James M. Joyce
C. H. Langford Collegiate Professor of Philosophy

Email: jjoyce@umich.edu
Phone: 734-330-6849

April 28, 2016

Report on *Information Theory and Partial Belief Reasoning* by Stefan Lukits

It was a pleasure to read Stefan Lukits's dissertation. In addition to being clearly written, it combines remarkable philosophical creativity with a deep understanding of controversies in Bayesian statistics as well as an uncommonly sophisticated grasp of information theory. Overall, I would rank this dissertation among the best five or five six I have seen as a member of nearly 30 Ph.D. dissertation committees in philosophy (and a few more in math, statistics and psychology) at institutions like Michigan, Princeton, Rutgers, LSE, ANU, Western Ontario, and Sydney, among others. When it comes to formal ability and expertise, and knowledge of the relevant technical literature, I can only think of one dissertation that compares to it (and that person holds now holds a full professorship at Harvard). Lukits has produced a terrific piece of work. It would merit the granting of a Ph.D. anywhere. This is a very easy case: Lukits should pass!

Below you will find some critical comments, which are intended to be shared with Lukits. Given the terrific philosophical sophistication, technical competence and scope of the dissertation, and the likelihood that it will one day be turned into a book, I have decided not to pull punches in my comments. I disagree with Lukits on a number of central issues, and I think it will be useful for him to hear the sorts of criticisms that those who oppose his approach are likely to offer. But, just to be clear, these criticisms should not mislead anyone into thinking that I have anything but the highest admiration for what Lukits has accomplished. This is a wonderful piece of work, that will almost surely develop into a landmark book.

I have included some questions that Lukits might be asked in his defense in **ALL CAPS BOLD**.

Chapters 1-3

In these introductory Lukits does a very good job describing the details of his project and the basic structure of information-theoretic approach. I particularly liked his way of thinking about prior probabilities, and his focus on information-theoretic justifications of belief change, rather than the on the use of MaxEnt methods as a means to selects prior probabilities.

p. 24, Caticha and Giffin. This should have been explained in more detail. I took a look at the cited article and was not sure I saw anything in it that has not been known since the 1980s. E.g., how do they generalize what's already in Diaconis and Zabell. Also, it should be noted that that the "locality" constraint in Caticha and Giffin seems to rely on a largely unanalyzed

notion of what it means for "new information to refer to a domain". Also, on the most natural reading of this phrase, the "locality" condition is really equivalent to Jeffrey's rigidity condition, so it's not a surprise that Jeffrey conditioning can be derived.

p. 33. "The inevitable... flower bed." Who would have said otherwise? And, how is this relevant? Imprecision in measurement usually does not lead to calls for imprecise probabilities. One often knows that measurement errors are normally distributed around a mean, and when one does the handling of measurement errors does not require imprecise probabilities. They are required when information is lacking (e.g., no measurements were made) or where the data is ambiguous (e.g., where one does not knows enough about the measuring apparatus to know how its errors might be distributed.

p. 35, "According to Jaynes....." There is a little too much by way of appeals to authority at crucial points in the book. I went back and looked at the Jaynes article mentioned. Aside from being snotty and nasty it ignores the point of Shimony's example (and spends lots of time correcting Shimony's terminology). I looked at the Cyranski too, which also seemed to be arguing against a straw-man. The important point about Shimony's example is that one can treat the specification of an expected value, like the expected number of spots that come up on a die (say Exp(spot) = 3), either as a constraint on the set of probability functions to which PME is applied, or one can first apply PME and then condition on the event Exp(spot) = 3. One can get different answers in the two cases, depending on what information one already has. What Jaynes needs, as he himself says in various places, is account of when it is appropriate to treat a piece of information as a constraint on a prior to be selected by PME and when it is appropriate to treat is as an item of data to be conditioned upon. He never really explained how this distinction was supposed to work, except by examples. For the record, I think there is such a distinction to be made, but I think Jaynes and his acolytes never really made it clear what it was. So, given that Shimony's argument was laid out in so much detail, further discussion of this point would have been useful.

p. 36, "Well posed problems..." As will be clear in my remarks on the Judy Benjamin problem I do not think Lukits succeeds in answering this challenge: the problem is not "well posed".

p. 37, "In this dissertation, I am limiting myself..." I am glad Lukits is up-front about this, but it is worth noting that many of the worst problems for PME, those in which the result of applying PME depends on the partitioning of the underlying event-space, arise with most force in uncountable infinite cases (e.g., the wine/water paradox that Lukits discusses briefly on p. 17).

p. 42, "If the constraint is non-affine...". This strikes me, and many others, as a very serious shortcoming of the PME approach. While it is fine to ignore the problem in a dissertation like this, this worry would need to be addressed in detail in a full defense of PME.

p. 46 "Like the likelihood principle... point of view." I was doubly puzzled by this remark. I'd have thought that some Bayesians would be fine with ORIGINAL PME as long as it is

understood as a method for choosing priors. Indeed, this is just what many "objective" Bayesians think. Also, all Bayesians *accept* the Likelihood Principle. There are also non-Bayesians who accept it (likelihoodists like Richard Royall, Ward Edwards or Elliott Sober, for example).

p. 48, Piecemeal learning. I was not convinced by Lukits' treatment of this important case, which many people see as a central problem with the PME program. The idea is that you start out knowing the following item of information about an urn that contains blue, red and yellow balls: $Item_1$ = at least half the balls are blue. On the basis of this PME tells you to use (½, ¼, ¼) as your prior for (blue, red, yellow). Then you learn $Item_2$ = at most 20% of the balls are red. If you then update by minimizing Kullback-Leibler divergence to get a posterior of (8/15, 1/5, 4/15). The problem is that if you *combine* the two items and apply PME to their conjunction $Item_{1+2}$ = at least half the balls are blue and at most 20% of the balls are red, then you get the posterior (1/2, 1/5, 3/10). Or, if you take the two items in the reverse order, $Item_2$ then $Item_1$, you get (1/2, 1/6, 1/3). So, the order in which the information is processed matters, and it matters whether the information is processed in a sequence or as a single unit. It does not seem like this should matter. Lukits answers by saying, first, claiming that "the information provided in a problem calling for a synchronic norm and the information provided in a problem calling for a diachronic norm is different, as temporal relations and their implications for dependency between variables clearly matter." In general, that's true, but it is not clear that it is true *in this particular case*. Leaning $Item_2$ after $Item_1$ does nothing to alter the informational content of $Item_1$: after learning that red ≤ 20 you still know exactly what you knew before about blue, viz., blue ≥ 50. Likewise, leaning $Item_1$ after $Item_2$ does no alter the informational content of $Item_2$: after learning that blue ≥ 50 you still know exactly what you knew before about red, viz., red ≤ 20. And, what you know at the end of either process is exactly what you would know it you just learned the conjunction of $Item_1$ and $Item_2$. So, I do not see why, in the case at hand, we should think that these differences are acceptable. Dynamic considerations can matter, but I do not see any reason what they should matter *in this case*. Lukits goes on to compare the non-commutativity of PME with the well-known non-commutativity of Jeffrey conditioning. But, there is a crucial difference. The only scenarios in which JC fails to commute are those in which a second update is what Joyce calls a "hard" shift (in the cited (Joyce 2009)) that destroys information contained in the first update. In such case one has a clear causal story to tell about why the updates should not commute. No such story seems to apply in this case: learning the one item leaves the information content of the other item intact. **HOW WOULD LUKITS RESPOND?**

Chapter 4 is solid, but somewhat uneven. The rhetorical framing of the chapter is problematic, whihc blunts its overall effect. Lukits sets up the dialectic as a contrast between "the geometry of reason" and "information theory". By the former he means a view, held by H. Leitgeb and R. Pettigrew, that epistemology for credences is just decision theory with accuracy understood as "distance from the truth" playing the role of utility, and with divergences among probability functions being measured by *metrics* (which are symmetric and obey the triangle inequality). While Leitgeb and Pettigrew do hold this view (in their

more extreme moments), it is probably not held by anyone else, e.g., Joyce. (Lukits should take a look at Joyce's (2009) Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief," as well as Predd, et. al., "Probabilistic Coherence and Proper Scoring Rules," neither of which he seems to have read.) The field has come a long way in the last 18 years, and the contrast between "the geometry of reason" and "information theory" now looks like a false dichotomy, and attacking the idea that accuracy must be a metric is beating a dead horse. Most people working in the field today agree that one should measure accuracy using *proper scoring rules*, and that one should gauge differences among credence functions using the Bergman divergences associated with these rules. These divergences are *not* usually metrics, and few metrics can be represented as divergences. From this perspective, "the geometry of reason" becomes the view that one should use the quadratic scoring rule and its associated *squared*-euclidean divergence, while "information theory" becomes the view that one should use the logarithmic scoring rule and its associated divergence (Kullback-Leibler). But, there are lots of other proper scores, and each has its own characteristic entropy function (for the quadratic score it is mean variance, for the logarithmic it's Shannon entropy), and each can be seen as encouraging its own kind of info-min. Some people, e.g., Joyce, like the quadratic score for some purposes, the logarithmic score for other purposes, but they prize conclusions most when they can be shown to hold for any way of measuring divergence from truth.

I largely agree with Lukits' criticisms of LP conditioning in §4.4, and found his arguments new and interesting. I was especially impressed by his discussions of invariance and expansibility (§4.4.4 and §4.4.5), which struck me as really insightful. My one misgiving is that Lukits neglects to consider an option which some proponents of scoring rules favor: namely, denying that Jeffrey conditioning should be justified in terms of minimum divergence. (Joyce has recently been arguing this way in talks.) Roughly, the reasoning would work like this: For any proper scoring rule, conditionalization can be justified as the posterior that minimizes expected divergence from the prior, when expectations are computed using the prior. (This is the gist of the well-known article by Greaves and Wallace, but it was really established earlier by David Lindley in a fantastic paper called "Scoring Rules and the Inevitability of Probability," though one needs to read between the lines a bit to find it.) As Lukits notes, it is well-known that Jeffrey conditioning can only be justified in a similar way if we use Kullback-Leibler as our measure of divergence among probabilities. But, one might suggest that there is a better justification for Jeffrey conditioning that works with any proper scoring rule. Brian Skyrms once gave a Dutch book argument for JC that made use of the idea that if one starts with a prior p and Jeffrey conditions relative to a partition $\{a1, a2,…, a_n\}$ then one should end up with a posterior p^* that would put you at $p(x|a_i)$ if you conditioned on a_i, so that $p^*(x|a_i) = p(x|a_i)$ for all i. Of course, p^* then has to be JC. This line of argument can be easily adapted to the scoring rule framework, since the framework can be used to justify both the conclusion that, upon learning any a_i, one should update p to $p(x|a_i)$ and one should update p^* to $p^*(x|a_i)$. This would be a justification for Jeffrey conditioning that would work with any proper score. It rules out LP conditioning, for example, because a person who first LP conditions and then conditions on the true member of a_i will not end up where she should, at p conditioned on a_i. **WHAT DOES LUKITS HAVE TO SAY ABOUT THIS POSSIBLITY?**

The discussion of the Horizion constraint (§4.4.5) is interesting. I especially liked the examples on p. 92, but I thought the intuitive motivations for Horizion could have been explained more fully. On its face, Horizion seems like a stability claim. This is suggested by the line "it ought to be more difficult to update as probabilities become more extreme (or less middling)" (p. 93) In light of the examples, that seems to mean that shifting a credence from, say, 0.9 to 0.91 should be harder than shifting it from 0.6 to 0.61 in the sense that the improvement in accuracy should be smaller for the second shift than for the first. I suspect that it should follow, **THOUGH YOU CAN ASK LUKITS ABOUT THIS**, that it would take strictly more evidence to justify a move from 0.9 to 0.91 than it would to justify a move from 0.6 to 0.61. This does not seem right to me.

I am not sure the detour through confirmation theory in (§4.4.5) was useful. Differences or ratios of accuracies are quite different from differences and ratios of priors to posteriors, or of odds ratios.

I really liked the discussion of symmetry in 4.5.3. But, it should be noted that, contrary to what is suggested on p. 105, that Joyce's Symmetry requirement does not require the divergence function itself to be symmetric in the sense Lukits has in mind.

p. 96. Skew asymmetry. One would not expect this to hold for non-additive measures of incremental change. It's a mistake to think of it as a non-negotiable requirement on all measures of incremental confirmation.

§4.5.1, Triangularity. If one takes the scoring rule approach the failure of triangularity does not seem like much of a problem. This is because there is a natural generalization of the triangle inequality which does hold for any proper scoring rule. This website gives a nice account:

http://mark.reid.name/blog/meet-the-bregman-divergences.html

p. 102. "The more stops… *x* and *y*." This would be right if we thought that the divergence between *x* and *y* can be computed by summing up a bunch of distances between points on a Euclidean straight line. If one is using the logarithmic divergence, however, one would not do this since, intuitively, Euclidean straight lines are not the paths that minimize divergence between points. When the divergence is different the geometry is different. So, this seem like mixing two things (log divergence and Euclidean geodesics) that should be kept apart.

p. 107, "in higher-dimensional… information theory." This is interesting. I would have liked to hear more about it.

 p. 108, "How counterintuitive… sense." But this does not sound that counterintuitive when one is thinking in terms of divergences, which are expectations of accuracy of the posterior computed using the prior. One would not expect such quantities to have the property in question.

p. 109, "Information geometry… re-established." I would have liked to hear more about this. It seems quite interesting. I would encourage Lukits to develop these thoughts further (in work beyond the dissertation).

CHAPTER 5 seeks to show that PME can be used to justify Jeffrey conditioning (which, as Lukits well recognizes, is old news), and also that it can be used to justify Carl Wagner's generalization of JC. I really liked this chapter, and am in agreement with most of it. Though I disagree with Lukits about the status of imprecise probabilities, I agree that Wagner's justification requires them, and that Lukits' new justification does not.

p. 114. "the narrowest event…". I might be missing something, but is it clear that such an event will always exist?

p. 120, "the answer is yes…". Why isn't this just another example of PME introducing a lot of new information into a problem, information that is not there originally? The conditional probabilities leave the marginals underdetermined, logically speaking. So, how can the latter be derived from the former without adding information? **LUKITS MIGHT BE ASKED TO RESPOND.**

p. 121, (5.12). It is useful to think of the quantity that follows β_j as a Bayes factor.

CHAPTER 6 is an attack on the idea that subjective probabilities should be imprecise. While I do not agree with Lukits about his ultimate conclusion, I did find the chapter to be quite well done and was very impressed by many of his arguments. He has given proponents of "imprecise probabilities" an awful lot to think about, and I am sure that this chapter will develop into an important paper. I'm offering more criticisms to this chapter than to others mainly because I want to give Lukits a sense of the issues he needs to think about in order to make the chapter publishable.

p. 133 "To compare…" This is the wrong way to think about assessing the informational content of instates. Since instates are imprecise or indeterminate their informational content will necessarily be imprecise or indeterminate as well. For example, one might measure *one aspect* of the informational content of an instate C by taking the entropy of each probability function in C and using the resulting *set* of values as a measure of informational content. Then, one will be able to say that one state definitely contains less information than another when the upper limit of the entropy of the first is lower that the lower limit of the second. But, it is also true that a numerical information measure of the kind imagined will miss other aspects of informativeness. For example, if one instate is a subset of another then the narrower state will take a stand on more issues than the former. This is an aspect of informativeness that will not be captured in a numerical measure like the one envisioned. The general point is that trying to find a single number that represents the information in an instate is hopeless. At most, one might be able to give a partial order of instates with respect to informativeness. This ordering would have to include both the informativeness of the probabilities in the instate as well as facts about the range of probabilities in the instate itself.

CAN LUKITS GIVE USE MORE REASON THAN HE HAS FOR THINKING THAT THERE SHOLD BE A PRECISE MEASURE OF INFO CONTENT FOR IMPRECISE STATES.

p. 134, (S5). Those who use Shannon entropy to measure information like to imagine that this is a substantive axiom that, presumably, can be justified by reflection on the concept of information itself. In fact, this is really nothing more than a measuring convention that, in effect, says that we want to use a logarithmic scale (in which products, that are used to express independence, are turned into sums).

p. 137, Skittles example. I am not sure anyone actually holds the Boole-A position that Lukits describes. If the pile of candies really is randomized (in the sense that the ratios given reflect sharp chances for various skittles getting into the bag), then everyone should agree that the subject should have sharp credences. Imprecision is inappropriate only in cases where the subject has incomplete information about the chances of skittles of various colors getting into the bag.

p. 144, Repetition example. I did not understand what Lukits means by a "sleight of hand". It was not clear whether he is endorsing the view that, in contexts of complete ignorance, the epistemic state should not narrow even after lots of evidence, or whether he was criticizing this position.

p. 146, (AC1). I think this captures Augustin's view of things quite well, but it captures Joyce's views imperfectly. Joyce emphasizes the fact that states can stand in *complementary* relations to one another, which is something like the idea behind (AC1), but he does not ever identify complementary states as the same "epistemic state".

p. 148, criticism of Joyce. This right, and surely a mistake by Joyce. The example would work if one replaces "=" by "≠".

p. 149. "The example licenses…" This is incorrect. See the comment on pp. 157-58

p. 150. "without ever referring to (AC1)." Joyce's version of (AC1) actually is invoked on p. 306. See the line "In fact, from your perspective…"

p. 152. "an ability Boole-B lacks." I'm not seeing why. After all, in certain contexts using considerations of information might be a perfectly good way to choose an instate, e.g., one's instate could be the set of probability functions whose Shannon entropy falls in some proper subset of [0, 1].

p. 153. "Joyce denies…" Joyce denies every premise of this argument, I think.

p. 153 (Aggregating expert opinion). The suggestion that the instate is [0.3, 0.6] is not mandatory. As to how to represent the difference in reliability of the two forecasters, one cannot do this properly (on either a precise or imprecise view) without discussing probabilities of the form p(rain | GPY says x & QCT says y). If these are determinate, say

always equal to 0.8·x + 0.2·y, then this will be true all across the instate. But, they might be indeterminate, in which case one will have p(rain | GPY says x & QCT says y) = m·x + (1 – m)·y, where m ranges over some subset of [0. 1]. **HOW DOES LUKITS RESPOND TO THIS CRITICISM?**

p. 156. "but seldom about imprecise utility." That's not true, though it may be that there is a little less discussion of imprecise utility because it is less controversial.

p. 157-158 (Example 36). This example is incorrectly handled in a number of ways. First, if the monkey's draw is really known to be random, then there is no imprecision here at all: everyone will agree that 0.5 is the right probability for Red. So, I am going to assume that Lukits meant to say something like "the monkey draws the ball by some process that we know nothing about, so that every chance distribution over (Red, Red), (Red, Yellow), (Yellow, Red) and (Yellow, Yellow) is consistent with our evidence. Second, the claim that "this instate… drawn" is not correct either. I am not sure what Lukits means by a bet being licensed, but he seems to be assuming that an agent can be licensed in taking any bet that has a positive expected value relative to any probability function in their instate. This is, to put it mildly, a very controversial claim among proponents of imprecise probabilities. Many would suggest that that is a necessary, but insufficient condition for a wager to be permissible. Finally, the urn/hand example is incorrect. The question is whether or not the ball is drawn randomly from the urn. If so, then then your subjective probability for Red is 0.5. This is the *natural* way to understand the case. To get an interval of [0, 1] for Red we would have to imagine a situation in which you have no information at all about how the contents of the urn relate to the contents of your hand, in which case the composition of the urn is genuinely irrelevant. **LUKITS COULD BE ASKED TO CLARIFY HIS POSITION ON THIS MATTER, SINCE SIMILAR CASES COME UP A NUMBER OF TIMES IN THE WORK.**

P. 158. "Walley's… was pardoned." This is not the right way to read Walley at all. There is no sense in which the prisoner is "given hope" of the possibility that the warden will say X_3 rather than X_2. Here is the relevant joint probability:

	X_1 & X_2	X_1 & X_2	X_2 & X_3
Say "X_2"	1/3	0	p
Say "X_3"	0	1/3	$1/3 – p$

When the agent updates on "X_3" her credence for death (X_1) dilates to cover the range from ½ to 1. I am not sure either how she is given any hope or how she makes any unfounded assumption. If she has been given hope (because part of her credal interval is below 2/3), it is mixed in with an equal measure of despair (because a larger part of the interval is above 2/3).

P. 158. "The fallacy... again included." This is correct, but it doesn't cut any ice. All parties know full well that one can have probabilistic independence at the level of precise credences without having independence among variables. Even so, it remains true that that PME introduces new information about dependencies among variables that is not found in the evidence. Take the three-prisoners example. Everyone recognizes that PME does not *directly* require that $p(\text{"}X_2\text{"} \mid X_2 \& X_3) = p(\text{"}X_2\text{"})$, i.e., it does not just ASSUME that the warden will report "X_2" or "X_3" with equal probability when $X_2 \& X_3$ by stipulating that $p/(p + {}^1/_3 - p) = {}^1/_3 + p$. But, PME does something just a problematic when it comes to adding information. Since the prisoner's has no evidence that discriminates between "X_2" and "X_3" it deems these equally probable, thereby requiring that ${}^1/_3 + p = {}^2/_3 - p$, which is not information that is found anywhere in the statement of the problem. This leads to the $p = 1/6$ answer by a different route, but it is going to be just as objectionable to those who want to resist the addition of new information into the problem. Whether independence at the level of individual credences is imposed directly or whether it arises via an application of PME, the fact is that potential correlations are being swept under the rug on the basis of no evidence. GIVEN THIS CRITICISM, HOW WOULD LUKITS DEFEND THE USE OF PME IN THIS CASE?

CHAPTER 7 seeks to rebut a standard objection to the information-theoretic approach (the "Judy Benjamin" problem). In bare form, the objection starts with a subject whose subjective probability is starts out as $p(R\&H) = p(R\&{\sim}H) = ¼$ and $p({\sim}R) = ½$. The subject then learns a piece of *conditional* information: if R is true, then H is three times as likely as ${\sim}H$. Proponents of PME treat this as a constraint on the posterior of the form $p^*(H|R) = ¾$, and minimize Kullback-Liebler divergence from the prior to a posterior that meets the constraint. This yields the posterior of $p^*(R\&H) = 0.35$, $p^*(R\&{\sim}H) = 0.117$ and $p^*({\sim}R) = 0.535$. The objection is that (i) the conditional learned does not seem to convey any information at all about R's probability, but (ii) PME treats it as evidence for R (since $p^*(R) > p(R)$). I found much to agree with in Lukits' criticism of alternate solutions to the JB problem, and was quite impressed by some of the nice technical work in the chapter, especially the interesting section on the "powerset" approach. In the end, however, I think Lukits does not give sufficient attention to the possibility that the JB problem is just underspecified, and that it is unsolvable as a result. *Nothing in the statement of the problem tells us anything about the possible messages Judy might receive or anything about how likely she is to receive these various messages when she is in various locations.* Until that is done the problem *has no solution*. One can pull various solutions out of thin air by making additional assumptions or by importing information that is not found in the statement of the problem. But, neither approach solves "the" JB problem. They might solve elucidations of the problem, but not the problem itself. **IN LIGHT OF THIS, CAN LUKITS DEFEND HIS CLAIM THAT THE JUDY BENJAMIN PROBLEM IS "WELL-POSED," I.E., THAT IT IS SPECIFIED IS A SUFFICIENTLY DETAILED WAY TO PREMIT A UNIQUE SOLUTION? SEE ALSO MY COMMENT ON P. 168.**

p. 165. "Under a... optimal." This will be true of every proper scoring rule and it associated entropy function. Also, "We do not want to... necessary." Evidence is not being "drawn out" it

is being *added*: the information is not already in the problem. Indeed, the JB problem is massively underspecified (see below).

p. 168.   The handling of T2 is muddled. The idea is that if the subject learns C_t = "if R is true, then H is t times as likely as $\sim H$", then as t goes infinite "the problem reduces to one of ordinary conditioning" and so the posterior probability of $\sim H$ should approach 2/3. Lukits is assuming that, in the limit, learning C_t is tantamount to learning $\sim(R \, \& \sim H)$. Lukits is not the first person to think this, and it is very tempting to think of C_∞ = "if R is true, then H is infinitely more likely than $\sim H$" as equivalent to the posterior constraint that $p^*(R \, \& \sim H) = 0$, but this begs the question. Because T2 does not place any constraints on R's posterior probability, there are many different sequences of learning experiences that can converge on C_∞. As Lukits points out, a number of his adversaries want to treat the learning of C_t as a case of Jeffrey conditioning in which the posterior is constrained to satisfy $p_t^*(R\&H) = t/2(t + 1)$, $p_t^*(R\&\sim H)$ = $1/2(t + 1)$ and $p_t^*(R\&H)$ = ½. But, if this is the information each C_t conveys then in the limit C_∞ does not merely convey $\sim(R \, \& \sim H)$. Rather, it conveys $p_t^*(R\&H) = 1/2$, $p_t^*(R\&\sim H) = 0$ and $p_t^*(R\&H)$ = ½. In light of this, appealing to a principle like T2 is not legitimate in the context of this discussion because it assumes that the message conveys only the information $R \, \& \sim H$, which is part of the question at issue.   HOW WOOULD LUKITS DEFEND T2 IN LIGHT OF THIS CRITICISM?

p. 170. There is a mistake in equations (7.6) and (7.8). When one plugs in t = 3 one gets the solution (0.1204, 0.3613, 0.5183), but the correct answer is (0.117, 0.350, 0.533). Even worse, when one sets t = 1 (i.e., so that the probabilities of H and $\sim H$ are equal conditional on R) one gets the solution (0.268, 0.268, 0.464). This would be deadly to Lukits' view since it would mean that receiving the *non*-information C_1 would change the probability of R. But, I am fairly sure it is just a mathematical error. Of course, it should be corrected in the final version of the thesis.

pp. 171-172. Lukits needs to be clear what sort of conditionals he is talking about here. On p. 171 he says that "you learn that A entails B." "Entails" usually means "logically entails", in which case a coherent agent could never learn this. Later on, in "Sundowners" he seems to means something like "entailment given what you know." Then later he cites Huisman's paper on simple indicative conditionals. I think the last thing is what he really means. This should be made clearer.

pp. 173-174. Lukits is right that nothing in the problem tells us whether the information Judy gets might depend on her location (i.e., whether she will get a different signal depending on whether R or $\sim R$). The examples I-III make the point nicely. He is also entirely right that Douven and Romeijn err in assuming that there is no correlation between the content of the message and her location vis-à-vis the Red or Blue area. Even so, this observation is a double-edged sword. The JB problem will always be underspecified until one (a) lists the possible messages that Judy might receive, (b) conjoins them with her various possible locations, *and (c) specifies (ratios of) conditional probabilities of the form p(message | location)*. (This is the "sophisticated space" that should be used to describe JB, contrary to what is suggested on p.

178.) Until the problem is formulated in this fuller way, JB is either unsolvable due to ambiguity or its solution will be some imprecise credal state. Of course, proponents of PME will claim to have a way of deciding how strongly to weight the various possible correlations that might obtain, but the result they get –– that the report is evidence about which side of the Red/Blue line Judy is on –– seems just as unmotivated as Douven and Romeijn's result.

P. 177. "Because A has no information… equal 0.5." Of course, the claim that the warden has no information TO privilege either case and the claim that their probabilities are therefore equal is just an application of PME (or insufficient reason). But, is there any plausible independent justification for it? Why think either that (if both A and B are to die) the warden is equally likely to say A or B, or that the probabilities of scenarios in which he says A are exactly balanced out, probabilistically speaking, by the probabilities of scenarios in which he says B? AGAIN, ISN'T THIS JUST ANOTHER CASE OF ADDING INFORMATION?

Chapter 8 is a very nice summing up of the dissertation. I particularly liked the final few paragraphs, and my reply is "*Ce que nous connaissons est peu de chose, et nous ne devrions pas raisonner comme si nous en savons plus.*"

Bottom Line: A wonderful dissertation that really engaged by attention. It should pass with flying colors!

Please let me know is you need more information.

Sincerely,

James M. Joyce
C. H. Langford Collegiate Professor of Philosophy and (by courtesy) of Statistics