Logic, Methodology and Philosophy of Science

Proceedings of the Fifteenth International Congress

Hannes Leitgeb
Ilkka Niiniluoto
Päivi Seppälä
Elliott Sober
Editors

18 Models and modeling in formal epistemology: Some thoughts on probability aggregation

ELEONORA CRESTO *

Abstract. In this paper I discuss the role of normativity in model building, particularly within formal epistemology. I begin by making some distinctions and clarifications, and then I focus on the problem of testing normative models. I suggest a novel way to think about putting normative models to the test, which consists in building meta-models for first order normative settings. I argue that a successful meta-modeling strategy should enable us to illuminate the mechanism that underlies a given normative structure, and in this sense it can further test or refine our intuitions concerning what ought to be the case.

Next I propose a model for probability aggregation that seeks to illustrate our prior discussion on the relevance and purpose of meta-model building for normative modeling in general. I suggest that, under certain circumstances, it can be rewarding to look at probability aggregation as a type of cooperative bargaining. Individual agents are assumed to hold utilities over possible probability assignments to propositions. Given such utilities, I show how to build an appropriate (pseudo)bargaining situation, such that points inside the bargaining set are correlated with sets of probability assignments (on a given proposition) by the individual agents. Solving the bargaining problem helps us figure out the probability that can be credited to the group as a whole. We then obtain a unified perspective on two seemingly disparate phenomena – probability aggregation and cooperative bargaining. The proposal illustrates how normative

*CONICET, Buenos Aires.

meta-models are meant to work: Bargaining models here act as meta-models that can help us elicit intuitions regarding probability aggregation (the first-order phenomenon) in an indirect way.

Keywords: normative models, probability aggregation, cooperative game theory, bargaining solutions, utilities.

1 Introduction

My aim in this paper is twofold. In the first part (Sections 2 to 4) I will offer some reflections about the role of normativity in model building, and in particular about normativity within formal epistemology. This topic remains largely unexplored, so I would like to establish some ground for future discussion. I will begin by making some distinctions and clarifications, and then I will focus on the problem of testing normative models. I will suggest a novel way to think about putting normative models to the test, which consists in building meta-models for first order normative settings. A successful meta-modeling strategy should enable us to illuminate the mechanism that underlies a given normative structure, and in this sense it can further test or refine our intuitions concerning what ought to be the case.

In the second part (Sections 5 and 6) I will develop one particular model for probability aggregation (a subject that squarely belongs to formal epistemology), which seeks to illustrate some of the general traits of satisfactory normative modeling, including model testing, discussed in the first part of the paper. More precisely, the proposal attempts to illustrate our prior discussion on the relevance and purpose of meta-model building for normative modeling in general. The proposed model might have some interest in itself, regardless of our prior discussion on modeling strategies. I will suggest that, under certain circumstances, it is rewarding to look at probability aggregation as a type of cooperative bargaining. Individual agents can be interpreted as holding utilities over possible probability assignments to propositions, such that, for a given proposition p, each agent gives maximum utility to the probability of p that each one takes to be 'correct' (i.e., to his or her actual credence on p); utility functions are assumed to decrease continuously from there. Given such utilities, I show how to build an appropriate (pseudo)bargaining situation (for proposition p), such that points inside the bargaining set are correlated with sets of probability assignments by the individual agents. I will argue that solving the bargaining problem helps us figure out the probability of p that can be credited to the group as a whole; traditional discussions on the adequacy and correctness of different bargaining solutions become relevant for our current setting as well. We then obtain a unified perspective on two seemingly disparate phenomena, probability aggregation and cooperative bargaining. The proposal illustrates how normative meta-models are meant to work: Bargaining models here act as meta-models that can help us elicit intuitions regarding probability aggregation (the first-order phenomenon) in an indirect way; as with most normative meta-models,

the legitimacy of Bargaining models springs from their role in a different context altogether.

2 Models in formal epistemology: The role of normativity

Formal Epistemology is an umbrella term, which includes many different disciplines and research lines. Correspondingly, we should expect to find many different types of models that can be said to belong to the formal epistemology realm, or to formal epistemology modeling. Consider, for example, the idealized setting that we take as a starting point to study the properties of certain voting procedure, within the quarters of social choice theory. Or consider how a system of dynamic epistemic logic explores the ways in which the knowledge of a group of agents should (ideally) evolve in response to new information; or how rational agents are supposed to accept hypotheses in agreement with a particular brand of cognitive decision theory, among many other problems. It is customary to use the word 'model' to refer to these and similar examples. Thus we typically read that a *S*5 system *models* knowledge in a way that allows for negative and positive introspection. Or we say that while AGM proposal in Alchourrón, Gärdenfors, and Makinson (1985) aims at *modeling* the way agents revise their beliefs when they learn new facts, Katsuno and Mendelzon (1991) is *a model* of belief update that is better understood as aiming to capture doxastic modifications that spring from environmental changes. Or we read that Bayesian conditionalization provides a *model* for (partial) belief change that applies to agents in particular contexts of uncertainty. One salient property of all these examples, compared to typical models in the empirical sciences, is that they exhibit strong normative features, in one way or another. We can even wonder whether the term 'model' does not have a different meaning once normativity enters so heavily into the picture. Indeed, the nature and features of normative models appear elusive, and not as well studied as their counterparts within empirical models. In what follows I will try to set some background for further discussion.

We can begin by considering a very broad distinction between normative and descriptive models, where "normative models attempt to fit normative facts" (Titelbaum, forthcoming; more on this later). Other authors would rather commit to a tripartite setting that classifies models into normative, prescriptive and descriptive. So-called prescriptive models are conceived of as tools to try to improve the extent to which practice differs from what it should be (according to some account of normativity); we find a paradigmatic example of this distinction in the work of Jonathan Baron:

> One task of our field is to compare judgments to normative models. We look for systematic deviations from the models. These are called biases. If no biases are found, we may try to explain why not. If biases are found, we try to understand and explain them by making descriptive models or

> theories. With normative and descriptive models in hand, we can try to find ways to correct the biases, that is, to improve judgments according to the normative standards. The prescriptions for such correction are called prescriptive models. (Baron, 2004, 19, in the context of a discussion about judgment and decision making.)

Not everybody understands the relation between the normative and descriptive realms in the same way. Some authors tend to conceive of whole areas of inquire normally thought of as normative as ultimately descriptive. Yap (2014) for example, argues that epistemic logic is in fact descriptive: it contains idealizations, in the same way scientific models do; thus "criticizing models of epistemic logic in which agents know all propositional tautologies as being unrealistic would be like criticizing frictionless planes in physics for being unrealistic." (Yap, 2014). More generally, other authors have argued that normative models should be really close to descriptive ones, to actual practice. Gabbay and Woods (2003) are a good example of this standpoint:

> how cognitive agents actually do behave is a (substantial approximation to) how they should behave. (Gabbay & Woods, 2003, 605; they make this remark regarding inference, rather than belief.)

And also:

> ...normativity inheres in how we act and behave... normativity is descriptively immanent, rather than transcendent. (Gabbay & Woods, 2003, p. 605.)

As it is apparent, this position defies a long and venerable tradition that would insist that one cannot derive 'ought' from 'is'. Regardless of the details, even if we embrace a 'descriptively immanent normativity', the case remains that idealized behavior is not actual behavior, and there will always be many different, incompatible ways to settle on what counts as ideal for each particular situation. Notice that, as opposed to other context in which we may find normative constraints, when applied to behavior, 'ideal' amounts to 'desirable' (for certain purposes and goals).

In any case, it might well be that no model is *purely* descriptive, prescriptive, or normative. To accommodate for this idea, let me try here a somewhat different taxonomy – one that seeks to capture the extent to which, and the sense in which, normativity can play a role within a particular model. In this sense the labels I will mention below are not meant to single out disjunctive types of models, but to identify non-exclusive features that might be present in any model. There are (at least) three ways in which normativity can enter into the picture – three ways in which models can be said to have normative traits:

(a) Normativity in the context of a (mostly) descriptive enterprise

It is important to acknowledge that normativity definitively plays a role within models from the empirical sciences. Empirical models can exhibit normative features, for example, at the time of reasoning *about* the model, or at the time of relying on various kinds of idealizations:

(i) No matter which discipline we are working on, eventually we will have to assess whether it is reasonable to assert certain claims within the model; for example, we will have to assess whether we have deductive or inductive reasons to draw certain inferences. This is just part of the business of carrying out what is customarily referred to as "investigation or research *about* the model" – as opposed to: from the model, to the target system.[1] At least in this minimal sense most models share some normative traits: if we accept a given starting point, other claims *should* be asserted too.

(ii) On the other hand, empirical models can be quite removed from reality, as we all know; many models involve what the literature dubbed either Galilean or Aristotelian idealizations (or both); many models are also said to be *caricatures*.[2] Idealizations sometimes seem to implement a sort of 'peeling off' process; at other times they may work by way of *adding* certain features, rather than just removing them.[3] As I see it, many idealizations incorporate normative features, to the extent that ideal entities of various kinds are conceptually required to possess certain traits. Normativity in this sense is not related to what someone should strive to do or achieve, nor is it related to considerations of better or worse performance. As Titelbaum (forthcoming), has rightly pointed out, "there's no sense in which a frictionless plane is *better* than a real plane". This is of course true; however, a frictionless plane *ought* to have certain properties – although this is not a moral 'ought'.

Focusing on idealization mechanisms can establish a link between empirical models and models within the formal epistemology realm (more on this later). However, typical models in formal epistemology include normative features in other senses as well, as we will see below.

(b) Constitutive Normativity

Some models seem to have the explicit goal of helping us acquire a better understanding of rationality, or personhood, among other possibilities – or, to be more precise: some models *can be used* with the explicit aim of bringing about such understanding, regardless of the original motivation of the authors who first developed them. We might conceive of them as aiming to represent normative facts (as Titelbaum would put it), or perhaps as aiming to represent our *intuitions* regarding the features of various types of non-empirical phenomena (fairness, rationality, correct language usage, or many others). Under the last conception, a big question at this juncture is whose in-

[1] Cf. also Williamson (forthcoming); models (including normative models) can give us both "vague unconditional knowledge" that compares one model to another, as well as precise conditional knowledge of the form "if a given case satisfies the model description, then it satisfies this other description too".

[2] For an overall perspective on idealized models cf. Frigg and Hartmann (2012).

[3] It is not always obvious which of the two procedures is actually in place in a given example. Cf. for example Mäki (2009): "In assuming perfect information on the part of economic agents, a model appears to add a feature that does not obtain in the real world: an excessively powerful mental capacity seems to be attributed to the model agents. However, it seems to me that the correct reading of the function of this idealizing assumption is that it is used to remove certain real-world features from the model world: the search, acquisition and processing of information."

tuitions we are seeking to model: those held by the layperson? By the expert? Which expert, anyway? The resulting model can be very different in each case.

(c) Normativity within a prescriptive setting

Some models depict ideal features of different phenomena with the explicit purpose of providing a guide to action; even if the full-fledged picture remains ultimately unattainable, it can function as a regulative ideal. One way of putting it could be to say that such models seek to capture 'prescriptive intuitions': intuitions *on what we ought to do*. In any case, we should avoid thinking of prescriptive and normative models as mutually exclusive types. Consider again an $S5$ system of epistemic logic with Common Knowledge. What is the target here? The model can be taken to represent certain aspects of the behavior of rational agents, and hence constitutive features of ideal reasoners. But at the same time, by way of modeling what perfect reasoners will do we may also prescribe what real agents *should* try to do, even if they will ultimately fail to do so, if they find it desirable to think of themselves as rational creatures. Consider now various voting models, within social choice theory. We could say that the goal here is mostly prescriptive, in the sense that it is possible to take conscious implementations seriously; at the same time, however, we can seek to use such models to have a better understanding of rational features of fair choice, or impartiality, among other things.

The present reflections point to the fact that normativity is not an all-or-nothing affair, which speaks in favor of treating empirical models and models in formal epistemology (and, in general, in philosophy) along similar lines. Indeed, several accounts try to assimilate normative to empirical models, by claiming that the most characteristic features of normative models are in fact idealizations akin to the ones we find in the empirical sciences (cf. Baron, 2004, p. 24; Williamson, forthcoming; or Colyvan, 2013). This has led some authors to claim, for example, that rationality models are not 'fully' normative.[4] I find this move curious, as I tend to think that the situation is exactly the opposite: the fact that idealizations are widespread shows that there are normative elements within empirical models, rather than non-normative elements in models from logic or formal epistemology. In any case, regardless of whether we take idealizations to inject normativity (to descriptive models) or to take it away (from models of rationality), it is not clear to me whether what I have called 'constitutive normativity' can be fully reduced to the presence of various idealizations of actual behavior – even though idealizations can of course play a role at the time of building such models.

[4]Cf. Colyvan (2013), or Yap (2014). Colyvan contends that theories of rationality within formal epistemology are normative, but "they are not normative through and through" (2013, p. 1338); he wants to separate the normative elements from other kinds of idealizations, which he takes to be non-normative. On the other hand, Yap (2014) follows Weisberg (2007) in identifying three kinds of idealizations in science (Galilean, minimalist and multiple-models); all three of them are present in models within epistemic logic. Yap suggests that, to the extent that we can show that the unrealistic features are in fact idealizations of various types, a given model is not truly normative, but descriptive.

3 A problem

In light of the above, we could be tempted to conclude that the distinction between models in the empirical sciences and models in formal epistemology is really a matter of degree. However, things are slightly more complicate. Even authors sympathetic to pragmatist conceptions of models, like Suárez (for whom the reference to agents and purposes is essential, and models do not 'mirror' the world in any interesting sense) are fast to point out that "scientific representations have cognitive value because they aim to provide us with specific information regarding their targets" (Suárez, 2004). As I see it, one of the main problems we face here is that, when constitutive normativity enters into the picture, the target system becomes elusive. What type of information are we aiming at, exactly? It is true that gathering data can always be a difficult business, regardless of the discipline. However, if our data are normative facts the problem is compounded. How can we ever ensure we are getting such facts right? We could try to overcome this problem, to some extent, by claiming that normative models target our intuitions on various normative matters. But this answer does not solve our worries.

The reason is the following. An assessment of the correctness of models in the empirical sciences can proceed, among other things, by way of keeping track of their predictions. How does the predictive enterprise work in the case of (mostly) normative models? Suppose we agree that such models seek to capture certain intuitions as a starting point; later on we draw consequences from there, which might force us to commit ourselves to claims we would not have thought of before building the model. Such consequences should be equally intuitive. To put it differently, normative models can be said to make predictions regarding which other claims we will find intuitive. So far so good. The problem is that, as I have already pointed out, intuitions on various phenomena are seldom universally shared. Therefore, even though we can still assess whether certain sets of intuitions have been more or less well captured by a given model, it is not clear how to assess whether the intuitions that lie at the heart of it are misguided in the first place, or perhaps whether the model itself is faulty. Unwanted consequences (i.e., consequences that are no longer intuitive, or which are at odds with the intuitions that constituted our starting point) can mean either than the model was not a good one, or that some of our intuitions should be re-educated: perhaps we should bite the bullet and accept the odd consequences, or perhaps we should reconsider the plausibility of our starting point altogether. In this sense, modeling intuitions fosters endogamy: the quality control mechanisms for adequate model building are not independent of the model we are using.

It should be clear that what we may call 'the endogamy problem' of normative models is not tied to our description of the target system in terms of intuitions. Consider, for example, Titelbaum's analysis (in Titelbaum, forthcoming). He endorses a well known distinction among the model framework, its interpretation, and the individual model that results from so interpreting the framework. In the empirical sciences, we accommodate the particular model (i.e., the particular interpretation) whenever the

predictions do not match the data; on the other hand, if there is a significant mismatch, we tend to question the framework itself. According to Titelbaum, normative models are analogous to empirical models in this respect.[5] Take once again the AGM model for belief change: "after the loss of belief, rationality requires the agent to believe such-and-such propositions. Now suppose that this prediction mismatches the data: It's not the case *that rationality requires the specified beliefs* [my emphasis]. This calls the combination of AGM and our interpretation into question." But the problem is how we can ever check whether rationality requires or does not require certain things. The very possibility of putting the model to the test requires our having some mechanism to gather reliable data which is at least partially independent of the modeling process itself. This is sometimes a problem for the empirical sciences as well, but it is not clear how normative models can ever avoid it. It is of course true, as Williamson points out, that "if we started in total ignorance about the target, we could hardly expect to learn much about it by modelling alone" (Williamson, forthcoming), but by itself this observation leaves us in the dark as to how to proceed when we face competing candidates to function as the *real* set of normative facts.

Endogamy, in the sense just explained, might be a trait of normative models we have to learn to live with. Is there anything we can do to make things better? In the next section I will make a suggestion that goes some way towards presenting a possible improvement, at least for some cases.

4 Modeling first-order normativity

The problem described in the previous section can be summarized as the problem of choosing among competing normative models. This is related (though not identical) to the problem of choosing among competing sets of intuitions. My suggestion here is that in order to choose among various normative models it could be helpful to count with a *second*-order model. In other words, sometimes we can profit from developing meta-models for first order normative settings. I do not want to say that every legitimate modeling strategy for normative models should rely on second-order considerations, but only that this is an interesting possibility, which enables us to think about the problem of testing normative models, in particular models within formal epistemology, under a somewhat different light.

Very briefly, this is what I have in mind. We begin the process, at the very first moment, by acknowledging that we have certain intuitions (on rationality, on equity, on grammatical correction, etc); if you prefer, we can take such intuitions to correspond

[5] Sometimes, when predictions are not satisfied, we can explain away the inconsistency by better specifying the intended domain of application (on this cf. also Colyvan, 2013); by contrast, *genuine* counterexamples (i.e., those that amount to a real threat for the model) are those that belong to the intended application domain.

in each case to a particular element from a normative space.[6] We are supposed to elicit our normative intuitions, or perhaps our intuitions on what we have to do, through a model encompassing constitutive normativity, or a model with prescriptive features. But very often we will find incompatible sets of intuitions, at level zero, which may give rise to different normative models at level one. For example, sometimes we have a list of criteria, each element of which sounds perfectly reasonable to us, but which cannot be jointly satisfied (think for example of the several desiderata we can use to fix of a voting procedure, or of the many ways to escape from Arrow's impossibility result). How shall we choose among different normative models, then?

A possible way to go is to try to find a model that was originally developed for a (possibly) very different target system (ideally, from the empirical sciences), and use it *to represent features of some of the first order normative models*.[7] The second-order model then represents, indirectly, a particular normative or prescriptive element from our normative space. Such a second-order model can only be said to represent such elements (or our level zero intuitions) indirectly, because the immediate goal is to explain the mechanism behind a given normative model. By doing so it can reveal not so obvious features of the normative model in question. *Then a first order normative model is worth selecting if we find a suitable second-order model that represents it*. In short, the tool to choose can be precisely another model, preferably some other model with roots in empirical phenomena.

Ultimately, this can be seen as a more sophisticated way to test our intuitions regarding what ought to be the case. By proceeding in this way we can notice that the original normative fact or the original prescriptive task (for which we are trying to elicit intuitions) shares structural features with other phenomena, so we can come to see the original problem as part of a more general question. In this sense the second-order model can provide *unification*: it can help us see apparently disparate phenomena in a unified way.

A short clarification: The way I am conceiving of it here, the second-order model is not meant to supersede the first-order model(s) that it represents. Rather, the aim of a particular meta-model is to provide *validation* for a given first-order model. When this happens, we can say that the given first-order model is robust. Sometimes none of the elements from our prior pool of first-order models gets so validated; instead, the second-order model points to the possibility of building an alternative first-order structure, of which the modelers where not aware before. In this case no robustness is revealed, which is in itself a kind of progress, in a sort of Popperian way (though of course I do not mean to say that first-order models which are not found robust are

[6] Concomitantly, we can take the target system to be a particular 'normative fact'. For the purposes of this section, it is irrelevant whether we take target systems to be sets of intuitions (on normative facts), or the normative facts themselves.

[7] Several authors acknowledged that models can be reinterpreted; in Titelbaum's terms, the same structure can be used in multiple applications, some of them normative and some of them descriptive (Titelbaum, forthcoming). Cf. also Weisberg (2013).

thereby *disproved*).

Finally, it is clear that not all (first-order) normative scenarios will allow for second-order modeling. Still, the fact that this maneuver is in principle available contributes to ease our worries regarding the testing of normative models.

In what follows I will illustrate this proposal with the problem of probability aggregation. I will suggest that we turn our attention to bargaining models; we will see how a single (meta)model (the bargaining model) can be profited to do more than one task at the same time. In this way we can come to see probability aggregation as a sub-problem of cooperative game theory. For our present purposes, the target system of the meta-model is a first order normative model, which will typically consist in sets of equations; our meta-model could eventually help us decide among an array of first-order models, and choose the "right" set of equations.

5 An illustration: A (meta)model for probability aggregation

Consider a set of probability functions. The individual measures may actually be the measures of the individual members of a given group; alternatively, they can also represent different attempts to capture experimental results, perhaps by a single agent, among many other possibilities. How shall we represent the probability *of the set* as a whole? A rather straightforward answer would be to represent it as a set as well – perhaps a convex set containing the individual measures. Others would argue that what we actually need is a bona fide aggregation method, a method that delivers one single probability function. I will not try to settle this discussion here. Rather, in what follows I will just *assume*, for the sake of the argument, that we are interested in finding *the* single measure that can be said to correspond to the probabilistic attitude of a group.

How shall we combine the individual functions in order to obtain a single measure, then? Notoriously, there are many aggregation rules we could follow here. Among the most common solutions we should include:[8]

- Linear opinion pools (e.g., the arithmetic mean)
  $F(P_1 ... P_n) = w_1 P_1 + ... + w_n P_n$
- Geometric opinion pools
  $F(P_1 ... P_n) \propto P_1^{w1} ... P_n^{wn}$

[8]For a classic overview on probability aggregation cf. Genest and Zidek (1986); cf. also Dietrich and List (forthcoming).

- Supra Bayesian approaches
  $F(P_1...P_n)(A) = F(A|P_1...P_n)$
- Multiplicative opinion pools:

  $$\pi \propto \frac{\pi_1...\pi_n}{P_1...P} F(P_1...P_n)$$

(where 'F' is our aggregation function; '$P_1...P_n$' are the individual probability functions, for some finite n; '$w_1...w_n$' are weights that add up to 1; and '$\pi_1...\pi_n$' refer to i's posterior probability, for each $i \in (1...n)$.)

How shall we proceed, then? This is a non-exhaustive list, of course, but even if we restrict our attention to these few options, there is no consensus regarding which strategy is best all things considered. The standard answer is to examine which properties are fulfilled in each case, and let such properties guide our choice of an aggregation method. For example, if we want aggregation to commute with conditionalization, we could pick a geometric opinion pool; if we want an analogous of the so-called independence of irrelevant alternatives to hold in this context, the linear opinion pool seems to be a good choice, whereas if we favor an independence preservation property, linear opinion pools no longer look promising.[9] Some sets of criteria may be more desirable than others for particular purposes and in particular contexts (for example, Supra Bayesian accounts seem appropriate for contexts in which the aggregation is performed by a single agent external to the group). But in many cases it may be unclear how to assess the properties, or how to decide which set of criteria is more relevant.

[9] A property such as:

> $T(P_1,...,P_n)(A) = F[(P_1(A),...,P_n(A)]$, for some arbitrary $F : [0,1]^n \to [0,1]$, and every event A in the algebra

implies that the aggregation is a linear opinion pool. On the other hand, linear opinion pools cannot satisfy the so-called 'independence preservation property':

> $T(P_1,...,P_n)(A \cap B) = T(P_1,...,P_n)(A)T(P_1,...,P_n)(B)$, whenever $P_i(A \cap B) = P_i(A)P_i(B)$ for some A and B in the algebra,

unless they are dictatorial or trivial (i.e., unless weights are 0 for all agents except for one). Cf. Genest and Zidek (1986, p. 117).

6 The proposal

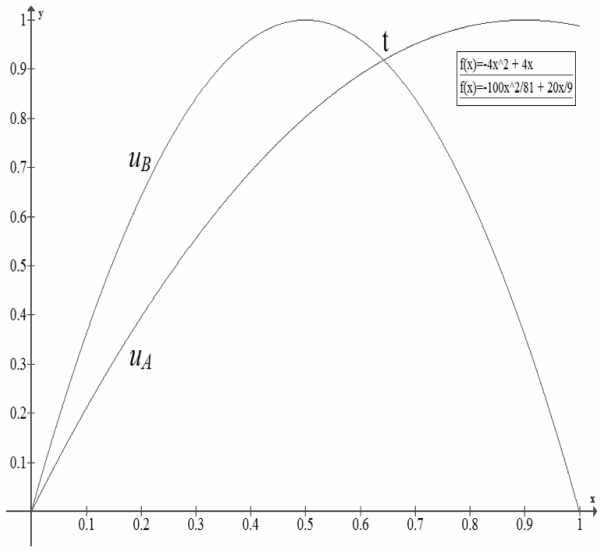
Let us look at the problem from a different perspective. Let me clarify at the outset that this proposal will not work for every possible aggregation problem. Still it can model quite well, I think, the attempt to arrive at a single function for a group of agents when there is no third party agent, external to the group, in charge of producing the aggregation. In other words, the present account attempts to capture the mechanism of probability aggregation from the inside, as it were – as experienced from the first person perspective. In addition, the present account is not meant to work for a complete probability distribution; rather, it seeks to determine which probability the group is entitled to have, qua group, on a particular proposition, or on a particular partition of rival propositions or hypotheses. The proposal is still programmatic, but I hope the broad picture is clear enough.

In what follows we assume that each agent thinks of herself as well suited to have a well-founded opinion – at any rate, each agent thinks of herself as being at least as much entitled to having a well-founded opinion as any of her peers. In this context, we will also assume that the members of the group have *utilities* over the possible probability assignments, such that each of them gives maximal utility to her actual credence; utilities are assumed to decrease continuously from there. Utilities here do not measure the desirability of a given proposition, of course, but *the desirability of adopting a particular (not necessarily precise) probability on that proposition*. Each agent would like to impose her view regarding that particular proposition on the group; as far as each individual agent is concerned, the closer the collective probability is to her own credence, the better.

By way of concreteness, let me illustrate with a simple example. For the easiest case, consider a group made of two people, agents A and B, with definite probability assignments on some proposition p. Let $P_A(p) = 0.9$ and $P_B(p) = 0.5$. Consider now each agent's utilities on the possible probability of p. We can take utility scales to be equivalent under positive linear transformation; to simplify we will consider values in the range $[0,1]$. Different utility functions may be appropriate here. In what follows I will illustrate with quadratic functions; this is not the only way to go, of course, but there are some technical and conceptual advantages in proceeding in this way. Agent A gives maximal utility to probability 0.9, and minimum utility to probability 0, whereas agent B gives maximal utility to probability 0.5, and minimum utility to both probabilities 0 and 1. Thus we have:

- For A : $u_A(x) = -100x^2/81 + 20x/9$
- For B : $u_B(x) = -4x^2 + 4x$

We can draw the two curves on the same graph, such that the probability of p is placed on the x-axis:



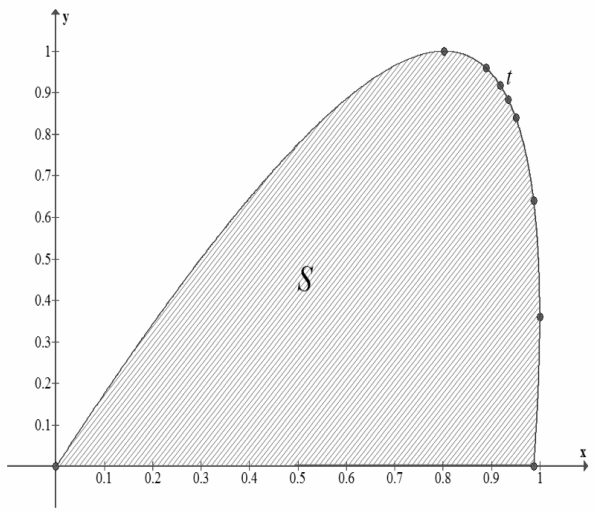
We can see that u_A and u_B intersect at $t = (0.643, 0.918)$, which maximizes the two functions simultaneously. Can we say that t captures the probability of the group? Ultimately, the answer is 'yes', but a proper justification for this claim requires a few more steps. Notice that point s corresponds neither to the arithmetic mean (0.7), nor to the geometric mean (0.671) between P_A and P_B. How shall we interpret t? Let us look at this problem more closely.

To have a better understanding of the task utility functions are performing, we shift to a second graph. Let us place utility scales of agents A and B along the x- and y-axes, respectively. Consider first the set of all points on the square $[0,1] \times [0,1]$. We will interpret each point as *a set of probabilities*: the probabilities that correspond to the utility level of the two agents at that point.[10] Thus, for our example, the point $(1,1)$ corresponds to $\{0.5\ ;\ 0.9\}$. We then build a (pseudo) bargaining situation $S \subseteq [0,1] \times [0,1]$. We define S as the smallest convex set that includes all points for which the utilities of the two agents agree on the same probability; we can take the defect point to be $(0,0)$.[11] Clearly, points on the boundary of S will represent singletons (unique probabilities); by way of illustration, consider the following points on the boundary of S, up to three decimal places:

[10] There is a possible alternative interpretation here, according to which such sets are taken to be convex. In what follows I will not adopt this path, but nothing changes substantially if we do.

[11] The defect point is not playing any substantial role in the present meta-model. This might motivate us to explore in the future the use of bargaining proposals that do not demand a defect point (thanks to Ted Seidenfeld for making this suggestion).

$(0,0)$	represents	$\{0\}$
$(0.802,1)$	represents	$\{0.5\}$
$(0.889,0.96)$	represents	$\{0.6\}$
$(0.918,0.918)$	represents	$\{0.643\}$ (as captured by s)
$(0.935,0.883)$	represents	$\{0.671\}$ (the geometric mean)
$(0.951,0.84)$	represents	$\{0.7\}$ (the arithmetic mean)
$(0.988,0.64)$	represents	$\{0.8\}$
$(0.988,0)$	represents	$\{1\}$
$(1,0.36)$	represents	$\{0.9\}$



It is easy to see that t will lie in the diagonal; the probability captured by t is represented by the furthest North-Eastern point for which (at least) one probability assignment of agent A coincides with (at least) one probability assignment of agent B, and hence the set is a singleton. This amounts to the Kalai-Smodorinsky and the Egalitarian bargaining solutions.[12] Notice that Nash solution does not coincide with s, and neither do the arithmetic or the geometric mean between $P_A(p)$ and $P_B(p)$, as we have already pointed out. In any case, both the arithmetic and the geometric mean lie within the borders, so neither of them is Pareto inferior. Notice, however, that no stan-

[12]For an overview cf. for example Gartner (2009, chapter 8).

dard game-theoretic solution provides justification for picking them, under the current utility functions.

It is also interesting to notice that, within this setting, point $(1,1)$ is not in S. This fact rules out the possibility that the probability of the group be represented by the (possibly convex) set containing the preferred credences of each agent. The decision to proceed in this way can be thought to be the result of a trade-off between maximizing utility and minimizing higher order uncertainty – where a set containing more than one probability measures can be said to carry more uncertain (at least in some sense of this word) than a singleton. The current setting, in this sense, favors uniqueness.

Is the Kalai-Smodorinsky solution the way to go? We need not take a definite stance on this question for the moment. Regardless of the details, the core idea is that we can look at probability aggregation *as a type of cooperative bargaining*. Thus, we can borrow well-known discussions on the adequacy and correctness of different bargaining solutions and apply them here. We then obtain a unified approach to seemingly disparate phenomena.

This approach is still programmatic, but we can think of many different ways to continue it in the future. To begin with, notice that the graphics might look very different if we change the scenario a bit. For example, one of the agents may be truly uncertain about her first order probability assignments, so she might not give utility 1 to any probability value whatsoever (in which case the $(1,1)$ point will not be feasible). Or she might have a (convex) set of probabilities herself, and assign utility 1 to a whole interval. Alternatively, for highly opinionated individuals utility functions may descend abruptly, and agents might feel compelled to give *utility* 0 to a whole interval. In this scenario u_A and u_B might fail to intersect. Once again, at the time of constructing the bargaining set we will have to make a decision as to which points to include; we can again think of a trade-off between the amount of utility captured by each point, and the size of the set of probabilities such points represent in each case (the smaller the better). It would also be nice to work out generalizations to take care of n-person groups.

Related to this, an interesting project would be to look for the precise conditions (such as the specific utility functions) that will enable us to represent traditional methods for probability aggregation as solutions to bargaining problems, such as the arithmetic or the geometric mean. Once we accomplish this, we are left with the empirical problem of finding out whether actual people can be credited with utility functions close to those that would validate traditional proposals from the point of view of the bargaining model. Last, but not least, we can also investigate the extent to which agents can be motivated to adopt the aggregated measure as their own, in which case we will find a natural connection with the peer disagreement literature.

7 Conclusions

In the previous section I have suggested a principled account for probability aggregation that puts the *value* of probabilities at the center of the discussion. The proposal seems particularly well suited to deal with cases in which there is no external agent in charge of producing the aggregation; in that sense it can be said to honor a first person perspective.

Let me go back to our starting point now – to the discussion of normative models in formal epistemology – to see what we have accomplished. The proposal discussed in the previous section constitutes a good illustration of how normative meta-models are supposed to work. Bargaining models as I have used them here can be taken to aim at two different target systems at the same time: (i) they seek to capture intuitions on how groups of agents should negotiate their differences, and (ii) they seek to represent (in the mathematical sense) models that in turn seek to capture our intuitions on probability aggregation in general. Notice that bargaining models can help us elicit intuitions regarding probability aggregation in an indirect way: Strictly speaking, different bargaining models represent intuitions *regarding the way to settle a particular type of group disagreement*. Now, as it happens, some such settlement strategies can *in addition* represent particular sets of equations, which in turn may capture (normative) intuitions on probability aggregation – where the legitimacy of the bargaining models is inherited from their role in a different context altogether. In our particular example, the settlement strategies we selected did not coincide with any of the traditional approaches on probability aggregation (none of the standard first-order models got validated by the meta-model), so in this case they actually *defined* a new aggregation proposal.

As I have anticipated, relying on meta-models can help us unify apparently disparate situations and, at the same time, it helps us offer an illuminating interpretation of the mechanism that underlies the normative phenomenon. This fosters a better understanding of the phenomenon under consideration, and gives us the opportunity to further test our intuitions concerning what ought to be the case.

Bibliography

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, *50*(2), 510–530.

Baron, J. (2004). Normative models of judgment and decision making. In D. J. Koehler (Ed.), *The Blackwell handbook of judgment and decision making* (pp. 19–36). Blackwell Publishing.

Colyvan, M. (2013). Idealisations in normative models. *Synthese*, *190*(8), 1337–1350.

Dietrich, F., & List, C. (forthcoming). Probabilistic opinion pooling. In C. Hitchcock, & A. Hajek (Eds.), *Oxford handbook of probability and philosophy*. Oxford: Oxford University Press.

Frigg, R., & Hartmann, S. (2012). Models in science. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Fall 2012 Edition)*. Retrieved from http://plato.stanford.edu/archives/fall2012/entries/models-science/

Gabbay, D. M., & Woods, J. (2003). Normative models of rational agency: The theoretical disutility of certain approaches. *Logic Journal of IGPL*, *11*(6), 597–613.

Gartner, W. (2009). *A primer in social choice theory*. Oxford: Oxford University Press.

Genest, C., & Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, *1*, 114–148.

Katsuno, H., & Mendelzon, A. O. (1991). On the difference between updating a knowledge base and revising it. In *Principles of knowledge representation and reasoning: Proc. Second international conference (KR '91)* (pp. 387–394). San Francisco, California: Morgan Kaufmann.

Mäki, U. (2009). MISSing the world. Models as isolations and credible surrogate systems. *Erkenntnis*, *70*(1), 29–43.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, *71*(5), 767–779.

Titelbaum, M. G. (forthcoming). Normative modeling. In J. Horvath (Ed.), *Methods in analytic philosophy: A contemporary reader*. Bloomsbury Academic Press.

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, *104*(12), 639–659.

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford: Oxford University Press.

Williamson, T. (forthcoming). Model building in philosophy. In R. Blackford, & D. Broderick (Eds.), *Philosophy's future: The problem of philosophical progress*. Oxford: Wiley.

Yap, A. (2014). Idealization, epistemic logic, and epistemology. *Synthese*, *191*(14), 3351–3366.

Author biography. Eleonora Cresto (Ph.D. in Philosophy, *Columbia University*, 2006). She is Philosophy Professor at Universidad Torcuato Di Tella and UNTREF

(Buenos Aires), as well as a Permanent Researcher at the CONICET (National Council for Scientific and Technical Research, Argentina). Her main interests lie in both formal and mainstream epistemology; she has also written on philosophy of science, decision theory, and philosophy of logic, among other topics. Currently she is Associate Editor of *Erkenntnis*, and Area Editor of *Ergo*, for Epistemology.