Abstract

Given a set of probabilities on an event space and a new observa-

tion, it is common to form updated probabilities by using rules of

conditioning. Some observations pose constraints that cannot be ad-

dressed by standard conditional probabilities or Jeffrey conditioning.

The principle of maximum entropy claims that it is then appropriate

to form an updated probability assessment by minimizing information

gain consistent with the observation. Opposition to the principle of

maximum entropy leans heavily on a counterexample: the Judy Ben-

jamin problem. The article shows that an intuitive approach, which on

account of independence assumptions should support the opponents,

turns out to support the principle of maximum entropy instead. The

main target of the article is independence assumptions that opponents

improperly apply to give the appearance that the solution provided

by the principle of maximum entropy is implausible.

1 Introduction

Probability kinematics is the field of inquiry asking how we should update

a probability distribution in the light of evidence. If the evidence comes

as an event, it is relatively uncontroversial to use conditional probabilities

(call this standard conditioning). Sometimes, however, the evidence may not

relate the certainty of an event but a reassessment of its uncertainty or its

probabilistic relation to other events (see Jeffrey, 1965, 153ff), expressible

in a shift in expectation (see Hobson, 1971). Bas van Fraassen has come

up with an example from the 1980 comedy film Private Benjamin (see van

Fraassen, 1981), in which Goldie Hawn portrays a Jewish-American woman

(Judy Benjamin) who joins the U.S. Army.

In van Fraassen's interpretation of the movie, Judy Benjamin is on an

assignment and lands in a place where she is not sure of her location. She is

on team Blue. Because of the map, her probability of being in Blue territory

equals the probability of being in Red territory, and being in the Red Second

Company area equals the probability of being in the Red Headquarters area.

Her commanders inform Judy by radio that in case she is in Red territory,

her chance of being in the Red Headquarters area is three times the chance

of being in the Red Second Company area.

At the heart of our investigation are two incompatible but independently

plausible intuitions regarding Judy's choice of updated probabilities for her

1

Figure 1: Judy Benjamin's map. Blue territory (A ) is

3

friendly and does not need to be divided into a Headquar-

ters and a Second Company area.

location (much more in the next section). First we note, however, that there

is no immediately obvious event space in which we can condition on an event

of which we are certain. Grove and Halpern have written an article (1997) on

how to construct such event spaces and then condition on the event that Judy

Benjamin receives the information that she receives from her commanders.

They admit, however, that the construction of such spaces (sometimes called

retrospective conditioning) is an exercise in filling in missing details and

supplying information not contained in the original problem.

If we assume that the attempt fails to define an event on which Judy

Benjamin could condition her probabilities, we are left with two possibilities.

Her new information (it is three times as likely to land in A than to land in

2

A , see figure 1 and the details of the problem in the next section) may mean

1

that we have a redistribution of a complete partition of the probabilities.

This is called Jeffrey conditioning and calls for Jeffrey's rule. Jeffrey's rule

is contested in some circles, but we will for this project accept its validity

in probability kinematics. We will see in what follows that some make the

case that Jeffrey conditioning is the correct way to solve the Judy Benjamin

problem.

2

The third possibility to solve this problem (after standard conditioning

and Jeffrey's rule) is to consult a highly contested procedure to find an

objective updating procedure in case the first two scenarios do not apply:

the principle of maximum entropy (from now on maxent for abbreviation).

maxent can be applied to any situation in which we have a completely

quantified probability distribution and an affine constraint (we will explain

the nature of affine constraints in more detail later). If our new evidence

is the observation of an event (or simply certainty about an event that we

did not have previously), then the event provides an affine constraint and

can be used for updating by means of standard conditioning. If our new ev-

idence is a redistribution of probabilities where we can apply Jeffrey's rule,

then the redistribution provides an affine constraint and can be used for up-

dating by means of Jeffrey's rule. These two possibilities, however, do not

exhaust affine constraints. The Judy Benjamin problem illustrates the third

possibility where the affine constraint only redistributes some groups of prob-

abilities and leaves open the question how this will affect the probabilities

not included in this redistribution.

Advocates of maxent claim that in this case the probabilities should be

adjusted so that they are minimally affected (we make this precise by using

information theory) while at the same time according with the constraint.

Opponents of this view grant that maxent is an important tool of probability

kinematics, but that results of maxent that are difficult to accept (such as

in the Judy Benjamin case) urge us to embrace a more pluralistic, situation-

specific methodology.

Joseph Halpern, for example, writes in Reasoning About Uncertainty that

"there is no escaping the need to understand the details of the application"

(Halpern, 2003, 423) and concludes that maxent is a valuable tool, but

should be used with care (see Grove and Halpern, 1997, 110), explicitly bas-

ing his remark on the counterintuitive behaviour of the Judy Benjamin prob-

lem. Diaconis and Zabell state "that any claims to the effect that maximum-

entropy revision is the only correct route to probability revision should be

viewed with considerable caution" (Diaconis and Zabell, 1982, 829). "Great

caution" (Howson and Franklin, 1994, 456) is also what Colin Howson and

Allan Franklin advise about the more basic claim that the updated prob-

abilities provided by maxent are as like the original probabilities as it is

possible to be given the constraints imposed by the data.

Igor Douven and Jan-Willem Romeijn agree with Richard Bradley that

"even Bayes' rule 'should not be thought of as a universal and mechanical

3

rule of updating, but as a technique to be applied in the right circumstances,

as a tool in what Jeffrey terms the art of judgment.' In the same way, deter-

mining and adapting the weights ... may be an art, or a skill, rather than

a matter of calculation or derivation from more fundamental epistemic prin-

ciples" (Douven and Romeijn, 2009, 16) (for the Bradley quote see Bradley,

2005, 362).

What is lacking in the literature is an explanation by maxent advocates

of the counterintuitive behaviour of the cases repeatedly quoted by their ad-

versaries. This is especially surprising as we are not dealing with an array

of counter-examples but only a handful, the Judy Benjamin problem be-

ing prime among them. In Halpern's textbook, for example, the reasoning

is as follows: maxent is a promising candidate which delivers unique up-

dated probability distributions; but, unfortunately, there is counterintuitive

behaviour in one specific case, the Judy Benjamin case (see Halpern, 2003,

110, 119); therefore, we must abide by the eclectic principle of considering

not only maxent, but also lower and upper probabilities, Dempster-Shafer

belief functions, possibility measures, ranking functions, relative likelihoods,

and so forth. The human inquirer is the final arbiter between these condi-

tionalization methods.

We will undermine the notion that maxent's solution for the Judy Ben-

jamin problem is counterintuitive. The intuition that maxent's solution for

the Judy Benjamin problem violates (call it T1) is based on fallacious inde-

pendence and uniformity assumptions. There is another powerful intuition

(call it T2) in direct contradiction to T1 which maxent obeys. Therefore,

Halpern does not give us sufficient grounds for the eclecticism advocated

throughout his book. We will show that another intuitive approach, the pow-

erset approach, lends significant support to the solution provided by max-

ent for the Judy Benjamin problem, especially in comparison to intuition

T1, many of whose independence and uniformity assumptions it shares.

We have no proof that maxent is the only rationally defensible objective

method to update probabilities given an affine constraint. The literature out-

lines many of the 'nice properties' (so-called by an anonymous reviewer) of

maxent: It seamlessly generalizes standard conditioning and Jeffrey's rule

where they are applicable (see Caticha and Giffin, 2006). It is essential to

the entropy concentration phenomenon described in Jaynes' standard work

Probability Theory: the Logic of Science, which contains other arguments in

favour of maxent, some of which you may recognize by family resemblance in

the rest of this paper. Entropy concentration refers to the unique property

4

of the maxent solution to have other distributions which obey the affine

constraint cluster around it. Shore and Johnston have shown that under

certain rationality assumptions maxent is the unique solution to problems

of probability update (see Shore and Johnson, 1980). When used to make

predictions whose quality is measured by a logarithmic score functions, pos-

terior probabilities provided by maxent result in minimax optimal decisions

(see Tops[U+00F8]e, 1979, Walley, 1991, and Gr"unwald, 2000). Under a logarithmic

scoring rule these posterior probabilities are in some sense optimal.

Despite all these nice properties, we want the reader to follow us in a more

simple line of argument. When new evidence is provided to us, it is rational

to adjust our beliefs minimally in light of it. We do not want to draw out

more information from the new evidence than necessary. There are numerous

problems here that need addressing. What do we mean by rationality? What

are the semantics of the word 'minimal'? What are the formal properties of

such posterior probabilities? Are they unique? Are they compatible with

other intuitive methods of updating? Are there counter-intuitive examples

that would encourage us to give up on this line of thought rather than live

with its consequences? Given persuasive answers to these questions, however,

maxent cuts a good figure as a first pass to provide objective solutions to

these types of problems. The burden on opponents who deny that there are

such objective solutions exist grows heavy.

One particular emphasis in this paper is that the reasoning of the oppo-

nents of maxent in the Judy Benjamin case is flawed because they make

independence assumptions that on closer inspection do not hold up. We

provide a number of scenarios which easily pass the information given in the

problem which violate these independence assumptions. That does not mean

that the information given in the problem suggests these scenarios, it only

means that we are not entitled to make independence assumptions. That, in

turn, does not privilege the maxent solution, although maxent does not

lean on independence assumptions that other solutions illegitimately make.

For maxent confronts us with a much stronger claim than merely providing

a passable or useful solution to the Judy Benjamin problem: it lays claim to

much greater generality and, abhorred by many formal epistemologists, to

objectivity. These claims must be motivated elsewhere---we are only show-

ing that opponents cannot claim an easy victory by pulling out old Judy

Benjamin, although this is still widely practised.

Dealing in stereotypes for a moment, there is a long-standing disagree-

ment between philosophers on the one hand and physicists on the other hand

5

whether (the philosophers) updating probabilities is irreducibly accompanied

by thoughtful deliberation choosing between approaches depending on indi-

vidual problems, or (the physicists) problems are ill-posed if they do not

contain the information necessary to let a non-arbitrary, objective process

arrive at a unique updated probability distribution. In the literature, Judy

Benjamin serves as an example to defend in favour of the philosophers what

I shall call the full employment theorem of probability kinematics.

The full employment theorem of probability kinematics claims that max-

ent is only one of many different strategies to update probabilities. In order

to decide which strategy is the most appropriate for your problem you need

a resident formal epistemologist to do the thinking and weigh the intuitions

for you. For a fee, of course. Thus formal epistemologists will always be

fully employed. (E.T. Jaynes makes similar observations when he derisively

talks about the statistician-client relationship as one between a doctor and

his patient, see Jaynes and Bretthorst, 1998, 492 and 506.) There is an anal-

ogous full employment theory in computer science about writing computer

programs which has been formally proven to be true. Our contention is that

no such proof is forthcoming in probability kinematics. The case rests in a

significant measure (see Halpern's book) on a counterexample to maxent,

the Judy Benjamin problem. We show in this article that our intuitions are

initially misguided about the Judy Benjamin problem because we make in-

dependence and uniformity assumptions that on closer examination are not

tenable.

2 Two Intuitions

There are two pieces of information relevant to Judy Benjamin when she

decides on her updated probability assignment. We will call them (MAP)

and (HDQ). As in figure 1, A is the Red Second Company area, A is

1 2

the Red Headquarters area, A is Blue territory. Judy presumably wants

3

to be in Blue territory, but if she is in Red territory, she would prefer their

Second Company area (where enemy soldiers are not as well-trained as in

the Headquarters area).

(MAP) Judy has no idea where she is. She is on team Blue. Because of the

map, her probability of being in Blue territory equals the probability

of being in Red territory, and being in the Red Second Company area

equals the probability of being in the Red Headquarters area.

6

(HDQ) Her commanders inform Judy that in case she is in Red territory, her

chance of being in their Headquarters area is three times the chance of

being in their Second Company area.

In formal terms (sloppily writing A for the event of Judy being in A ),

i i

2 [U+00B7]P(A ) = 2 [U+00B7]P(A ) = P(A ) (MAP)

1 2 3

3

q = P(A |A [U+222A]A ) = (HDQ)

2 1 2 4

(HDQ) is partial information because in contrast to the kind of evidence we

are used to in Bayes' formula (such as 'an even number was rolled'), and to

the kind of evidence needed for Jeffrey's rule (where a partition of the whole

event space and its probability redistribution is required, not only A [U+222A]A , but

1 2

see here the objections in Douven and Romeijn, 2009), the scenario suggests

that Bayesian conditionalization and Jeffrey's rule are inapplicable. We are

interested in the most defensible updated probability assignment(s) and will

express them in the form of a normalized odds vector (q ,q ,q ), following

1 2 3

van Fraassen (1981). q is the updated probability Q(A ) that Judy Benjamin

i i

is in A . Let P be the probability distribution prior to the new observation

i

and p the individual 'prior' probabilities. These probabilities are not to be

i

confused with prior probabilities that precede any kind of information. In the

spirit of probability update, or probability kinematics, we will for the rest of

the article refer to prior probabilities as probabilities prior to an observation

and the subsequent update. The q sum to 1 (this differs from van Fraassen's

i

canonical odds vector, which is proportional to the normalized odds vector

but has 1 as its first element). We define

q

t =

1 -q

t is the factor by which (HDQ) indicates that Judy's chance of being in A

2

is greater than being in A . In Judy's particular case, t = 3 and q = 0.75.

1

Van Fraassen found out with various audiences that they have the following

intuition:

T1 (HDQ) does not refer to Blue territory and should not affect P(A ):

3

q = p (= 0.50).

3 3

There is another, conflicting intuition (due to Peter Williams via personal

communication with van Fraassen, see van Fraassen, 1981, 379):

7

T2 If the value of q approaches 1 (in other words, t approaches infinity)

then q should approach 2/3 as the problem reduces to one of ordinary

3

conditioning. (HDQ) would turn into 'if you are in Red territory you are

almost certainly in the Red Headquarters area.' Considering (MAP),

q should approach 2/3. Continuity considerations pose a contradiction

3

to T1. (These considerations are strong enough that Luc Bovens uses

them as an assumption to solve Adam Elga's Sleeping Beauty problem

by parity of reasoning in Bovens, 2010.)

To parse these conflicting intuitions, we will introduce several methods to

provide G, the function that maps q to the appropriate normalized updated

odds vector (q ,q ,q ). The first method is extremely simple and accords

1 2 3

with intuition T1: G (q) = (0.5(1 -q),0.5q,0.5). In Judy's particular case

ind

with t = 3 the normalized odds vector is (ind stands for independent):

G (0.75) = (0.125,0.375,0.500)

ind

Both Grove and Halpern (1997) as well as Douven and Romeijn (2009) make

a case for this distribution. Grove and Halpern use standard conditioning on

the event of the message being transmitted to Judy. Douven and Romeijn

use Jeffrey's rule (because they believe that T1 is in this case so strong that

Q(A ) = P(A ) is as much of a constraint as (MAP) and (HDQ), yielding

3 3

a Jeffrey partition). T1, however, conflicts with the symmetry requirements

outlined in van Fraassen et. al. (1986).

Van Fraassen introduces various updating methods which do not con-

flict with those symmetry requirements, the most notable of which is max-

ent. Shore and Johnson have already shown that, given certain assumptions

(which have been heavily criticized, however, e.g. Uffink, 1996), maxent

produces the unique updated probability assignment according with these

assumptions. The minimum information discrimination theorem of Kullback

and Leibler (see for example Csisz[U+00B4]ar, 1967, section 3) demonstrates how

Shannon's entropy and the Kullback-Leibler Divergence formula can provide

the least informative updated probability assignment (with reference to the

prior probability assignment) obeying the constraint posed by the evidence.

The idea is to define a space of probability distributions, make sure that the

constraint identifies a closed, convex subset in this space, and then determine

which of the distributions in the closed, convex subset is least distant from

the prior probability distribution in terms of information (using the mini-

mum information discrimination theorem). It is necessary for the uniqueness

8

of this least distant distribution that the subset be closed and convex (in

other words, that the constraint be affine, see Csisz[U+00B4]ar, 1967).

For Judy Benjamin, maxent suggests the following normalized odds vec-

tor:

G (0.75) [U+2248] (0.117,0.350,0.533) (1)

max

The updated probability of being on Blue territory (A ) has increased from

3

50% to approximately 53%. Grove and Halpern find this result "highly coun-

terintuitive" (Grove and Halpern, 1997, 2). Van Fraassen summarizes the

worry:

It is hard not to speculate that the dangerous implications of being in the

enemy's Headquarters area are causing Judy Benjamin to indulge in wishful

thinking, her indulgence becoming stronger as her conditional estimate of

the danger increases. (Van Fraassen, 1981, 379)

There are two ways in which we can arrive at result (1).

We may use Jaynes' constraint rule and find the updated probability dis-

tribution that is both least informative with respect to Shannon's entropy

and in accordance with the constraint (using Dempster's Rule of Combina-

tion, which together with the constraint rule is equivalent to the principle of

minimum cross-entropy, see Cover and Thomas, 2006, 409, exercise 12.2.).

Alternatively, if circumstances are favourable (as they are in Judy Benjamin's

case), we may use the Kullback-Leibler Divergence and differentiate it to ob-

tain where it is minimal.

Figures 2 and 3 show in diagram form the distribution of (q ,q ,q ) de-

1 2 3

pending on the value of q (between 0 and 1), respectively following intuition

T1 and maxent. Notice that in accordance with intuition T2, maxent

provides a result where q [U+2192] 2/3 for q approaching 0 or 1.

3

3 Epistemic Entrenchment and Coarsening at

Random

Even though T1 is an understandably strong intuition, it does not take into

account that the information given to Judy by her commanders may be

dependent on whether she is in Blue or in Red territory. To underline this

objection to intuition T1 we want to consider three scenarios, any of which

may form the basis of the partial information provided by her commanders.

9

Figure 2: Judy Benjamin's updated proba-

bility assignment according to intuition T1.

0 < q < 1 forms the horizontal axis, the

vertical axis shows the updated probability

distribution (or the normalized odds vector)

(q ,q ,q ). The vertical line at q = 0.75 shows

1 2 3

the specific updated probability distribution

G (0.75) for the Judy Benjamin problem.

ind

I Judy is dropped off by a pilot who flips two coins. If the first coin lands

H, then Judy is dropped off in Blue territory, otherwise in Red territory.

If the second coin lands H, she is dropped off in the Headquarters area,

otherwise in the Second Company area. Judy's commanders find out

that the second coin is biased q : 1-q toward H with q = 0.75. The nor-

malized odds vector is G (0.75) = (0.125,0.375,0.500) and agrees with

I

T1, because the choice of Blue or Red is completely independent from

the choice of the Red Headquarters area or the Red Second Company

area.

II The pilot randomly lands in any of the four quadrants and rolls a die.

10

Figure 3: Judy Benjamin's updated probabil-

ity assignment using maxent. 0 < q < 1

forms the horizontal axis, the vertical axis

shows the updated probability distribution (or

the normalized odds vector) (q ,q ,q ). The

1 2 3

vertical line at q = 0.75 shows the specific up-

dated probability distribution G (0.75) for

max

the Judy Benjamin problem.

If she rolls an even number, she drops off Judy. If not, she takes her to

another (or the same, the choice happens with replacement) randomly

selected quadrant to repeat the procedure. Judy's commanders find

out, however, that for A , the pilot requires a six to drop off Judy, not

1

just an even number. The normalized odds vector in this scenario is

G (0.75) = (0.1,0.3,0.6) and does not accord with T1.

II

III Judy's commanders have divided the map into 24 congruent rectangles,

A into twelve, and A and A into six rectangles each (see figures 4 and

3 1 2

5). They have information that the only subsets of the 24 rectangles

in which Judy Benjamin may be located are such that they contain

11

three times as many A rectangles than A rectangles. The normalized

2 1

odds vector in this scenario is G (0.75) [U+2248] (.108,.324,.568) (evaluating

III

almost 17 million subsets).

I--III demonstrate the contrast between scenarios when independence is true

and when it is not. Douven and Romeijn's capital mistake in their pa-

per is that they assume that the Judy Benjamin problem is analogous to

their example of Sarah and the sundowners at the Westcliff (see Douven and

Romeijn, 2009, 7). Sarah, however, knows that whether it rains or not is in-

dependent of her activity the next night, whereas in Judy Benjamin we have

no evidence of such independence, as scenario II makes clear. This is not to

say that scenario II is the scenario that pertains in Judy Benjamin's case. It

only says that there is no natural assumption in Judy Benjamin's case that

the probabilities are independent of each other in light of the new evidence,

for scenario II is perfectly natural (whether it is true or not is a completely

different question) and reveals how dependence can naturally flow from the

information that Judy Benjamin receives.

Douven and Romeijn's strong independence claim relying on intuition T1

leads them to apply Jeffrey's rule to the Judy Benjamin problem with the

additional constraint q = p . They claim that in most cases "the learning

3 3

of a conditional is or would be irrelevant to one's degree of belief for the

conditional's antecedent ... the learning of the relevant conditional should

intuitively leave the probability of the antecedent unaltered" (Douven and

Romeijn, 2009, 9).

This, according to Douven and Romeijn, is the usual epistemic entrench-

ment and applies in full force to the Judy Benjamin problem. They give

an example where the epistemic entrenchment could go the other way and

leave the consequent rather than the antecedent unaltered (Kate and Henry,

see Douven and Romeijn, 2009, 13). The idea of epistemic entrenchment is

at odds with maxent and seems to imply just what the full employment

theorem claims: judgments so framed "will depend on the judgmental skills

of the agent, typically acquired not in the inductive logic class but by subject

specific training" (Bradley, 2005, 349). To pursue the relation between the

semantics of conditionals, maxent, and the full employment theorem would

take us too far afield at present and shall be undertaken elsewhere. For

the Judy Benjamin problem, it is not clear why Douven and Romeijn think

that the way the problem is posed implies a strong epistemic entrenchment

for Adams conditioning (Adams conditioning is the kind of conditioning that

12

will leave the antecedent alone). Scenarios II-III provide realistic alternatives

where Adams conditioning is inappropriate.

Judy Benjamin may also receive (HDQ) because her informers have found

out that Red Headquarters troops have occupied the entire Blue territory

(q = 3p ,q = p ,q = 0, the epistemic entrenchment is with respect to q );

1 1 2 2 3 2

because they have found out that Blue troops have occupied two-thirds of

the Red Second Company area (q = p ,q = (1/3)p ,q = (4/3)p , the epis-

1 1 2 2 3 3

temic entrenchment is with respect to q ); or because they have found out

1

that Red Headquarters troops have taken over half of the Red Second Com-

pany area (q = (1/2)p ,q = (3/2)p ,q = p , the epistemic entrenchment is

1 1 2 2 3 3

with respect to q and what Douven and Romeijn take to be an assumption

3

in the wording of the problem). There is nothing in the problem that sup-

ports Douven and Romeijn's narrowing of the options. The table reiterates

these options, with the third, shaded line representing intuition (T1) and the

epistemic entrenchment defended by Douven and Romeijn.

Epistemic entrenchment q q q

1 2 3

with respect to A 1/4 3/4 0

1

with respect to A 1/12 1/4 2/3

2

with respect to A 1/8 3/8 1/2

3

Another at first blush forceful argument that maxent's solution for the Judy

Benjamin problem is counterintuitive has to do with coarsening at random,

or CAR for short. It is spelled out in Gr"unwald and Halpern (2003), where

the authors see a parallel between the Judy Benjamin problem and Martin

Gardner's Three Prisoners problem (see Gardner, 1959, 180f). In the Three

Prisoners problem, three men (A, B, and C) are under sentence of death

when the governor decides to pardon one of them. The warden of the prison

knows which of the three men is pardoned, but none of the men do. In a

private conversation, A says to the warden, Tell me the name of one of the

others who will be executed---it will not give anything away whether I will be

executed or not. The warden agrees and tells A that B will be executed. For

the puzzling consequences, see the wealth of literature on the Three Prisoners

problem or the Monty Hall problem.

According to Gr"unwald and Halpern, for problems of this kind (Judy

Benjamin, Three Prisoners, Monty Hall) there are naive and sophisticated

spaces to which we can apply probability updates. If A uses the naive space,

for example, he comes to the following conclusion: of the three possibilities

13

that (A,B), (A,C), or (B,C) are executed, the warden's information excludes

(A,C). (A,B) and (B,C) are left over, and because A has no information

about which one of these is true his chance of not being executed is 0.5. His

chance of survival has increased from one third to one half.

Gr"unwald and Halpern show, correctly, that the application of the naive

space is illegitimate because the CAR condition does not hold. More gener-

ally, Gr"unwald and Halpern show that updating on the naive space rather

than the sophisticated space is legitimate for event type observations always

when the set of observations is pairwise disjoint or, when the events are ar-

bitrary, only when the CAR condition holds. For Jeffrey type observations,

there is a generalized CAR condition which applies likewise. For affine con-

straints on which we cannot use Jeffrey conditioning (or, a fortiori, standard

conditioning) maxent "essentially never gives the right results" (Gr"unwald

and Halpern, 2003, 243).

Gr"unwald and Halpern conclude that "working with the naive space,

while an attractive approach, is likely to give highly misleading answers"

(246), especially in the application of maxent to naive spaces as in the

Judy Benjamin case "where applying [maxent] leads to paradoxical, highly

counterintuitive results" (245). For the Three Prisoners problem, Jaynes'

constraint rule would supposedly proceed as follows: the vector of prior prob-

abilities for (A,B), (A,C), and (B,C) is (1/3,1/3,1/3). The constraint is that

the probability of (A,C) is zero, and a simple application of the constraint

rule yields (1/2,0,1/2) for the vector of updated probabilities. The CAR

condition for the naive space does not hold, therefore the result is mislead-

ing.

By analogy, using the constraint rule on the naive space for the Judy Ben-

jamin problem yields (0.117,0.350,0.533), but as the CAR condition fails in

even the simplest settings for affine constraints ("CAR is (roughly speaking)

guaranteed not to hold except in 'degenerate' situations" (251), emphasis

in the original), it certainly fails for the Judy Benjamin problem, for which

constructing a sophisticated space is complicated (see Grove and Halpern,

1997, where the authors attempt such a construction by retrospective condi-

tioning).

The analogy, however, is misguided. The constraint rule has been for-

mally shown to generalize Jeffrey conditioning, which in turn has been shown

to generalize standard conditioning (the authors admit as much in Gr"unwald

and Halpern, 2003, 262). We can solve both the Monty Hall problem and

the Three Prisoners problem by standard conditioning, not using the naive

14

space, but simply using the correct space for the probability update. For the

Three Prisoners problem, for example, the warden will say either 'B' or 'C' in

response to A's inquiry. Because A has no information that would privilege

either answer the probability that the warden says 'B' and the probability

that the warden says 'C' equal each other and therefore equal 0.5. Here is

the difference between using the naive space and using the correct space, but

either way using standard conditional probabilities:

P('A is pardoned'|'B will be executed') =

P('A is pardoned') 1

= (incorrect)

P('A is pardoned') + P('C is pardoned') 2

P('A is pardoned'|'warden says B will be executed') =

P('A is pardoned' and 'warden says B will be executed') 1/6 1

= = (correct)

P('warden says B will be executed') 1/2 3

Why is the first equation incorrect and the second one correct? Information

theory gives us the right answer: in the first equation, we are conditioning on

a watered down version of the evidence (watered down in a way that distorts

the probabilities because we are not 'coarsening at random'). 'Warden says B

will be executed' is sufficient but not necessary for 'B will be executed.' The

former proposition is more informative than the latter proposition (its prob-

ability is lower). Conditioning on the latter proposition leaves out relevant

information contained in the wording of the problem.

Because maxent always agrees with standard conditioning, maxent

gives the correct result for the Three Prisoners problem. For the Judy Ben-

jamin problem, there is no defensible sophisticated space and no watering

down of the evidence in what Gr"unwald and Halpern call the 'naive' space.

The analogy between the Three Prisoners problem and the Judy Benjamin

problem as it is set up by Gr"unwald and Halpern fails. A successful criti-

cism would be directed at the construction of the 'naive' space: this is what

we just accomplished for the Three Prisoners problem. There is no parallel

procedure for the Judy Benjamin problem. The 'naive' space is all we have,

and maxent is the appropriate tool to deal with this lack of information.

15

4 The Powerset Approach

In this section, we will focus on scenario III and consider what happens

when the grain of the partition becomes finer. We call this the powerset

approach. Two remarks are in order: First, the powerset approach has little

independent appeal. The reason behind using maxent is that we want our

evidence to have just the right influence on our updated probabilities, i.e.

neither over-inform nor under-inform. There is no corresponding reason why

we should update our probabilities using the powerset approach.

Second, what the powerset approach does is lend support to another

approach. In this task, it is persuasive because it tells us what would happen

if we were to divide the event space into infinitesimally small, uniformly

weighed, and independent 'atomic' bits of information. It would be especially

interesting if the powerset approach did not support the independence and

uniformity assumptions of intuition T1, because both of these features are

strongly represented in the powerset approach. On its own the powerset

approach is just what Gr"unwald and Halpern call a naive space, for which

CAR does not hold. Hence the powerset approach will not give us a precise

solution for the problem, although it may with some plausibility guide us in

the right direction---especially if despite all its independence and uniformity

assumptions it significantly disagrees with intuition T1.

Let's assume a partition of the Blue and Red territories into sets of equal

measure (this is the division into rectangles of scenario III). (MAP) dictates

that the number of sets covering A equals the number of sets covering A [U+222A]

3 1

A . Initially, any subset of this partition is a candidate for Judy Benjamin

2

to consider. The constraint imposed by (HDQ) is that now we only consider

subsets for which there are three times as many partition sets (or rectangles,

although we are not necessarily limiting ourselves to rectangles) in A than

2

there are in A . In figures 4 and 5 there are diagrams of two subsets. One

1

of them (figure 4) is not a candidate, the other one (figure 5) is.

Let X be the random variable that corresponds to the ratio of the number

of partition elements (rectangles) that are in A and the total number of

3

partition elements (rectangles) for a randomly chosen candidate. We would

now anticipate that the expectation of X (which we will call EX) gives us an

indication of the updated probability that Judy is in A (so EX [U+2248] q ). The

3 3

powerset approach is often superior to the uniformity approach (Grove and

Halpern use uniformity, with all the necessary qualifications): if you have

played Monopoly, you will know that the frequencies for rolling a 2, a 7, or

16

Figure 4: This choice of rectangles is not a

candidate because the number of rectangles in

A is not a t-multiple of the number of rectan-

2

gles in A , here with s = 2,t = 3 as in scenario

1

III.

a 10 with two dice tend to conform more closely to the binomial distribution

(based on a powerset approach) rather than to the uniform distribution with

P(rolling i) = 1/11 for i = 2,...,12.

Calculations not provided here give us G and a graph of the normal-

pws

ized odds vector (see figure 6), a bit bumpy around the middle because the

t-values are discrete and farther apart in the middle, as t = q/(1-q). Com-

paring the graphs of the normalized odds vector under Grove and Halpern's

uniformity approach (G ), Jaynes' maxent approach (G ), and the pow-

ind max

erset approach suggested in this paper (G ), it is clear that the powerset

pws

approach supports maxent.

Going through the calculations, it seems at many places that the pow-

erset approach should give its support to Grove and Halpern's uniformity

approach in keeping with intuition T1. It is unexpected to find out that in

the mathematical analysis the parameters converge to a non-trivial factor

and do not tend to negative or positive infinity, enabling a graph of the nor-

malized odds vector that is not of the simple nature of the graph suggested

by Grove and Halpern. Most surprisingly, the powerset approach, prima facie

unrelated to an approach using information, supports the idea that a set of

17

Figure 5: This choice of rectangles is a candi-

date because the number of rectangles in A

2

is a t-multiple of the number of rectangles in

A , here with s = 2,t = 3 as in scenario III.

1

events about which nothing is known (such as A ) gains in probability in the

3

updated probability distribution compared to the set of events about which

something is known (such as A and A ), even if it is only partial informa-

1 2

tion. Unless independence is specified, as in Sarah and sundowners at the

Westcliff, the area of ignorance gains compared to the area of knowledge.

We now have several ways to characterize Judy's updated probabilities

and updated probabilities following upon partial information in general. Only

one of them, the uniformity approach, violates van Fraassen, Hughes, and

Harman's five symmetry requirements in Van Fraassen, 1986 and intuition

T2. The uniformity approach, however, is the only one that satisfies intuition

T1, an intuition which most people have when they first hear the story. Two

arguments attenuate the position of the uniformity approach in comparison

with the others.

First, T1 rests on an independence assumption which is not reflected in

the problem. Although there is no indication that what Judy's commanders

tell her is in any way dependent on her probability of being in Blue territory,

it is not excluded either (see scenarios I--III earlier in this paper). maxent

takes this uncertainty into consideration. Second, when we investigate the

problem using the powerset approach it turns out that a division into equally

18

Figure 6: Judy Benjamin's updated probabil-

ity assignment according to the powerset ap-

proach. 0 < q < 1 forms the horizontal axis,

the vertical axis shows the updated probabil-

ity distribution (or the normalized odds vec-

tor) (q ,q ,q ). The vertical line at q = 0.75

1 2 3

shows the specific updated probability distri-

bution G for the Judy Benjamin problem.

pws

probable, independent, and increasingly fine bits of information supports not

intuition T1 but rather intuition T2. maxent, for now, is vindicated. We

need to look for full employment not by cleverly manipulating prior proba-

bilities, but by making fresh observations, designing better experiments, and

partitioning the theory space more finely.

Bovens, L., 2010. Judy Benjamin is a Sleeping Beauty. Analysis, 70(1):23--26.

Bradley, R., 2005. Radical Probabilism and Bayesian Conditioning. Philos-

ophy of Science, 72(2):342--364.

Caticha, A. and Giffin, A., 2006. Updating Probabilities. In MaxEnt 2006,

19

the 26th International Workshop on Bayesian Inference and Maximum

Entropy Methods.

Cover, T. and Thomas, J., 2006. Elements of Information Theory, volume 6.

Wiley, Hoboken, NJ.

Csisz[U+00B4]ar, I., 1967. Information-Type Measures of Difference of Probability

Distributions and Indirect Observations. Studia Scientiarum Mathemati-

carum Hungarica, 2:299--318.

Diaconis, P. and Zabell, S., 1982. Updating Subjective Probability. Journal

of the American Statistical Association, pages 822--830.

Douven, I. and Romeijn, J., 2009. A New Resolution of the Judy Benjamin

Problem. CPNSS Working Paper, 5(7):1--22.

Gardner, M., 1959. Mathematical Games. Scientific American, 201(4):174--

182.

Grove, A. and Halpern, J., 1997. Probability Update: Conditioning Vs.

Cross-Entropy. In Proceedings of the Thirteenth Conference on Uncertainty

in Artificial Intelligence. Citeseer, Providence, Rhode Island.

Gr"unwald, P., 2000. Maximum Entropy and the Glasses You Are Looking

Through. In Proceedings of the Sixteenth conference on Uncertainty in

artificial intelligence, pages 238--246. Morgan Kaufmann Publishers.

Gr"unwald, P. and Halpern, J., 2003. Updating Probabilities. Journal of

Artificial Intelligence Research, 19:243--278.

Halpern, J. Y., 2003. Reasoning About Uncertainty. MIT Press, Cambridge,

MA.

Hobson, A., 1971. Concepts in Statistical Mechanics. Gordon and Beach,

New York, NY.

Howson, C. and Franklin, A., 1994. Bayesian Conditionalization and Prob-

ability Kinematics. The British Journal for the Philosophy of Science,

45(2):451--466.

Jaynes, E. and Bretthorst, G., 1998. Probability Theory: the Logic of Science.

Cambridge University Press, Cambridge, UK.

20

Jeffrey, R., 1965. The Logic of Decision. McGraw-Hill, New York, NY.

Shore, J. and Johnson, R., 1980. Axiomatic Derivation of the Principle of

Maximum Entropy and the Principle of Minimum Cross-Entropy. IEEE

Transactions on Information Theory, 26(1):26--37.

Tops[U+00F8]e, F., 1979. Information-Theoretical Optimization Techniques. Kyber-

netika, 15(1):8--27.

Uffink, J., 1996. The Constraint Rule of the Maximum Entropy Principle.

Studies In History and Philosophy of Science Part B: Studies In History

and Philosophy of Modern Physics, 27(1):47--79.

Van Fraassen, B., 1981. A Problem for Relative Information Minimizers in

Probability Kinematics. The British Journal for the Philosophy of Science,

32(4):375--379.

Van Fraassen, B., 1986. A Problem for Relative Information Minimizers,

Continued. The British Journal for the Philosophy of Science, 37(4):453--

463.

Walley, P., 1991. Statistical Reasoning with Imprecise Probabilities. Chapman

and Hall London.

21