Iterated Abduction and Conditional Coherence

Horacio Arló Costa
CMU

Abstract

The paper offers a complete probabilistic characterization of the abductive operator axiomatized by Maurice Pagnuco in [11]. The model assumes the axiomatic characterization of Conditional Coherence presented by Lester Dubins in section 3 of [8]. It follows, as a philosophical consequence, that a form of probabilism based on Dubins' conditional probability not only permits ampliative rules of inference, but that it also mandates them. This result is contrasted with various arguments provided by Bas van Fraassen in [19] showing that the adoption of any form of Inference to the Best Explanation as a rule for belief revision must make one's belief system incoherent.

In the second part of the paper a property of iterated probability change, called Cumulativity, is presented as well as the corresponding axiomatization for iterated abductive operators (in the style of [17] and [7]). Finally we consider non-cumulative methods of change constrained by the general principles of Conditional Coherence.

1 Conditional Probability and Conditional Coherence

Lester E. Dubins presents in [8] an axiomatization of finitely additive conditional probability. The axioms provide minimal conditions of conditional coherence, which will result useful throughout the article. So, I shall start by providing background on Dubins' axioms .[1]

A *probability measure* on a Boolean algebra $\mathcal{A}$ with unit Ω is a finitely additive, non-negative function, normalized so as to assume the value 1 on Ω.

If $\mathcal{H}$ is a sub-algebra of a Boolean algebra $\mathcal{A}$, then P is a *conditional probability* on $(\mathcal{A}, \mathcal{H})$, and $(\mathcal{A}, \mathcal{H}, P)$ is a *conditional probability space*, if P is a function whose domain is $\mathcal{A} \times \mathcal{H}^0$ ($\mathcal{H}^0$ is $\mathcal{H}$ without the null element of $\mathcal{H}$), satisfying:

(1) $P(.|A)$ is a probability measure on $\mathcal{A}$, for each $A \in \mathcal{H}^0$.

(2) $P(H|H) = 1$ for all $H \in \mathcal{H}^0$.

---

[1] The term 'conditional coherence' is first used by Schervish et al. in [14], page 213. Nothing will be assumed in the paper about conditional coherence aside from the constraints provided by Dubins' axioms. The eventual addition of countable additivity for restricted examples will be clarified in footnotes.

(3) $P(B \cap C|A) = P(B|A)P(C|B \cap A)$ for all A, B, C with $A \in \mathcal{A}$, $B \in \mathcal{H}^0$, $C \in \mathcal{H}^0$.

A conditional probability P on $(\mathcal{A}, \mathcal{H})$ is *full* (on $\mathcal{A}$) if $\mathcal{H} = \mathcal{A}$. Dubins shows that for every conditional probability space $(\mathcal{A}, \mathcal{H}, P)$, there is an extension Q of P which is a full conditional probability on $\mathcal{A}$. When $P(A) = 0$, there might not be a unique such extension, but each of the extensions provided by Dubins provides a meaningful manner of conditioning with events of measure zero.

Axiom (2) is a mandatory constraint on any reasonable notion of conditional probability. The axiom seems to be constitutive of the notion of supposition which conditional probability intends to encode. Axiom (1) might be seen as restrictive for anyone familiar with the Kolmogorovian notion of probability, where countable additivity plays a crucial role. Nevertheless, countable additivity is far from being an uncontroversial axiom of probability. On the other hand, I shall review below arguments suggesting that the price of adopting the Kolmogorovian view of conditional probability is to abandon the full generality of Axiom (2). Axiom (3) has a long pedigree, which can be traced back at least to [10], where Jeffreys gave it the name of 'W. E. Johnson's product rule'.[2]

There is, of course, a well-known tradition initiated by Kolmogorov capable of handling conditioning with events of measure zero. This well-known alternative to the former view operates as follows. Let $\langle \Omega, \mathcal{B}, P \rangle$ be a measure space where Ω is a set of points, $\mathcal{B}$ a σ-field of sets of subsets of Ω, with points w. Then when $P(A) > 0$, $A \in \mathcal{B}$, the conditional probability over $\mathcal{B}$ given $\mathcal{A}$ is defined by: $P(.|A) = P(. \cap A) \ / \ P(A)$. Of course, this does not provide guidance when $P(A) = 0$. For that the received view implements the following strategy. Let $\mathcal{A}$ be a sub-σ-field of $\mathcal{B}$. Then $P(.|\mathcal{A})$ is a *regular conditional distribution* [rcd] of $\mathcal{B}$, given $\mathcal{A}$ provided that:

(4) For each $w \in \Omega$, $P(.|\mathcal{A})(w)$ is a probability on $\mathcal{B}$.
(5) For each $B \in \mathcal{B}$, $P(B|\mathcal{A})(.)$ is an $\mathcal{A}$-measurable function.
(6) For each $A \in \mathcal{A}$, $P(A \cap B) = \int_A P(B|\mathcal{A})(w)dP(w)$

Kolmogorov shows that $P(.|\mathcal{A})$ is probability not given an event, but given a σ-field. Blackwell and Dubins discuss in [8] conditions of *propriety* for rdcs. An rcd $P(.|\mathcal{A})(w)$ on $\mathcal{B}$ given $\mathcal{A}$, is proper at w, if $P(.|\mathcal{A})(w) = 1$, whenever $w \in A \in \mathcal{A}$. $P(.|\mathcal{A})(w)$ is improper otherwise. Recent research has shown that when rcds exist and $\mathcal{A}$ is countably generated, almost surely the rcds are proper. Nevertheless when rcds exist but $\mathcal{A}$ is not countably generated, there are circumstances when the rcds are maximally improper [16]. This is so in two senses. On the one hand the set of points where propriety fails has measure 1 under P. On the other hand $P(a(w)|\mathcal{A})(w) = 0$, when propriety requires that $P(a(w)|\mathcal{A})(w) = 1 - a(w)$ denotes here the $\mathcal{A}$-atom containing the point w.

It seems that failures of propriety conspire against any reasonable epistemological understanding of probability of the type commonly used in various

---

[2]The epistemologist W. E. Johnson of Cambridge seemed to have proposed it for the same time in a seminar in epistemology attended by both Jeffreys and Keynes.

branches of mathematical economics, philosophy and computer science. To be sure finitely additive probability obeying Conditional Coherence is not free from foundational problems,[3] but, by clause 2 of Conditional Coherence, each coherent finitely additive probability is proper.

So, in this article I shall adopt Conditional Coherence, which captures some aspects of De Finetti's idea of conditional probability *given an event*, rather than *given a σ-field*.[4] I shall add the axiom of Countable Additivity only to restricted applications where the domain Ω, when infinite, is at most countable. [5] My first goal will be to show that a form of probabilism based on Conditional Coherence is tightly connected to recent axiomatizations of ampliative rules of inference. In particular I shall offer a complete probabilistic characterization of the abductive operator axiomatized in [11]. This will require defining qualitative belief from conditional probability by appealing to a procedure studied in [20], [3], [1].

2 Conditional Probability and Belief

I shall introduce the notion of *probability core*. I follow here ideas presented in [1], which, in turn, slightly modify the schema first proposed in [20]. A notion that plays an important role in these works is the notion of *normality*. Here this notion only has a marginal value. The basic idea is that an event A is normal for P as long as $P(\emptyset|A) = 0$. Conditioning on abnormal events would lead to incoherence. In this paper the underlying Boolean algebra will always be tailored in such a way that abnormal events will never be available for conditioning.

A probability *core* for P is an event $[K]$ which is normal and satisfies the *strong superiority condition* (SSC) i.e. if A is a nonempty subset of $[K]$ and B is disjoint from $[K]$, then $P(B|A \cup B) = 0$ (and so $P(A|A \cup B) = 1$). Thus any non-empty subset of $[K]$ is more "believable" than any set disjoint from K.

Two important facts can now be established. First one can show that the family of cores induced by a two-place function P is nested [20]. Second, it is possible to show (for countable spaces) that every countable additive function P is such that the chain of cores induced by it cannot contain an infinitely descending chain of cores. A proof of this fact is given in [2]. A similar result was proven in a different context in [18].

---

[3]Perhaps the most poignant result is presented in [14], where it is shown that each finitely additive probability fails the property that De Finetti called *conglomerability* in some denumerable partition. See also [8].

[4]Conditional Coherence does not capture, nevertheless, important aspects of De Finetti's ideas about primitive conditional probabilities. Unlike probabilists, De Finetti uses a primitively given notion of information in order to define conditional probability. Such notion of information set does not admit, in his construction, a probabilistic account. Some repercussions of this aspect of De Finetti's notion will be discussed below.

[5]The addition of Countable Additivity will be specifically clarified. Otherwise it will be assumed that we are working with finitely countable measures.

2.1 Probability Cores, Full Belief and Expectation

When the universe of points is at most countable, very nice properties hold. In general it can be shown that (when countable additivity holds) for each function P there is a smallest as well as a largest core and that the smallest core has measure 1. In addition, when the universe is countable we can add Countable Additivity without risking failures of propriety. In this case we have that the smallest core is constituted exactly by the points carrying positive probability [1].

All cores carry probability one, but, of course, only the innermost core lacks subsets of zero measure. There is, in addition, a striking difference between the largest and the smallest core (and between the largest and any other core). In fact, any set S containing the largest core is robust with respect to suppositions in the sense that, $P(S \mid X) = 1$ for all X, and the complement of S is abnormal. So, the largest core encodes a strong doxastic notion of certainty or *full belief*, while the smallest encodes a weaker notion of 'almost certainty'. Both formal and conceptual reasons are provided in [1] for understanding the weaker notion as a notion of qualitative *expectation*.[6] So, when the universe is countable and countable additivity is imposed we can define two main attitudes as follows: *an event A is expected if it contains the smallest core, whereas it is fully believed if it contains the largest.* I should point out here in passing that Bas van Fraassen prefers to identify the smallest core for P with the full beliefs for P. We will comment on this option below.

In the general case there is still enough structure to define both attitudes. In fact, in this case the existence of the innermost core cannot be guaranteed. But the definition of full belief needs no modification and the notion of expectation can be characterized as follows: *an event A is expected if it is entailed by some core.*

In the following sections I shall show how the notions just presented can be used in order to provide a model of operators implementing ampliative rules of inference. The following section is devoted to an axiomatic characterization of such operators.

3 Ampliative inference and theory change operators

Let a theory K encode the epistemic commitments for full belief of an agent at some time t.[7] I shall call such theory a *belief set*. Then, in the presence of an epistemic input α consistent with K, the resulting (abductively) expanded belief set will be a superset of K containing α. Intuitively one is looking for an

---

[6]The term 'expectation' is not decision theoretically motivated. Its motivation comes from the field of non-monotonic logic, where 'expectation' models of defeasible reasoning are usual – see [3] for details.

[7]In the sense that the proposition [K] corresponding to K is such that an event A is a full belief when $[K] \subseteq A$.

(eventual) extension of Cn(K $\cup$ $\{\alpha\}$), which can explain the incoming data.[8] In order to analyze these expansions it is useful to focus on the partition of maximally consistent supersets of K implying α. Let K$\perp$ α denote the elements of this partition.

Following a well-known strategy often used in the so-called theory of theory change (see [9]), one can define an ampliative operator $\oplus$ as follows. Let γ be a selection function on K$\perp$ α, picking the best elements of K$\perp$ α. Define then K$\oplus\alpha$ as $\cap\gamma($ K$\perp$ $\alpha)$, when K$\perp$ α is nonempty and as Cn($\perp$) = $K_\perp$, otherwise. Any such operator will be called a *partial meet* abductive expansion operator.

This characterization avoids inferring too little or too much. If one just takes $\gamma($ K$\perp$ $\alpha)$, the resulting ampliative inference 'jumps' to a maximal and consistent state, and, on the other hand, if one takes $\cap($ K$\perp$ $\alpha)$ this just collapses into Cn(K $\cup$ $\{\alpha$ $\}$),which is a non-ampliative outcome, given the background theory and incoming data.

Pagnuco [11] has axiomatized an operator of abductive expansion, which is characterizable via the previous constructions in terms of selections on K$\perp$ α. Here are the postulates he proposed.

(1)For any sentence α, and belief set K, K$\oplus\alpha$ is a closed belief set.

(2) $\alpha \in$ K$\oplus\alpha$ (success)

(3) K $\subseteq$ K$\oplus\alpha$ (inclusion)

(4) If $\neg$ $\alpha \in$ K, then K$\oplus\alpha = K_\perp$ (inconsistency)

(5) If $\neg$ $\alpha \notin$ K, then K$\oplus\alpha \neq K_\perp$ (consistency)

(6) If $\alpha \leftrightarrow \beta \in$ Cn(K), then K$\oplus\alpha =$ K$\oplus\beta$ (irrelevance of syntax)

(7) K$\oplus\alpha \subseteq$ Cn(K$\oplus(\alpha\vee\beta) \cup \{\alpha$ $\}$)

(8) If $\neg\alpha \notin$ K$\oplus(\alpha\vee\beta)$, then K$\oplus(\alpha\vee\beta) \subseteq$ K$\oplus\alpha$

The first two postulates are self-explanatory. The third postulate indicates that the operation is ampliative, it cannot lead to the retraction of beliefs. The fourth postulate indicates, again, that the operation $\oplus$ is an expansion, not a *revision* of the current data. The fifth postulate encodes the idea that the explanatory process of the data that $\oplus$ intends to represent does not introduce inconsistencies. In other words, whenever it is possible to find an explanation of the data, it is supposed to be consistent. Of course, this is not so for consistent data in general, but only for data consistent with the current view. The first five postulates provide a natural notion of abduction as follows. Pagnuco proves that the function $\oplus$ satisfies the first five postulates if and only if K$\oplus\alpha$ is identical to Cn(K $\cup$ $\{\beta\}$), for some β such that (i) α is deducible from K $\cup$ $\{\beta\}$ and (ii) K $\cup$ $\{\beta\}$ is consistent. On the other hand K$\oplus\alpha$ is set to K if such a β does not exist. So, the idea behind the operator $\oplus$ is as follows: the agent attempts to find an abduction of the epistemic input with respect to the current epistemic state, and adds this abduction to the current epistemic state, taking the deductive closure. If an abduction cannot be found - because the information contradicts the current belief - the epistemic state is not changed.

---

[8]Cn is the standard (Tarskian) notion of logical consequence.

177

The sixth postulate is a strong condition for the irrelevance of syntax, common in theories of belief revision. Concerning the last two postulates, they can be presented in a more familiar form for people acquainted to the theory of theory change. In fact, given well-known results in belief revision theory (see Appendix A of [9]) one can show that the following two postulates are equivalent to (7) and (8) in the presence of the other basic postulates:

(9) $K \oplus (\alpha \wedge \beta) \subseteq Cn((K \oplus \alpha) \cup \{\beta\})$
(10) If $\neg \alpha \notin K \oplus \beta$, then $Cn((K \oplus \beta) \cup \{\alpha\}) \subseteq K \oplus (\alpha \wedge \beta)$.

Let a selection function γ be called relational if and only if: $\gamma(\ K \bot\ \alpha)$ = $\{K' \in K \bot\ \alpha: K'' \leq K' \text{ for all } K'' \in K \bot\ \alpha\ \}$. When $\leq$ is transitive, the corresponding relation is a *transitively relational* selection function. Pagnuco [11] shows a full characterization of transitively relational partial meet expansion operators in terms of axioms (1)- (8). There are, of course, other types of characterization results capable of fully representing this family of operators. In the coming section we will show a probabilistic characterization in terms of primitive conditional probability.

4 A Probabilistic reconstruction of ampliative rules

Given a theory K, let [K] denote the proposition expressed by K. The letters A, B, etc denote the propositions expressed by sentences α, β, etc. Let $F(P)$ be the outermost core of P. If [K] is the proposition expressed by the background theory K, let $\mathcal{A}_{[K]}$ be the quotient algebra modulo [K]. In addition, let $Q_{[A]}(.|.)$ = $Q(.|. \cap A)$, defined on $\mathcal{A}_{[K] \cap A}$, for all A compatible with [K].

Theorem 4.1 *Let $Q(.|.)$ be a full conditional measure with outermost core [K] (characterized relative to a countable $(\mathcal{A}_{[K]}, \mathcal{H}, P))$. For all consistent sentences α such that $[\alpha] \in \mathcal{A}_{[K]}$, define $[K \oplus \alpha] \subseteq [\beta]$ if and only if B is entailed by some core for $Q[A]$. Then the corresponding operation satisfies the postulates (1)-(8).*

In addition we can completely characterize $\oplus$ in probabilistic terms. Some precisions about models of $\oplus$ are needed first. For every $\oplus$ satisfying (1)-(8) and for every belief set K, it is possible to prove that there is a nested collection C_K, of subsets of Ω satisfying;

(1.) C_K is totally ordered by $\subseteq$
(2.) [K] is in C_K and it is the $\subseteq$-maximal element of C_K.
(3.) C_K is closed under unions and non-empty intersections.

This internal grading on [K] is such that for every $[\alpha] \cap [K] \neq \emptyset$, $[K \oplus \alpha] \subseteq [\beta]$ if and only if $[\beta]$ is entailed by the intersection of $[\alpha]$ with some element of C_K. Following Pagnuco's notation we will call C_K an *internal system of spheres centered* on [K]. I do not include this part of the proof here due to space

178

limitations. The first concept we need is the concept of *rank system* for a regular system of spheres C_K induced by $\oplus$. Construct an enumeration of spheres, such that the outermost sphere $[K] = K_0$. For any n: 0, 1, ...; define a rank system r_n for C_K as follows:

(ranks) $r_n = \{w \in \Omega: w \in K_n - K_{n+1}, \text{ where } K_i \in C_K \}$

Theorem 4.2 *Let $\oplus$ be any function satisfying (1)-(8) for the belief set K and let A be a non-empty proposition. There is a countable probability space ($\mathcal{A}_{[K]}$, $\mathcal{H}$, P) which can be extended to a full conditional probability $Q(.|.)$ on $\mathcal{A}_{[K]}$, such that $[K \oplus \alpha] \subseteq [\beta]$ if and only if B is entailed by some core for $Q[A]$.*

Sketch of the proof: The proof proceeds along similar lines than the proof of Theorem 3 in [8]. I shall focus first on the case when $\mathcal{A}_{[K]}$ is finite.[9] In this case we can define a sequence of nonzero elements of $\mathcal{H}$, $\{r_i: 1,...,n\}$ partitioning the space - by appealing to the ranks defined above. Introduce now a sequence $\{p_i: 1,...,n\}$ of probability measures such that $p_n(u) > 0$ for every point u in the universe Ω, such that $u \in r_n$ and $p_n(u) = 0$, for every other point in Ω. We can then work with a P such that $P(r_1) = p_1(r_1) = 1$, and $P(r_2|r_1^c) = p_2(r_2) = 1$; and in general $P(r_{i+1}|(r_1 \cup ... \cup r_i)^c) = p_{i+1}(r_{i+1}) = 1$.

Define now $Q(A|B)$ as $p_i(A \cap B) / p_i(B)$, where i is the largest integer such that $B \cap r_i \neq \emptyset$. It is not difficult to check that $Q(A|H) = P(A|H)$ for nonzero H in $\mathcal{H}$, and that the axioms (1), (2) and (3) are obeyed. It can also be checked that the cores for Q are exactly the system of spheres C_K used in order to define the ranks r_i.

Assume now $[K \oplus \alpha] \subseteq [\beta]$, and $[\alpha] \cap [K] \neq \emptyset$. So, there is a core C for Q such that $C \cap A$ entails B. Therefore since $C \cap A$ is a core of $Q[A]$ – the proof of this appears in [5] – we have that there is a core for $Q[A]$ entailing B. On the other hand, if there is a core for $Q[A]$ entailing B, we need to show that $[K \oplus \alpha] \subseteq [\beta]$. This is also immediate now, given that - again, via the argument in [5] - any core C' for $Q[A]$ can be represented as the intersection $C' = C \cap A$, where C is a core for Q.

Finally we can appeal to (b) of Lemma 8 in [8] in order to get a proof for general $\mathcal{A}_{[K]}$. The idea is to start with an (countably) infinite sequence of ranks providing the nonzero elements of $\mathcal{H}$, and define P along the lines exhibited above. So we have a P defined on a subset $\mathcal{D}$ of $\mathcal{A}_{[K]} \times \mathcal{A}_{[K]}^0$, i.e. on $\mathcal{D} = \mathcal{A}_{[K]} \times \mathcal{H}^0$. Let $\mathcal{F}$ be a finite sub-algebra of the general $\mathcal{A}_{[K]}$. $\mathcal{H}_1 = \mathcal{F} \cap \mathcal{H}$ is a sub-algebra of $\mathcal{F}$ and P restricted to $\mathcal{F} \times \mathcal{H}_1^0$ is a conditional probability. Since $\mathcal{F}$ is finite there exist a full conditional probability which agrees with P on $\mathcal{F} \times \mathcal{H}^0$. Now, by the aforementioned lemma, the existence of this full measure agreeing with P on the intersection of $\mathcal{D}$ and $\mathcal{F} \times \mathcal{F}^0$ is equivalent to the existence of a full conditional probability on the general $\mathcal{A}_{[K]}$, with agrees with P on $\mathcal{D}$. The rest of the proof proceeds as in the finite case.●

---

[9] A very similar method of proof can be used for countable additive conditional probability defined over a countable space $\mathcal{A}_{[K]}$,

179

The previous method of proof does not give a direct idea of how to proceed when the core system contains infinitely descending chains of cores, which are permitted by the previous definitions. When $[K \oplus \alpha]$ has finitely many elements, we can provide a more direct definition as follows. Assume the space is countable and start with an infinite partition of the space given by the ranks $\{r_i: 1,...,n, ...\}$. Given the assumptions, each rank contains finitely many elements. Then, if there is a smallest integer i such that $A \cap C_i \neq \emptyset$, define $Q(B|A)$ as before: $p_i(A \cap B) / p_i(A)$. Otherwise set $Q(B|A)$ to 1 if there is r_i such that $A \cap C_i \neq \emptyset$, and $A \cap C_i \subseteq B$ and $Q(B^c|A)$ to 0 for Bs satisfying the given conditions. For the remaining infinite sets such that both B and B^c are infinite, arbitrarily set one of them to 1 and the complement to 0. Finally set $Q(B|A)$ to 0 for every other event in the space.

So, as an example, let the space be the (power set of the) positive integers and consider $E = \{2n: n = 1, ...\}$ be the even integers in the space, and let $O = \{2n-1: n = 1, ...\}$ be the odd integers in the space. Let $P(i) = \{1/2^n : if\, i = 2n\}$ and let $P(i) = 0$, otherwise. So the unconditional P is countably additive, whose support is the even integers E. But $P(.|O)$ might be uniform on the odd integers. This can be reflected by the fact that there is a core system over O defined as follows: Let the outermost core $C_1 = O$. Then $C_2 = \{1\}^c$, $C_3 = \{1,3\}^c$, etc. Each rank contains exactly one odd number with $r_1 = \{1\}$. According to the definition proposed in the previous paragraph each co-finite set in O has measure 1 (because it is entailed by at least a core) and each number in O carries zero probability. Of two infinite but not co-finite sets, say $S = \{1, 5, 9,...\}$ and $H = \{3, 7, 11, ...\}$, we can assign 1 to the set of lower rank (so S carries measure 1 and H zero). Notice that there are events that carry measure one even when they are not 'believed', in the sense that they are not entailed by any core (for example, S). This is a concrete case where we have an infinitely descending chain of cores and finitely additive probability defined for those cores. Different core systems for O determine different conditional extensions $P(.|O)$. The doxastic structure provided by the given set of cores has in this case some salient features. For example, for each finite sequence of odd numbers the largest is expected and the complement of each particular number is expected, which can be seen as entailing that each number is expected to be a 'loser' in this transfinite lottery. Nevertheless, it is not the case that all odd numbers are expected to be losers. The result of defining expectations in terms of what is entailed by some core determines a logical closure for expectations which does not satisfy the rule of Adjunction (or AND rule).

The following (more concrete) example adapts one example proposed by Pagnuco in [11]. Let O stand for 'the restaurant is open', L for 'lights on in restaurant', and C for 'the restaurant is being cleaned'. Let w_1 be the point $(\neg O, \neg C, \neg L)$, $w_2 = (O, C, L)$, $w_3 = (O, \neg C, L)$, $w_4 = (\neg O, C, L)$, $w_5 = (\neg O, \neg C, L)$, $w_6 = (O, \neg C, \neg L)$, $w_7 = (O, C, \neg L)$, $w_8 = (\neg O, C, \neg L)$.

An agent can be represented by a P-function which concentrates all the probability mass in w_1. Nevertheless P determines a system of cores C_P with outermost core 3, according to the following picture. So, the agent fully believes

only the logical consequences of $\{O \rightarrow L, C \rightarrow L\} = F(C_P)$. But he can still make distinctions among points receiving zero measure. For example, $P(\{w_4\}|\{w_3\} \cup \{w_4\}) = 0$. So, if the light is on the agent will expect that the restaurant is open, even when he is 'almost sure' that the restaurant is now closed, it is not being cleaned and the lights are off.

C_P	Possible worlds
0	w1
1	w2, w3
2	w4
3	w5
C_K	w6, w7

Notice that a standard 'expansion' of $F(C_P)$ does not provide much information. In fact $Cn(\{O \rightarrow L, C \rightarrow L\} \cup \{L\}) = Cn(\{L\})$.[10] That is to say the agent only knows that the light is on. The internal grading of full beliefs permits to 'jump' to the conclusion that the restaurant is open when it is supposed or learned that the lights are on. Notice also that the definitions we are using separate the notions of full belief and probability one. There are events of measure one, of which the agent is not completely certain. One can see the background certainties as law-like generalizations that guide the agent's reasoning process. They are not open to revision in the probabilistic representation, which provides no guidance for (full) belief revision. Nevertheless the agent is capable of 'jumping' to an abductive conclusion, given data compatible with the background certainties. And this capacity is encoded probabilistically by the comparisons the agent makes among events carrying zero measure. Of course, all fully believed items carry measure one. But, separating certainty (full belief) from probability one events helps with paradoxes like the so-called paradox of the lottery. It is clear form the construction method used to prove theorem 2 that postulate (5) only holds when countable additivity holds. Intuitively the abandonment of CA permits the existence of infinitely descending chains of cores, even for propositions logically compatible with the largest core. Together with (5) various important consequences of the postulates also fail, like the following condition, which follows from (1)-(8) (specifically from 10 and 5): (11) If $\beta \in K \oplus \alpha$ and $\gamma \in K \oplus \alpha$, then $\gamma \in K \oplus \alpha \wedge \beta$.

4.1 Are abduction rules 'bona fide' rules of inference?

Bas van Fraassen intends to show in [19] that: 'adopting any form of Inference to the Best Explanation (IBE) as a rule for belief revision must eventually make one's belief system incoherent' (see [19], ch. 7, section 4). The central idea of the argument is the adoption of a variant of David Lewis's 'dynamic' Dutch Book, showing that upon learning A at $t+j$ ($j > 0$) one leaps to incoherence if one does not assigns credence r to any proposition B such that at t: $P(B|A) = r$. This is tantamount to adopt what Isaac Levi calls *temporal credal conditionalization*.

---

[10]Notice also that as long as a tiny probability is spread among the points inside the cores, aside from w_1, we have an equally uninformative response after conditioning with the event that the lights are on.

I do not intend to discuss here the issue of whether such dynamic books are or not legitimate extensions of their synchronic counterparts – see [13] or [6] for arguments against the force of dynamic Books. I would like to focus here instead on using the previous theorems in order to construct the following argument. Even if temporal credal conditionalization were assumed for a form of probabilism based on positing conditional probability as the main epistemological primitive, this would entail commitment to the use of ampliative rules of inference as rules for belief revision.

I would like to stress that van Fraassen does allow for 'practices of ampliation beyond evidence' as long as they are relegated to the realm of pragmatics, or what in philosophy of science is called 'the context of discovery' rather than the 'context of justification'. It is the implementation of an abductive rule as a rule for belief revision, what van Fraassen finds incorrect. But the rules formalized by $\oplus$ are exactly an attempt to formalize ampliative rules as rules for belief change.

In the case of countable domains it is clear that the adoption of temporal credal conditionalization puts a constraint on updates. Notice that in the countable case any function, P having [K] as its largest core is such that, according to van Fraassen's characterization of full belief, the agent's full beliefs after the update are given exactly by $K \oplus A$. So, as long as temporal credal conditionalization holds, there is a direct commitment to a rule of abduction as a rule for belief revision.

Our own understanding of the role of cores is more liberal. The new full beliefs after the update with A are given by the intersection of A with the largest core for P. It is the smallest core (if it exists) the one that encodes the abductive jump. In our interpretation the content of the abduction is expected, not fully believed.

Notice in addition that (in the finite case) if P is a probability function with largest core [K], [$K \oplus T$] encodes the function's smallest core, which, in this case, is constituted by the points of positive probability. So, updates with propositions compatible with this smallest core should yield as output exactly the intersection of the input with the said core. This is indeed a consequence of Pagnuco's axioms:

(12) If $\neg \alpha \notin K \oplus T$, then $Cn((K \oplus T) \cup \{\alpha\}) = K \oplus \alpha$.

So, data A compatible with the smallest core is considered, one should take the intersection of this core and the data. Taking any strict subset of this intersection leads to incoherence in the traditional Bayesian sense. But when one shifts perspective by adopting as primitive a function $P(.|.)$, and one considers data compatible with the largest core of P (but incompatible with the smallest core for P), the cores of P provides guidance towards implementing an 'ampliative extension' of the held certainties. This is so 'ex ante', without even considering the problem of temporal shifts. The type of probabilism defended by van Fraassen in [20] (based on assuming conditional probability as the only epistemological primitive) seems to be capable of accommodating ampliative

182

rules of inference - unlike the garden variety of probabilism defended in [19].

5 Cumulativity

This section and the following section will be devoted to analyze iteration. First we will focus on the consequences of the following principle (K below is background knowledge for P):

(Cumulativity) Let $P_K[A](.|.) = Q(.|.)$ be a conditional probability function extending P. Then the update of Q with B, written $Q[B](.|.)$ is given by $P_K[A \cap B](.|.)$.

In order to see which is the scope of this principle, it is useful to notice that even if we do not impose it for every $Q(.) = P(.|A)$ there is always a Dubins' extension P' of P representing $Q(.|B)$, in the sense that $Q(.|B) = P'(.|A \cap B)$. Cumulativity is instead a condition on iteration which requires that $Q[B](.)$ should be represented by updating the original measure P with $A \cap B$.

Given every function P and associated core C, we can consider a function Ex yielding the innermost core of C (when it exists) and a function F yielding C's outermost core. These functions apply to the core system C (this appeals to an approach first proposed by W. Spohn in [17] and then extended axiomatically in [7]). By the same token we can consider an operation $\oplus$ on core systems, such that given a core system and a proposition compatible with $F(C)$, yields a new core system. In [5] it is shown that, as long as Cumulativity is adopted, the result of this operation is the system of cores obtained from the first by taking the intersection of each initial core with the incoming proposition [α]. The following axioms are then motivated (and provable from the properties of countable P – [3], [4] elaborate on some applications):

Entailment: $Ex(C) \subseteq F(C)$
Full belief expansion: $F(P) \cap A = F(C \oplus A)$
Success: $Ex(C \oplus A) \subseteq A$
Preservation: If $Ex(C) \cap A \neq \emptyset$, then $Ex(C) \cap A = Ex(C \oplus A)$
Restricted consistency preservation: If $A \cap F(P) \neq \emptyset$, then $Ex(C \oplus A) \neq \emptyset$.
Entertainability: If $F(P) \cap A = \emptyset$, then $Ex(C \oplus A) = F(C \oplus A) = \emptyset$
Cumulativity: $Ex((C \oplus A) \oplus B) = Ex(C \oplus (A \cap B))$

There is no easy translation between the principles of change stated in the former way and the principles stated in the manner used by Pagnuco. One can mention as a heuristics that $F(C)$ in the new setting corresponds intuitively to [K] in the previous setting; and that $Ex(C \oplus A)$ is the equivalent of [$K \oplus \alpha$]. So, Success is in this new notation the counterpart of Pagnuco's postulate 2. By the same token, the counterpart of postulate 3: $Ex(C \oplus A) \subseteq F(C)$ is derivable from the given postulates. And the counterpart of condition 11 follows easily from Cumulativity and Preservation. Of course, Restricted consistency preservation fails when CA does not apply. Now, it is quite evident that Cumulativity is

a rather strong postulate for belief dynamics. Data incorporated in a data sequence is never forgotten in subsequent stages of learning or supposing. The following section is devoted to present a view of conditional probability, which manages to circumvent Cumulativity.

6   Conditional Coherence without Cumulativity

In this final section I shall sketch how to proceed in order to maintain the constraints imposed by Dubins' axioms without assuming Cumulativity. The gist of the idea is to define extended conditional probability as coherent degree of belief on qualitatively given suppositions. So, one can start with a qualitative notion of supposing. Given background knowledge K and a proposition A, let K*A denote the suppositional scenario, relative to K, where A is assumed as true. Assume as well that for each K and K*A, we have probability measures corresponding to them and attributing probability only to events compatible with background knowledge. So P_K and P_{K*A} are the measures corresponding to K and K*A. We also assume that if K entails A, $P[K](A) = 1$ and that the function '*' obeys the axioms of the so-called AGM notion of belief revision (for an overview see [9]).

Now one can define $P_K[A](.|.) = P_{K*A}(.|.)$. Notice that while $P_K(.|Y)$ is only defined for events Y compatible with the background knowledge K, $P_K[A](.|Y)$ is defined for events Y compatible with K*A. This strategy permits defining conditioning on events of measure zero and fails to obey Cumulativity as long as the underlying notion of supposing fails to obey (K*A)*B = K*(A ∩ B) – as in AGM. Nevertheless we are supposing that Dubins' axioms are obeyed. The proposed definition permits focusing on a particular extension $P_{K*A}(.|.)$ of P_K. The following example can be used to illustrate that other possible extensions of P_K can be obtained by appealing to a different function *', as long as *' obeys the AGM axioms.

Consider a case where the space is finite, constituted by 3 atoms (a, b, c) and 8 points. Let (a, b, c) = w1; (a, -b, -c) = w2, and let K = w1, w2. The initial P_K is such that $P_K(w1) = .3$; $P_K(w2) = .7$. Now let (a, b, -c) = w3, (-a, b, -c) = w4; and (a, -b, c) = w'; (-a, -b, c) = w. In addition let A = {w3, w4, w', w} and B = {w, w'}. Consider now the following three revisions which are compatible with the AGM axioms: K*A = {w3, w4} and (K*A)*B = {w}, and well as K*(A ∩ B) = {w'}.

Now $P_K[A](.|.)$ is given by $P[K * A](.|.)$. This function is determined as follows: $P[K * A](\{w3\}|\Omega) = .5$ and $P[K * A](\{w4\}|\Omega) = .5$. Now the function $Q[B](.|.) = P_K[A, B](.|.) = P[(K * A) * B](.|.)$ is determined by the underlying revision in the sense that $P[(K * A) * B](\{w\}|\Omega) = 1$. And finally $P_K[A \cap B](.|.)$ is such that $P_K[A \cap B](\{w'\}|\Omega) = 1$. So, Cumulativity is violated in the sense that $P_K[A \cap B](.|.) \neq Q[B](.|.)$.

Notice, nevertheless, that Conditional Coherence guarantees the existence of some extension P' of P_K such that $Q[B](.|.) = P'_K[A \cap B](.|.)$, even when the latter measure is not obtained via our proposed definition of $Q[B]$ in terms of

$P[K * A]$, for a fixed function $*$. In other words, $P'_K[A \cap B](.|.) \neq P_{K*(A \cap B)}(.|.)$. Consider for example the belief revision function $*'$ such that $K^{*'}(A \cap B) = \{w\}$. This function is permitted by the AGM axioms, even when it is not the one we used in order to construct the example. By using this function we can show that there is indeed a manner of representing $Q[B](.|.)$ via $P'_K[A \cap B](.|.)$, namely by the function $P_{K*'(A \cap B)}(.|.)$.[11]

Recently T. Williamson considered as well transitions where the background knowledge is *contracted* - similar ideas were first proposed by Isaac Levi in [12]. This permits that an event of measure 1, relative to K, ceases to have measure 1, relative to background knowledge $K \div A$, where $\div$ is a contraction operation on qualitative belief (see [21], chapter 10). Defining probability kinematics for a notion of finitely additive probability without Cumulativity (as we did in this section) seems to require the use of notions of background knowledge and qualitative supposition for which there is no probabilistic account, denying therefore radical probabilism.

References

[1] Arló Costa, H. and Parikh R.: 'Two-place probabilities, beliefs and belief revision,' *Proceedings of the Twelfth Amsterdam Colloquium*, Paul Dekker (ed.), 1-6, 1999.

[2] Arló Costa, H. 'Qualitative and probabilistic models of full belief,' *Proceedings of Logic Colloquim'98, Lecture Notes on Logic 13*, S. Buss, P.Hajek, P. Pudlak (eds.), ASL, A. K. Peters, 1999.

[3] Arló Costa, H.: 'Bayesian epistemology and epistemic conditionals: On the Status of the Export-Import Laws,' *Journal of Philosophy*, Vol. XCVIII, 11, 555-598, 2001.

[4] Arló Costa, H. and Thomason, R. 'Iterative probability kinematics', *Journal of Philosophical Logic* 00, 46, 479-524, 2001

[5] Arló Costa, H. 'Hypothetical revision and matter-of-fact supposition,' *Journal of Applied Non-Classical Logic* Vol. 11, No 1-2, 1-14, 2001.

[6] Christensen, D. 'Clever bookies and Dutch strategies,' *Philosophical Review*, 100, 229-47, 1991.

---

[11]When the probability function $P_K(.|.)$ is interpreted as $Q_\Omega(.|.)$ and $P[K * A](.|.)$ as $Q'_{\Omega+A}(.|.)$, where $\Omega + A$ is the expansion of Ω with A: $\{w3, w4, w, w'\}$ then the principle of *confirmational conditionalization* advanced by Levi in [12] requires that $Q''_{\Omega+(A \cap B)}(.|\Omega) = Q'_{\Omega+A}(.|B) = Q_\Omega(.|A \cap B)$. But here $P_K[A](.|Y)$ abbreviates the result of assigning a credal state to the corpus of knowledge obtained after *contracting* K with $\neg$ A and then expanding it with A.

[7] A. Darwiche, J. Pearl, 'On the logic of iterated belief revision,' *Artificial Intelligence*, 89(1–2):1–31, 1997.

[8] Dubins, L.E. 'Finitely additive conditional probabilities, conglomerability, and disintegrations,' *Ann. Prob.* 3:89-99, 1975.

[9] Gärdenfors, P. *Knowledge in Flux*, Bradford Book, MIT Press, Cambridge, Mass, 1988.

[10] Jeffreys, H. *Theory of Probability*, New York, Oxford, 3rd edition, 1961.

[11] Pagnuco, M. *The Role of Abductive Reasoning within the Process of Belief Revision*, Ph.D.Thesis, Basser Department of Computer Science, University of Sydney, Australia, 1996.

[12] Levi, I. *The Enterprise of Knowledge*, MIT Press, Cambridge, Mass., '1980.

[13] Levi, I. 'The Demons of Decision,' *The Monist*, Vol. 70, No. 2: 193-211, 1987.

[14] Schervish, M.J., Seidenfeld, T. and Kadane J. 'The extent of non-conglomerability of finitely additive probability,' *Z. War.* 66:205-226, 1984.

[15] Seidenfeld, T. 'Remarks on the Theory of Conditional Probability: Some Issues of Finite versus Countable Additivity,' *Probability Theory: Philosophy, Recent History and Relations to Science* eds. V. F. Hendricks, S. A. Pedersen and K.F. Jorgensen, Kluwer Academic Publishers, Dordreccht:167-178, 2001.

[16] Seidenfeld, T.; Schervish, M. and Kadane J. 'Improper regular conditional distributions,'*Ann. Prob.*, 1-13, 2001.

[17] Spohn, W. 'A general non-probabilistic theory of inductive inference,' *Causation in Decision, Belief Change and Statistics*, eds. W. Harper and B. Skyrms, Dordrecht:Reidel, 105-134, 1988.

[18] Spohn, W. 'The representation of Popper Measures,' *Topoi*, V: 69-74, 1986.

[19] van Fraassen, B. C. *Laws and Symmetry*, Oxford, New York, Oxford University Press, 1989.

[20] van Fraassen, B. C. 'Fine-grained opinion, probability, and the logic of full belief,' *Journal of Philosophical Logic*, XXIV: 349-77, 1995.

[21] Williamson, T. *Knowledge and its Limits*, Oxford, 2000.