explains that in their study on social mobility, Peter Blau and Otis Dudley Duncan (1967) generated data on occupational status by first defining occupational status as a function of income and education. They defined social mobility as a function of the father's occupational status and of the status of his first occupation. By contrast, previous studies gathered data on respondent ranking of the standing of over sixty occupations on a scale from 1 (excellent) to 5 (poor). Occupational ratings were measured as a weighted average score. The difference had important consequences: whereas Blau and Duncan concluded that well educated African-Americans suffer greater discrimination than less educated African Americans, the contrary had been concluded on the basis of the previous results.

That sort of case is apparently very different from a case such as the study of rape. Khalifa argues for the existence of some criteria for the selection of relevant factors that, according to him, justify the epistemic superiority of the data generating procedure used by Blau and Dudley: "one data generating procedure is superior to another if it has greater coherence with theoretical claims as well as greater ability to replicate old data and to generate new data". Given the listed criteria, it seems an empirical matter, by contrast with the previous case, whether a particular data generating procedure satisfies them or not. And, even though Khalifa, unfortunately, does not explain what he means by 'greater coherence with theoretical claims', it seems doubtful that this consideration plays a role in the selection of a data generating procedure for the study of rape; likewise for replicating old data since one would not even think of replicating data that are deemed, at best, incomplete, and, at worst, biased because their ignore exclude marital context.

But the difference is far from being as great as it seems. Khalifa notes that Blau and Duncan have a plausible account for basing their procedure on income and education: “A man qualifies himself for occupational life by obtaining an education; as a consequence of pursuing his occupation, he obtains income.” (Duncan 1961: 116-117 in Khalifa 2004). But one may have an equally good, equally ‘empirical’, account for a data generating procedure that integrates marital contexts to the investigation of rape incidence: for instance, the nature of the physical and psychological damages. In neither case do these empirical considerations, in and by themselves, make some factors and data relevant. As Khalifa himself remarks, “the superiority of a data-generating procedure and the soundness of topics [the proposition that is to be explained] founded on it may not have universal agreement across the field.” Typically, the lack of agreement on the soundness of the topic will come from disagreement on the soundness of theoretical assumptions. But then, why should the new data generating procedure be constrained by theoretical claims whose rejection might be one of the motivations for a new data generating procedure? The application of the criteria of ‘greater coherence with theoretical claims’ will depend on what is meant by ‘coherence’ and, even more crucially, on which theoretical claims are selected. Similarly for the old data criterion: how could replication necessarily function as a norm when these old data might have been generated by a procedure that is regarded as inappropriate?

The choice of a data generating procedure to measure social mobility might well, then, be supported by empirical considerations but it is not a merely empirical judgment. It is a normative judgment. In the next sub-section we will go further: the role of relevance judgments is not restricted to investigations dealing with things that, as Dupré says, we care about, be it for their moral or social dimension.

II.2 The role of relevance judgments in natural science

There is a function for relevance judgments in natural science. The neurophysiological study of the behavior of visual cells and their so-called receptive field offers a case in point. The experimental study of the response of the visual cells used to be conducted in an artificial environment with anesthetized animal and isolated luminous stimuli (Hubel & Wiesel 1962) to record the response of visual cells to luminous stimuli of different shapes located in different positions in the visual field. The cell was then thought of as a spatio-temporal filter and the receptive field was construed as a property of the cells, an invariant characteristic of a certain type of cell identified with the kind of stimulus triggering the strongest response. It was also thought that the receptive field of the cells located on the visual path from the retina to the visual cortex were a combination of the receptive fields of the cells located earlier on the visual path.

The behavior of visual cells in *this* type of condition is now regarded as informative at best about what is called the ‘classical’ receptive field, offering only a small inkling of the complex behavior of visual cells. An anesthetized animal and isolated luminous stimuli are no longer regarded as representative of the conditions in which the receptive field is realized.

Justifications for changes in the conception of the receptive field generally appeal to new empirical evidence. For instance, empirical evidence that a stimulus that does not trigger a response of the cell by itself can nevertheless influence the response of the cell to its ‘preferred’ stimulus. That means that the visual context may make a difference to the reaction of the cell to a given stimulus, which difference is overlooked by the use of isolated stimuli. There is also empirical evidence that interpretation of the stimulus may make a difference to the response of cells located very early on the visual path, which difference is overlooked by the use of anesthetized animals (Albright and Stoner 2002; Albright et al. 2003). Increasingly, therefore, in contemporary research, the complexity of the natural environment, by contrast with the artificial conditions of isolated stimuli, is thought to be essential to the behavior of visual cells (Haslinger et al. 2012). Visual cells, it has been argued, are adapted through both evolutionary and developmental processes to the type of signals to which they are naturally exposed.

It may thus seem that scientists have simply made progress by discovering new empirical effects and trying to take into account a new causal factor, the complexity of the visual context, so as to arrive at a more complete conception of the phenomenon under study, the receptive field. On that view, relevance judgments would be nothing more than causal, empirical judgments: the complexity of the visual context is relevant because of empirical evidence showing its causal influence. But this interpretation becomes dubious when one looks more closely at how a certain approach developed at the expense of alternative approaches.

The idea of taking into account the visual context was actually not new. It had already been suggested even before the experimental studies that were to map the receptive fields of the cells by taking artificial conditions as canonical. In particular, Barlow (1961) argued that the anatomical or neurophysiological study of sensory structures had to be guided by a hypothesis about the function of these structures. He proposed the hypothesis, that is now investigated, that there are sensory relays which take advantage of the redundancy of information present in a natural environment and "recode sensory messages so that their redundancy is reduced but comparatively little information is lost". The exploration of that hypothesis, though, is recent.

The increasing popularity of making the natural environment part of the investigation of the receptive field is correlated with the development of statistical models of the environment and engineering non-linear models of the visual cells. Drawing on work and concepts from information theory and engineering disciplines, the complexity of the environment is represented in terms of statistical properties, as a function of spatio-temporal correlation, and the function of the visual system is construed as a recoding of the incoming, highly correlated and redundant, signals so as to improve efficiency.

For the complexity of the environment to become a relevant factor, there had to be a way of quantifying this complexity. But the dependence on the availability of a technique does not make the relevance of this factor an empirical matter, far from it. To begin, the method needs itself to be first regarded as fruitful, as providing *valuable* kinds of data. Then further normative judgments are still required at several levels to specify the experimental features of the conditions of application of the statistical method:

> Despite widespread agreement that neural processing must be influenced by environmental statistics, it has been surprisingly difficult to make the link quantitatively precise. [Even with the idea of efficient coding] the hypothesis is not fully specified. One needs also to state which environment shapes the system. (…). In addition, one must specify a timescale over which the environment should shape the system. Finally, one needs to state which neurons are meant to satisfy the efficiency criterion, and how their responses are to be interpreted. (Simoncelli and Olshausen, 2001: 1194)

Furthermore, even if there is a "widespread agreement that neural processing must be influenced by environmental statistics", there is actually no clear agreement on how the receptive field should now be conceived and "how revisionary or conservative neuroscientists could now be about the concept RF [receptive field] in the light of these findings" (Chirimuuta and Gold 2009). What is at issue is what type of stimuli should be elected as canonical—which is an issue about the typical conditions of realization of the concept.

One of the possibilities listed by Chirimuuta and Gold (2009) is to take into account the effect of visual context by electing natural stimuli as canonical type. Another possibility is to accommodate the influence of neuronal connections by modeling the visual system at the level of circuitry. In the former case, the cost may be to have to dismiss the huge amount of work done in artificial conditions; in the latter case, it may be a "dramatic rethinking of neural computation" which is usually based on neuronal units. Alternatively, however, as we will see later, the difference between the two types of experimental conditions may be seen in terms of a difference in the value of *a new relevant parameter* representing the complexity of the visual context. This parameter would take different values for different degrees of complexity of the environmental context, with the minimal value being for artificial conditions of isolated stimuli.

Still another approach emphasizes the embodiment of visual processes. The representation of the complexity of the environment in statistical terms is based on the analysis of images of the environment with the visual input being the output of a statistical computation of images.[3] And the receiver of the input is generally a cell or group of cells. On that approach, the difference between a cell receiving inputs from pre-analyzed images and an embodied visual system, embedded in the environment, is not regarded as relevant to the characterization of the activity of visual cells and their receptive field. This difference has, however, been explored by studies using a mobile robot equipped with a camera. Instead of receiving the result from pre-analyzed images the robot has the possibility to explore the environment and actively select some part of it for further exploration. It has been argued that the possibility of active visual exploration makes a difference to the development of the receptive field (Floreano et al. 2005; Suzuki and Floreano 2008).

There is little doubt that the actual exploration of the new dimension, the dimension of active vision, owes much, again, to the development of new techniques that make possible the systematic and quantified gathering of new empirical evidence. But the difference between the studies based on computation of images of the environment and those based on robotic exploration is not merely technical. It is also a difference in a larger perspective on perception and cognition that motivates taking advantage of certain tools and techniques to conduct certain types of exploration and investigate the effect of certain factors. The computational approach is grounded in the computer metaphor that views the cognitive system as passively receiving bits of information. The robotic approach is grounded and motivated by the enactive metaphor where the cognitive system is embodied and embedded in the environment, while action and perception inform and constrain one another (Varela 1993). The adoption of a metaphor is not an empirical judgment; it is what we may call 'a normative perspective' in that it highlights what is

---

[3] Note, however, that if the notion of 'environment' is specified in terms of statistics of natural images, there may be little room to accommodate the psychologists' concern that interpretation makes a difference to the activity of the cells located early on the visual path.

important to cognitive activity, what needs to be explored, what sorts of cognitive performances are most representative, what sorts of components need to be part of an account of cognition. What factors have a *causal* effect on the activity of visual cells does not change depending on which metaphor guide the investigation. What changes is what factors are relevant, that is, should be taken into account and their effect be accounted for by a model of the receptive field.

In the next sections, we will make the concept of relevance judgment more precise. We begin by discussing some similarities and differences between this concept and others that have been proposed to address what may look like related concerns.

III. Relevance vs. plurality and idealization

The concepts of social mobility or receptive field are not morally evaluative, in contrast to those of rape or violence. But claims about rape, violence, social mobility, or receptive field, are all similar in how they differ from claims about what we have called concrete models. Whereas concrete models specify by and in themselves the kind of DGP that is appropriate for testing them, claims about rape, violence, social mobility, or receptive field do not. Some non-empirical judgments must be made as to what kinds of data are relevant, and what kinds of data generating procedures are able to produce them. Those deemed relevant are the ones that will be used to guide, inform and constrain the construction of concrete models. This view on modeling is not completely unrelated to some other views that emphasize the pluralist or idealizing dimension of scientific modeling. Some similarities and differences will enable us to clarify further central aspects of the concept of relevance.

III.1 Relevance and plurality

The role of relevance judgments appears also in Helen Longino's pluralist view of scientific knowledge, to which the present account is especially indebted. It is important, however, to distinguish the pluralist thesis from the pluralist stance. According to the plurality thesis, "a plurality of adequate and epistemically acceptable explanations or theories can be generated by a variety of different factors in any situation of inquiry" (Longino 2002: 184). A pluralist *stance*, on the other hand, is just an acknowledgment of the possibility that, in some cases of investigation, there is will be a plurality of accounts which cannot be integrated with one another (Kellert and Longino 2006: xv).

Longino accounts for the possibility of a plurality of theories in terms of different epistemologies. A form of inquiry is constituted (in part) by assumptions of two kinds: substantive and methodological. Substantive assumptions concern the way the world is: for instance, the assumption that the material world is constituted of particles or that any biological development is controlled by genes. Methodological assumptions have to do with the means we have of developing and acquiring knowledge: whether one should do field observation or laboratory experiments; what kinds of data are relevant; commitment to unification or simplicity; standards of reliability.

Such a set of methodological choices, commitments, or standards forms what Longino calls a community's epistemology and it is differences in epistemologies that account for differences in explanations or theories. For instance, if different models are able to account for the same data, differences in so-called epistemic values may lead to the selection of one rather than another model. Or different ontological commitments, say, to waves or quanta, may generate different interpretations of the same observations. In such cases, much discussed in philosophy of science, differences in explanation or models arise without differences in the relevant data.

We are specifically interested in diversity of models linked to differences in what counts as relevant data, data that need to be accounted for. In that context, not any difference in substantive assumptions will be pertinent: in the study of the receptive field, commitments to different theoretical frameworks, computationalist or enactivist, will engender, as we saw, differences in relevance judgments, but a disagreement as to whether there are waves or particles will probably be of no consequence. Likewise for the methodological assumptions: the difference between field observation and laboratory experiment is not pertinent if the same factors are investigated in the field and in the laboratory

Differences in relevance judgments, then, will do all the work of accounting for plurality, while the various elements that form a 'local epistemology' offer some keys as to what motivates (i.e. theoretical commitment) or is motivated (i.e. methodology), by relevance judgments.

Even with this restriction, we are not (no more than Longino herself) defending the pluralist *thesis*. Whether there will always be a plurality of relevance judgments and whether a plurality of relevance judgments will always lead to an irreducible plurality of explanations is not obvious at all. As we will see in the next section, it might be possible in some cases to see different relevant judgments, leading to different concrete models, as belonging to a larger, unifying explanation. But neither are we committed to the view that this should always be the case. The role of relevance judgments in experimental modeling rather invites the adoption of a pluralist *stance*.

A difference between Longino's pluralist approach and the view that is proposed here is the role we ascribe to the distinction between concrete models and hypotheses that involve abstract concepts. Longino explains some need for non-empirical judgments by noting that such judgments are needed to bridge the gap between the unobservable and the empirical (Longino 1990: 58) The same point holds, according to her, for any hypothesis, even empirical generalizations, in virtue of the fact that they "are or consist of statements whose content always exceeds that of the statements describing the observational data." (Longino 1990: 58) A claim about the empirical accuracy of a concrete model should then be treated on a par with a claim using concepts of unobservable entities. This entails that the assumption we displayed earlier, that the evaluation of claims about a concrete model is a purely empirical matter, is not realistic.

But to relate the need for non-empirical judgments solely to the gap between data and hypothesis would obscure the difference between what is involved in the evaluation of a claim about a concrete model and in the evaluation of a claim that uses abstract concepts. A claim that uses abstract concepts does not go beyond our evidence. Rather it, so to speak, goes nowhere, yet. The need for non-empirical judgments in determining the conditions of realization does not stem from un-observability in general; after all, non-empirical relevance judgments are needed for the empirical investigation of rape, and rape, certainly, is not unobservable. The question is where, under what conditions, it is there to be observed. The use of concepts referring to unobservable entities does create the need for relevance judgments, but that is because, just as for abstract concepts, they require specification of the conditions of measurement in which they are realized and can be empirically investigated.

III.2 Relevance and idealization

The notion of relevant factor is commonly identified with that of causally relevant factors: factors that make a causal difference on a given variable of interest (see, for instance, Craver 2007), *causally efficient factors*. On our view relevant factors are not equated with causally efficient factors. Relevant factors are causally efficient but not all causally efficient factors are selected as relevant. Even if causal judgments are strictly empirical (however that may be), judgments of relevance are not. As we saw, there are other considerations that contribute to the selection of some data and factors as relevant to the construction of a concrete model.

A view that appears to be similar has been proposed by Michael Weisberg on the basis of a distinction between different desiderata that may govern the practice of modeling. On Weisberg's account, in a slightly reconstructed form, the need for selection of relevant factors will generate a tension between two ideals that can be seen to govern the construction of models: the ideal of precision or completeness and that of generality (Weisberg 2006, 2007). The ideal of completeness is to achieve a representation that would include all the factors that are capable, in principle, of contributing causally to the evolution of the quantity of interest.[4] Completeness, however, brings along the promise of a loss of sense in the models: our cognitive limitations would make us unable to grasp the models in their entirety because of their complexity and because of the lack of generality of these models. Such models, Weisberg writes, "are tailored to very specific phenomena. They will often not generalize beyond the particular phenomenon" (2006: 632). In fact, in the limit of completeness, the model could only be of a phenomenon uniquely instantiated. The loss of sense would not just affect models, but the world too. Experimental modeling helps make sense of the world by creating the experimental conditions not merely for instantiations but for exemplification (Rouse 2007). If the model does not go beyond what happens in some very specific conditions, what happens there does not exemplify anything.

An alternative strategy in modeling is, according to Weisberg, the strategy of idealization: "We no longer even aim at producing complete models. When we build a model of some phenomenon, we will try to include the most important or the most *relevant* aspects, which depends on the interests of the theorist." (Weisberg 2006: 633, italics added). On Weisberg's account, modelers will construct different models by selecting different factors to be taken into account and the basis for these different selections is a commitment to different desiderata. Generality, precision, and prediction are three desiderata that govern the practice of modeling and whose maximal satisfaction regulates the procedure of idealization.

Weisberg identifies an important, pervasive, source of plurality of models and there is a sense in which this plurality comes from a difference in judgments of relevance. That is not the kind of relevance judgments that we have been focusing on, though. On Weisberg's account, the aim of the modeling process is to account for a certain effect. Assuming this aim, there will be different ways of reaching it with models maximally satisfying different desiderata.

By contrast, the relevance judgments that we are discussing in this paper are judgments about what factors and what *data* are relevant. The source of plurality that we are targeting is located earlier in the process of model construction, before there is a well-defined effect to be accounted for. It is the phase of the process where the abstract concepts used to formulate the problem, or to loosely refer to the phenomenon, have to be specified in terms of their conditions of realization. Relevance judgments are involved in this specification.

---

[4] By "capable in principle of contributing causally" we mean, following the interventionist account of causation, that a change in this factor would result in a change in the quantity of interest as long as its effect is not neutralized.

Weisberg's form of plurality, in terms of different forms of idealization resulting from different interests of the modelers, is perhaps more pragmatic than epistemic. By contrast, on the view we are presenting, the selection of different relevant factors is not the formulation of different ways of representing the same phenomenon for different purposes. Rather, more basically, this selection consists in the formulation of a conception of what the phenomenon under study is. The question that arises again, just as for a pluralist account, is whether the different models that are produced, in that case, can 'peacefully' coexist, either in the way different models serving different purposes can do so, or in that they can be somehow unified as parts of a larger account. This is the topic for our next section.

IV. The scope of relevance judgments

Can differences in relevance judgments, and the different forms of investigation they generate, peacefully co-exist? Or would one interpretation preclude any other? Or could they be articulated, combined, into a more encompassing account of the phenomenon under study?

Three Suggestive Examples

Different examples seem to suggest different answers to the questions above. Inquiries into the incidence of rape that include marital sexual violence and those that exclude them will be exclusive of one another. Similarly, Khalifa's claim that one can establish the epistemic superiority of the data generating procedure used by Blau and Dudley over the one used by previous studies implies that the former supplants, and excludes, the latter.

A different answer is suggested by Longino's study of the investigation of aggressive behavior, in terms of behavior genetics, developmental systems theory, neurophysiology, and anatomy. Longino (2006) concludes that each of these approaches provide only a partial account, and that one cannot replace or even be compared (as more or less correct) to the others, and in addition, that they cannot be unified into a single account: "[These approaches] constitute a nonunifiable plurality of partial knowledges" (p.127). The reason for this is not that these different approaches identify different causes of aggressive behavior, but rather that they focus on different causes and, as a result, explore different causal spaces, or more precisely, differently structured causal spaces. This analysis echoes the point made earlier concerning computational and enactive studies of the receptive field, that it is not which factors have a *causal* effect that distinguish different approaches but which factors are relevant, in the sense that they should be taken into account and their effect be accounted for by a model of the phenomenon under study.

The discussion of the receptive field case study suggested, however, still another answer. We noted that one might not have to choose exclusively between natural or artificial conditions of investigation, that is, between artificial and impoverished or natural and complex visual context. Although these different approaches focus on different causes, and elicit different factors as relevant, there might be a way to see them as unifiable.

To understand how different case studies can lead to these different answers, we need to characterize relevance judgments more precisely by highlighting two distinctions. One, already introduced, is between 'relevant factors' and 'causally efficient factors'; the other is among the relevant factors, between locally and generally relevant factors.

General vs. Local Relevance

Causally efficient factors comprise relevant and non-relevant factors. Factors that are causally efficient but not relevant are potentially interfering and need to have their effect neutralized. By contrast, relevant factors are factors that need to be taken into account by the model, factors whose effect needs to be systematically investigated and accounted for.

But given a set of generally relevant factors, some experimental set-up may be designed to systematically explore only some of them. For instance, engineers are well equipped to investigate how the visual system deals with the complexity of natural images because they can draw on techniques of image analysis. In this experimental context, the receptive field is often investigated with anesthetized animals. By contrast, psychologists interested in how the interpretation of the stimulus affects the activity of visual cells need the animal to be awake. But they will prefer the use of stimuli with a fixed degree of complexity so as to improve the control of the interpretive dimension. The use of anesthetized animal can be seen as fixing the value of interpretation or awareness to zero. To keep fixed the value of this factor, or the value of the factor 'complexity', doesn't imply that these factors are taken to be non-relevant. They may simply be taken as *locally* non-relevant.

The distinction between general and local relevance is required to account for the fact that different experimental set-ups might be designed and used to explore separately the effects of different relevant factors. Whereas generally relevant factors form the dimensions of the total space in which the phenomenon under study develops, the *space of general relevance*, locally relevant factors are those among the generally relevant factors that are actually investigated in a given experimental context. Selecting a set of locally relevant parameters among a set of generally relevant factors amounts to selecting a *subspace* of the space in which the phenomenon develops. Such a subspace is, more precisely, characterized by the fixed values of the generally relevant factors that are not explored, which is to say, the locally non-relevant factors.

Locally non-relevant factors are treated in a way apparently similar to potentially interfering factors (generally non-relevant factors): their effects need to be neutralized. As for generally non-relevant factors, failure to neutralize the effect of locally non-relevant factors leads to lack of reliability due to undetermined interference with the factors that are being explored. But, in the case of locally non-relevant factors, more is needed than mere neutralization. For, by contrast to generally non-relevant factors, locally non-relevant factors are recognized as factors that will need to be (eventually) explored in order to arrive at a complete representation of the phenomenon. Explorations of different subspaces of the space of general relevance have to be viewed as complementary. Locally non-relevant factors have then to be neutralized in such a way that the results from measurements can be located within the space of the phenomenon and that the results from different subspaces can be related to one another. The condition is ensured by not only keeping fixed the values of the locally non-relevant factors, but also keeping track of these values.

Analysis of the Three Suggestive Examples

With this framework in hand, we are now in a position to offer an account of the different answers listed above. Two data generating procedures will be exclusive of one another if they are not located within the same space of general relevance, as in the rape and the social mobility case studies, and probably also in the receptive field case study regarding the computational and the enactive approaches.

The opposite situation occurs when two data generating procedures are complementary, and their results can be unified to form a richer representation of the phenomenon. That happens when these procedures are exploring different well-specified subspaces of the same space of general relevance. This situation could be illustrated, again, with the study of the receptive field, with a procedure investigating the effect of the complexity of the visual context and another proceeding with a fixed value for this factor.

The study of aggression satisfies neither of these two conditions. As Longino notes, the procedures in question are, in this case, not exclusive of one another. That suggests that they are located within a single space of general relevance: they are not only recognizing the same causally efficacious factors but the same relevant factors. However, they are exploring different subspaces within that space of general relevance. Why, then, are they deemed not unifiable? That will happen if the subspaces that are explored cannot be related to one another. The reason might be that these were not well located within the space of general relevance, in terms of the fixed values of the locally non-relevant factors. To use a simple example: that would be like exploring the relation between pressure and volume without keeping track of the value of the temperature and exploring the relation between volume and temperature without keeping track of the pressure. It is clear that these two procedures would not be exclusive of one another, but their results would not be unifiable either.

V. The normative character of relevance

Can focusing on *modeling* in scientific practice, then, make a difference in the appreciation of the role of norms and values in science? If modeling is understood as an activity consisting in testing what we have called 'concrete models', the discussion of the role of norms and values will just be a discussion about cases of hypothesis testing where it is taken for granted, as it generally is, what data are relevant for the test.

But testing concrete models is not all there is to modeling. In this paper we focused on model construction conducted hand in hand with experimentation, what we have called 'experimental modeling'. The aim of experimental modeling is not just to produce a concrete model. When the investigation starts with an abstract concept of the phenomenon under study, experimental modeling also consists, in part, in specifying the kind of conditions in which the concept is realized, and thus to specify what kind of data is relevant to the testing of a model of the phenomenon. That sets a norm, for this determines what is to be accounted for, or on the contrary, need not be taken into account.

Focusing on this aspect of modeling does therefore make a difference to the appreciation of the role of norms and values in science. The difference comes from how this focus makes visible their role in specifying what the relevant data are, through what we have called 'relevance judgments'. Relevance judgments are the sort of judgments that we saw at play in determining what data are relevant not only to the study of rape or violence but also to the study of social mobility or receptive fields. They are no more empirical in the latter cases than in the former and no less normative.

REFERENCES

Albright, Thomas and Gene Stoner. (2002) Contextual Influences on Visual Processing. *Annual Review of Neuroscience*, 25: 339-79.

Albright, T. D.; Croner, L.J.; Duncan, R. O.; and Stoner, G. R. (2003) "Neuronal bases of perceptual organization in the primate visual system." Pp. 305-336 in R. Kimchi, M. Behrmann, and C. R. Olson (eds.) *Perceptual Organization in Vision: Behavioral and Neural Perspectives,* Mahwah, NJ: Lawrence Erlbaum Associates,

Barlow, Horace B. (1961) "Possible principles underlying the transformation of sensory messages". Pp. 217–234 in Rosenblith 1961.

Bickle, J. (ed.) (2009) *Oxford Handbook of Philosophy and Neuroscience*. Oxford: Oxford University Press.

Blau, Peter and Otis Dudley Duncan (1967) *The American Occupational Structure*. New York: John Wiley and Sons.

Boumans, Marcel (1999) "Built-in justification". Pp. 66-96 in Morgan and Morrison 1999.

Boumans, Marcel (2005). *How Economists Model The World Into Numbers.* London and New York: Routledge.

Cartwright, Nancy (1999). *The Dappled World: A Study Of The Boundaries Of Science.* New York: Cambridge University Press.

Cartwright, Nancy D., Shomar, T., Suarez, M. (1995) "The Tool Box of Science: Tools for the Building of Models with a Superconductivity Example'" in Herfel et al. 1995.

Chirimuuta M. and I. Gold (2009) "The embedded neuron, the enactive field?" Pp. 200-225 in Bickle 2009.

Dorling, Jon (1971) "Einstein's Introduction of Photons: Argument by Analogy or Deduction from the Phenomena?" *British Journal for the Philosophy of Science* 22: 1-8.

Douglas, Heather (2000). "Inductive Risk and Values in Science." *Philosophy of Science* 67: 559-579.

Douglas, Heather (2006) "Norms for Values in Scientific Belief Acceptance", *philsci-archive.pitt.edu/3024/1/Values_Beliefs_PSA_06_Final.pdf*

Duncan, Otis Dudley (1961) "A Socioeconomic Index for All Occupations". In A. Reiss (ed.) *Occupations and Social Status*. New York: Free Press of Glencoe.

Dupré, John (2004) "Science and values and values in science: comments on Philip Kitcher's *Science, Truth, and Democracy*". *Inquiry* 47: 505-514.

Floreano, Dario; Suzuki, M.; and Mattiussi, C. (2005) "Active vision and receptive field development in evolutionary robots". *Evolutionary Computation* 13: 527-544.

Haslinger, Robert; Gordon Pipa; Bruss Lima; Emery Brown; and Sergio Neuenschwander (2012) "Context matter: the illusive simplicity of macaque V1 receptive field". PLOS One 7(7): e39699.

Herfel, W. E.; Krajewski, W.; Niiniluoto, I.; and Wojcicki, R. (eds.) (1995) *Theories and Models in*

Scientific Processes. Amsterdam-Atlanta: Rodopi.

Hubel, D. H & T. N. Wiesel (1962) "Receptive fields, binocular interactions and functional architecture in the cat's visual cortex". *The Journal of Physiology* 160: 106-154.

Kellert, Stephen H. and Helen Longino (2006) *Scientific Pluralism.* Minneapolis: University of Minnesota Press.

Khalifa, Kareem (2004) "Erotetic contextualism, data generating procedures and sociological explanations of social mobility", *Philosophy of the Social Sciences* 34: 38-54.

Kitcher, Philip (2001) *Science, Truth and Democracy.* New York: Oxford University Press.

Kitcher, Philip (2011) *Science in a Democratic Society*. Amherst, NY: Prometheus Books.

Longino, Helen (1990) *Science as Social Knowledge*. Princeton: Princeton University Press.

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

Morgan, Mary S. and Margaret Morrison (Eds.) (1999) *Models as Mediators: Perspectives on Natural and Social Science.* Cambridge: Cambridge University Press.

Morrison, Margaret (2007). "Where Have All the Theories Gone?". *Philosophy of Science* 74: 195-228.

Rolin, Kristina (2011) "Values in Science". ms. 10/26/2011. http://www.uta.fi/yky/oppiaineet/filosofia/opetusohjelma/index/Feminist%20epistemology%20I.pdf

Rosenblith, W. A. (ed.) (1961) *Sensory Communication.* Cambridge, MA: MIT Press.

Joseph Rouse (2008) "Laboratory Fictions". Pp. 37-55 in Suarez 2008.

Rudner, Richard (1953). "The Scientist *Qua* Scientist Makes Value Judgments." *Philosophy of Science* 20: 1-6.

Simoncelli, Eero P. and Bruno A. Olshausen (2001) "Natural image statistics and neural representation". *Annual Review of Neuroscience* 24: 1193-1216.

Suarez, Mauricio (ed.) (2008) *Fiction in Science: Philosophical Essays on Modeling and Idealization.* New York: Routledge.

Suzuki, Mototaka and Dario Floreano (2008) "Enactive Robot Vision". *Adaptive Behavior* 16: 122-128.

Weisberg, Michael (2006) "Forty Years of ' The Strategy': Levins on Model Building and Idealization", *Biology and Philosophy* 21: 623–645

Weisberg, Michael. (2007) "Three Kinds of Idealization", *The Journal of Philosophy* 104: 639-659.