

# Proposition-valued Random Variables as Information

Richard Bradley

Department of Philosophy, Logic and Scientific Method  
London School of Economics

Houghton Street  
London WC2A 2AE  
r.bradley@lse.ac.uk

September 28, 2009

## Abstract

The notion of a proposition as a set of possible worlds or states occupies central stage in probability theory, semantics and epistemology, where it serves as the fundamental unit both of information and meaning. But this fact should not blind us to the existence of prospects with a different structure. In the paper I examine the use of random variables - in particular, proposition-valued random variables - in these fields and argue that we need a general account of rational attitude formation with respect to them.

## 1 Introduction

The notion of a proposition as a set of possible worlds or states occupies central stage in probability theory, semantics, epistemology and decision theory, where it serves as the fundamental unit both of information and meaning. In its former role, a proposition represents the way the world is or might be; in its latter role, as the content of sentences and of agents' attitudes of belief and desire. A single possible world space thus serves both to model what is the case and how agents perceive their environment, learn from it and shape it.

But the elegance and simplicity of this model should not blind us to its limits. There are features of the interaction of agent and world that are better modelled by postulating the existence of prospects with a structure somewhat more complicated than a set of possible worlds. The sorts of prospects that I have in mind include such things as actions, observations, causal devices, states of knowledge of agents, conditionals and modal prospects. I shall argue that they are best modelled, not as propositions, but as proposition-valued random variables.

Proposition-valued random variables are functions mapping possible worlds to propositions. In simple cases these functions take only a finite number of values, in which case they can be conveniently represented by a matrix of the

following kind:

$$\begin{bmatrix} \textit{Events} & \textit{Values} \\ E_1 & V_1 \\ \vdots & \vdots \\ E_n & V_n \end{bmatrix}$$

with the second column displaying the propositional values and the first the set of possible worlds (the event propositions) taking those values. Intuitively the matrix represents a kind of dependency in the world between these two classes of propositions. The scope for applying the idea is as wide as the possible interpretations of the notion of dependency (causal, evidential, symbolic, etc.), as well as of the objects falling under the two classes.

If such prospects exist then it is natural to ask what attitudes agents should take to them. What do I learn when I find out that such a prospect is the case? How should I adjust my attitudes to other prospects in the light of this information? How probable or desirable should I regard the prospect in question? An attempt at answering these central questions will be made later in the paper, but first a case must be made for the existence of prospects having the structure of proposition-valued random variables. This I do in the next section, by looking at some examples of theories in which proposition-valued random variables are either already in use (possibly implicitly) or could be usefully employed. In the section that follows, proposition-valued random variables are treated more formally and the notions of expectation and conditional expectation of a random variable introduced. Section 4 addresses the question of the desirability and probability of proposition-valued random variables, section 5 the question of how to revise our beliefs in response to learning that the state represented by a random variable holds.

## 2 Proposition-valued Random Variables: Some examples

**Savage Acts** In Richard Jeffrey’s decision theory [10], acts are modelled as propositions. But in the older theory of Leonard Savage [18], still the industry-standard in many disciplines, acts are modelled as functions from states of the world to consequences. Although Savage sometimes suggests that consequences are hedonic states of the agent, his theory is quite compatible with a more naturalist reading of them as states of the world, making their identification with propositions appropriate. This being so, Savage’s acts can be represented by random variables taking worlds as arguments and outcome propositions as values. For example, for Savage, the act of breaking a final, sixth egg into a bowl containing the other five, in the course of making an omelette, is a function that determines the proposition ‘a six-egg omelette’ in all states in which the sixth egg is good and ‘eggs wasted’ in all states in which it is bad. In matrix form we have:

$$\begin{bmatrix} \textit{Events} & \textit{Outcomes} \\ \text{Good egg} & \text{6-egg omelette} \\ \text{Bad egg} & \text{No omelette. Eggs wasted} \end{bmatrix}$$

That a Savage act carries information about the world not expressible by an event proposition is visible in the differences between his theory and the

more parsimonious one of Richard Jeffrey that works only with attitudes to propositions; in particular, in the fact that in Savage's theory agents' preferences for acts suffice to determine a unique measure of their degrees of belief, while in Jeffrey's it does not.<sup>1</sup> A more formal demonstration of this fact is given below.

**Jeffrey's observations** It is common in statistical theory to represent an observation or an experiment by a partition of propositions, each element of which represents both a possible state of the world and a possible item of information that the observer will glean. Thus peeking at the colour of the ball in the urn is represented by a partition - e.g. {red, green, yellow} - of the possible ball colours. This representation might be thought of as a simplification of the Savage-style act of observation, which would take the form of a function from states of the world to propositions expressing what is observed at that world:

$$\left[ \begin{array}{ll} \textit{Events} & \textit{What is observed} \\ \text{Red ball} & \text{Red} \\ \text{Green ball} & \text{Green} \\ \text{Yellow ball} & \text{Yellow} \end{array} \right]$$

However, as Richard Jeffrey [11] observes, the partition representation is adequate only when observation is both veridical and conclusive; that is, when an observation maps every world to a proposition containing it. But in cases of non-veridical or inconclusive observation the full structure of the random variable representation is required. For instance, in cases of colour blindness, the mapping may be many to one:

$$\left[ \begin{array}{ll} \textit{Events} & \textit{What is observed} \\ \text{Red ball} & \text{red-green} \\ \text{Green ball} & \text{red-green} \\ \text{Yellow ball} & \text{Yellow} \end{array} \right] \text{ or } \left[ \begin{array}{ll} \textit{Events} & \textit{What is observed} \\ \text{Red or green ball} & \text{red-green} \\ \text{Yellow ball} & \text{Yellow} \end{array} \right]$$

reflecting the observer's inability to distinguish red from green.<sup>2</sup> In cases where the observer fails to observe anything (e.g. an infrared light source) the random variable representing the observation will take no value at all.

Our colour blind individual may make no error in their own terms if the observation of red-green is reliably correlated with the presence of red and green balls. But if they always report 'red' when making a red-green observation because of a learnt association between the English word 'red' and red-green experiences, their reports will be defective but not uninformative. For third-parties, the niceties of the observer's mental states matter far less than the mapping from states of the world to (possibly non-veridical) reports on this state, that their observation induces. So for third parties the observation is

<sup>1</sup>This is to recklessly simplify a complicated matter, but the point is essentially true nonetheless.

<sup>2</sup>In one of the more common forms of colour-blindness what the person sees is what those with normal vision would classify as yellowish-green. When we represent the colour-blind person's observation as red-green what we mean is that they make an observation which allows them to place the object in the same category that the normally sighted would place red and green objects.

better represented as:

$$\left[ \begin{array}{cc} \textit{Events} & \textit{Reported state} \\ \text{Red ball} & \text{red} \\ \text{Green ball} & \text{red} \\ \text{Yellow ball} & \text{Yellow} \end{array} \right] \text{ or } \left[ \begin{array}{cc} \textit{Events} & \textit{Reported state} \\ \text{Red or green ball} & \text{Red} \\ \text{Yellow ball} & \text{Yellow} \end{array} \right]$$

a fact that they must take into account when drawing inferences from the observer's reports.

**Aumann's partitions/ Modalities** In the interactive epistemology of Aumann an agent's state of knowledge is represented not by the set of propositions that she knows to be true, but by a partition of the set of possible worlds. The knowledge partition is generated by a function which maps each world  $\omega$  to a set of worlds, interpreted as the set of worlds consistent with what the agent knows at  $\omega$ , satisfying the property that  $\omega \in f(\omega)$  and the property that if  $\omega \in f(\omega')$  then  $f(\omega) = f(\omega')$ . As in the case of observations, with which it bears considerable affinity,<sup>3</sup> the partition representation of knowledge is only adequate when the mechanism producing knowledge is both reliable and complete in the sense that the agent never acquires false beliefs or fails to develop a belief. Not all agents are so epistemically well-endowed however. The credulous agent, for instance, may be represented by the matrix:

$$\left[ \begin{array}{cc} \textit{Events} & \textit{Belief} \\ \text{True testimony that X} & X \\ \text{False testimony that X} & X \end{array} \right]$$

for which the corresponding knowledge matrix will be:

$$\left[ \begin{array}{cc} \textit{Events} & \textit{Knowledge} \\ \text{True testimony that x} & X \\ \text{False testimony that x} & \emptyset \end{array} \right]$$

This matrix cannot be reduced to a knowledge partition for it violates the condition that  $\omega \in f(\omega)$ . But being credulous is not inconsistent with being rational in the decision-theoretic sense and game theory ought to consider strategic interactions in which credulousness is not ruled out, so that, for instance, lying is not automatically excluded as a rational strategy.

The formal similarities between the epistemic attitudes such as belief and knowledge and modalities such as possibility and necessity are well known and so it is to be expected that the points made here regarding the former will carry over to the latter. In the treatment of modalities given by Kripke and others,<sup>4</sup> modal statements have truth conditions that depend not only the possible world in which they are evaluated but also on an 'accessibility' relation between worlds. Letting  $P(\omega)$  denote the set of worlds accessible from world  $\omega$ , it is straightforward to define the proposition-valued random variable that corresponds to a particular accessibility relation; it is the mapping from worlds  $\omega$  to  $P(\omega)$ . When the accessibility relation is reflexive, symmetric and transitive it partitions the space of possible worlds and the corresponding matrix is

<sup>3</sup>Indeed it is plausible that perceptual knowledge was the case that Aumann had in mind.

<sup>4</sup>For instance in Kripke [12].

formally equivalent to that induced by Aumann’s knowledge partitions. As we shall see later, weaker assumptions about the accessibility relation correspond to other kinds or proposition-valued random variables.

**Conditionals** If it is true that there exist prospects that take the form of proposition-valued random variables then one would expect that ordinary language would have the means to refer to them. And it does; namely through conditional sentences. And many theories of conditionals, either explicitly or implicitly, depart from the traditional view that the semantic contents of sentences are propositions, to a large extent motivated by dissatisfaction with the formalisation of ordinary language conditionals as material conditionals.

In general the content of a sentence may be regarded as a function from worlds to the proposition expressed by the sentence at that world. In the simplest cases, what a sentence asserts does not vary with the world in which it is asserted and so this function will determine a constant proposition. But in the case of sentences containing indexicals, for instance, this will not be true and the proposition expressed by a sentence will depend on the context of its utterance. Thus the sentence ‘I am here’ will have a matrix of the following kind:

$$\left[ \begin{array}{cc} \textit{Utterance context} & \textit{Proposition} \\ \text{Uttered by John in London} & \{\omega \in \Omega : \text{In } \omega \text{ John is in London}\} \\ \vdots & \vdots \\ \text{Uttered by Luis in Madrid} & \{\omega \in \Omega : \text{In } \omega \text{ Luis is in Madrid}\} \end{array} \right]$$

Bradley [5] applies this treatment to indicative conditionals, treating them as sentences with variable propositional content. The sentence ‘If  $A$  then  $B$ , and if not  $A$  then  $C$ ’, for instance, is treated as expressing the proposition that  $B$  in all worlds in which  $A$  is true and the proposition that  $C$  in all worlds in which  $A$  is false. It has the following matrix:

$$\left[ \begin{array}{cc} \textit{Event} & \textit{Proposition} \\ A & B \\ A' & C \end{array} \right]$$

A rather different approach is taken by Lewis and Stalnaker in their treatment of counterfactual conditionals. On this view a conditional sentence is a function from either similarity relations (Lewis [13]) or selection functions (Stalnaker [16]) to propositions. Intuitively the similarity relations or selection functions pick out, for each world  $\omega$  and proposition  $A$ , the set of worlds most similar to  $\omega$  in which  $A$  is true. For instance the conditional sentence ‘If  $A$  then  $B$ ’ can be represented by the matrix:

$$\text{If } A \text{ then } B = \left[ \begin{array}{cc} \textit{selection functions} & \textit{propositions} \\ s_1 & \{\omega \in \Omega : s_1(\omega, A) \in B\} \\ \vdots & \vdots \\ s_n & \{\omega \in \Omega : s_n(\omega, A) \in B\} \end{array} \right]$$

So on their account too, conditionals can be viewed as proposition-valued random variables. (Given this is it somewhat surprising that Stalnaker and Lewis

both still assume that agents' attitudes to conditionals depend only on their propositional content and not on the full structure of the random variable determining this content - in effect assuming that there is no uncertainty about the selection function itself.)

### 3 Proposition-valued Random Variables: More Formally

Let  $\langle \Omega, \mathcal{F}, p \rangle$  be a probability space with  $\Omega = \{\omega_1, \omega_2, \dots\}$  a set of possible worlds,  $\mathcal{F} = \wp(\Omega)$  the set of propositions (i.e. sets of worlds), and  $p$  a probability measure on  $\mathcal{F}$ . A *proposition-valued random variable* is a function on  $\Omega$  taking values in  $\mathcal{F}$ . If a random variable assumes only a finite number of values then it is said to be *discrete*. If it assumes the same value at every world it is called *constant*. Constant random variables can very naturally be identified with their constant value. Discrete proposition-valued random variables also have a natural representation (called its Normal Form), namely by a finite matrix of the following kind:

$$f = \begin{bmatrix} A_1 & X_1 \\ \vdots & \vdots \\ A_n & X_n \end{bmatrix}$$

where the  $X_i \in \mathcal{F}$  are the various values taken by the random variable  $f$  and  $A_i = f^{-1}(X_i)$ . Note that it follows immediately that the  $A_i$  are disjoint and exhaustive of  $\Omega$ .

The informal discussion in the preceding section suggests a further set of distinctions. Let us call a proposition-valued random variable:

1. *Refining* iff  $\forall \omega \in \Omega, f(\omega) \subseteq f^{-1}(f(\omega))$
2. *Coarsening* iff  $\forall \omega \in \Omega, f^{-1}(\omega) \subseteq f(\omega)$
3. *Partitional* iff  $\forall \omega \in \Omega, f(\omega) = f^{-1}(\omega)$

Actions and knowledge were modelled by refining random variables, for instance, veridical observation and veridical belief by coarsening ones, and non-veridical observation and belief by random variables that are neither.

If we derive an accessibility relation,  $A$ , on worlds from a proposition-valued random variable,  $f$ , by means of the condition that  $\omega A \omega' \Leftrightarrow \omega' \in f(\omega)$  then these distinctions can be seen to have counterparts in properties of the accessibility relation. In particular whenever  $f$  is:

1. Refining the accessibility relation will be *euclidean* and *transitive*,
2. Coarsening the accessibility relation will be *reflexive* and *symmetric*,
3. Partitional the accessibility relation will be an *equivalence* relation.

The relation between the two suggests that, very informally, coarsening functions may be thought of as expressing possibility and refining partitions as expressing necessity (where these possibilities and necessities may be physical, epistemic, logical, etc.,).

### 3.1 Algebraic Structure

Let  $\Phi = \{f, g, \dots\}$  be the set of all (total) proposition-valued random variables on  $\Omega$ . In particular, corresponding to the supremum  $\Omega$  and infimum  $\emptyset$  of  $\mathcal{F}$ ,  $\Phi$  contains the functions 1 and 0 defined as follows;  $\forall \omega \in \Omega$  :

$$\begin{aligned} 1(\omega) &= \Omega \\ 0(\omega) &= \emptyset \end{aligned}$$

For any proposition-valued random variable  $f$ , let  $\delta(f)$  be the domain of  $f$  and let  $[f]$  be the set of worlds  $\omega \in \Omega$  such that  $\omega \in f(\omega)$ .

We now define a number of function operations on the members of  $\Phi$ , that are the natural counterparts of the Boolean operations on sets. Let  $A \in \Omega$  and  $f, g, h \in \Phi$  be such that  $\delta(f) = \delta(g)$  and  $\delta(f) \cap \delta(h) = \emptyset$ . Then:

1.  $f.g(\omega) = f(\omega) \cap g(\omega)$
2.  $f + g(\omega) = f(\omega) \cup g(\omega)$
3.  $-f(\omega) = \Omega - f(\omega)$

Here we will consider only total functions on  $\Omega$  and so no other function operations will be required.<sup>5</sup>

**Constant Functions** Let  $\mathcal{F}^* = \{\bar{f}, \bar{g}, \dots\} \subseteq \Phi$  be the set of constant proposition-valued random variables. Note that  $\mathcal{F}^*$  is closed under the operations  $+$ ,  $.$  and  $-$ , i.e. if  $f, g \in \mathcal{F}^*$ , then  $f + g, f.g, -f \in \mathcal{F}^*$ , and that the constant value of any random-variable  $\bar{f} \in \mathcal{F}^*$  is just the proposition  $[\bar{f}]$ . The set  $\mathcal{F}^*$  is in fact essentially a copy of the set  $\mathcal{F}$  of propositions; more exactly, there is a one-to-one correspondence between the elements of the Boolean algebra of sets of worlds  $\langle \mathcal{F}, \cup, \cap, ', \Omega, \emptyset \rangle$  and the structure  $\langle \mathcal{F}^*, +, ., -, 1, 0 \rangle$ . To see this let  $\xi$  be a mapping from  $\mathcal{F}^*$  to  $\mathcal{F}$  such that  $\xi(\bar{f}) = [\bar{f}]$  and note that it is structure preserving since:

$$\begin{aligned} \xi(\bar{f}.\bar{g}) &= [\bar{f}] \cap [\bar{g}] = \xi(\bar{f}) \cap \xi(\bar{g}) \\ \xi(\bar{f} + \bar{g}) &= [\bar{f}] \cup [\bar{g}] = \xi(\bar{f}) \cup \xi(\bar{g}) \\ \xi(-\bar{f}) &= \Omega - [\bar{f}] = \xi(\bar{f})' \\ \xi(1) &= \Omega \\ \xi(0) &= \emptyset \end{aligned}$$

It immediately follows that the structure  $\langle \mathcal{F}^*, +, ., -, 1, 0 \rangle$  is a Boolean algebra and that  $\langle \Omega, \mathcal{F}^*, P \rangle$  a probability space with  $P$  a probability on  $\mathcal{F}^*$  defined by:

$$P(\bar{f}) = p([\bar{f}])$$

---

<sup>5</sup>Both Jeffrey [11] and Bradley [5] consider non-Boolean operations that allow functions to be restricted and extended through concatenation.

**Variable Functions** It is obvious that  $\Phi$ , like  $\mathcal{F}^*$ , is closed under the operations  $+$ ,  $\cdot$  and  $-$ , i.e. if  $f, g \in \Phi$ , then  $f + g, f \cdot g, -f \in \Phi$ . But does it form a the basis of a Boolean algebra? In order to answer this question it is useful to consider a partial order,  $\leq$ , on  $\Phi$  defined as follows;  $\forall f, g \in \Phi$  :

$$f \leq g \Leftrightarrow \forall \omega \in \Omega, f(\omega) \subseteq g(\omega)$$

In the case of constant functions this order simply mirrors that between their constant values i.e.  $\forall \bar{f}, \bar{g} \in \mathcal{F}^*$ :

$$\bar{f} \leq \bar{g} \Leftrightarrow \forall \omega \in \Omega, [\bar{f}] \subseteq [\bar{g}]$$

Now note that:

1.  $\leq$  is transitive, reflexive and anti-symmetric; hence a partial order on  $\Phi$ .
2.  $f + g$  and  $f \cdot g$  are respectively the supremum and infimum of  $\{f, g\}$  with respect to  $\leq$ , so that  $(\Phi, \leq)$  is a lattice.
3. This lattice it is bounded above and below by 1 and 0 respectively.
4. Every element  $f \in \Phi$  contains a unique complement; namely  $-f$ .

It follows that the structure  $\langle \Phi, +, \cdot, -, 1, 0 \rangle$  is also a Boolean algebra. It is thus possible to define standard probability and desirability measures on the space of proposition-valued functions.

### 3.2 Measures on random variables

Let  $m$  be any real-valued function on either  $\Omega$  or  $\mathcal{F}$  and let  $f$  be a proposition-valued random-variable. We define a corresponding real-valued random variable  $m \circ f$ , the composition of  $m$  and  $f$ , by:  $\forall \omega \in \Omega$ ,

$$(m \circ f)(\omega) = m(f(\omega))$$

Now relative to the reference probability  $p$  on  $\Omega$ , the expectation of  $m \circ f$ ,  $E(m \circ f)$ , is given by:

$$E(m \circ f) = \int m(f(\omega)) \cdot dp(\omega)$$

When  $f$  is the discrete random variable with Normal Form:

$$\begin{bmatrix} A_1 & X_1 \\ \vdots & \vdots \\ A_n & X_n \end{bmatrix}$$

then  $m \circ f$  is represented by the matrix<sup>6</sup>:

$$m \circ f = \begin{bmatrix} A_1 & m(X_1) \\ \vdots & \vdots \\ A_n & m(X_n) \end{bmatrix}$$

---

<sup>6</sup>This may not be in its Normal Form as the function  $m$  could assign equal numbers to distinct values of  $f$ .



and  $E(m \circ f) = \sum_i m(X_i) \cdot p(A_i)$ .

Another useful notion is that of the conditional expectation of the random variable  $m \circ f$  given an event  $X$  denoted  $E(m \circ f|X)$ . When  $f$  is discrete, this quantity is defined as follows:

$$E(m \circ f|X) = \sum_i m(X_i) \cdot p(A_i|X).$$

i.e. it is the expectation of the random-variable  $m \circ f$  relative to the conditional probability, given  $X$ , of the arguments of  $m \circ f$ . This allow us to define a new real-valued random variable, the conditional expectation of  $m \circ f$  given random variable  $g$ , where  $g$  is any random variable. When  $g$  is discrete with Normal Form:

$$g = \begin{bmatrix} B_1 & Y_1 \\ \vdots & \vdots \\ B_m & Y_m \end{bmatrix}$$

the conditional expectation of  $m \circ f$  given  $g$  is represented by:

$$E(m \circ f|g) = \begin{bmatrix} B_1 & E(m \circ f|g = Y_1) \\ \vdots & \vdots \\ B_m & E(m \circ f|g = Y_m) \end{bmatrix}$$

This random variable too has an expectation: it is given by:

$$\begin{aligned} E(E(m \circ f|g)) &= \sum_j E(m \circ f|g = Y_j) \cdot p(B_j) \\ &= \sum_j (\sum_i m(X_i) \cdot p(A_i|g = Y_j)) \cdot p(B_j) \end{aligned}$$

## 4 Attitudes to Random Variables

If it is the case that a state of affairs represented by a random variable is in existence then it is natural to ask what attitudes agents should take to this fact, and how they should adjust their attitudes to other prospects in the light of it. For instance:

1. You learn that an action has been chosen, but not what the outcome of it is e.g. a friend phones and says that he will be making omelettes and that you should join him for dinner. Decisions must be made at once (do you accept the invitation?) and this requires forming attitudes in the light of the information that the action will be performed.
2. You receive an observation report from a less than totally reliable observer e.g. someone who has some colour-blindness reports that the lights ahead are green. What credence should you attach to the lights actually being green given your knowledge of the mapping between colour states and the observer's reports?
3. In situations of strategic interaction, it matters to agents what others know or believe. Suppose, for instance, that Bob is informed of the belief

function of Anne but not what the state of the world is. In this case Bob does not know for sure what Anne believes, but he should be able to form some expectation about her state of belief based on his probabilities for the various possible states of the world upon which her beliefs depend.

4. In a famous example of van Fraassen [9], a fictional soldier, Judy Benjamin is parachuted into a territory that it divided into two parts - Red and Blue territory - each of which is further divided into Headquarters and Second Company areas. She receives radio message telling her that if she is in Red territory then she is most likely in Headquarters area. Judy's problem is: what should she infer about her location from this message?

If actions, observations, knowledge states and conditionals are not adequately represented by propositions, then we cannot draw on propositional semantics and decision theory in a simple-minded way for answers to these types of questions. So in this section I will attempt to extend the theory of attitudes to propositions to the more general case of proposition-valued functions and then consider the extent to which it solves these problems.

## 4.1 Desirability

In investigating how desirable an agent should regard the prospect of a proposition-valued random variable we can draw on well worked out answer provided by decision theory to the question as applied to the particular case of actions. Let  $u$  be a real-valued utility function on worlds and define the (Jeffrey-Bolker) desirability,  $v$ , of a proposition  $X$  as its conditional expectation of utility relative to the underlying probability measure  $p$ , i.e.  $v(X) = E(u|X)$ .<sup>7</sup> Note that Jeffrey's averaging condition follows from this definition; namely:

**Condition 1 (Averaging)**  $\forall X, Y \in \mathcal{F}$  such that  $X \cap Y = \emptyset$ :

$$v(X \cup Y) = v(X).p(X|X \cup Y) + v(Y).p(Y|X \cup Y)$$

How do we extend desirability to action? The way I propose to do so is to take the value,  $V$ , of an action to be its expected desirability, i.e. the expectation of the conditional expectation of utility, given the acts. Recall that a Savage-style act  $f$  can be represented by a proposition-valued random variable of the form:

$$f = \begin{bmatrix} \text{Events} & \text{Outcomes} \\ A_1 & X_1 \\ \vdots & \vdots \\ A_n & X_n \end{bmatrix}$$

where the  $A_i$  are possible states of the world and the  $X_i$  the consequences of performing the action in those states. Now we define the corresponding real-valued random variable  $v \circ f$  where:

$$v \circ f = \begin{bmatrix} A_1 & v(X_1) \\ \vdots & \vdots \\ A_n & v(X_n) \end{bmatrix}$$

---

<sup>7</sup>See Bolker [2] and [3] and Jeffrey [10].

Finally  $V(f) := E(v \circ f)$  where:

$$E(v \circ f) = \sum_i v(X_i) \cdot p(A_i) = E(u|f) = \sum_i E(u|X_i) \cdot p(A_i)$$

The defined quantity  $V(f)$  is then the proposed measure of the desirability of the act  $f$ . Indeed the conjecture is that the proposal can be generalised to extended to proposition-valued random variables in general, i.e.:

**Proposition 2** *The desirability of a proposition-valued random variable  $f$  is its expected desirability. Formally:  $V(f) := E(v \circ f)$*

The most direct support for this claim, as applied to acts, is the fact that it offers a way of unifying the decision theories of Ramsey, Savage, Jeffrey and others. Its relation to Jeffrey's theory is the most direct, for the latter can be viewed as a restriction of our theory to the special case of constant proposition-valued random variables. This does however raise a question of central importance to this paper, namely whether, given some proposition-valued function  $f \in \Phi$ , there exists a proposition  $X_f \in \Omega$  such that the desirability of  $X_f$  must always equal that of  $f$ , i.e. that  $v(X_f) = E(f_v)$ . The answer is that we cannot generally do so: for most proposition-valued random variables, *it is not the case that there exists a proposition whose desirability always equals that of the random variable*. This fact is proved in the appendix as Theorem 5. The theorem provides a general argument for the importance of studying proposition-valued random variables as forms of information and as prospects towards which agents can take attitudes. In the particular contexts of the study of actions, it also formally vindicates the claim that the introduction of the view of prospects taking the form of proposition-valued random variables strictly enriches the Jeffrey-Bolker framework.

Savage's theory stands in a slightly more complicated relation to ours. In the first place, while for Savage consequences are not logically constrained by the states of the world in which they are realised, it is implicit in the account given here that the values of the actions are a subset of their arguments, i.e. that actions are refining random variables. Consequently ours, unlike Savage's, gives a state-dependent version of expected utility theory in which the desirability of the consequences of the action are conditioned by the states of the world in which they occur. Secondly, while Savage requires that the prospects that serve as consequences be maximally specific with regard to any features of concern to the agent, the account here is much more permissive and allows for rather coarse-grained specification of the consequences. As such they are much more plausibly the kind of prospects that agents can choose amongst, in full knowledge of what they are choosing.

So much for acts. The question of remains whether this treatment of the desirability holds up for interpretations of proposition-valued random variables other than the decision-theoretic one. It seems to me that it applies unproblematically both to a much wider class of causal devices than just actions and to the conditional sentences which describe them. Other applications are less straight-forward. While it is plausible that the desirability of making an observation or of someone being in a particular belief state is measured by the expected desirability of the observations they will make in each circumstance or of the beliefs that they will develop, some care must be taken with the choice

of random variable to represent the belief state or observation. The desirability of a belief or an observation depends on the state of the world in which it is developed or made (normally it being desirable to believe or observe  $X$  only if  $X$  is the case). To capture this formally we must work with the desirabilities, not of the contents of the observations or beliefs, but of the observing or believing of these contents. For instance, consider the belief matrix for a dogmatic agent:

$$f = \begin{bmatrix} \text{Events} & \text{Belief} \\ X & X \\ X' & X \end{bmatrix}$$

Its desirability will be the expected desirability of the proposition  $B(X)$  that the agent believes that  $X$  i.e.  $V(f) = V(X \cap B(X)).p(X) + V(X' \cap B(X)).p(X')$  which will, in normal cases, vary positively with the probability of  $X$ . To put it somewhat differently its desirability is the expected desirability, not of  $f$ , but of the derived proposition-valued random variable  $f'$  given by:

$$f' = \begin{bmatrix} X & B(X) \\ X' & B(X) \end{bmatrix}$$

where it is implicit that  $\omega \in f'(\omega)$ .

The general point is this: when the desirabilities of the prospects under consideration depend on the states in which they are realised, this dependence must be reflected in the associated proposition-valued random variable in a way which permits sensible application of Proposition 2.

## 4.2 Probability

Unlike the case of desirability, we have no default theory of the probability of proposition-valued random variables to turn to for guidance. There are in fact a number of different ways of extending probability measures to random variables, which may serve different purposes. We consider two of them here.

**Indicator Functions** One common way of defining the probability of a proposition  $X$  is as the expectation of its indicator function,  $I_X$ , where the latter is a special type of value function that assigns the value 1 to every world in the proposition  $X$  (i.e. at which  $X$  is true) and 0 elsewhere. Intuitively the indicator function for  $X$  is the value function associated with the mere fact of  $X$  being true. Such a value function underlies the Dutch Book arguments for the probability axioms, for instance. Indeed, we can think of  $I_X$  as being the value function associated with the act  $B$  of betting on the truth of  $X$  when the bet pays out  $\$u^{-1}(1)$  in case  $X$  is true and  $\$u^{-1}(0)$  otherwise (taking  $u$  to be utility measure on dollar amounts). Given that this act has matrix representation:

$$B = \begin{bmatrix} X & \$u^{-1}(1) \\ X' & \$u^{-1}(0) \end{bmatrix}$$

then it follows that  $I_X = u \circ B$ .

This way of characterising the probability of events is very naturally extended to the random variables, by defining the indicator function,  $I_f$ , for

proposition-valued random variable  $f$  to be a real-valued random variable from  $\Omega$  to  $\{0, 1\}$  such that:

$$I_f(\omega) = \begin{bmatrix} 1 & \text{if } \omega \in f(\omega) \\ 0 & \text{if } \omega \notin f(\omega) \end{bmatrix}$$

The probability of  $f$ ,  $P(f)$ , may now be identified with the  $E(I_f)$  i.e.  $P(f) := \int I_f(\omega) dp(\omega) = p([f])$ . Note that  $P$ , so defined, is a genuine probability measure on  $\Phi$ . For  $P(1) = p(\Omega) = 1$ ,  $P(f) = p([f]) \geq 0$  and if  $f.g = 0$  then  $P(f + g) = p([f + g]) = p([f]) + p([g]) = P(f) + P(g)$ .

**Probability Functions** A second strategy for extending probabilities to proposition-valued random variables starts with the real-valued random-variable  $p \circ f$ , associated with  $f$ , where:

$$p \circ f = \begin{bmatrix} A_1 & p(X_1) \\ \vdots & \vdots \\ A_n & p(X_n) \end{bmatrix}$$

The probability of  $f$  is then identified with the expectation of  $p \circ f$ , by defining  $P(f) := E(p \circ f)$ . Then in the discrete case then  $P(f) = \sum_i p(X_i).p(A_i)$ .

So defined,  $P(\cdot)$  is also a genuine probability measure on  $\Phi$  since:

1.  $E(p \circ f) \geq 0$
2.  $E(p \circ 1) = 1$
3.  $E(p \circ (f + g)) = E(p \circ f) + E(p \circ g)$  if  $f.g = 0$

This is proved in the appendix as Theorem 6.

**Discussion** The two strategies lead to different assignments of probabilities to random variables because  $E(p \circ f)$  and  $E(I_f)$  are not generally the same quantity. The quantity  $E(I_f)$  measures the probability that the world is truly as  $f$  says or that  $f$  correctly identifies the actual world, while  $E(p \circ f)$  gives a measure of the average probability of  $f$  or the probability that  $f$  is expected to have. The former measure is useful for interpretations of random variables as proposition-like entities (e.g. indexical sentences or indicative conditionals) and perhaps for other prospects represented by refining random variables. For instance our uncertainty about someone's knowledge state is plausibly measured by the probability of the worlds in which what they believe is true at that world. On the other hand it has little or no application to prospects not represented by refining random variables. In these cases, average probability may do a better job: for instance, our uncertainty about whether someone believes truly is plausibly measured by the expected probability of what he believes, where the latter depends on the state of the world.

The two strategies are similar in one respect: they both are reductionist in the sense of associating the probability of a random-variable with that of a particular proposition or with the average of a set of propositions.<sup>8</sup> Arguably however there is uncertainty associated with  $f$  that is not reducible to the uncertainty of any proposition. We may be sure which event the actual world

<sup>8</sup>Indeed if  $\mathcal{F}$  is atomless it will contain a proposition whose probability is this average probability.

belongs to, but not be sure that  $f$  correctly describes the dependency between this event and the proposition it determines. That is, we can be uncertain not just about which world is actual but also about the relevant dependency between worlds and propositions: whether for instance an action really determines the action specified by  $f$  or whether a set  $X_i$  is really possible/accessible at  $A_i$ . If this is right then the approach taken here to measuring the uncertainty of a proposition-valued random variable is at best only part of the story.

## 5 Conditioning on Random Variables

When we learn a random variable, we do not generally learn that any particular proposition is true. But our expectations regarding what is true will change. Consider for instance the proposition-valued random variable:

$$\begin{bmatrix} A & X \\ A' & Y \end{bmatrix}$$

This random variable does not carry the information that  $X$  is the case, or that  $Y$  is. Nor does it carry the information that  $A$  is true or false. Rather it carries the information that  $X$  is the case if  $A$  is and  $Y$  is the case if  $A$  is not. So to the extent that we regard  $A$  as probable, we should now regard  $X$  as being so, and to the extent that we regard it as not being so, we should regard  $Y$  as being probable rather than  $X$ . More precisely, my new probability for  $X$  after learning this random variable should equal my old probability for  $A$  and my new probability for  $Y$  should equal my old probability for  $A'$ . On the other hand, learning the random variable gives me no reason to change my probability for  $A$ .

To state this both more formally and more generally, let  $p$  be a probability on  $\wp(\Omega)$  and let  $f$  be the discrete proposition-valued random variable represented by:

$$f = \begin{bmatrix} A_1 & X_1 \\ \vdots & \vdots \\ A_n & X_n \end{bmatrix}$$

In the standard Bayesian theory of probability revision, an agent who learns a proposition  $X$  (and nothing more than this), should adopt as her new probabilities of truth her old conditional probabilities of truth, given  $X$ , i.e.,

$$\begin{aligned} p^{*X}(\cdot) &= p(\cdot|X) \\ p^{*X}(\cdot) &= p(\cdot|X) \end{aligned}$$

This theory is not uncontroversial, but I will not question it here. Instead, I will propose a definition for the conditional probability of a proposition given the discrete proposition-valued random variable  $f$  and then consider whether revising your beliefs upon learning that  $f$  by adopting these conditional probabilities as your new unconditional ones is the right thing to do.

Recall that the probability of proposition  $Y$  was defined earlier as the expectation of its indicator function. Its conditional probability given  $X$  may similarly be associated with the conditional expectation of its indicator function, given  $X$ . For then:

$$E(I_Y|X) = \int I_Y(\omega).dp(\omega|X) = p(Y|X)$$

It follows that the conditional probability of any proposition  $X$  given the discrete proposition-valued random variable  $f$  is a random variable  $I_Y|f$  which assigns to each world  $\omega \in A_i$  the conditional expectation of  $I_Y$  given that  $f(\omega) = X_i$ . By definition this expectation is just  $p(Y|f = X_i)$ , so the random variable in question has the following matrix:

$$I_Y|f = \begin{bmatrix} A_1 & p(Y|f = X_1) \\ \vdots & \vdots \\ A_n & p(Y|f = X_n) \end{bmatrix}$$

Now the natural proposal is that an agent who learns that  $f$  is the case should adopt as her new probability for any proposition  $Y$  the expectation of  $I_Y|f$ . More formally:

**Proposition 3 (*Updating on Random Variables*):** *If an agent with prior probabilities  $p$  learns that discrete proposition-valued random variable  $f$  is the case, she should adopt  $p^{*f}(\cdot)$  as her posterior probabilities, where  $\forall Y \in \Omega$ :*

$$p^{*f}(Y) := E(I_Y|f) = \sum_{i=1}^n p(Y|f = X_i) \cdot p(A_i)$$

In many cases of interest to us the expectation of  $Y$  given that  $f(\omega) = X_i$  is just  $p(Y_i|X_i)$ . For instance if  $f$  represents an action with consequence  $X_i$  in event  $A_i$  (where once again  $X_i \subseteq A_i$ ) then the probability of a prospect, given that the world is in a state in which the action yields consequence  $X_i$  (i.e.  $\omega \in A_i$ ), is just the conditional probability of the prospect given that  $X_i$ . So too if  $f$  represents veridical observation or the content of a conditional sentence. In these cases, it follows from Proposition 3 that:

$$p^{*f}(Y) = E(I_Y|f) = \sum_{i=1}^n p(Y|X_i) \cdot p(A_i) \quad (1)$$

From which it follows that:

$$p^{*f}(A_i) = \sum_{i=1}^n p(A_i|X_i) \cdot p(A_i) = p(A_i) \quad (2)$$

$$p^{*f}(X_i) = \sum_{i=1}^n p(X_i|X_i) \cdot p(A_i) = p(A_i) \quad (3)$$

so that updating one's probabilities in this way can be seen as a matter of placing all the probability mass that was on  $A_i - X_i$ , onto  $A_i \cap X_i$ , leaving the total mass on  $A_i$  the same as before.

The first question that needs to be addressed is whether updating on proposition-valued random variables in accordance with Proposition [?], or even with equation 1, can be reduced to classical Bayesian conditioning. The answer is that it cannot generally be so, a fact established by Theorem 8, which shows that even the effects of updating in accordance with equation 1 on a very simple random variable cannot be reduced to classical conditioning on any proposition. This establishes that the informational content, or more exactly the belief-change inducing properties, of proposition-valued random variables are not expressible propositionally.

Although the proposed rule for updating on proposition-valued is not a form of classical Bayesian conditioning it is closely related to one of its near cousins, namely Adams conditioning.<sup>9</sup> Adams conditioning is a rule for belief revision that applies in cases when the conditional probabilities, given some condition, of some set of prospects change without the unconditional probabilities for the conditions themselves changing. The exact relation between Adams conditioning and the proposal given in Proposition 3 is the following: the result of probability revision in accordance with equation 1 is exactly that achieved by Adams conditioning on the constraint that the posterior conditional probabilities of the  $X_i$  given the  $A_i$  should equal one. This is proved in the appendix as Theorem 7.

How does proposition 3 fare as an answer to the problem of how to revise belief on learning a proposition-valued random variable? The merits of Adams' conditioning as a rule for revising on conditionals has already received some discussion in the literature and much of it is relevant here. Douvens and Romeijn [8], for instance, argue that it provides a solution to the problem Judy Benjamin faces in inferring her location. They and others also identify examples in which Adams conditioning works less well.<sup>10</sup> The main conclusion to be drawn from this discussion is, in my opinion, that Adams conditioning on a particular conditional (like classical Bayesian conditioning on an ordinary proposition) is only the right thing to do if a crucial constraint is met, namely that the conditional (or proposition in the Bayesian case) that serves as the conditioning basis represents everything that has been learnt by the agent. Sometimes it is hard to tell whether this is the case, but when it is, then Adams conditioning and hence proposition 3 give the right revision rule for information in the form of a proposition-valued random variable.

However instead of recapitulating the discussion surrounding the particular case of conditionals, I would like to look at the question in the context of the interpretation of random variables as actions, for it turns out that similar lessons apply in this case. Suppose that one were told that an action had been performed, whose description was given by say:

<i>Events</i>	<i>Outcomes</i>
Rain	Go to the cinema
Sun	Go to the beach

then it would, I claim, be quite correct to revise one's beliefs in the manner required by Proposition 3 i.e. one should come to believe in the prospect of a trip to the cinema to extent that one believes that it is going to rain, without revising one's beliefs in the prospect of rain. But in other cases this doesn't seem right. Suppose, for instance, that you know learn that your friend has taken the car to come and see you, an action that might be represented by:

<i>Events</i>	<i>Outcomes</i>
Heavy traffic	Arrives late
Light traffic	Arrives on time

Then knowing that one's friend hates to arrive late, one might reasonably infer that the traffic cannot be too heavy or else he would have chosen to walk instead. Clearly doing so would lead to a violation of Proposition 3. But only on

<sup>9</sup>Adams conditioning is introduced and discussed in [6].

<sup>10</sup>See, for instance, Bradley [6] and Bovens [4].



the assumption that we have correctly represented the information we received about action. And in this case, it would seem that we have not and that his action is better represented by:

<i>Events</i>	<i>Outcomes</i>
Heavy traffic and known to be so	Arrives on time
Light traffic and known to be so	Arrives on time
Heavy traffic and not known to be so	Arrives late
Light traffic and not known to be so	Arrives on time

Once his action is represented in this way, revising in accordance with the proposed rule seems correct i.e. I should revise our beliefs about my friend arriving on time in accordance with my probabilities for the traffic being heavy and him not knowing that it is, but not revise my probabilities for this latter condition.

## 6 Concluding Remarks

The main aim of this paper was to motivate the study of proposition-valued random variables as a general type of prospect with applications to semantics, epistemology, decision theory and philosophical logic. The argument for their significance has two planks: firstly the observation that objects of this kind are already playing a role (though often only implicitly) in theories in these fields and that their use could be productively extended; and secondly the demonstration(s) that the roles played by these objects is not one that can be assumed by propositions (this be carried out formally in Theorems 5 and 8).

A second aim was to contribute to such a study by exploring the characteristics of both proposition-valued random variables and the real-valued random variables that can be derived from them and to apply the fruits of this study to various problems concerning the attitudes that rational agents should take to prospects having the form of a proposition-valued random variable (in some cases building on theory that already exists). In the interests of showing the full scope for application, I have necessarily only given the briefest of sketches of what seem to me to be the most fruitful paths of enquiry. By no stretch of the imagination would I claim that the beginnings made here settle all the interesting questions that can be asked about the role these prospects should play in theoretical and practical enquiry. But I would hope that enough has been done to motivate the interest in the questions themselves.

## 7 Appendix: Proofs

Throughout, if  $X$  and  $Y$  are any two sets, we denote  $X \cap Y$  by  $XY$ .

**Lemma 4** *Let  $X, Y, Z \in \mathcal{F}$  be such that  $Z \neq \emptyset$ . Then  $XZ = YZ$  iff for all pairs of probability and desirability functions  $(p, v)$  on  $\mathcal{F}$  such that  $p(Z) = 1$ , it is the case that  $v(X) = v(Y)$ .*

**Proof.** Note that if  $p(Z) = 1$ , then by the Averaging condition:

$$v(X) = v(XZ).p(Z|X) + v(X - Z).p(Z'|X) = v(XZ) \quad (4)$$

$$v(Y) = v(YZ).p(Z|Y) + v(Y - Z).p(Z'|Y) = v(YZ) \quad (5)$$

Assume that  $XZ = YZ$ . Then it follows immediately that for all pair of probability and desirability functions  $(p, v)$  on  $\mathcal{F}$  it is the case that  $v(XZ) = v(YZ)$ . Hence by equations 4 and 5 above, if  $p(Z) = 1$  then  $v(X) = v(Y)$ .

Now assume that  $XZ \neq YZ$ . Then there are two cases to consider. (i) If  $XZ - Y \neq \emptyset$ , then there exists a pair of probability and desirability functions  $(p^*, v^*)$  on  $\mathcal{F}$  such that  $p^*(Z) = 1$ ,  $p^*(Y'|XZ) > 0$ ,  $p^*(X'|YZ) = 0$ ,  $v^*(XYZ) = 0$  and  $v^*(XZ - Y) > 0$ . But then by equations 4 and 5,  $v^*(X) = v^*(XZ)$  and  $v^*(Y) = v^*(YZ)$ . And by Averaging:

$$\begin{aligned} v^*(XZ) &= v^*(XYZ).p(Y|XZ) + v^*(XZ - Y).p(Y'|XZ) > 0 \\ v^*(YZ) &= v^*(XYZ).p(X|YZ) + v^*(YZ - X).p(X'|YZ) = 0 \end{aligned}$$

Hence  $v(X) \neq v(Y)$ .

(ii) If  $YZ - X \neq \emptyset$ , then there exists a pair of probability and desirability functions  $(p^*, v^*)$  on  $\mathcal{F}$  such that  $p^*(Z) = 1$ ,  $p^*(X'|YZ) > 0$ ,  $p^*(Y'|XZ) = 0$ ,  $v^*(XYZ) = 0$  and  $v^*(YZ - X) > 0$ . But then by equations 4 and 5,  $v^*(X) = v^*(XZ)$  and  $v^*(Y) = v^*(YZ)$ . And by Averaging:

$$\begin{aligned} v^*(XZ) &= v^*(XYZ).p(Y|XZ) + v^*(XZ - Y).p(Y'|XZ) = 0 \\ v^*(YZ) &= v^*(XYZ).p(X|YZ) + v^*(YZ - X).p(X'|YZ) > 0 \end{aligned}$$

Hence again  $v(X) \neq v(Y)$ . ■

**Theorem 5** Let  $\langle \mathcal{F}, \subseteq \rangle$  be a Boolean algebra of sets with  $\mathcal{F}$  a set containing distinct elements  $A$  and  $B$  such that  $\Omega \neq A \neq \emptyset$ . Then there exists no  $X \in \mathcal{F}$  such that for all pairs,  $(p, v)$ , of probability and desirability functions on  $\mathcal{F}$ , it is the case that  $v(X) = v(AB).p(A) + v(A').p(A')$ .

**Proof.** Suppose, contrary to hypothesis, that there exists  $X \in \mathcal{F}$  such that for all pairs  $(p, v)$  of probability and desirability functions on  $\mathcal{F}$ ,  $v(X) = v(AB).p(A) + v(A').p(A')$ . Then if  $p$  is any probability function on  $\mathcal{F}$  such that  $p(A) = 0$  and  $v$  any desirability function consistent with  $p$ , it must be the case that  $v(X) = v(A')$ . Then by Lemma 4,  $X - A = A'$ . Similarly if  $p$  is any probability function on  $\mathcal{F}$  such that  $p(A) = 1$  and  $v$  any desirability function consistent with  $p$ , it must be the case that  $v(X) = v(AB)$ . Then by Lemma 4  $XA = AB$ .

Now if  $XA = AB$  and  $X - A = A'$ , then  $X = AB \cup A'$ . But then, for all pairs  $(p, v)$  of probability and desirability functions on  $\Omega$ :

$$\begin{aligned} v(X) &= v(AB).p(A|AB \cup A') + v(A').p(A'|AB \cup A') \\ &= v(AB).p(A) + v(A').p(A') \end{aligned}$$

by assumption. But then for all probabilities  $p$ ,  $p(A|AB \cup A') = p(A)$  and  $p(A'|AB \cup A') = p(A')$ . But consider probability  $p'$ , such that  $p'(AB) = 0$  and  $p'(A) > 0$  ( $p'$  exists since by assumption  $\Omega \neq A \neq \emptyset$ ). Then  $p'(A'|AB \cup A') = 1 \neq p'(A')$ . ■

**Theorem 6**  $E(p \circ f)$  is a probability measure on  $\Phi$ .

**Proof.** Since  $p$  is a probability measure on  $\mathcal{F}$ , it follows that  $\forall \omega \in \Omega$ ,  $(p \circ f)(\omega) \geq 0$ . By definition,  $\forall \omega \in \Omega$ ,  $1(\omega) = \Omega$ . So  $(p \circ 1)(\omega) = 1$ ,  $\forall \omega \in \Omega$ . Hence

$(p \circ 1)(\omega) = 1$ . Finally if  $f.g = 0$ , then

$$\begin{aligned}
E(p \circ (f + g)) &= \int (p \circ (f + g))(\omega).dp \\
&= \int p((f(\omega) \cup g(\omega)).dp \\
&= \int p(f(\omega)) + p(g(\omega)).dp \\
&= \int p(f(\omega)).dp + \int p(g(\omega)).dp \\
&= E(p \circ f) + E(p \circ g)
\end{aligned}$$

■

**Theorem 7** *Adams conditioning on a redistribution of probability over the partition  $\{A_i X_i\}$  subject to the constraint that  $p^*(X_i|A_i) = 1$ , is equivalent to adopting new probabilities  $p^{*f}(\cdot)$  where  $f$  is the discrete random variable:*

$$\begin{bmatrix} A_1 & X_1 \\ \vdots & \vdots \\ A_n & X_n \end{bmatrix}$$

**Proof.** By Theorem 5 of Bradley [6] Adams conditioning on a redistribution of probability over the partition  $\{A_i X_i\}$  is equivalent to picking a unique probability  $p^*$  such that  $p^*(X_i|A_i) = 1$ ,  $p^*(A_i) = p(A_i)$  and  $p^*(\cdot|A_i X_i) = p(\cdot|A_i X_i)$ . The second condition has already been established for  $p^{*f}$  as equation 2 while the first is an immediate consequence of equation 3. Now for all  $Y \in \Omega$  and  $A_i X_i \in \{A_i X_i\}$

$$\begin{aligned}
p^{*f}(Y|A_i X_i) &= \frac{p^{*f}(A_i X_i)}{p^{*f}(A_i X_i)} \\
&= \frac{\sum_{i=1}^n p(Y A_i X_i|A_i \cap X_i).p(A_i)}{\sum_{i=1}^n p(A_i X_i|A_i \cap X_i).p(A_i)} \\
&= \frac{p(Y|A_i X_i).p(A_i)}{p(A_i)} \\
&= p(Y|A_i X_i)
\end{aligned}$$

■

**Theorem 8** *Let  $\langle \mathcal{F}, \subseteq \rangle$  be a Boolean algebra of sets with  $\mathcal{F}$  a set containing distinct elements  $A$  and  $B$  such that  $\Omega \neq A \neq AB \neq \emptyset$ . Then there exists no  $X \in \mathcal{F}$  such that for all probability functions  $p$  on  $\mathcal{F}$ ,  $p(\cdot|X) = p(\cdot|AB).p(A) + p(\cdot|A').p(A')$ .*

**Proof.** Suppose, contrary to hypothesis, that there exists  $X \in \mathcal{F}$  such that for all probability functions  $p$  on  $\mathcal{F}$ ,  $p(\cdot|X) = p(\cdot|AB).p(A) + p(\cdot|A').p(A')$ . Then in particular, for all  $p$ :

$$p(X|X) = p(X|AB).p(A) + p(X|A').p(A') = 1 \quad (6)$$

$$p(A'|X) = p(A'|AB).p(A) + p(A'|A').p(A') = p(A') \quad (7)$$

But since  $p(A) + p(A') = 1$ , it follows from (5) that it must be the case that  $p(X|AB) = 1 = p(X|A')$ . But from (6):

$$\begin{aligned} \frac{p(A'X)}{p(X)} &= p(A') \\ \Leftrightarrow \frac{p(X|A').p(A')}{p(X)} &= p(A') \\ \Leftrightarrow p(X|A') &= p(X) \end{aligned}$$

Hence  $p(X) = 1$ . But this can only be the case for all  $p$  if  $X = \Omega$ . But this impossible because, for example, for all  $p$  such that  $p(A - B) > 0$ :

$$p(B|AB).p(A) + p(B|A').p(A') = p(A) + p(A'B) > p(B|\Omega) = p(B)$$

■

## References

- [1] Aumann, Robert (1976) "Agreeing to Disagree" *Annals of Statistics* 4(6): 1236–1239
- [2] Bolker, E. (1966) "Functions Resembling Quotients of Measures", *Transactions of the American Mathematical Society* 124: 292-312
- [3] Bolker, E. (1967) "A Simultaneous Axiomatisation of Utility and Subjective Probability", *Philosophy of Science* 34: 333-340
- [4] Bovens, L. (forthcoming) "Judy-Benjamin is a Sleeping-Beauty, modulo Monty Hall", *Analysis*
- [5] Bradley, Richard. 2002. Indicative Conditionals, *Erkenntnis* 56: 345-378, 2002
- [6] Bradley, Richard. 2005. Radical Probabilism and Mental Kinematics. *Philosophy of Science* 72: 342-364
- [7] Bradley, Richard. 2007. The Kinematics of Belief and Desire. *Synthese* 156: 513-535
- [8] Douvens, I. and Romeijn, J-W. Forthcoming. A New Resolution of the Judy-Benjamin Problem, *Mind*
- [9] van Fraassen, B. C. 1981. "A Problem for Relative Information Minimizers in Probability Kinematics," *British Journal for the Philosophy of Science* 32:375–379.
- [10] Jeffrey, R. C. (1983) *The Logic of Decision*, 2nd ed, Chicago, University of Chicago Press
- [11] Jeffrey, R. C. (unpublished manuscript) Proposition-valued Random Variables
- [12] Kripke, S. (1963). Semantical Considerations on Modal Logic. *Acta Philosophica Fennica* 16: 83-94.

- [13] Lewis, D. K. (1973). *Counterfactuals*, Blackwell Publishers
- [14] McGee, V. (1989). "Conditional Probabilities and Compounds of Conditionals." *Philosophical Review* 48: 485-541.
- [15] Ramsey, F. P. (1926) "Truth and Probability" in D. H. Mellor (ed) *Philosophical Papers*, Cambridge University Press, Cambridge New York
- [16] Stalnaker, R. 1968: "A Theory of Conditionals" in *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series, 2. Oxford: Blackwell, pp. 98-112. Reprinted in Jackson, Frank ed. 1991.
- [17] Stalnaker, R. and Jeffrey, R. 1994: "Conditionals as Random Variables", in Eells, E. and Skyrms, B. eds., *Probability and Conditionals*. Cambridge: Cambridge University Press.
- [18] Savage, L. J. (1954/1972) *The Foundations of Statistics*, 2nd ed, Dover, New York