Miklós Rédei and Zalán Gyenis

Measure theoretic analysis of consistency of the Principal Principle

Article (Accepted version) (Refereed)

Title	Measure theoretic analysis of consistency of the Principal Principle
Article type	PSA 2014 Symposium Paper Submission
Author #1	Miklós Rédei
Affiliation #1	Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, m.redei@lse.ac.uk
Author #2	Zalán Gyenis
Affiliation #2	BUTE Department of Algebra, Budapest, Hungary, gyz@renyi.hu
Abstract	Weak and strong consistency of the Abstract Principal Principle are defined in terms of classical probability measure spaces. It is proved that the Abstract Principal Principle is both weakly and strongly consistent. The Abstract Principal Principle is strengthened by adding a stability requirement to it. Weak and strong consistency of the resulting Stable Abstract Principal Principle are defined. It is shown that the Stable Abstract Principal Principle is weakly consistent. Strong consistency of the Stable Abstract Principal principle remains an open question.
Acknowledgement	Research supported in part by the National Research, Development and Innovation Office, K 115593 and K 100715. M. Rédei thanks the Institute of Philosophy of the Hungarian Academy of Sciences, with which he was affiliated as Honorary Research Fellow while this paper was written.

1 The claims

This paper investigates the measure theoretic consistency of what we call the "Abstract Principal Principle". The consistency expresses that the Abstract Principal Principle is in harmony with the basic structure of measure theoretic probability theory. This type of consistency is tacitly assumed in the literature on the Principal Principle, although we will see that the consistency in question is not trivial. The main philosophical significance of proving such a consistency is that without making sure that such a consistency obtains, the Abstract Principal Principle would be inconsistent as a general norm that guides forming subjective degrees of belief (credences): Without such consistency a Bayesian Agent would not always be able to adjust his degrees of belief to objective probabilities (e.g. chances) in a Bayesian manner, via Bayesian conditionalization.

After stating the Abstract Principal Principle informally in section 2, we define formally the *weak* and *strong* consistency of the Abstract Principal Principle (Definitions 3.1 and 3.3) in section 3, and state weak and strong consistency of the Abstract Principal Principle (Propositions 3.2 and 3.4). We will then argue that it is very natural to strengthen the Abstract Principal Principle by requiring it to satisfy a *stability* property, which expresses that conditional degrees of belief in events *already* equal (in the spirit of the Abstract Principal Principle) to the objective probabilities of the events do not change as a result of conditionalizing them further on knowing the objective probabilities of *other* events (in particular of events that are independent with respect to their objective probabilities). We call this amended principle the *Stable* Abstract Principal Principle (if stability is required only with respect to further conditionalizing on values of probabilities of *independent* events: *Independence-Stable* Principal Principle). This stability requirement leads to suitably modified versions of both the weak and strong consistency of the (*Independence-*)Stable Abstract Principal Principle (Definitions 5.1 and 5.4). We will prove that the Stable Abstract Principal Principle is *weakly* consistent (Proposition 5.2). This entails weak consistency of the Independence-Stable Abstract Principal Principle (Proposition 5.3). The *strong* consistency of both the Stable and the Independence-Stable Abstract Principal Principle remain open problems however; we conjecture that both consistencies hold[1].

Until section 6 no references are given. Section 6 puts the results into context, here we discuss the relevance of strong consistency of the Stable Abstract Principal Principle from the perspective of Lewis' Principal Principle and its "debugged" versions. The details of all the proofs are in the Appendix.

2 The Abstract Principal Principle informally

The Abstract Principal Principle regulates probabilities representing the subjective degrees of belief $p_{subj}(A)$ of an abstract Bayesian agent by stipulating that $p_{subj}(A)$ are related to the objective probabilities $p_{obj}(A)$ as

$$p_{subj}(A|\ulcorner p_{obj}(A) = r\urcorner) = p_{obj}(A) \tag{1}$$

where $\ulcorner p_{obj}(A) = r\urcorner$ denotes the proposition "the objective probability, $p_{obj}(A)$, of A is equal to r".

The formulation (1) of the Abstract Principal Principle presupposes that both p_{subj} and p_{obj} are probability measures: additive maps defined on a σ-algebra taking values in $[0,1]$. p_{obj} is supposed to be defined on a σ-algebra $\mathcal{S}_{obj}$ of random events; and p_{subj} is supposed to be a map with a domain of definition being a σ-algebra $\mathcal{S}_{subj}$.

It is crucial to realize that the σ-algebras $\mathcal{S}_{obj}$ and $\mathcal{S}_{subj}$ cannot be unrelated: for the

[1] G. Bana, in his contribution to the symposium and to the present volume proved this conjecture.

conditional probability $p_{subj}(A|\ulcorner p_{obj}(A)=r\urcorner)$ in eq. (1) to be well-defined via Bayes' rule, the σ-algebra $\mathcal{S}_{subj}$ must contain *both* the σ-algebra $\mathcal{S}_{obj}$ of random events *and* with every random event A also the proposition $\ulcorner p_{obj}(A)=r\urcorner$ — otherwise the formula $p_{subj}(A|\ulcorner p_{obj}(A)=r\urcorner)$ cannot be interpreted as an expression of conditional probability specified by Bayes' rule.

It is far from obvious however that, given *any* σ-algebra $\mathcal{S}_{obj}$ of random events with *any* probability measure p_{obj} on $\mathcal{S}_{obj}$, there exists a σ-algebra $\mathcal{S}_{subj}$ meeting these algebraic requirements in such a way that a probability measure p_{subj} satisfying the condition (1) also exists on $\mathcal{S}_{subj}$. If there exists a σ-algebra $\mathcal{S}^*_{obj}$ of random events with a probability measure p^*_{obj} giving the objective probabilities of events for which there exists *no* σ-algebra $\mathcal{S}_{subj}$ on which a probability function p_{subj} satisfying (1) can be defined, then the Abstract Principal Principle would be inconsistent as a general norm: In this case the agent, being in the epistemic situation of facing the objective facts represented by $(\mathcal{S}^*_{obj}, p^*_{obj})$, cannot have degrees of belief satisfying the Abstract Principal Principle for fundamental structural reasons inherent in the basic structure of classical probability theory. We say that the Abstract Principal Principle is *weakly consistent* if it is *not* inconsistent in the sense described. (The adjective "weakly" will be explained shortly.)

Remark 2.1. One can construe the Principal Principle differently: taking it as a norm that regulates *internal consistency* of the Agent.[2] Under this construal the subjective degrees of belief should satisfy

$$p_{subj}(A|\ulcorner p_{obj}(A)=r\urcorner)=r \qquad \text{for all } r\in[0,1] \tag{2}$$

Here $\ulcorner p_{obj}(A)=r\urcorner$ is the proposition that the Agent believes that the objective probability

[2]We thank C. Hoefer and G. Bana for pointing this out in the discussion in the symposium.

of A is equal to r, and (2) requires that the Agent's subjective degrees of belief conditional on this belief should be equal to r – otherwise the Agent is inconsistent in his thinking. The difference between (1) and (2) is that r on the right hand side of (2) need not be equal to the real objective probability $p_{obj}(A)$. The difference between these two interpretations plays no role however from the perspective of the consistency problem we investigate here: Because of the universal quantification over p_{obj} in the consistency definitions and because of the universal quantification over r in (2) the two construals lead to the same consistency problem.

3 Weak and strong consistency of the Abstract Principal Principle

$(X, \mathcal{S}, p)$ denotes a classical probability measure space, where $\mathcal{S}$ is a σ-algebra of (some) subsets of X and p is a probability measure on $\mathcal{S}$. Given two σ-algebras $\mathcal{S}$ and $\mathcal{S}'$, the injective map $h\colon \mathcal{S} \to \mathcal{S}'$ is a σ-algebra embedding if it preserves all Boolean-σ-operations. The probability space $(X', \mathcal{S}', p')$ is called an extension of $(X, \mathcal{S}, p)$ with respect to h if h is a σ-algebra embedding of $\mathcal{S}$ into $\mathcal{S}'$ that preserves the probability measure p:

$$p'(h(A)) = p(A) \qquad A \in \mathcal{S} \tag{3}$$

Definition 3.1. The Abstract Principal Principle is called *weakly consistent* if the following hold: Given any probability space $(X_{obj}, \mathcal{S}_{obj}, p_{obj})$, there exists a probability space $(X_{subj}, \mathcal{S}_{subj}, p_{subj})$ and a σ-algebra embedding h of $\mathcal{S}_{obj}$ into $\mathcal{S}_{subj}$ such that

(i) For every $A \in \mathcal{S}_{obj}$ there exists an $A' \in \mathcal{S}_{subj}$ with the property

$$p_{subj}(h(A)|A') = p_{obj}(A) \tag{4}$$

(ii) If $A, B \in \mathcal{S}_{obj}$ and $A \neq B$ then $A' \neq B'$.

Definition 3.1 says: Given the "objective" probability space $(X_{obj}, \mathcal{S}_{obj}, p_{obj})$, the σ-algebra $\mathcal{S}_{subj}$ in $(X_{subj}, \mathcal{S}_{subj}, p_{subj})$ contains the "copies" $h(A)$ of all the random events $A \in \mathcal{S}_{obj}$ and also an element A' to be interpreted as representing the proposition "the objective probability, $p_{obj}(A)$, of A is equal to r" (this proposition we denoted by $\ulcorner p_{obj}(A) = r \urcorner$). If $A \neq B$ then $A' \neq B'$ must hold because $\ulcorner p_{obj}(A) = r \urcorner$ and $\ulcorner p_{obj}(B) = s \urcorner$ are different propositions – this is expressed by (ii) in the definition. The main content of the Abstract Principal Principle is then expressed by condition (4), which states that the *conditional* degrees of beliefs $p_{subj}(h(A)|A')$ of an agent about random events $h(A) \leftrightarrow A \in \mathcal{S}_{obj}$ are equal to the objective probabilities $p_{obj}(A)$, where the condition A' is that the agent knows the values of the objective probabilities.

Proposition 3.2. *The Abstract Principal Principle is weakly consistent.*

The above proposition follows from Proposition 5.2 stating the weak consistency of the *Stable* Abstract Principal Principle, which we state later.

Definition 3.3. The Abstract Principal Principle is defined to be *strongly consistent* if, in addition to conditions (i)-(ii) in Definition 3.1, the following hold:

(iii) The probability space $(X_{subj}, \mathcal{S}_{subj}, p_{subj})$ is an extension of the probability space $(X_{obj}, \mathcal{S}_{obj}, p^0_{subj})$ with respect to h; i.e. we have

$$p_{subj}(h(A)) = p^0_{subj}(A) \qquad A \in \mathcal{S}_{obj} \tag{5}$$

The content of this additional requirement is that the agent's prior probability function p_{subj} restricted to the random events can be equal to probability measure p^0_{subj} on $\mathcal{S}_{obj}$ that can differ from the objective probabilities of the random events given by p_{obj}.

Proposition 3.4. *The Abstract Principal Principle is strongly consistent if p_{obj} is absolutely continuous w.r.t. the agent's prior degrees of beliefs p^0_{subj}.*

4 The Stable Abstract Principal Principle

Once the agent has adjusted his subjective degree of belief by conditionalizing, $p_{subj}(h(A)|\ulcorner p_{obj}(A) = r\urcorner) = r$, he may then learn the value of another objective probability, $\ulcorner p_{obj}(B) = s\urcorner$, in which case he must conditionalize again. What should be the result of this second conditionalization? Since the agent's conditional degrees of belief $p_{subj}(h(A)|\ulcorner p_{obj}(A) = r\urcorner)$ in A are already correct (equal to the objective probabilities), it would be irrational to change his already correct degree of belief about A upon learning an additional *truth*, namely the value of the objective probability $p_{obj}(B)$. So a *rational* agent's conditional subjective degrees of belief should be *stable* in the sense of satisfying the following condition:

$$p_{subj}\big(h(A)|\ulcorner p_{obj}(A) = r\urcorner\big) = p_{subj}\big(h(A)|\ulcorner p_{obj}(A) = r\urcorner \cap \ulcorner p_{obj}(B) = s\urcorner\big) \quad (\forall B \in \mathcal{S}_{obj}) \quad (6)$$

If A and B are independent with respect to their objective probabilities $p_{obj}(A \cap B) = p_{obj}(A)p_{obj}(B)$, then, if the conditional subjective degrees of belief are stable in the sense of (6), then (assuming the Abstract Principal Principle) one has

$$\begin{aligned}
p_{subj}(h(A)\cap h(B)|\ulcorner p_{obj}(A)=r\urcorner\cap\ulcorner p_{obj}(B)=s\urcorner\cap\ulcorner p_{obj}(A\cap B)=t\urcorner) \qquad (7)\\
=p_{subj}(h(A\cap B)|\ulcorner p_{obj}(A)=r\urcorner\cap\ulcorner p_{obj}(B)=s\urcorner\cap\ulcorner p_{obj}(A\cap B)=t\urcorner)\\
=p_{subj}(h(A\cap B)|\ulcorner p_{obj}(A\cap B)=t\urcorner)\\
=p_{obj}(A\cap B)\\
=p_{obj}(A)p_{obj}(B)\\
=p_{subj}(h(A)|\ulcorner p_{obj}(A)=r\urcorner)p_{subj}(h(B)|\ulcorner p_{obj}(B)=s\urcorner)\\
=p_{subj}(h(A)|\ulcorner p_{obj}(A)=r\urcorner\cap\ulcorner p_{obj}(B)=s\urcorner\cap\ulcorner p_{obj}(A\cap B)=t\urcorner) \qquad (8)\\
\cdot p_{subj}(h(B)|\ulcorner p_{obj}(A)=r\urcorner\cap\ulcorner p_{obj}(B)=s\urcorner\cap\ulcorner p_{obj}(A\cap B)=t\urcorner) \qquad (9)
\end{aligned}$$

Equations (7) and (8)-(9) mean that if the conditional subjective degrees of belief are stable, then, if A and B are objectively independent, then they (their isomorphic images $h(A),h(B)$) are also *subjectively* independent: independent also with respect to the probability measure that represents *conditional* subjective degrees of belief, where the condition is that the agent knows the objective probabilities of *all* of A, B and $(A\cap B)$. In this case the conditional subjective degrees of beliefs properly reflect the objective independence relations of random events – they are *independence-faithful*. Note that for the subjective degrees of belief to satisfy the independence-faithfulness condition expressed by eqs. (7) and (8)-(9), it is sufficient that stability (6) only holds for the restricted set of elements B in the σ-subalgebra $\mathcal{S}_{obj}^{A,ind}$ of $\mathcal{S}_{obj}$ generated by the elements in $\mathcal{S}_{obj}$ that are independent of A with respect to p_{obj}.

This motivates to amend the Abstract Principal Principle by requiring stability of the subjective probabilities, resulting in the "Stable Abstract Principal Principle":

Stable Abstract Principal Principle The subjective probabilities $p_{subj}(A)$ are related to the objective probabilities $p_{obj}(A)$ as required by equation (1); furthermore, the subjective probability function is *stable* in the sense that the following holds:

$$p_{subj}\big(h(A)|\ulcorner p_{obj}(A)=r\urcorner\big) = p_{subj}\big(h(A)|\ulcorner p_{obj}(A)=r\urcorner \cap \ulcorner p_{obj}(B)=s\urcorner\big) \qquad (\forall B \in \mathcal{S}_{obj}) \tag{10}$$

If the subjective probability function is only *independence-stable* in the sense that (10) above holds for all $B \in \mathcal{S}^{A,ind}_{obj}$, then the corresponding Stable Abstract Principal Principle is called the *Independence-Stable* Abstract Principal Principle.

5 Is the Stable Abstract Principal Principle strongly consistent?

Definition 5.1. The Stable Abstract Principal Principle is defined to be *weakly consistent* if it is weakly consistent in the sense of Definition 3.1 and the subjective probability function p_{subj} is stable: it satisfies condition (10). The *Independence-Stable* Abstract Principal Principle is defined to be weakly consistent if it is weakly consistent in the sense of Definition 3.1 and the subjective probability function p_{subj} is *independence-stable*: it satisfies (10) for all $B \in \mathcal{S}^{A,ind}_{obj}$.

Proposition 5.2. *The Stable Abstract Principal Principle is weakly consistent.*

The above proposition entails

Proposition 5.3. *The Independence-Stable Abstract Principal Principle is weakly consistent.*

Definition 5.4. The Stable Abstract Principal Principle is defined to be *strongly consistent* if it is strongly consistent in the sense of Definition 3.3 and the subjective probability function p_{subj} is stable. The *Independence-Stable* Abstract Principal principle is strongly consistent if it is strongly consistent in the sense of Definition 3.3 and the subjective probability function p_{subj} satisfies (10) for all $B \in \mathcal{S}_{obj}^{A,ind}$.

Problem 5.5. Is the (Independence-)Stable Abstract Principal Principle strongly consistent?

The problem of strong consistency of both the Stable and the Independence-Stable Abstract Principal Principle remain open[3].

6 Relation to other works

Lewis (1986) introduced the term "Principal Principle" to refer to the principle linking subjective beliefs to chances. In the context of the Principal Principle $p_{subj}(A)$ is called the "credence", $Cr_t(A)$, of the agent in event A at time t, $p_{obj}(A)$ is the chance $Ch_t(A)$ of the event A at time t, and the Principal Principle is the stipulation that credences and chances are related as

$$Cr_t(A|\ulcorner Ch_t(A) = r\urcorner \cap E) = Ch_t(A) = r \tag{11}$$

where E is any *admissible* evidence the agent has at time t in addition to knowing the value of the chance of A.

Proposition $\ulcorner Ch_t(A) = r\urcorner$ is clearly admissible evidence for (11), and, substituting $E = \ulcorner Ch_t(A) = r\urcorner$ into equation (11), we obtain

$$Cr_t(A|\ulcorner Ch_t(A) = r\urcorner) = Ch_t(A) = r \tag{12}$$

which, at any given time t, is an instance of the Abstract Principal Principle if we make the identifications $p_{obj}(A) = Ch_t(A)$, $p_{subj}(A) = Cr_t(A)$. By Proposition 3.4 we know that, for any time parameter t, relation (12) is consistent with probability as measure.

[3]See footnote 1.

If, however, admissibility of evidence E is defined in such a way that propositions stating the values of chances of other events B at time t (i.e. propositions of the form $\ulcorner Ch_t(B) = s \urcorner$) are admitted as E, then (11) together with (12) entail that we also should have

$$Cr_t(A|\ulcorner Ch_t(A) = r \urcorner \cap \ulcorner Ch_t(B) = s \urcorner) = Ch_t(A) = r \tag{13}$$

The relation (13) together with equation (12) is, at any given time t, an instance of the *Stable* Abstract Principal Principle if we make the identifications $p_{obj}(A) = Ch_t(A)$, $p_{subj}(A) = Cr_t(A)$ and $p_{obj}(B) = Ch_t(B)$. Thus whether relations (13) and (12) can hold at all is exactly the question of whether the *Stable* Abstract Principal Principle is strongly consistent. If one allows as evidence E in (13) only propositions stating the value of objective chances of events B that are *objectively independent* of A, then the question of whether relations (13) and (12) can hold in general is exactly the question of whether the *Independence*-Stable Abstract Principal Principle is strongly consistent. Since Lewis regarded admissible all propositions containing information that is "irrelevant" for the chance of A (Lewis 1986, 91), for Lewis, admissible evidence should include propositions about values of chances of events that are independent of A with respect to the probability measure describing their chances. Under this interpretation of "irrelevant" information, the consistency of Lewis' Principal Principle as a general norm needs proving consistency of the *Independence*-Stable Abstract Principal Principle. It should be emphasized that this kind of consistency has nothing to do with any metaphysics about chances or with the concept of natural law that one may have in the background of the Principal Principle; in particular, this inconsistency is different from the one related to "undermining" (see below). This consistency expresses a simple but fundamental compatibility of the Principal Principle with the basic structure of probability theory.

Lewis himself saw a consistency problem in his Principal Principle (he called it the "Big

Bad Bug"): If A is an event in the future of t that has a non-zero chance $r > 0$ of happening at that later time but we have knowledge E about the future that entails that A will in fact not happen, $E \subset A^{\perp}$, then substituting this E into (11) leads to contradiction if $r > 0$. Such an A is called an "unactualized future that undermines present chances" – hence the phrase "undermining" to refer to this situation. Since certain metaphysical arguments led Lewis to think that one is forced to admit such an evidence E, he tried to "debug" the Principal Principle (Lewis 1994); the same sort of debugging was proposed simultaneously by Hall (1994) and Thau (1994). Other debugging attempts have followed (Black 1998; Roberts 2001; Loewer 2004; Hall 2004; Hoefer 2007; Ismael 2008; Meacham 2010; Glynn 2010; Nissan-Rozen 2013; Pettigrew 2013; Frigg–Hoefer 2015), and to date no consensus has emerged as to which of the debugged versions of the Principal Principle is tenable: Vranas (2004) claims that there was no need for a debugging in the first place; Briggs (2009) argues that none of the modified principles work; Pettigrew (2012) provides a framework that allows to choose the correct Principal Principle depending on one's metaphysical concept of chance.

Papers aiming at "debugging" Lewis' Principal Principle typically combine the following three moves (a), (b) or (c):

(a) Restricting the admissible evidence in (11) to a particular class $\mathcal{A}_A$ of propositions in order to avoid "undermining" (Hoefer 2007).

(b) Modifying the Principal Principle by replacing $Ch_t(A)$ on the right hand side of (11) with a value $F(A)$ given by a function F different from the objective chance function (New Principle by Hall (1994); General Principal Principle by Lewis (1980) and by Roberts (2001)).

(c) Modifying the Principal Principle by replacing the conditioning proposition $\ulcorner Ch_t(A) = r \urcorner \cap E$ on the left hand side of (11) by a different conditioning proposition

C_A, which is a conjunction of some propositions from $\mathcal{S}_{obj}$, $\mathcal{A}_A$, and propositions of form $\ulcorner p_{obj}(B) = r \urcorner$ (Conditional Principle and General Principle by Vranas (2004)); General Recipe by Ismael (2008)).

To establish a theory of chance along a debugging strategy characterized by a combination of (a), (b) and (c), it is not enough to show however that undermining is avoided: one has to prove that the debugged Principal Principle is consistent in the sense of Definition 6.1 below, which is in the spirit of the notion consistency investigated in this paper:

Definition 6.1. We say that the "$(\mathcal{A}_A, C_A, F)$-debugged" Principal Principle is *strongly* consistent if the following hold:
Given any probability space $(X_{obj}, \mathcal{S}_{obj}, p_{obj})$ and another probability measure p^0_{subj} on $\mathcal{S}_{obj}$, there exists a probability space $(X_{subj}, \mathcal{S}_{subj}, p_{subj})$ and a σ-algebra embedding h of $\mathcal{S}_{obj}$ into $\mathcal{S}_{subj}$ such that

(i) For every $A \in \mathcal{S}_{obj}$ the set $\mathcal{A}_A$ is in $\mathcal{S}_{subj}$, and for every $A \in \mathcal{S}_{obj}$ there exists a $C_A \in \mathcal{S}_{subj}$ with the property

$$p_{subj}(h(A)|C_A) = F(A) \tag{14}$$

(ii) If $A, B \in \mathcal{S}_{obj}$ and $A \neq B$ then $C_A \neq C_B$.

(iii) The probability space $(X_{subj}, \mathcal{S}_{subj}, p_{subj})$ is an extension of the probability space $(X_{obj}, \mathcal{S}_{obj}, p^0_{subj})$ with respect to h; i.e. we have

$$p_{subj}(h(A)) = p^0_{subj}(A) \qquad A \in \mathcal{S}_{obj} \tag{15}$$

(iv) For all $A \in \mathcal{S}_{obj}$ and for all $B \in \mathcal{A}_A$ we have

$$p_{subj}(h(A)|C_A) = p_{subj}\big(h(A)|C_A \cap B\big) \tag{16}$$

We say that the "$(\mathcal{A}_A, C_A, F)$-debugged" Principal Principle is *weakly* consistent if (i),(ii) and (iv) hold.

Taking specific C_A, and F, one obtains particular consistency definitions expressing the consistency of specific debugged Principal Principles. For instance, stipulations

$$C_A = B \cap \ulcorner p_{obj}(A|B) = r \urcorner \tag{17}$$

$$F(A) = p_{obj}(A) \tag{18}$$

yield Vranas' Conditional Principle (Vranas 2004, 370); whereas Hall's New Principle (Hall 1994, 511) can be obtained by

$$C_A = H_{t,w} \cap T_w \tag{19}$$

$$F(A) = p_{obj}(A|T_w) \tag{20}$$

where $H_{t,w}$ is "the proposition that completely characterizes w's history up to time t" (Hall 1994, 506) and T_w is the "proposition that completely characterizes the laws at w" (Hall 1994, 506) (w being a possible world).

Proving consistency of the $(\mathcal{A}_A, C_A, F)$-debugged Principal Principles is necessary for the respective debugged Principal Principles to be compatible with measure theoretic probability theory. To our best knowledge such consistency proofs have *not* been given: it seems that this type of consistency is tacitly assumed in the works analyzing the modified Principal Principles, although, as the propositions and their proofs presented in this paper show, the truth of these types of consistency claims are far from obvious.

The problem of strong consistency of the Stable Abstract Principle is also relevant from the perspective of existence of particular models of the axioms of higher order probability theory (HOP) suggested by Gaifman (1988). If one regards the theory of HOP as an axiomatic theory, then the question arises whether models of the theory exist. Gaifman provides a few specific examples that are models of the axioms (Gaifman 1988, 208–10) but he does not raise the general issue of what kind of models exist. What one would like to know is whether any objective probability theory can be made part of a HOP in such a way that the objective probabilities are related to the subjective ones in the manner required by the HOP axioms. Proving the existence of such HOPs entail that the Stable Abstract Principal Principle is strongly consistent.

7 Appendix

7.1 Proof of strong consistency of the Abstract Principal Principle (Proposition 3.4)

The statement follows from Proposition 7.1 below if we make the following identifications:

- $(X_{obj}, \mathcal{S}_{obj}, p_{obj}) \leftrightarrow (X, \mathcal{S}, \hat{p})$
- $(X_{obj}, \mathcal{S}_{obj}, p^0_{subj}) \leftrightarrow (X, \mathcal{S}, p)$
- $(X_{subj}, \mathcal{S}_{subj}, p_{subj}) \leftrightarrow (X', \mathcal{S}', p')$

Proposition 7.1. *Let $(X, \mathcal{S}, p)$ be a probability space and let $\hat{p}$ be another probability measure on $\mathcal{S}$ such that $\hat{p}$ is absolutely continuous with respect to p. Then there exists an extension $(X', \mathcal{S}', p')$ of $(X, \mathcal{S}, p)$ with respect to the embedding $h\colon \mathcal{S} \to \mathcal{S}'$ having the following properties:*

(i) For all $A \in \mathcal{S}$ there is $A' \in \mathcal{S}'$ such that

$$p'\big(h(A)|A'\big) = \hat{p}(A)$$

(ii) $A \neq B$ implies $A' \neq B'$

Proof. We distinguish two cases: (i) the σ-algebra $\mathcal{S}$ is finite (ii) non-finite.

When $\mathcal{S}$ is finite, the proof consist of two steps. In the first step we choose an arbitrary element $A \in \mathcal{S}$ and construct an extension $(X^*, \mathcal{S}^*, p^*)$ of $(X, \mathcal{S}, p)$ with respect to an embedding h^* in such a manner that in this extension this particular event A has a pair $A' = A^*$ with the required properties. In step 2 we repeat this step $n-1$ times, choosing each time another element from $\mathcal{S}$ until we exhaust $\mathcal{S}$ and obtain the extension $(X', \mathcal{S}', p')$ of $(X, \mathcal{S}, p)$.

Step 1. Take any $A \in \mathcal{S}$. We wish to construct a space $(X^*, \mathcal{S}^*, p^*)$ and a function $h^* : \mathcal{S} \to \mathcal{S}^*$ such that

- $h^* : (\mathcal{S}, p) \to (\mathcal{S}^*, p^*)$ is a measure preserving, injective Boolean algebra homomorphism.
- There is $A^* \in \mathcal{S}^*$ such that $p^*\big(h^*(A)|A^*\big) = \hat{p}(A)$.

Let let $(X^1, \mathcal{S}^1)$ and $(X^2, \mathcal{S}^2)$ be two disjoint copes of $(X, \mathcal{S})$, and fix the algebra isomorphisms $h^1 : (X, \mathcal{S}) \to (X^1, \mathcal{S}^1)$ and $h^2 : (X, \mathcal{S}) \to (X^2, \mathcal{S}^2)$. Put $X^* = X^1 \cup X^2$ and define

$$\mathcal{S}^* = \big\{h^1(A) \cup h^2(B) : A, B \in \mathcal{S}\big\} \tag{21}$$

It is a routine task to verify that $\mathcal{S}^*$ is a Boolean algebra of subsets of X^* with respect to the usual set theoretical operations $\cup$, $\cap$, $\setminus$ (below we also use the notation $A^{\perp}$ to refer to the set theoretical complement of an element A with respect to a set which is fixed by the context).

Define the map $h^* : \mathcal{S} \to \mathcal{S}^*$ by

$$h^*(A) = h^1(A) \cup h^2(A) \qquad\qquad A \in \mathcal{S} \tag{22}$$

h^* is a homomorphism between $\mathcal{S}$ and $\mathcal{S}^*$.

Let $0 \le \alpha \le 1$ be any number and define p^* on $\mathcal{S}^*$ by

$$p^*\big(h^1(A) \cup h^2(B)\big) \doteq \alpha \cdot p(A) + (1-\alpha) \cdot p(B) \qquad A, B \in \mathcal{S} \tag{23}$$

For each $A \in \mathcal{S}$ we have then

$$p^*\big(h^*(A)\big) = \alpha \cdot p(A) + (1-\alpha) \cdot p(A) = p(A) \tag{24}$$

Consequently, $h^* : (\mathcal{S}, p) \to (\mathcal{S}^*, p^*)$ is a measure preserving, injective Boolean algebra homomorphism.

For any fixed $A \in \mathcal{S}$ define A^* by

$$A^* \doteq h^1(A) \cup h^2(A^\perp) \tag{25}$$

Our aim now is to choose α in such a way that the following is true:

$$p^*(h^*(A)|A^*) = \hat{p}(A) \tag{26}$$

Some basic algebra shows that

$$p^*(h^*(A)|A^*) \quad = \quad \frac{\alpha \cdot p(A)}{\alpha \cdot p(A) + (1-\alpha) \cdot (1 - p(A))} \tag{27}$$

Thus in order to satisfy (26) we have to choose α to guarantee

$$\frac{\alpha \cdot p(A)}{\alpha \cdot p(A) + (1-\alpha)\cdot(1-p(A))} = \hat{p}(A) \tag{28}$$

By assumption, if $p(A) = 1$ then $\hat{p}(A) = 1$, and thus any $\alpha \neq 0$ makes (28) true. Similarly, if $p(A) = 0$, then $\hat{p}(A) = 0$, which means that any $\alpha \neq 1$ will do. Also, if $\hat{p}(A) = 0$, then $\alpha = 0$ will do. Therefore we may assume $0 < p(A) < 1$ and $0 < \hat{p}(A) \leq 1$. By re-ordering equation (28) and using the notation $p = p(A)$, $r = \hat{p}(A)$ we obtain

$$\alpha = \frac{rp - r}{rp - r + pr - p} \tag{29}$$

To guarantee (28) we only have to show that α in equation (29) is between 0 and 1. Since $0 < p < 1$ and $0 < r \leq 1$ we have $rp < r$ and $pr \leq p$. This means that both the numerator and the denominator of the fraction in (29) is negative, whence α is positive. On the other hand, we have

$$\begin{aligned} 0 &\geq pr - p \\ rp - r &\geq rp - r + pr - p \\ \frac{rp - r}{rp - r + pr - p} &\leq 1 \end{aligned}$$

Thus $0 \leq \alpha \leq 1$ can always be chosen so that equation (26) holds.

Step 2. We obtain $(X', \mathcal{S}', p')$ by iterating Step 1. Let $A_1, \ldots, A_n$ be an enumeration of $\mathcal{S}$. Applying Step 1. with A_1 in place of A, one finds a space $(X_1, \mathcal{S}_1, p_1) = (X^*, \mathcal{S}^*, p^*)$, an event $A_1^* \in \mathcal{S}_1$ and an embedding h_1

$$(X, \mathcal{S}, p) \xrightarrow{h_1} (X_1, \mathcal{S}_1, p_1),$$

such that

$$p_1\big(h_1(A_1)|A_1^*\big) = \hat{p}(A_1) \tag{30}$$

Continuing in this way, we get elements $\big(h_{i-1}\cdots h_1(A_i)\big)^* \in \mathcal{S}_i$ and a chain of extensions

$$(X,\mathcal{S},p) \xrightarrow{h_1} (X_1,\mathcal{S}_1,p_1) \xrightarrow{h_2} (X_2,\mathcal{S}_2,p_2) \xrightarrow{h_3} \cdots \xrightarrow{h_n} (X_n,\mathcal{S}_n,p_n)$$

such that

$$p_n\left(h_n\cdots h_2h_1(A_i)\middle| h_n\cdots h_{i+1}\Big(\big(h_{i-1}\cdots h_1(A_i)\big)^*\Big)\right) = \hat{p}(A_i)$$

holds for all A_i. Therefore we can complete the proof by letting

$$\begin{aligned}
(X',\mathcal{S}',p') &= (X_n,\mathcal{S}_n,p_n)\\
h &= h_nh_{n-1}\cdots h_1\\
A_i' &= h_n\cdots h_{i+1}\Big(\big(h_{i-1}\cdots h_1(A_i)\big)^*\Big)
\end{aligned}$$

One has to verify that the extension in step j does not destroy the result of the previous one. But this is a consequence of h_j being an embedding that preserves the probability.

When the σ-algebra $\mathcal{S}$ is not finite, we take the extension $(X',\mathcal{S}',p')$ to be the product space

$$(X,\mathcal{S},p)\circledast([0,1],\mathcal{L},\lambda) = (X\circledast[0,1],\mathcal{S}\circledast\mathcal{L},p\circledast\lambda)$$

where $([0,1],\mathcal{L},\lambda)$ is the standard Lebesgue space over the unit interval, and where $\circledast$ denotes the special product of two probability spaces introduced in (Gyenis–Rédei 2011). The elements of $\mathcal{S}\circledast\mathcal{L}$ are certain $[0,1]\to\mathcal{S}$ functions, the embedding $h:(X,\mathcal{S},p)\to(X',\mathcal{S}',p')$ is via the constant function

$$h(A)(x) = A \quad (x\in[0,1])$$

The extension of p:

$$p'(h(A)) = \int_0^1 p \circ h(A) d\lambda = \int_0^1 p(A) d\lambda = p(A).$$

Fix a real number $\alpha \in [0,1]$ and take any Lebesgue-measurable subset $B \subseteq [0,1]$ with measure $\lambda(B) = \alpha$. Write A' for the function $A' : [0,1] \to \mathcal{S}$

$$A'(x) = \begin{cases} A & \text{if } x \in B \\ A^{\perp} & \text{otherwise.} \end{cases}$$

Then $A' \in \mathcal{S}'$ and one can verify easily that

$$p'\left(h(A)|A'\right) = \frac{\alpha \cdot p(A)}{\alpha \cdot p(A) + (1-\alpha) \cdot (1 - p(A))}. \tag{31}$$

It follows that if we choose α such that

$$\frac{\alpha \cdot p(A)}{\alpha \cdot p(A) + (1-\alpha) \cdot (1 - p(A))} = \hat{p}(A), \tag{32}$$

then we get

$$p'\left(h(A)|A'\right) = \hat{p}(A)$$

That we can choose α to satisfy (32) is contained in the proof of the finite case.

∎

7.2 Proof of weak consistency of the Stable Abstract Principal Principle (Proposition 5.2)

The statement of weak consistency of the Stable Abstract Principal Principle follows from Proposition 7.2 below if we make the following identifications:

- $(X_{obj}, \mathcal{S}_{obj}, p_{obj}) \leftrightarrow (X, \mathcal{S}, p)$
- $(X_{subj}, \mathcal{S}_{subj}, p_{subj}) \leftrightarrow (X', \mathcal{S}', p')$

Proposition 7.2. *Let $(X, \mathcal{S}, p)$ be a probability space. Then there exists an extension $(X', \mathcal{S}', p')$ of $(X, \mathcal{S}, p)$ with respect to a σ-algebra homomorphism $h\colon \mathcal{S} \to \mathcal{S}'$ such that*

(i) For all $A \in \mathcal{S}$ there is $A' \in \mathcal{S}'$ such that

$$p'\big(h(A)|A'\big) = p(A)$$

(ii) $A \neq B$ implies $A' \neq B'$

(iii)

$$p'(h(A)|A') = p'\big(h(A)|A' \cap B'\big) \qquad (\forall B' \in \mathcal{S}) \tag{33}$$

Proof. Let $(X, \mathcal{S}, p)$ be a probability space and Y_0 be a set disjoint from $\mathcal{S}$ and having the same cardinality as the cardinality of $\mathcal{S}$. We can think of Y_0 as having elements y_A labeled by elements $A \in \mathcal{S}$. Consider the set

$$Y \doteq Y_0 \cup \{y\} = \{y_A : A \in \mathcal{S}\} \cup \{y\}$$

where y is an auxiliary element different from every y_A. Take the powerset $\mathcal{P}(Y)$ and let q be any probability measure on $\mathcal{P}(Y)$ such that $q(\{y\}) \neq 0$. Then $(Y, \mathcal{P}(Y), q)$ is a probability space and we can form the product space

$$(X', \mathcal{S}', p') = (X \times Y, \mathcal{S} \otimes \mathcal{P}(Y), p \times q)$$

with $p' = (p \times q)$ being the product measure on $\mathcal{S} \otimes \mathcal{P}(Y)$. The map $h : \mathcal{S} \to \mathcal{S}'$ defined by $h(A) \doteq A \times Y$ is an injective, measure preserving σ-algebra embedding. For each $A \in \mathcal{S}$ put

$$A' \doteq X \times \{y_A, y\}$$

It is clear that (ii) in the proposition holds for A', B' so defined. Utilizing that p' is a product measure one can verify by explicit calculation that both (i) and (iii) hold. ∎

References

Black, R. (1998). Chance, credence, and the Principal Principle. *The British Journal for the Philosophy of Science*, 49:371–385.

Briggs, R. (2009). The anatomy of the Big, Bad Bug. *Noûs*, 43:428–449.

Frigg, R. and Hoefer, C. (2015). The best Humean system for statistical mechanics. *Erkenntnis*, 80:551–574.

Gaifman, H. (1988). A theory of higher order probabilities. In B. Skyrms and W.L. Harper, editors, *Causation, Chance, and Credence. Proceedings of the Irvine Conference on Probability and Causation, Volume 1*, volume 41 of *The University of Western Ontario series in philosophy of science*, pages 191–219. Kluwer Academic, Dordrecht.

Glynn, L. (2010). Deterministic chance. *The British Journal for the Philosophy of Science*, 61(51-80).

Gyenis, Z. and Rédei, M. (2011). Characterizing common cause closed probability spaces. *Philosophy of Science*, 78:393–409.

Hall, N. (1994). Correcting the guide to objective chance. *Mind*, 103:505–518.

——— (2004). Two mistakes about credence and chance. *Australasian Journal of Philosophy*, 82:93–111.

Hoefer, C. (2007). The third way on objective probability: A sceptic's guide to objective chance. *Mind*, 116:449–596.

Ismael, J. (2008). Raid! Correcting the Big Bad Bug. *Noûs*, 42:292–307.

Lewis, D. (1980). A subjectivist's guide to objective chance. In R.C. Jeffrey, editor, *Studies in Inductive Logic and Probability*, volume II, pages 263–293. University of California Press, Berkely. Reprinted in (Lewis 1986).

——— (1986). *Philosophical Papers*, volume II. Oxford University Press.

——— (1986). A subjectivist's guide to objective chance. In *Philosophical Papers, vol. II*, pages 83–132. Oxford University Press, Oxford.

——— (1994). Humean supervenience debugged. *Mind*, 103:473–490.

Loewer, B. (2004). David Lewis' Humean theory of objective chance. *Philosophy of Science*, 71:115–1125.

Meacham, C.J.G. (2010). Two mistakes regarding the Principal Principle. *The British Journal for the Philosophy of Science*, 61:407–431.

Nissan-Rozen, I. (2013). Jeffrey conditionalization, the Principal Principle, the desire as belief thesis and Adam's thesis. *The British Journal for the Philosophy of Science*, 64:837–850.

Pettigrew, R. (2012). Accuracy, chance and the Principal Principle. *Philosophical Review*, 121:241–275.

——— (2013). A new epistemic utility argument for the Principal Principle. *Episteme*, 10:19–35.

Roberts, J.T. (2001). Undermining undermined: Why Humean supervenience never needed to be debugged. *Philosophy of Science*, 68:S98–S108. Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I Contributed Papers.

Thau, M. (1994). Undermining and admissibility. *Mind*, 103:491–504.

Vranas, P.B.M. (2004). Have your cake and eat it too: The Old Principal Principle reconciled with the New. *Philosophy and Phenomenological Research*, LXIX:368–382.