Imaging all the People

Hannes Leitgeb
Hannes.Leitgeb@lmu.de

1. Introduction

A standard proposal of aggregating the degree-of-belief functions or subjective probability measures of subjects $S_1,..., S_k$ into a joint "group probability measure" P is *linear pooling* (or *averaging*):[1] let $\alpha_1,\ldots, \alpha_k$ be real numbers in the closed interval [0,1], such that $\alpha_1 + \ldots + \alpha_k = 1$. If P_i is the probability measure of subject S_i, then the result of applying linear pooling to $P_1,\ldots, P_k$ with respect to $\alpha_1,\ldots, \alpha_k$ is the probability measure P, such that for all propositions A:

$$P(A) = \alpha_1\, P_1(A) + \ldots + \alpha_k\, P_k(A).$$

Intuitively, and geometrically, such a social probability measure P lies "between" the individual probability measures $P_1(A),\ldots, P_k(A)$, where the exact "position" of P depends on the degree of influence that is exerted on the prospective social measure by the individual subjects, as expressed by their corresponding coefficients or weights $\alpha_1,\ldots, \alpha_k$ (which might reflect their expertise or the social respect that they enjoy in the group). Mathematically speaking, P is simply a convex combination of $P_1,\ldots, P_k$.

In spite of its simplicity, naturalness, and various other virtues (as discussed in the corresponding literature), it is well known that linear pooling is also subject to a concern

[1] For a standard overview of probabilistic pooling in general (that is, the combining or aggregating or amalgamating of probability measures), see Genest and Zidek (1986), where also linear pooling is discussed in detail (see e.g. p.117), and Dietrich and List (forthcoming), in which Sections 4 and 5 are devoted especially to linear pooling. For the discussion and criticism of linear pooling in particular, see e.g. Wagner (1985) and Steele (2012). Dietrich (2010) includes a recent criticism of all classical pooling methods, including linear pooling.

regarding update or learning:[2] other than trivial cases, conditionalization on evidence followed by linear pooling differs in outcome from linear pooling followed by conditionalization on evidence. So it will make a difference whether, at first, each individual in the group updates in the standard Bayesian manner, after which their posterior beliefs are linearly aggregated, or whether first their prior beliefs are linearly aggregated, and only then the emerging group belief is updated by the standard Bayesian method of learning. (I am going to simplify the discussion by assuming, for the sake of the argument, that we are dealing with a situation in which such combinations of update and aggregation would be desirable.)

More formally: let some numbers $\alpha_1,\ldots, \alpha_k$ be given as stated above, and let us assume that the subjects $S_1,..., S_k$ are confronted with a new piece of evidence E that they are supposed to accommodate as a group, and to which they all assign positive probability individually; so, for all i, $P_i(E)>0$. Then, except for trivial cases (e.g., where $\alpha_1 = 1$ and $\alpha_2,\ldots, \alpha_k = 0$, or where $P_1,\ldots, P_k$ are identical), the outcomes of the following two natural sequences of epistemic actions will *differ*: (i) each individual subject follows the standard diachronic norm of Bayesian rationality of first updating his or her degree-of-belief function P_i by conditionalizing on E, after which the resulting individual probability measures $P_i(\cdot/E)$ (that are given by $P_i(A/E) = P_i(A \;\&\; E)/P_i(E)$) are aggregated by linear pooling with respect to $\alpha_1,\ldots, \alpha_k$. (ii) First, linear pooling is applied to $P_1,\ldots, P_k$ with respect to $\alpha_1,\ldots, \alpha_k$, after which the resulting social probability measure P is updated to $P(\cdot/E)$ by conditionalization.

For instance, for two subjects ($k = 2$), it is generally *not* the case that

$$\text{(i) } \alpha_1 P_1(A/E) + \alpha_2 P_2(A/E) =$$

$$\text{(ii) } [\alpha_1 P_1 + \alpha_2 P_2](A/E)$$

[2] See Genest and Zidek (1986), p.117, and Dalkey (1975).

where '$[\alpha_1 P_1 + \alpha_2 P_2]$' denotes the linear pool of P_1, P_2 with respect to α_1, α_2. In mathematical terms: the underlying diagram does not commute. Or in terms of the literature on probabilistic aggregation: linear pooling is not "externally Bayesian".[3]

Clearly, this leads any group of Bayesian subjects who insist on linear pooling as their favored method of aggregation into a normative dilemma: how should they update *qua group*? For there seem to be equally good reasons for them to update by means of (i) as there are for (ii). But the group cannot do both (except for trivial cases), since the resulting group measures in the two cases will differ. Various reactions to this dilemma can be found in the relevant literature,[4] such as: dropping linear pooling in favor of some alternative pooling mechanism; changing the coefficients $\alpha_1,\ldots, \alpha_k$ in the course of the update; and more.

In the following, we want to suggest a different kind of response, which, to the best of our knowledge, has not been discussed before: keeping both linear pooling and the coefficients intact but *dropping conditionalization in favor of an alternative rational update mechanism*. As we are going to see, this can be done so that the new updating scheme is itself well-known, natural, and well-behaved; if each individual learns by means of it, then linearly pooling the so-updated probability measures indeed leads to the same outcome as linearly pooling their prior probability measures and then applying this alternative update scheme to the group probability measure; and all of that is going to follow from a result that has already been established in a different area (the study of conditionals) but which, it seems, has not been applied before in the area of probabilistic aggregation. This alternative probabilistic updating mechanism is: (general) *imaging*.

Section 2 will explain the essentials of this alternative learning scheme. Section 3 will show, based on an existing formal result by Gärdenfors (1982), that general imaging

---

[3] External Bayesianity was studied first by Madansky (1964). Genest et al. (1986) provide a characterization of External Bayesianity in terms of properties that are closely related to so-called logarithmic pooling methods.

[4] See Genest and Zidek (1986), pp.117-120.

commutes with linear pooling, and therefore the dilemma from above does not apply to that alternative update method. In Section 4 we will interpret this result in less abstract terms, and we will illustrate our findings in terms of a concrete toy example. The final Section 5 will summarize what has been achieved, and it will point to a consequence that these findings may have for rational group decision-making.

2. General Imaging

Let $W = \{w_1,\ldots, w_n\}$ be a sample space of finitely many possible worlds (for simplicity). Different interpretations of the members of W are permissible, but for our purposes it will be most convenient to regard them not so much as full metaphysical specifications of what the actual world might have been like in the past, present, and future, but rather as merely synchronous specifications of what the *present state* of the actual world might be like. For instance, grabbing a banana at a world and eating it will correspond to a change from that world (in which, say, there is precisely one banana in the fruit basket) to another one (in which no banana is in the basket). Such an interpretation of worlds is often employed in decision theory or in deontic logic, and it will be presupposed in what follows.

As usual, we are going to identify propositions with subsets of W. The probability measures that we will consider will always be defined on all, and only, the subsets of our given, finite set W.

In his famous article on "Probabilities of Conditionals and Conditional Probabilities", David Lewis (1976) suggested a novel kind of probabilistic update method called 'Imaging' which works like this: when imaging a probability measure P on a non-empty or consistent proposition E,[5] the probability of each possible world w' is being transferred to a uniquely determined "image" world w in E that may be interpreted as the unique world in E that is closest or most similar to w' (assuming, with Robert Stalnaker's classical semantics for conditionals, that such a uniquely determined world exists). Once all such transfers have taken place, the result of imaging P on E is the image function $P(\cdot\backslash E)$ that is given by defining $P(A\backslash E)$ to be the sum of all these new probabilities of the worlds w that are members of A.[6] As Lewis argued, both conditionalizing P on E (which results in $P(\cdot/E)$) and imaging P on E (which results in $P(\cdot\backslash E)$) may be regarded as "minimal" revisions of P by E, it is just that the term 'minimal' must be assigned different meaning in each case. Lewis also showed that Stalnaker's thesis on conditionals (a version of the Ramsey test for conditionals) holds if it is formulated for imaging, while famously it fails if formulated for conditionalization. In the literature on conditionals, both conditionalization and imaging have been studied as probabilistic methods of *supposing* E rather than *learning* E, but in what follows their learning interpretation will be more salient.

Not much later both Gärdenfors (1982) and Lewis (1981)[7] proposed a generalization of imaging in which it is no longer assumed that for each "source" world w' there will be a unique "image" world w to which all of the probabilistic mass of w' is being transferred. Instead, the probabilistic mass of w' may be distributed over various "image" worlds, where each such world w will only receive $T_E(w', w) \cdot 100$ percent of the original probability $P(\{w'\})$ of w', as being determined by what we are going to call a transition or transfer function T. Lewis' (1976) original "sharp" version of general imaging is then given by the

[5] Updating on the empty (contradictory) set of worlds will not play a role in what follows.

[6] We borrow the notation '\' from Joyce (1999).

[7] Lewis (1981) refers to an unpublished manuscript by Gärdenfors (which turned later into Gärdenfors 1982); Lewis' paper also includes references to further formalisms in the literature that coincide with, or are closely related to, general imaging. We should note that our own presentation will in some respects be closer to Lewis' than to Gärdenfors'; in particular, Lewis presents general imaging in a possible worlds framework just as we do, while Gärdenfors aims to avoid reference to possible worlds in his presentation.

special case in which for every world w' the quantity $T_E(w', w)$ is equal to 1 for precisely one world w (and equal to 0 for all other worlds).

Stated more formally, *general* imaging amounts to this: let T be a mapping which takes as arguments an arbitrary non-empty proposition E, a world w' in W, and another world w in W, and which maps them to a number $T_E(w', w)$ in the closed interval [0,1], such that for all w' in W it holds that $\sum_{w \text{ in } W} T_E(w', w) = 1$.[8] The probability measure $P(\cdot \backslash\backslash_T E)$ that results from general imaging of a given probability measure P on a proposition E (relative to the transfer function T) can then be defined on worlds w in W in this manner:

$$P(\{w\} \backslash\backslash_T E) = \sum_{w' \text{ in } W} P(\{w'\}) \cdot T_E(w', w).$$

Once the updated probability for each world w has been determined as such, one can define for all propositions A: $P(A \backslash\backslash_T E) = \sum_{w \text{ in } A} P(\{w\} \backslash\backslash_T \mathrm{E})$. By the assumption that for all w' in W, $\sum_{w \text{ in } W} T_E(w', w) = 1$, it follows that $P(\cdot \backslash\backslash_T \mathrm{E})$ is a probability measure again, since the fractions $T_E(w', w)$ for fixed w' sum up to 1 (or 100 percent), and hence in the course of transferring probabilities no probabilistic mass is destroyed or created.[9] The measure $P(\cdot \backslash\backslash_T \mathrm{E})$

[8] In fact, one might want to impose the more stringent requirement that $\sum_{w \text{ in } E} T_E(w', w) = 1$ where the sum in question is taken just over the members of E instead of all worlds whatsoever: assuming that requirement (e.g. with Lewis 1981) has the natural consequence that general imaging of P on E (relative to T) will determine a probability measure that assigns probability 1 to E. Since this feature of updating on E will not play a role, however, for any of the arguments later in the paper, we will not take up this requirement here.

[9] Since for fixed E and w', $T_E(w', w)$ gives rise to a probability measure $T_{E,w'}$ that is both evidence-relative and world-relative (by setting $T_{E,w'}(A) = \sum_{w \text{ in } A} T_E(w', w)$), the result of imaging P on E relative to T can also be expressed in the following alternative manner,

$$\begin{aligned} P(A \backslash\backslash E) &= \sum_{w \text{ in } A} P(\{w\} \backslash\backslash \mathrm{E}) \\ &= \sum_{w \text{ in } A} \sum_{w' \text{ in } W} P(\{w'\}) \cdot T_E(w', w) \\ &= \sum_{w' \text{ in } W} \sum_{w \text{ in } A} P(\{w'\}) \cdot T_E(w', w) \\ &= \sum_{w' \text{ in } W} [P(\{w'\}) \cdot \sum_{w \text{ in } A} T_E(w', w)] \\ &= \sum_{w' \text{ in } W} P(\{w'\}) \cdot T_{E,w'}(A), \end{aligned}$$

in the last line of which $P(A \backslash\backslash E)$ may be interpreted as a convex combination of these evidence&world-relative probabilities $T_{E,w'}(A)$ of A, or, if one likes to, as the "expected chance" of A

is then called the result of general imaging of P on E (relative to the given transfer function T), and the mapping $\backslash\backslash_T$ that sends a probability measure P and a non-empty proposition E to $P(\cdot\backslash\backslash_T \mathrm{E})$ is the very update procedure of general imaging (relative to T). Clearly, given W, any such update function $\backslash\backslash_T$ is determined completely by its underlying transfer function T.

Since its introduction, general imaging has been applied and explored in different areas. In particular, Joyce (1999, chapters 5 and 6) suggests general imaging to express, in probabilistic terms, subjunctive or counter-to-the-facts supposition, just as conditionalization expresses supposition as a matter of fact. Accordingly, he uses general imaging as a means of making the tenets of Causal Decision Theory precise, much as conditionalization is employed in Evidential Decision Theory. More recently, Baratgin and Politzer (2011) defend the thesis that general imaging is a plausible description of actual human belief revision processes in dynamic environments, based on a series of empirical findings.

One of the crucial features of general imaging, if compared to conditionalization, is this:[10] what the fraction $T_E(w', w)$ ($\cdot$ 100 percent) that a world w' transfers to a world w is like depends only on w', w, and E; in particular, it does not depend on the probability measure P that is to be updated. Whatever one's degree-of-belief function P may be, it is solely a matter of the evidence (that is, E) and the world(s) in question (that is, w', w) what fraction of the probabilistic mass that P supplies to w' will be moved to w.

This contrasts with conditionalization: if conditionalization is presented in a similarly additive format, that is,

$$P(\{w\}/\mathrm{E}) = \textstyle\sum_{w' \text{ in } W} P(\{w'\}) \cdot T_{E,P}(w', w),$$

the underlying transition function will have to be defined so that $T_{E,P}(w', w) = P(\{w\}/E)$. In other words: T will also depend on P (while it will not actually depend on w' at all).[11] In the

where the corresponding "chances" are determined by $T_{E,w'}(A)$ relative to worlds w' and the evidence E. We will return to this point later in this section.

[10] This feature is emphasized, discussed, and ultimately criticized by Joyce (2010), pp.149f.

[11] The same point is discussed by Pearl (1994), p.205.

case of conditionalization, each degree-of-belief function will thus determine its "own" corresponding transfer function, unlike the case of general imaging in which one and the same probability-independent transition function is used for all probability measures on W whatsoever.

For instance: the transfer function T that determines uniquely the corresponding general imaging function $\backslash\backslash_T$ might be defined such that the set of worlds w for which $T_E(w', w)>0$ holds coincides with the set of worlds in E that are most similar or close to w', and $T_E(w', w)$ might be chosen to be uniform over these worlds, in which case T would be determined solely by something like *world-relative similarity orderings* (along the lines of the Stalnaker-Lewis semantics for counterfactuals). Or $T_E(w', w)$ might be identified with the conditional objective chance at w' of the proposition $\{w\}$ given the proposition E, in which case T would be fully determined by *worldly conditional chance measures*.[12] Or $T_E(w', w)$ might be defined as the *objective chance of ending up in w given that one acts in w' such that E becomes true* (if that proposal differs at all from the conditional chance ascription in the previous case);[13] and so on. In each of these cases, T would be given in a manner that is *independent of P*, and the corresponding instance $\backslash\backslash_T$ of general imaging would inherit this feature, unlike the case of conditionalization. If spelled out in terms of *learning*: while conditionalizing on the evidence corresponds to learning something new about the present state of the actual world (where each world w counts as a candidate of what the actual world might be like), general imaging $\backslash\backslash_T$ corresponds to learning that the previously present state of the world (w') has changed, or has been changed by someone or something, into a new one (w), to an extent that is measured by $T_E(w', w)$.

---

[12] Compare footnote 8. This choice corresponds to Skyrms' (1980a, 1980b, 1984) proposal of determining degrees of acceptability for counterfactuals in terms of their corresponding expected conditional chances.

[13] Pearl (1994), p.205, discusses imaging as a possible method of transforming a probability measure by means of actions.

While this point is not discussed very much in the philosophical literature on (generalized) imaging itself—as most of that literature is concerned with subjunctive *supposition* rather than with, as it were, subjunctive *learning*—theoretical computer scientists have been discussing this difference between *learning about properties of the present state of the world* vs. *learning of the world being changed into another one having certain properties* a lot:[14] in that area of research, learning is studied primarily in the context of qualitative or all-or-nothing belief, and consequently probabilities do not necessarily play a role; however, analogous points apply to the opposition between learning in terms of *belief revision*—in the well-known "AGM sense" of Alchourron et al. (1985), which may be viewed as the qualitative counterpart of conditionalization—and learning in terms of so-called *belief update*—as introduced by Katsuno and Mendelzon (1992), which may be viewed as a qualitative counterpart of (generalized) imaging. In belief *update*, a given belief state is modified along a similarity ordering of worlds much in the way in which general imaging modifies a probability measure along a transition function, and indeed Katsuno and Mendelzon (1992) discuss the analogy between belief update and imaging explicitly. And belief update has been proposed to be the right form of learning if, and only if, what is at stake is learning of *change*: "We make a fundamental distinction between two kinds of modifications to a knowledge base. The first one, *update*, consists of bringing the knowledge base up to date when the world described by it changes… The second type of modification, *revision*, is used when we are obtaining new information about a static world" (see Katsuno and Mendelzon 1992, Introduction).

If translated back into our discussion of general imaging, the corresponding suggestion would be: general imaging might be a subject's adequate response to evidence if, and only if, the evidence expresses a change of the actual world (say, from the actual world w' to another world w), where the extent $T_E(w', w)$ to which w can be expected to be the outcome of that

[14] Katsuno and Mendelzon (1992) is the standard reference, which triggered this debate.

kind of change only depends on w', w and E, while it does not depend on the agent's degree-of-belief function P. So the change in question must be one in which the actual world changes independently of the subject's beliefs, which means it must be some kind of "worldly" (and perhaps evidence-related) change. And since the subject does not necessarily know which world w' is the actual one, she will have distributed her degrees of beliefs initially over possible candidates w' for being the actual world, and she will start from these initial degrees of beliefs in $\{w'\}$ when transferring them to the worlds w into which w' is being changed. In other words: her new degree-of-belief in w will be given by $\sum_{w' \text{ in } W} P(\{w'\}) \cdot T_{E,P}(w', w)$, which is nothing but the definition of general imaging. For similar reasons, and as mentioned before, general imaging has been applied in accounts of causal decision theory in which the required probabilities of 'if action E, then state S' are determined by imagining on the proposition that action E is carried out. (Obviously, actions constitute the standard manner of bringing about change.)

The relevance of these differences between general imaging and conditionalization to our intended context of rational belief aggregation will follow from a formal result by Gärdenfors (1982) on general imaging, to which we will turn in the next section.

3. General Imaging Commutes with Linear Pooling

Here is the crucial result:[15]

[15] Unlike Gärdenfors, we formulate the theorem relative to a (fixed finite) set W of possible worlds. Our presentation of the theorem will also differ from Gärdenfors' insofar as we suppress his original reference to "belief systems". See Joyce (2010), p.149, and Pearl (1994), p.205, for reformulations of Gärdenfors' theorem that are similar to ours.

Theorem (Gärdenfors 1982): Let U be an arbitrary update function that takes a probability measure and a non-empty proposition as input and which maps them to another probability measure. The probability measure that results from applying U to P and E is denoted by: $P(\cdot\backslash_U E)$; accordingly, the probability of A after any such an update will be denoted by: $P(A \backslash_U E)$.

Then for every such U, the following two statements are equivalent:

(I) *U preserves mixing* (in Gärdenfors' terminology) or preserves convex combinations (in our terminology), that is:

For all natural numbers k, for all probability measures P, $P_1,\ldots, P_k$, if there are real numbers $\alpha_1,\ldots, \alpha_k$ be in the closed interval [0,1], such that $\alpha_1 + \ldots + \alpha_k = 1$, and where for all propositions A,

$$P(A) = \alpha_1 P_1(A) + \ldots + \alpha_k P_k(A),$$

then it also holds for propositions A that

$$P(A\backslash_U E) = \alpha_1 P_1(A\backslash_U E) + \ldots + \alpha_k P_k(A\backslash_U E).$$

(II) There is a transfer function T, such that U is determined by applying general imaging of probability measures on non-empty propositions relative to T in the sense explained before; that is: for all probability measures P and for all non-empty propositions E, it holds that for all worlds w,

$$P(\{w\}\backslash_U \mathrm{E}) = \sum_{w' \text{ in } W} P(\{w'\}) \cdot T_E(w', w),$$

and for all propositions A,

$$P(A\backslash_U E) = \sum_{w \text{ in } A} P(\{w\}\backslash_U \mathrm{E}).$$

Or more briefly: there is a T, such that $\backslash_U$ is identical to $\backslash\backslash$ as determined by T.

We will not state the proof of this theorem here, which can be found in the appendix of Gärdenfors (1982). But its underlying thought is simple enough: general imaging is itself defined in terms of a convex combination of a fixed family of probability measures (given by T—recall our footnote 8), which is also why general imaging commutes with any form of linear pooling in the intended manner. And if an update function commutes with any form of linear pooling in the intended manner, then one can show that it must be given by a convex combination of a fixed family of probability measures, which means that it must coincide with the instance of general imaging that corresponds to (the fixed T function given by) this family of probability measures.

While Gärdenfors' theorem was motivated by Lewis' triviality results on probabilities of conditionals, what it tells us in our present context of social epistemology is: *update by general imaging (with respect to fixed transfer function T) is the unique update mechanism that commutes with linear pooling with respect to arbitrary coefficients*. For, by the theorem above, the result of first updating given individual degree-of-belief functions P_i by general imaging and then linearly pooling them with respect to numbers $\alpha_1,\ldots, \alpha_k$, that is, determining for each proposition A the number

$$\alpha_1 \, P_1(A\backslash\backslash E) + \ldots + \alpha_k \, P_k(A\backslash\backslash E),$$

leads to the same result as first linearly pooling the given individual degree-of-belief functions P_i with respect to numbers $\alpha_1,\ldots, \alpha_k$, which gives $P(A) = \alpha_1 \, P_1(A) + \ldots + \alpha_k \, P_k(A)$, and then updating the so determined social probability measure P by general imaging:

$$P(A\backslash\backslash E) = [\alpha_1 P_1 + \alpha_2 P_2](A\backslash\backslash E).$$

That is, by Gärdenfors' theorem, it holds that

$$\alpha_1 P_1(A\backslash\backslash E) + \alpha_2 P_2(A\backslash\backslash E) =$$

$$[\alpha_1 P_1 + \alpha_2 P_2](A\backslash\backslash E).$$

As promised initially, we have thus determined a natural mechanism of probabilistic update other than conditionalization that is not affected by the dilemma concerning group update that had been explained in the introductory section.

In the next section we will interpret this finding, we will at least partially determine a class of situations for which it will be highly relevant, and we will illustrate it by means of a concrete toy example.

4. Interpreting the Result and a Toy Example

What are the consequences of the formal result from the last section for rational group learning? Will subjects $S_1,...,S_k$ ever rationally employ general imaging so that Gärdenfors' theorem becomes applicable to them?

In the following, we will give a partial answer to these questions by focusing just on the (II)->(I) direction of the theorem from the last section.[16] If our subjects find themselves in a situation in which update by general imaging is the appropriate response to evidence, and indeed—what is more—update by general imaging with respect to a *joint* or socially *shared* transfer function T is the right response (in the sense that the transfer function that is underlying \\ as denoted in the k expressions '$P_1(A\backslash\backslash E)$',…, '$P_k(A\backslash\backslash E)$' from the last section is always one and the same), then the (II)->(I) direction will guarantee that linear pooling of the individual beliefs will *not* be subject to the initial concern regarding update; which is clearly a

[16] The other direction (I)->(II) is interesting, too, but probably less salient: it expresses that if one and the same learning method is supposed to commute with linear pooling whatever the group of subjects and whatever the subjects' weights $\alpha_1,…, \alpha_k$ of social influence are like, then that learning method must coincide with an instance of general imaging that is determined by some fixed transfer function. So if a learning method is meant to be stable across groups and across changes in the social weights of the members of a group, and if linear pooling is the preferred method of aggregation, then general imaging is the only update scheme that is up for the job.

good feature. The remaining question is: what are such situations like in which general imaging with respect to a joint transfer function delivers the right group method of accommodating evidence?

So let us suppose that our k subjects do in fact learn that the state of the actual world has changed in a particular manner, so that updating by means of general imaging does seem to be the appropriate response. While being a first step into the right direction, if just taken by itself, this is still not quite what we are looking for: because, as explained before, we need to characterize a type of situation in which additionally our subjects' updating by general imaging ought to be given with respect to one and the same *joint* transfer function T. So when will that be plausibly the case?

Returning to our previous discussion of probability-independent transition functions, let us thus additionally assume T to express inferential commitments that reflect solely world-related knowledge (given the evidence) about, say, objective similarities between worlds, or conditional chances at worlds, or the chances of bringing about certain worlds (or states), or the like; then our group of subjects will apply imaging with respect to a *joint* such T, if, and only if, they rely on the *same* such information (whether implicitly or explicitly) concerning similarity or chances or the like when updating by general imaging. We might summarize what these types of information have in common in the following rough (and admittedly vague) terms: *all of them derive from relevant bits of causal (and perhaps statistical) information* that our subjects must share, where the reasons why they share them might be manifold: perhaps because they had all acquired the relevant bits of commonsense experience concerning causal and statistical relationships before (as would be sufficient for many everyday contexts); or because they had all been presented with the same set of causal data before (as would be the case, e.g., for jury members in a legal trial); or because before updating on the new piece of evidence they had already exchanged and discussed the relevant causal data on which their update would be based (as would be the case maybe in scientific panels of experts); and so on.

In approximate terms, therefore, we have: *if subjects* $S_1,\ldots, S_k$ *receive a piece of evidence E concerning some worldly change* (and hence ought to apply general imaging)*, if additionally they are in possession of the same causal and statistical background knowledge that is relevant for the facts in question, and if they also put the same pragmatic-epistemic emphasis on the same features of these facts when accommodating E*, then also the similarity orderings on which their resulting applications of general imaging might be based should be the same, the conditional chance measures that might feed into their applications of general imaging should be the same, and so forth. In which case, *they should also invoke one and the same transition mapping T in the course of updating by general imaging.*

Summing up: if a group of subjects finds itself in a situation that is characterized by

- The relevant evidence being about changes (of the state) of the actual world,

in combination with

- the subjects sharing the relevant inferential dispositions that derive from sharing the same causal-statistical background knowledge on which their updates will be be based,

then update by general imaging with respect to a shared transfer function *T* will seem to be the normatively right thing for them to do. And by the right-to-left direction of the theorem from the last section, general imaging with respect to such a joint transfer function will not be subject to the commutativity dilemma concerning group update with linear pooling. Therefore, in circumstances as described above, the commutativity dilemma loses its bite, and ‘group update’ has indeed a unique referent. The claim is not at all that linear averaging and general imaging would be anything like *the* “universally valid” methods of aggregation and update. They are certainly more ideally targeted to some problems of belief aggregation and updating than others (e.g., they are more plausible in cases where several subjects are learning how the world is changing, rather than learning new information about how the world once was). The claim is merely that there exist problems to which it is reasonable to apply linear averaging and general imaging, and that in such cases the commutativity dilemma will not arise.[17]

Let me illustrate this now in terms of a very simple toy example[18].

Say, we are dealing with two subjects S_1, S_2 who have different information about the contents of a basket with fruit. Both of them can rule out all possibilities except for these three worlds (distributions of fruit in the basket):

w_1, in which there are precisely one apple and one banana in the basket,

w_2, in which there is precisely one pear in the basket,

w_3, in which there is precisely one apple in the basket.

So all worlds other than these three have prior probability 0 according to both subjects. However, S_1, S_2 differ in terms of the degrees of belief that they assign to w_1, w_2, w_3:

S_1's degree-of-belief function is such that

$$P_1(\{w_1\}) = 1/6, P_1(\{w_2\}) = 2/3, P_1(\{w_3\}) = 1/6,$$

which means that S_1 is pretty sure that there is just one pear in the basket and nothing else, while S_2's degree-of-belief function satisfies

$$P_2(\{w_1\}) = 1/3, P_2(\{w_2\}) = 1/3, P_2(\{w_3\}) = 1/3$$

and hence S_2 is completely impartial concerning the three possibilities.

[17] I am grateful to an anonymous reviewer and to an editor of this journal for very helpful comments on this point.

[18] The example is similar to one discussed by Baratgin and Politzer (2011), though some of the formal details differ.

Next we can determine the group belief for our two subjects as recommended by linear pooling. For that purpose, we need to choose the coefficients or weights that correspond to the subjects: let us assume that the two of them are peers with equal influence on the intended social outcome; so $\alpha_1 = 1/2$, $\alpha_2 = 1/2$. With that in place, we can determine the social probability measure P for S_1 and S_2 viewed as a group, which is given by

for all propositions A: $P(A) = \alpha_1 P_1(A) + \alpha_2 P_2(A)$,

and in particular,

$P(\{w_1\}) = 1/12 + 1/6 = 1/12 + 2/12 = 3/12 = 1/4$,

$P(\{w_2\}) = 2/6 + 1/6 = 3/6 = 1/2$,

$P(\{w_3\}) = 1/12 + 1/6 = 1/12 + 2/12 = 3/12 = 1/4$.

The outcome reflects S_1's preference for w_2, although the preference is weakened from 2/3 to 1/2 in view of S_2's indifference.

Accordingly, for instance, the corresponding degrees of belief for there being an apple in the basket are: P_1(*apple in basket*) = 1/3, P_2(*apple in basket*) = 2/3, P(*apple in basket*) = 1/2. So linear pooling makes it more likely for S_1 but less likely for S_2 that there is an apple in the basket.

Now let us assume that some new piece of evidence comes along: *there is actually no banana in the basket*. How should the subjects update on this new available information *qua* group?

In line with our previous discussion, this depends on whether the evidence is meant to express some additional information about an unchanged world (*the banana is not in the basket, i.e., it has not been there*) or rather information about how the world has changed (*the banana is not in the basket, i.e., it was removed if it had been there at all*). And as explained before, we are going to analyze this in the way that two different kinds of update functions will be applied to one and the same piece of evidence, that is, to the proposition *the banana is* *not in the basket* (or the set $\{w_2, w_3\}$ of worlds): in the first case standard Bayesian update on this proposition will be the rational response, while in the other case general imaging on the proposition will be called for.[19]

Let us consider the Bayesian option first: for instance, the new piece of evidence might have been communicated to S_1 and S_2 by someone whom they know to be precise, epistemically trustworthy, and uninterested in eating or stealing fruit; and there is no other reason either for thinking that anyone was causally interfering with the situation. Hence conditionalization seems to be the appropriate update mechanism.

(i) Conditionalizing first and linearly pooling afterwards leads to: Each of our two subjects updates individually on the evidence by means of conditionalization, which yields

$P_1(\{w_1\}$ / *banana not in basket [i.e., it has not been there]*) = 0,

$P_1(\{w_2\}$ / *banana not in basket [i.e., it has not been there]*) = 4/5,

$P_1(\{w_3\}$ / *banana not in basket [i.e., it has not been there]*) = 1/5.

$P_2(\{w_1\}$ / *banana not in basket [i.e., it has not been there]*) = 0,

$P_2(\{w_2\}$ / *banana not in basket [i.e., it has not been there]*) = 1/2,

$P_2(\{w_3\}$ / *banana not in basket [i.e., it has not been there]*) = 1/2.

Linear pooling of these degree-of-belief functions after conditionalization gives then:

α_1 $P_1(\{w_1\}$ / *banana not in basket [i.e., it has not been there]*) + α_2 $P_2(\{w_1\}$ / *banana not in basket, i.e., it has not been there*) = 0.

[19] It is an interesting question whether one might instead analyze the propositional evidence differently in the two cases and then apply the *same* update mechanism to the two corresponding *distinct* propositions, rather than applying two *distinct* update mechanisms to one and the *same* proposition. We cannot explore this any further here, but there would be at least two downsides to any such alternative analysis: It might be that *semantically* indeed the same message is delivered to the two subjects in the two situations and that it is only the *pragmatic* connotations of the respective acts of delivery that make it clear to them how they ought to react to the message; in which case applying two distinct update functions to the same proposition would seem to be the more appropriate reconstruction. And: the underlying space of possibilities would need to be much more sophisticated if the differences between *the banana has not been there* and *the banana was removed if it had been there at all* ought to be captured in terms of propositional content.

α_1 P_1($\{w_2\}$ / *banana not in basket [i.e., it has not been there]*) + α_2 P_2($\{w_2\}$ / *banana not in basket, i.e., it has not been there*) = 4/10 + 1/4 = 8/20 + 5/20 = 13/20.

α_1 P_1($\{w_3\}$ / *banana not in basket [i.e., it has not been there]*) + α_2 P_2($\{w_3\}$ / *banana not in basket, i.e., it has not been there*) = 1/10 + 1/4 = 2/20 + 5/20 = 7/20.

In particular, as far as the proposition *an apple is in the basket* (the set $\{w_1, w_3\}$) is concerned, the outcome will be:

α_1 P_1(*apple in basket / banana not in basket [i.e., it has not been there]*) + α_2 P_2(*apple in basket / banana not in basket [i.e., it has not been there]*) = 7/20,

which is less than the prior P(*apple in basket*) = α_1 P_1(*apple in basket*) + α_2 P_2(*apple in basket*) = 1/2. So this first kind of Bayesian group update disconfirms the thesis that the apple is in the basket.

On the other hand, (ii) first applying linear pooling and then conditionalizing leads to: as the group probability measure P has been determined above already, it only remains to update P by conditionalizing it on the evidence, which yields

P(*apple in basket / banana not in basket [i.e., it has not been there]*) = 1/3,

which is also less than the prior P(*apple in basket*) = α_1 P_1(*apple in basket*) + α_2 P_2(*apple in basket*) = 1/2. However, 1/3 is distinct from the 7/20 that had been calculated in (i). As expected, conditionalization and linear pooling do not commute:

$$\alpha_1 P_1(\textit{apple in basket/banana not in basket}) + \alpha_2 P_2(\textit{apple in basket/banana not in basket}) \neq$$
$$[\alpha_1 P_1 + \alpha_2 P_2](\textit{apple in basket / banana not in basket}),$$

and it is not clear anymore whether 'Bayesian group update' ought to refer to (i) or (ii) (if to either of them at all).

Now let us turn to the other type of situation: for instance, the new piece of evidence might have been communicated to S_1 and S_2 by someone whom they know to be precise, epistemically trustworthy, but also highly interested in eating banana. They interpret this person's message therefore as conveying that the person would have removed the banana from the fruit basket had it been there initially. So they ought to apply general imaging instead of conditionalization (or so I am going to assume).

Based on their shared everyday knowledge of causal relationships—as in: taking the banana out of a fruit basket does not change the basket and its contents except for the banana being gone—the underlying transfer function T will be one and the same for both subjects: in particular, if 'E' denotes the proposition *the banana is not in the basket [i.e., it was removed if it had been there at all]*, then $T_E(w_1, w_1) = 0$, $T_E(w_1, w_2) = 0$, and $T_E(w_1, w_3) = 1$, since removing the banana from the basket will change w_1 into w_3, whilst $T_E(w_2, w_1) = 0$, $T_E(w_2, w_2) = 1$, $T_E(w_2, w_3) = 0$, $T_E(w_3, w_1) = 0$, $T_E(w_3, w_2) = 0$, $T_E(w_3, w_3) = 1$, as there had not been any banana in the basket at either of w_2 and w_3 in the first place. Because the values of T_E are crisp, it is clear that we will be dealing with a case of imaging *simpliciter* here instead of general imaging proper. For instance, in more complex circumstances, grabbing the banana from the basket might come just with a high chance of succeeding in its removal (say, $T_E(w_1, w_3) = 0.9$) so that failing to remove the banana from the basket could not be ruled out completely ($T_E(w_1, w_1) = 0.1$), in which case plain imaging would not be sufficient anymore. However, for the sake of simplicity, let us stick to the binary T_E values from before.

Let \\ (or, if we like to, \) be the update by (generalized) imaging that is given by T, where we will only be interested in the context in which the proposition *the banana is not in the basket* is the relevant piece of evidence.

(iii) First applying imaging and only then linear pooling leads to: Each of our two subjects updates individually on the evidence by means of imaging, which yields (by "moving" probabilities from w_1 to w_3)

$P_1(\{w_1\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) = 0,

$P_1(\{w_2\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) = 2/3,

$P_1(\{w_3\}$ \\ *banana not in basket [i.e., it is removed if there at all]*) = 1/3.

$P_2(\{w_1\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) = 0,

$P_2(\{w_2\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) = 1/3,

$P_2(\{w_3\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) = 2/3.

Linear pooling of these individual probability measures after imaging gives then:

α_1 $P_1(\{w_1\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) + α_2 $P_2(\{w_1\}$ \\ *banana not in basket, i.e., it is removed if there at all*) = 0.

α_1 $P_1(\{w_2\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) + α_2 $P_2(\{w_2\}$ \\ *banana not in basket, i.e., it is removed if there at all*) = 2/6 + 1/6 = 3/6 = 1/2.

α_1 $P_1(\{w_3\}$ \\ *banana not in basket [i.e., it was removed if there at all]*) + α_2 $P_2(\{w_3\}$ \\ *banana not in basket, i.e., it is removed if there at all*) = 1/6 + 2/6 = 3/6 = 1/2.

In particular, as far as the proposition *an apple is in the basket* (the set $\{w_1, w_3\}$) is concerned, the outcome is:

α_1 P_1(*apple in basket* \\ *banana not in basket [i.e., it was removed if there at all]*) + α_2 P_2(*apple in basket* \\ *banana not in basket [i.e., it was removed if there at all]*) = 1/2.

On the other hand, (iv) first applying linear pooling and then imaging leads to: as the group probability measure P has been determined already, it only remains to update P by imaging on the evidence, which yields

P(*apple in basket* \\ *banana not in basket [i.e., it was removed if there at all]*) = 1/2,

which coincides in value with the 1/2 that had been calculated in (iii) before. (Generalized) imaging and linear pooling commute:

$$\alpha_1 P_1(\textit{apple in basket} \setminus\!\!\setminus \textit{banana not in basket}) +$$

$$\alpha_2 P_2(\textit{apple in basket} \setminus\!\!\setminus \textit{banana not in basket})$$

$$=$$

$$[\alpha_1 P_1 + \alpha_2 P_2](\textit{apple in basket} \setminus\!\!\setminus \textit{banana not in basket}).$$

Furthermore, the value of both of these expressions coincides with the original group probability P(*apple in basket*) = 1/2, and it does so for good reasons: after all, the apple's lying in the basket is causally independent of removing the banana, which is why the probability of the apple being in the basket should remain unaffected by the corresponding instance of imaging, and that is indeed the case.

Accordingly, our subjects S_1 and S_2 do not face a commutativity dilemma concerning group update with linear pooling in this kind of situation in which they learn about change (in the present case, change caused by an action), and in which their updates will be based on a shared set of relevant pieces of causal-statistical background knowledge. Whether 'group update' refers to (iii) or (iv) above simply does not matter, as the outcomes of the two strategies of group update are guaranteed to coincide.

5. Summary and Consequences

We started with the familiar observation that conditionalization or standard Bayesian update does not commute with the linear pooling of subjective probability measures. This led to a dilemma: How should a group of subjects update? The response to that dilemma that we were exploring in this paper was to replace conditionalization by an alternative updating scheme: general imaging. We explained that method of update in some detail, we stated Gärdenfors' characterization result for it, and we derived from that result that general imaging was not subject to a similar kind of dilemma with respect to linear pooling as conditionalization was. Afterwards we argued that in a situation in which a group of subjects is meant to learn about a changing world and in which their update captures a common set of relevant pieces of causal-statistical background knowledge, general imaging might actually constitute the right method of group learning. Therefore, by the previous finding, no variant of the original commutativity dilemma concerning group update with linear pooling is going to affect the right method of group update in these situations.

Various extensions of these findings come to mind. Most importantly, it should be possible to extend both the original dilemma and our way out of it to the case of *group decision-making with linear pooling*: e.g., assume two subjects trying to determine as a group the expected utility of carrying out an action E. Let us assume that the intended social utility measure is the function u. Finally, for simplicity, let us assume that their focus is just on two possible outcomes O_1 and O_2 that the relevant possible courses of action might have. By first individually supposing the proposition that E is carried out and then linearly pooling the resulting probability measures (assuming that linear pooling is their preferred method of aggregation again), they might determine

$$[\alpha_1 P_1(O_1 \text{ if } E) + \alpha_2 P_2(O_1 \text{ if } E)]\, u(O_1 \,\&\, E) + [\alpha_1 P_1(O_2 \text{ if } E) + \alpha_2 P_2(O_2 \text{ if } E)]\, u(O_2 \,\&\, E).$$

Alternatively, our two subjects might first linearly pool their individual probability measures and only then suppose E to be carried out, which gives an expected utility of the form

$$[\alpha_1 P_1 + \alpha_2 P_2](O_1 \text{ if } E)\, u(O_1 \,\&\, E) + [\alpha_1 P_1 + \alpha_2 P_2](O_2 \text{ if } E)\, u(O_2 \,\&\, E),$$

or more briefly, where P is the linear pool of P_1 and P_2,

$$P(O_1 \text{ if } E)\, u(O_1 \,\&\, E) + P(O_2 \text{ if } E)\, u(O_2 \,\&\, E).$$

The ensuing questions should be familiar by now: Will these two procedures lead to the same outcome? If not, how should these subjects determine expected utilities as a group: in the first manner or in the second one? Now we are facing a potential commutativity dilemma concerning *group decision-making* and linear pooling. Unsurprisingly, it should be possible to answer these questions and to reply to the potential dilemma in ways that should be familiar by now, and the answers and the reply will be sensitive to the mode of supposition again (matter of fact vs subjunctive). In particular, if causal decision theory formulated by means of generalized imagining were to be valid (at least in a certain type of situations),[20] and hence expressions of the form '$P(O \text{ if } E)$' were to be analyzed by means of general imaging, then group decision-making with linear pooling could be shown in the very same manner not to be affected by the corresponding commutativity dilemma. There would be, extensionally, a unique method for a group to determine socially expected utilities of actions by means of averaging. But we leave such applications to social (causal) decision theory to another paper.[21]

References

Alchourron, Carlos E., Peter Gärdenfors, and David Makinson. 1985. "On the Logic of Theory Change. Partial Meet Contraction and Revision Functions." *Journal of Symbolic Logic*, 50: 510-530.

Baratgin and Politzer. 2011. "Updating: A Psychologically Basic Situation of Probability Revision." *Thinking and Reasoning*, 16: 253-287.

[20] Joyce (1999) gives an overview and defense of causal decision theory in general, and of causal decision theory stated in terms of general imaging in particular. Lewis (1981) argues that his own version of causal decision theory can be reformulated in terms of general imaging. Skyrms et al. (1999) present the essentials of a closely related causal decision theory that is formulated in terms of expected conditional chances.

[21] Acknowledgements: I am very grateful to Kurt Norlin, Martin Rechenauer, Jan-Willem Romeijn, Thomas Boyer-Kassem, Richard Pettigrew, an anonymous reviewer, and an editor of this journal for their comments on an earlier version of this paper.

Dalkey, Norman C. 1975. "Toward a Theory of Group Estimation." In H. Linstone and M. Turoff (eds), *The Delphi Method: Techniques and Applications*. Reading, MA: Addison-Wesley, 231-256.

Dietrich, Franz. 2010. "Bayesian Group Belief." *Social Choice and Welfare*, 35: 595-626.

Dietrich, Franz and Christian List. Forthcoming. "Probabilistic Opinion Pooling." In C. Hitchcock and A. Hajek (eds), *Oxford Handbook of Probability and Philosophy*. Oxford: Oxford University Press.

Gärdenfors, Peter. 1982. "Imaging and Conditionalization." *The Journal of Philosophy*, 79/12: 747-760.

Genest, Christian, Kevin J. McConway, and Mark J. Schervish. 1986. "Characterization of Externally Bayesian Pooling Operators." *The Annals of Statistics*, 14/2: 487-501.

Genest, Christian and James V. Zidek. 1986. "Combining Probability Distributions: A Critique and an Annotated Bibliography." *Statistical Science*, 1/1: 114-148.

Joyce, James M. 1999. *The Foundation of Causal Decision Theory*. Cambridge: Cambridge University Press.

Joyce, James M. 2010. "Causal Reasoning and Backtracking." *Philosophical Studies*, 147: 139-154.

Katzuno, Hirofumi and Alberto O. Mendelzon. 1992. "On the Difference Between Updating a Knowledge Base and Revising It." In P. Gärdenfors (ed), *Belief Revision*. Cambridge: Cambridge University Press, 183-203.

Lewis, David K. 1976. "Probabilities of Conditionals and Conditional Probabilities." *The Philosophical Review*, 85: 297-315.

Lewis, David K. 1981. "Causal Decision Theory." *Australasian Journal of Philosophy*, 59/1: 5-30.

Madansky, Albert. 1964. "Externally Bayesian Groups." Technical Report RM-4141-PR, RAND Corporation.

Pearl, Judea. 1994. “From Imaging and Stochastic Control to a Calculus of Actions.” *AAAI Technical Report*, SS-94-06: 204-209.

Skyrms, Brian. 1980a. “The Prior Propensity Account of Subjunctive Conditionals.” In W. Harper, R. Stalnaker, and G. Pearce (eds), *Ifs. Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: Reidel, 259-265.

Skyrms, Brian. 1980b. *Causal Necessity*. New Haven, CT: Yale University Press.

Skyrms, Brian. 1984. *Pragmatics and Empiricism*. New Haven, CT: Yale University Press.

Skyrms, Brian, Gary D. Bell, and Peter Woodruff. 1999. “Theories of Counter-Factual and Subjunctive Conditionals in Contexts of Strategic Interaction.” *Research in Economics*, 53: 275-291.

Steele, Katie. 2012. “Testimony as Evidence: More Problems for Linear Pooling.” *Journal of Philosophical Logic*, 41: 983-999.

Wagner, Carl. 1985. “On the Formal Properties of Averaging as a Method of Aggregation.” *Synthese*, *62*: 97-108.