

Irrationality Re-Examined: A Few Comments on the Conjunction Fallacy

Michael Aristidou

Mathematics and Natural Sciences, American University of Kuwait, Safat, Kuwait
Email: maristidou@auk.edu.kw

Received December 3rd, 2012; revised January 5th, 2013; accepted January 19th, 2013

Copyright © 2013 Michael Aristidou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, I argue that the probability model used to infer irrationality for the subjects in the famous Linda problem is not appropriate, and I suggest different approaches based on fuzzy reasoning models. My line of argument is two-fold: 1) If the term “probability” is understood properly (mathematically), then the experimenters used the wrong model. 2) If the term “probability” is understood casually (non-mathematically), then alternative models perhaps should be used to justify the subjects’ responses. The objective is to experiment with new ways of looking at irrationality and raise a discussion regarding the relation between irrationality, reasoning errors and logical models that are used as frameworks to study irrationality.

Keywords: Probability; Rationality; Irrationality; Fuzzy; Conjunction; Fallacy

Introduction

To examine how humans reason, psychologists have conducted a series of experiments, the results of which demonstrate a high degree of deviation in “rationality” from the so called normative standards. Some of these experiments are: 1) The Wason’s Card Selection Task¹ (Wason and Johnson-Laird, 1972) and 2) Linda’s Conjunction Fallacy (Tversky & Kahneman, 1982). The former was reformulated by Oaksford and

¹In this experiment, subjects have to pick those cards out of four that makes a certain rule true or false. The four cards on the side that the subjects can see have a “4”, “7”, “A” and “D” written on them. The rule is “If a card has a vowel on the one side then it has an even number on the other.” On the question of which cards’ other side one has to see in order to falsify the rule, around 90% of the subjects responded incorrectly “A” and “4” where the right answer is “A” and “7”. This controversial experiment revealed some flaw in the deductive capacities of humans, but it surprised also researchers much more when a more concrete version of it (Robertson, 1999: pp. 69-70), dealing with everyday matters dramatically increased the correct responses that people gave. For example, subjects did far better when they were asked to play the role of a policeman that is trying to find whether the law of under aged drinking has been violated. The subject has to go in a bar, where at one of the tables four people are sitting. The first person is 15 years old, the second is 45 years old, the third is drinking lemonade and the fourth person is drinking beer. The rule is “If a person is under 18 years old, then she is not allowed to drink alcohol.” Note that this task is nothing but the Wason’s selection task stated above, where “lemonade” corresponds to “4”, “beer” to “7”, “A” to “15” and “45” to “D.” On the question of what one has to do in order to check the rule, most of the subjects found the right answer which is check the 15-year old, and check the beer to see by who is being drunk. The reason why subjects found the concrete version of Wason’s Task easier is, according to Robertson, that “certain contexts made the selection task much easier because the task fitted in with people’s previously constructed schemas about those contexts”(Robertson, 1999: p. 71). People are more familiar with permission/obligation contexts, therefore it was easier for them find the correct answer, namely “15” and “beer”, since the reasoning that generates those answers is closer to the permission/obligation schema.

Chater (1998) using Probability Theory and ultimately the “error” in judgment was justified, and further it was actually claimed that it was a “good” error. Here, I will be concerned with the latter, and I will argue that although the conjunction experiment is quite useful in giving an insight into human cognitive capacities, it is *not* sufficient to establish irrationality on behalf of humans. It is good enough perhaps to cast a doubt on the claim that humans are a priori, or mostly, rational (Davidson, 1984: pp. 183-198, 1982a: pp. 302-303). Not to be accused that I operate in the context of an exclusivist true/false dilemma, that only considers extreme cases (either fully rational or completely irrational), I claim from the beginning that surely rationality admits of degrees, and humans display a high degree of rationality on many occasions, and low degree on many others.

The Conjunction Fallacy

If the probabilistic approach was effective on selection tasks, such as Wason’s Task, it does not necessarily imply that the same probabilistic techniques might work in other reasoning tasks too. Proponents of the probabilistic approach think perhaps that probabilistic reasoning is at the heart of human reasoning (carried away by its success on the Wason’s Task), but, as Tversky and Kahneman (1982) revealed, this is not the case. People tend to violate standard probability rules in many circumstances and “irrationalities” are apparent, even when adopting a probabilistic framework to explain reasoning.

For example, people seem to conform to the conjunction fallacy in probability judgments. The most famous example perhaps is this:

Linda is 31 years old, single, outspoken and very bright.

She majored in Philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations (Tversky & Kahneman, 1982: p. 361).

The subjects were asked to state which is the most probable from the following five statements:

- 1) Linda is a teacher in an elementary school.
- 2) Linda is a bank teller.
- 3) Linda works in a bookstore and takes yoga classes.
- 4) Linda is active in the feminist movement.
- 5) Linda is a bank teller and is active in the feminist movement.

The majority of the subjects chose 5) as more probable than 2), violating a fundamental law of probability: that the probability of a conjunction must be less (or equal) than the probability of its conjuncts. This, according to the experimenters, is an “irrationality” case. But is it? One could say that the above “irrationalities” are irrationalities with respect to the approach one chooses to model human reasoning. Oaksford and Chater, commending on Tversky and Kahneman’s results, argue that:

If this [mismatch of probability and rationality]² is correct, then the whole idea of rational models of cognition is misguided: cognition simply is not rational (Oaksford & Chater, 1998: p. 18).

One could easily disagree with this. If human reasoning fails to be modeled or formalized by probability theory, does this make human reasoning irrational? Humans might seem “irrational” with respect to probabilistic systems of inference, but we have no reason to believe that these are the only systems that can give an account of human reasoning. There can be alternative systems that come closer to how humans reason and explain many of the alleged irrationalities that resulted from the experiments. For example, a Fuzzy Reasoning System³ could be, in the words of Stich, a “pragmatically superior alternative” to the probability system mentioned above. In other words, there could be better systems of inference relative to a problem, as some systems (given a problem) might be more general and provide more explanatory frameworks. They could also take into consideration the weight of some information and its relevance. For example, words like “discrimination” and “social justice”, included in the information about Linda, are words that indicate relevance or irrelevance to choices 2) and 5), a fact that probability theory ignores. Also, words like “very” and “deeply” indicate weight, or a degree, that certain information about Linda bear in relation to choices 2) and 5), a fact that probability theory also ignores.

Stich’s Theory

Davidson’s Principle of Charity states that “one should interpret an agent’s utterances in such a way that most of her assertions turn out to be true and most of her inferences turn out to be rational” (Davidson, 1984: pp. 183-198, 1982a, pp. 302-

²My italics.

³By *fuzzy reasoning system* I mean broadly any system of inference that takes into consideration relevance of data, imprecision, subjectivity, perception-based information, and is closer to how humans actually reason, and not how they “ought” to reason. For more on Fuzzy Reasoning see Siler and Buckley (2004). Harten (2008) in a Master’s thesis examines a variety approaches in solving the conjunction fallacy and he concludes that fuzzy theory made progress on the issue but did not resolve it in all cases. Instead he proposes a quantum probability approach.

303). In other words, when we are interpreting someone and we find that his reasoning deviates too much from what people usually hold as “rational”, then it is more likely, according to Davidson, that we are interpreting the subject incorrectly, and less likely that the subject is irrational (Davidson, 1982a: p. 303).

Inference that is frequently irrational, according to Davidson, is conceptually impossible. That is because inference—a process that generates beliefs—must have high levels of rationality and truth. But, how could one link rationality with beliefs? The keyword is interpretation, and Davidson’s argument (Davidson, 1982a: pp. 302-303, 1982b: p. 327; 1984: pp. 195-198, p. 170) revolves around the following idea: the meaning of a word cannot be fixed if what the agent utters by that word is not in accord with what he means by that word and what he believes about that word.

According to Davidson, an agent is irrational if he somehow generates beliefs that do not cohere, or are not consistent, with the pattern of his beliefs. Davidson’s approach to rationality, through interpretation, demands high degrees of consistency. Hence, any approach that allows much inconsistency is not acceptable to Davidson. That is clear when he says that: “... inconsistency breeds unintelligibility” (Davidson, 1982a: p. 303), and when he says:

If we are intelligently to attribute attitudes and beliefs ..., then we are committed to finding in the pattern of behavior, belief, and desire a large degree of rationality and consistency (Davidson, 1980: p. 237).

Stich argues (Stich, 1990: pp. 15,17) that if Davidson’s position is true then the following problems arise:

- 1) Our ability to explore irrationality empirically is undermined; hence we are losing a good insight into human cognition, since the interpreters are the ones most probably mistaken.
- 2) The claim that bad reasoning is conceptually impossible leads to a normative theory of rationality of no practical importance, since the theory turns its back on the empirical results.

Stich thinks that Davidson is right to say that content and rational reasoning are linked, but wrong in the assumption that there is only one type of rational reasoning. The latter assumption is something that Stich opposes, since he claims that it is *not* the case that there are no alternative systems of reasoning that are *all* rational. This is something that Stich calls Normative Cognitive Pluralism.

But how could one decide then that a system S is a rational cognitive system? Stich mentions Goodman’s attempts to describe a procedure (or a test) that a system of inferential rules should pass, in order to count as rational. But as Stich observes, it is very difficult to discern the relation between rationality and “appropriate” test, because we basically assume a priori that the test itself is appropriate. Hence one needs a more realistic account of what makes a system rational. Perhaps a middle ground could be found. Davidson recognizes high degrees of consistency, but nevertheless speaks of “degrees” and not absolute consistency. Also he alerts us on the conceptual difficulties of inference being frequently irrational. Stich demands a more realistic system of inference, which possibly considers relevance of data, and is more explanatory on how actually humans reason in praxis and not normatively. A fuzzy system could be such a system. I do not claim that this is the only correct way to go, but based on the nature of the problem at hand (Linda’s Problem, which contains fuzzy parameters) it seems that it is at least more explanatory.

A Fuzzy Model Perhaps?

I mentioned previously that assuming a probabilistic way of thinking, the “irrationality” appeared in Wason’s Cards Task was in a way explained away. That is, humans seem quite rational if we ascribe to them a probabilistic way of thinking rather than a propositional-logic way of thinking. But how about the “irrationality” appearing in Tversky and Kahneman’s Linda problem? Could that be set in a framework in which, if not to ascribe rationality to the responders, at least explain their error in a rational manner? It is true that the subjects erred (probabilistically?), but could one at least, suggest a formal model that would explain their reasoning? As Oaksford and Chater did in the Wason Task, to some extent, one could justify people’s choice of premise 5) over 2) in the Linda Problem by considering a non-classical approach to the problem.

It is important to clarify that the reason I suggest another model is not because I want to render humans a-priori rational or defend their judgment. It is simply because the (classical) probability model that was used here seems to be completely inappropriate for the problem at hand. Hertwig and Gigerenzer (1999: pp. 276-278), tried to explain the fallacy indicating the ambiguity of the word “probability”. I agree to a large extent, but I actually also claim that if indeed by “probability” the subjects understood the (classical) mathematical probability, as the examiners clearly intended, then the fault still lies more with the examiners as classical probability is not the right model to formulate this particular problem. The reason is because a good amount of data that refer to Linda are fuzzy data, including the conjunctive statement 5).

For example, from the description of Linda, it is clear that her being “outspoken”, “very bright”, and “deeply concerned with discrimination and justice issues” are all fuzzy data. Also, one of the conjuncts in statement 5), namely “active in the feminist movement”, is also fuzzy, which makes the whole statement 5) a fuzzy statement. So, it is not just that the data did not enter the probability calculus in this problem, as Hertwig and Gigerenzer (1999: p. 276) correctly pointed out, it is also that it is the wrong probability model to account for those data. The appropriate model for the problem is Fuzzy Probability (Zadeh, 1968, 1984; Buckley, 2010) which extends classical probability as it accounts also for fuzzy events. It is also sensitive to the degrees of relevance and membership of certain elements in certain sets. I will not get into the mathematical technicalities, as it is beyond the scope of this paper, but it should be noted that the conjunction law should also apply in the fuzzy probability case (as the latter generalizes the classical case), and the experimenter should be aware of that when using the appropriate model.

So, the only problem, as we said, is that fuzzy probability is supposed to generalize, not to reject classical probability, so any rule true in the later (such as $P(A \text{ and } B) < P(A)$) *must* be also true in the former. Hence, in fuzzy probability $P_f(A \text{ and } B) < P_f(A)$ also applies. It seems that the experimenters might be correct after all. But, it also seems that they were unaware of the following three cases: 1) $P(A \text{ and } B) < P(A)$ where A, B crisp, 2) $P(A \text{ and } B) < P(A)$ where A, B fuzzy, 3) $P_f(A \text{ and } B) < P_f(A)$ where A, B fuzzy. Randomness is distinguished from fuzziness as a form of uncertainty (Kosko, 1990). Obviously, the experimenters used case 1) but this is not the right case (considering the problem) despite of the fact that all three cases render the same result. So, it is possible that humans are sus-

ceptible to the conjunction fallacy, probability errors, or as Gould (1992: p. 469) said that “our minds are not built to work by the rules of probability”, as the numerous experiments show⁴. But, the original Linda problem has not showed that yet, as it is essentially a fuzzy problem and is using the wrong probability calculus. It is worth noticing, by the way, that most variations of the Linda problem in the literature that came to my attention are crisp versions, or do not have the degree of fuzziness as the original Linda problem (Tversky & Kahneman, 1982, 1983; Bar-Hillel & Nater, 1993; Tentori et al., 2004).

One could get into the technicalities and certainly show that indeed $P_f(A \text{ and } B) < P_f(A)$, where A, B fuzzy, but we will not do it here, as it is beyond the scope of this paper. It seems also a bit pointless, as even if it is true, what exactly would this fact say on the issue of rationality? Would it necessarily imply that humans are irrational? Leaving aside broader objections to the notion of probability per se, the question still remains: How exactly probability, crisp or fuzzy, relates to human reasoning? Is even the term “probability” the appropriate term to be analyzed here, in the example of Linda? Is this term what the subjects understood and interpreted?

It is possible though, as Hertwig and Gigerenzer (1999: pp. 276-278) suggest, that mathematical probability was not what was understood in the experiment, assuming that the relevance maxim applied. Then, I would like to suggest the following approach (“*Fuzzy Membership/Relevance Interpretation*”) as an alternative to the mathematical interpretation of probability. It is possible that the subjects interpreted “Linda most probably is” as “Linda most likely belongs” (or “most likely member of”)⁵, and if that is the case, then the following (fuzzy, but non-probabilistic) scenario might be plausible:

Let us view Linda as a (fuzzy) set of properties, i.e. as the set $L = \{\text{Phil, Discr, Int, Outsp, ...etc}\}$, where Phil = majored in Philosophy, Discr = concerned with discrimination issues and social justice, etc. The crucial properties of “being a bank teller” and “being a bank teller and active in the feminist movement” denote them by B and BF, respectively. The point is what is the relation of the B and BF to L. In other words, qualitatively and quantitatively speaking, how, which, and how many of the properties that characterize L relate to the B and BF. Now, based on the relation “relevant things go with relevant things”⁶ which seems consistent also with neuroscientific evidence (Nielsen, 2003: pp. 118-119), it is more likely that BF relates more to L, casually, on the basis that usually less Philosophy majors become bank tellers, and usually more people that are “deeply” concerned with issues of discrimination, etc, are active members of feminist movements. Do B and BF relate to any of the properties of Linda? In virtue of “being active in the feminist movement”, yes, to some degree BF does relate.

⁴Also, in Manktelow (2012: p. 245), even though he advocates the probabilistic paradigm, he states that “these formal systems ask too much of ordinary minds”.

⁵Or ignored the “probability” term altogether as irrelevant, and simply looked for relevance (likelihood, membership, etc). After all, these assessments about people in everyday life are not done by throwing dice.

⁶That’s a very realistic classification, since humans tend to classify things that way. As Nielsen says: “For example, if the first three words (the assumed facts of the consensus building process) are down the garden, consensus building is then carried on the fourth (answer) region to yield an answer token, which turns out to be (in our system), path. This four-word sequence actually occurred in our training corpus. If we then blank out the token for the word path and repeat the thought process, the next word obtained is lane. (Nielsen, 2003: pp. 118-119).

On the other hand, B does not seem to relate much. This determination probably strikes one as too subjective, but, that is exactly the point on why a fuzzy model (not necessarily fuzzy probabilistic) might be better. It allows for subjective judgment and accounts for perception-based information and fuzzy data.

The subjects in the experiment had exactly this to do: a subjective judgment given fuzzy data. Therefore, one could claim that there was some rational basis for people choosing statement 5) over 2) in the Linda problem.

More precisely, and without getting too technical, one could suggest the following model of fuzzy reasoning (generalized modus ponens) which incorporates how humans tend to classify things based on membership and relevancy:

If x is A , then y is B
 x is A'

 y is B'

where x, y are variables taking values from the universal sets X and Y , A and A' are fuzzy sets in X , and B, B' are fuzzy sets on Y .

The above is quite useful for modeling human common sense reasoning, which is usually reasoning in fuzzy environments. For example:

If a book is *large*, then *it is expensive*
 Book x is *fairly large*

 Book x is *fairly expensive*

Conclusion B' is calculated for any y in Y by the formula:

$$B'(x) = \max_{x \in X} \min \{A'(x), I(A(x), B(y))\}$$

usually referred to as the *compositional rule of inference* (Klir et al., 1997: p. 212), where I is an appropriate fuzzy implication⁷. For simplicity and economy in calculations, we usually adopt the matrix format:

$$R_{B'} = R_A \circ R_I$$

in order to compute the conclusion B' , where “ \circ ” is the max-min operation⁸.

In regards to the Linda problem, we basically follow the “guessing game” application⁹ in Tanaka (1996: p. 76) mod

⁷Say Lukasiewicz’s or Mandami’s implication, etc. But, as known, there is no a priori justification on how to interpret an implication, or any logical connective for that matter. At best, one tries to capture some basic intuition of what a true or partially true proposition is, with respect to certain phenomena or applications at hand. One could actually push the argument all the way and say that experience would be the only arbitrator of truth to an implication. For example, consider the implication: “If I am a Philosophy major, then I am an active feminist.” The truth or partial truth of this could only be assessed by experience. In other words if actually, if to our knowledge, there are philosophy majors who are active feminists, to what degree they are active, how Philosophy relates or not to feminism so we can put weight to future assessments, etc. Probability here has little to tell us.

⁸It works as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \circ \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} (a \wedge a') \vee (b \wedge c') & (a \wedge b') \vee (b \wedge d) \\ (c \wedge a') \vee (d \wedge c') & (c \wedge b') \vee (d \wedge d) \end{pmatrix}$$

⁹“Sally is nearsighted and colorblind. When she goes to the fruit market, where fruits are placed on high shelves, she cannot see them very well. She can only recognize the size and blurred shape of the fruits. Sally lived in such a world for 20 years and now she has some knowledge about the features of the fruits. For example, tangerines are round and relatively small. When Sally says ‘quite round and quite large’, can you guess what fruit she sees?” (the solution in Tanaka, 1996: p. 76).

ifying things slightly. Instead of implication I will generalize to relevancy (they are both types of relations). In other words, instead of whether Phil implies B and to what degree, for example, I will check whether Phil is relevant to B and to what degree. The way I will produce the relation R_I (hereafter R) will be empirically (i.e. experimentally), and I will only be concerned with two of the crucial properties of Linda, namely “Philosophy major” and “deeply concerned with discrimination issues and social justice”. By “empirically”, I mean how subjects would actually infer (not how they supposedly ought to infer) based on their life knowledge and experience, and the specific data given. I explain:

Let $X = \{B, BF\}$, where B = bank teller, BF = bank teller and feminist, and $Y = \{\text{Phil, Discr}\}$ where Phil = majored in Philosophy, Discr = concerned with discrimination issues and social justice. Then, given the data and people’s prior knowledge and experience we have the following relation R:

$$R = \begin{matrix} & \text{Phil} & \text{Discr} \\ \begin{matrix} B \\ BF \end{matrix} & \begin{pmatrix} 0 & 0.2 \\ 0.3 & 0.8 \end{pmatrix} \end{matrix}$$

where each entry shows the degree of relevance between two parameters and was determined¹⁰ practically as follows:

A group of 10 college students were asked, to their knowledge, to correlate with a line and a percentage degree (degree of relevancy) the following two columns:

B	Phil
BF	Int
	Outsp
	Discr
	Prot

where the right column clearly contains some of the properties of Linda. The average rounded percentage (normalized to 1) correlating Phil and Discr to B and BF are shown in R above. Given the description of Linda, a fair assessment regarding properties Phil and Discr would be:

$$L = \begin{matrix} & \text{Phil} & \text{Discr} \\ & (1 & 0.9) \end{matrix}$$

¹⁰More rigorously, one could define a *relevance relation* $R_i(X)$ between two sets X and Y (“ X is relevant to Y ”) to be the matrix whose entries are given by the *relevance operator* $R_i(X)_y = \mu_y(x_i)$, where $\mu_y(x_i)$ is the degree of relevance of x_i to y_j with respect to some attributes and based on data. This is actually what in praxis the students in the experiment used when asked to correlate with a percentage the two columns in p. 7, the answers of which filled in the matrix entries for the relation R. This relevance relation (neither reflexive, nor symmetric, nor transitive—the implication is reflexive and transitive but not symmetric) is a slight generalization of the usual equivalence and compatibility relation in fuzzy logic. Also, given the above, one could also define a *relevancy measure* by:

$$|R_i(X)| = \frac{\sum_{y \in Y} \sum_{x \in X} \mu_y(x_i)}{|Y|}, \text{ where } \mu_y(x_i) \text{ is the degree of relevance of } x_i \text{ to } y_j$$

and $|X|, |Y|$ are the cardinalities of X and Y . Notice that for the special cases $Y = \{B\}$ and $X = \{\text{Phil, Discr}\}$, and $Y = \{BF\}$ and $X = \{\text{Phil, Discr}\}$, we have: $|R_B(X)| = 0.1 < 0.55 = |R_{BF}(X)|$, which once again shows why the subjects would consider BF as more relevant to the description of Linda.

as she did major in Philosophy (so degree of membership = 1) and she is “deeply” concern on issues of discrimination (hence high degree of relevancy). Therefore, the matrix formula for the conclusion (or “guess”) gives:

$$C = LoR = (1 \quad 0.9) \circ \begin{pmatrix} 0 & 0.2 \\ 0.3 & 0.8 \end{pmatrix} = (0.3 \quad 0.8)$$

B BF

i.e. $C = (0.3 \quad 0.8)$, which says that the subjects ranked BF higher than B, and saw it more possible given the data.

What about the probabilities then? Our model does not require any. And even though the question posed to the Linda problem contain the term “probability” the subjects do not, and apparently did not, have to interpret it literally. The problem is a real life problem that involves human experiences and knowledge; it is not a problem to be decided by throwing dice. The question posed was not “what is more probable to have in general or theoretically/mathematically speaking”, but “what is more probable Linda is”, given the data. Have the data been taken into consideration? And how were they interpreted, in a Bayesian or a frequentist manner?¹¹ And if they have, and subjects failed the “test”, why not just simply say they made a mistake, or they are bad in probabilities, just like millions of perfectly rational people are bad in basic mathematics? There is a difference on how we actually reason and how we ought to reason. And the evidence that we somehow ought to reason probabilistically is not quite convincing.

Again, the experimenters ignored the degrees of membership of B and BF to L, i.e. to L’s “relevant” properties. Their subjects though did not consider all properties of L as equivalent, and it seems that they considered certain properties having more relevance and weight to the presented state of affairs that they were asked to make a choice. For example, what if the choices were:

- 1) Linda is an alien.
- 2) Linda is an alien and is active in the feminist movement.

Most likely, from the description, she is not an alien, and we do not know whether aliens care or not for women’s rights. But, the description does point to some relevancy of Linda to feminism. So the subjects could have thought that Linda sounds like a feminist, and whether or not she is an alien is secondary. Because *just* being an alien doesn’t necessarily imply that she is a feminist, but being an alien and a feminist *certainly* implies that she is a feminist. So, the crucial quality or property of Linda being a feminist was “locked” first, to be consistent or relevant to the description, and the rest of the properties are either irrelevant or are examined second. And, so what if in general there are less aliens feminists than aliens? In this particular case, the particular subject L is more likely to be a feminists whether she is an alien or not, in virtue of her properties/qualities.

Some Objections

Before I turn into some possible objections, it is important to

¹¹It is well known that in probability the two approaches do not agree. For a discussion on the different definitions of probability see Manktelow (2012: p 2-5). It is worth noticing that in the same book, that the author in more than one occasion (p. xii, p.247) distinguishes it from binary logic, or calls it non-binary, but it is not clear to me what he means, as probability is also based on binary logic. He also never mentions fuzziness or fuzzy probability. Fuzzy logic can be distinguished from binary logic, but even that on the level of theorems is based on binary logic.

once again clearly state what are the main points that I am arguing for. First, I am criticizing the methodology of Tversky and Kahneman, the way they arrive to their conclusions, and what those conclusions entail for human rationality. Obviously, there is a vast literature devoted to this discussion and I could not comment or refer to all of it. It is important to note, however, in what my criticism differs from other criticisms of the Linda case. It differs to the point that it indicates a mistake on behalf of the experimenters: The classical probability model used in the Linda problem is *not* the right model. One has to acknowledge that at first, and then continue with any further discussion. Would a fuzzy probability model render different probability outcomes? No, but that is not my point of concern. My point is not to find an alternative model which will judge the subjects rational. My point is to not to judge them irrational on the basis of models they do not know, and they are not obligated to know. It is easy for me to give you an algebra problem, of a level you have not cover, and because you failed to solve it to judge you irrational. It is seems unfair. But, this is what the mathematical interpretation of probability in the Linda problem amounts to. Is correct responding to probability questions a pre-requisite to rational thinking? This is what the experimenters seem to assume, but I am not sure yet as to why anyone should accept such a prerequisite.

Second, as I stated earlier in this paper, I think a non-classical reasoning model is, perhaps, a more promising model in describing how humans reason. Perhaps ... this is the key word here. I do not currently have an alternative model, and I am not aware of anyone who does. The argument behind exploring an alternative non-classical model is based on the fact that the current one, that is classical probability, failed. People do not use probability calculus to make decisions; most do not even know what that is. Even in matters that involve chance most do not use probability to make decisions. Poker, for example, is a good example. Furthermore, there is a big distinction between chance and imprecision (two different types of uncertainty) and the problem of Linda is not a clear cut chance problem. The argument behind the suggestion for a fuzzy model is simply the nature of the data in the description of Linda. It is only reasonable to request that the probability of fuzzy data to be accounted with the correct model, if one is interested in linking further those results to rationality. Yet, my point still stands. The subjects are not obligated to know probability, classical or fuzzy, to count as rational. Unless, someone (the experimenters in this case) makes an argument on why should it be so. Finally, say that the people went through a quick probability seminar, they were all taught the conjunction rule, and they applied it correctly and responded as they should. Would they all count as rational then? How so? Or would that just simply mean that they know probability, applied it correctly, and solved a math problem successfully?

Now, a few comments on some specific objections:

1) I mentioned earlier that humans might seem “irrational” with respect to probabilistic systems of inference, but we have no reason to believe that these are the only systems that can give an account of human reasoning. One perhaps could object as follows: *Suppose that we changed the system, and hence make better sense of peoples’ responses. Then, the fact that people contradict basic rules of probability theory does seem problematic to rationality.*

I disagree. I believe that the last statement begs the question. How exactly does probability relate to rationality? One cannot

simply refer to the Linda test. The fact that they failed it does not render them irrational, as the link between probability and rationality has not been established yet. Neither their success to the Linda test would render the subject rational. When Gigerenzer et al. claim that “we should not care about being irrational in that sense”, this I interpret I as a quiet admission that probability does not have much to tell us about rationality. It is indeed not quite clear how introducing another logic one could shed more light on the Linda problem. But, considering the fact that the logic used is at least controversial, if not the wrong one, then it is only reasonable for one to explore other venues.

2) When I spoke of “alternative systems that could give an account of human reasoning”, one could quickly respond: *What does it mean to “give an account of human reasoning”? Would not this simply amount to specify the rules that we follow and the calculus of those rules? If this is so, then such a project would be simply descriptive, and therefore the rationality question would still remain untouched.*

Regarding the last statement, I believe the same applies for probability theory. The model in the Linda experiment gave at best a descriptive account, but has not been related to the rationality question in and of itself. Furthermore, specifying the rules that we follow and the mechanics and unfolding of those rules is one issue, but there are other issues too. For example, those rules have to be empirical too, and have some connection to how humans actually reason. Most humans are not aware of the specified probabilistic rules, their calculus, etc, and they certainly do *not* follow them even in matters of chance. So, I am not sure how much it can inform us about how humans reason and whether their reasoning is rational.

3) We noted that words like “very” and “deeply” indicate weight, or a degree, that certain information about Linda bear in relation to choices 2) and 5), a fact that probability theory also ignores. Question: *How does being “very bright” relate to Linda being a bank teller or a feminist bank teller?*

This is not to say that feminists are smarter or that if you are smart you must be a feminist. It is more to show that being a feminist adds another dimension to Linda’s qualities. Positive or negative, the point is that she thought about it. To put it bluntly, she is not simply counting money; she is concerned with other issues too, and exposed herself, as she also majored in Philosophy, to more interesting and difficult issues such as human rights, etc. It could be that one is very bright but has no other interests, but here it is clear that the description of Linda points to the fact that she is involved in other topics, and has opinion on matters. Such a person is perhaps more rare to find, but it is not going to be decided by throwing dice.

4) I tried to emphasize and clarify that the reason I suggest another model is not because I want to render people a-priori rational or defend their judgment. It is simply because the (classical) probability model that was used here seems to be completely inappropriate for the problem at hand. One then could ask: *Inappropriate how and to whom? If people were using the model correctly, then perhaps they would not have failed the test this way. Why is it inappropriate for us to judge them based on that model? Is the model still inappropriate if people had been reminded of the relevant rules of probability theory?*

It is inappropriate tool to model the experiment. This is simply a mistake on behalf of the experimenters, and I am indicating it. It might not affect the conclusion of the experiment, but the experimenters have the obligation to apply the right tools.

People have *no* obligations to use, or know, the probability model at all, just because the experimenters included the word “probability” in the question. Certain issues are not settled by throwing dice, and the distinction between randomness and fuzziness (imprecision) needs to be drawn. And probability, even on everyday matters of chance, gives us an idea and a hint, but we certainly do not know or explicitly compute rules of probability in order to make a decision. And, say that the people went through a quick probability seminar, as we mentioned above, and they were all taught the conjunction rule, and they applied it correctly and responded as they should. Would they all count as rational then? Or would it just simply mean that they know probability, applied it correctly, and solved a math problem successfully?

5) I noted that “a good amount of data that refer to Linda are fuzzy data, including the conjunctive statement 5).” One could object: *In what sense this data is fuzzy? People are receiving a description that involves no numbers, so how is this “fuzzy data”? For example, Linda being “outspoken”, “very bright”, and “deeply concerned with discrimination and justice issues” these are supposed to be fuzzy data”, but this is not quite clear what this is supposed to mean.*

These are fuzzy data in the sense that the sets of all “bright” people, or “outspoken” people, are inherently fuzzy. The membership or not in that those sets is not a yes or no situation, but there is a degree of membership in those sets for each case we examine, which is determined subjectively. Now, these membership degrees in the corresponding fuzzy sets are supposed to be taken into considerations when one consideration when one understands probability in the mathematical sense and does the calculations explicitly. Is this what the expectation of the experimenters was?

6) Earlier we mentioned that “based on the relation ‘relevant things go with relevant things’ which seems consistent also with neuroscientific evidence, it is more likely that y_2 relates more to L, casually on the basis that usually less Philosophy majors become bank tellers, and usually more people that are ‘deeply’ concerned with issues of discrimination, etc, are active members of feminist movements.” One could say: *What neuroscientific evidence have to do with the question at issue? Also, where are the claims about philosophy majors becoming bank tellers or not coming from?*

I am simply saying that humans tend to classify and group things based on relevancy. For example, if I tell you “ice-cream” you could reply back to me some relevant things that your mind related to it by experience. Such as “cone”, “cold”, “cup”, etc. You probably would not replied “calculator” or “fish” to me. And that is what I think happened with the Linda case. The subjects most I think saw more relevancy between “feminism” and “philosophy major” than between “bank teller” and “philosophy major”. If this is backed by experience, which I think it does, then the experimenters should take a gallop and actually count whether bank tellers or feminist bank tellers had philosophy degrees, or took philosophy classes, or have broader, or casual, philosophical concerns in general, etc.

7) Finally, I mentioned that “the question posed was not ‘what is more probable to have in general, but what is more probable Linda is’, given the data.” One could say: *Well, this is not the question posed to the subjects after all.*

Yes, but why was the description given to the subjects then? Was it given to mislead them? Was not given to them so they could weigh, and take into consideration, the information about

Linda and compute the probability? Because if not, and the description was irrelevant, then we basically test the subjects on whether they were aware of a certain math rule, namely the conjunction rule, and because of the lack of that knowledge we rendered them irrational. I do not believe that was reasonable, or fair.

Concluding Remarks

Davidson argues that inference that is frequently irrational is conceptually impossible since inference (a process that generates beliefs) must have high levels of rationality and truth. Therefore, humans must be rational. Davidson's argument is a plausible argument if one relaxes the demands for consistency. That will make it more pragmatic and bring it closer to Stich's position and experimental results. But then, one needs a new (more general) framework and model to study rationality.

Just as Oaksford and Chater have suggested that committing the fallacy in Wason's Task was, actually, the rational thing to do, other authors (such as Dulany & Hilton, 1991, Slugoski & Wilson, 1998) claim that committing the fallacy in the conjunction problem was a rational response. On the other hand, other authors (such as Stanovich & West, 1998, 2000) conducted studies which showed that subjects that committed the conjunction fallacy also scored lower (in general) in SAT tests than subjects that did not commit the fallacy. This fact is an important empirical piece of evidence that comes in the discussion, although it is still not clear how it translates in terms of rationality. Also, it is uncertain what SAT has to do with rationality anyway, considering the extensive criticism¹² against it on being an unfair, and biased test, and irrelevant to measuring reasoning skills. Perhaps if we tested the subjects in chess playing the ones who committed the conjunction fallacy might have been better chess players than the ones who avoided the fallacy. But that is not the point. Because whatever the test is, it still falls under Goodman's project to describe a test that tests rationality, which in turn falls into Stich's criticism who says that it is very difficult to establish a relation between rationality and "appropriate" test, since we always assume a priori that the test itself is appropriate.

In conclusion, a non-classical reasoning model is, perhaps, a more promising model in describing how humans reason. It is not the first time that an alternative to probabilistic models has been suggested. Cohen (1979) suggested a Baconian model of judgment as an alternative to Bayesian ones. Perhaps, a fuzzy reasoning model (not necessarily fuzzy probabilistic) could be a better way to model humans' natural way of communicating and reasoning. As Hilton says:

Understanding conversational inference may help clarify the question as to which normative model is appropriate in a given situation (Hilton, 1995: p. 266).

Indeed, since an important aspect of the conversational inferential model is the effect of conventional and conversational implicatures in which fuzzy words such as "intelligent", "ac-

tive" and "few" are frequently used (see Hilton, 1995: p. 265), then we need a model that is appropriate to deal with such vague concepts and respects implicatures.

Whether there was an error on the part of the subjects (in responding the way they did to a probability problem) or an error on the part of the experimenters (in testing subjects on probabilities to account for rationality), the lesson is that alternative models of reasoning perhaps need to be considered. Models that are not blind to relevance or imprecision. The nature of the testing, perhaps, needs also to be inverted. The objective should not be how well human reasoning approximates a reasoning model, but how well a reasoning model approximates human reasoning.

Acknowledgements

I would like to thank Andrei Zavalij for his comments on an earlier draft as well as the anonymous referee for his comments and productive criticism.

REFERENCES

- Buckley, J. J. (2010). *Fuzzy probability and statistics (studies in fuzziness and soft computing)*. New York: Springer.
- Cherniak, C. (1981). Minimal rationality. *Mind*, *XC*, 161-183. doi:10.1093/mind/XC.358.161
- Cohen, L. J. (1979). On the psychology of prediction: Whose is the Fallacy? *Cognition*, *8*, 385-407. doi:10.1016/0010-0277(79)90023-4
- Davidson, D. (1980). Psychology as philosophy. In D. Davidson (Ed.), *Essays on actions and events*. New York: Oxford University Press.
- Davidson, D. (1982a). Paradoxes of irrationality. In R. Wollheim, & J. Hopkins (Eds.), *Philosophical essays on Freud*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511554636.018
- Davidson, D. (1982b). Rational animals. *Dialectica*, *36*, 317-327. doi:10.1111/j.1746-8361.1982.tb01546.x
- Davidson, D. (1984). On the very idea of a conceptual scheme. In D. Davidson (Ed.), *Inquiries into truth and interpretation*. New York: Oxford Press.
- Davidson, D. (1984). Radical interpretation. In D. Davidson (Ed.), *Inquiries into truth and interpretation*. New York: Oxford Press.
- Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, *9*, 85-110. doi:10.1521/soco.1991.9.1.85
- Fisk, J. E. (2004). Conjunction fallacy. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook of fallacies and biases in thinking, judgment, and memory*. Hove: Psychology Press.
- Gould, S. J. (1992). *Bully for brontosaurus: Reflections in natural history* (p. 469). New York: Norton.
- Harten, D. S. (2008). How to solve the conjunction fallacy? A discussion of alternative approaches. Master's Thesis, Amsterdam: University of Amsterdam.
- Bar-Hillel, M., & Nater, E. (1993). How alike is it vs. how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology*, *65*, 1119-1131. doi:10.1037/0022-3514.65.6.1119
- Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275-305. doi:10.1002/(SICD)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, *118*, 248-271. doi:10.1037/0033-2909.118.2.248
- Johnson-Laird, P., & Wason, P. C. (1972). *Psychology of reasoning*. Cambridge: Harvard University Press.
- Klir, G., St. Clair, U., & Yuan, B. (1997). *Fuzzy set theory: Foundations and applications*. New Jersey: Prentice Hall.
- Kosko, B. (1990). Fuzziness vs. probability. *International Journal of*

¹²J. Katzman (founder of *The Princeton Review*) in an interview in *Frontline* (PBS, 1999) he says regarding SAT: "The SAT is said to predict freshman year grades in college, a little. And it does. It measures it a little. But the point is that it *doesn't* measure intelligence." In the same lines, from *Ten Myths About The SAT* (The National Center for Fair & Open Testing), we have: "The SAT cannot measure writing ability, strategic reasoning, higher order thinking skills, experience, persistence and creativity. See also: *Achievement Versus Aptitude Tests in College Admissions*, by R. C. Atkinson, 2001.

- General Systems*, 17, 211-240. doi:10.1080/03081079008935108
- Manktelow, K. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgment, and decision making*. New York: Psychology Press.
- Morier, D. M., & Borgida, E. (1984). The conjunction fallacy: A task specific phenomenon? *Personality and Social Psychology Bulletin*, 10, 243-252. doi:10.1177/0146167284102010
- Nielsen, R. H., & McKenna, T. (2003). A theory of thalamocortex. In R. H. Nielsen, & T. McKenna (Eds.), *Computational models for neuroscience*. London: Springer. doi:10.1007/978-1-4471-0085-0_4
- Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world*. East Sussex: Psychology Press. doi:10.4324/9780203345955
- Robertson, S. I. (1999). *Types of thinking*. London: Routledge.
- Shier, D. (2000). Can human rationality be defended a priori? *Behavior and Philosophy*, 28, 67-81.
- Siler, W., & Buckley, J. J. (2004). *Fuzzy expert systems and fuzzy reasoning*. New Jersey: Wiley-Interscience. doi:10.1002/0471698504
- Slugoski, B. R., & Wilson, A. E. (1998). Contribution of conversation skills to the production of judgmental errors. *European Journal of Social Psychology*, 28, 575-601. doi:10.1002/(SICI)1099-0992(199807/08)28:4<575::AID-EJSP882>3.0.CO;2-9
- Stanovich, K. E., & West, R. F. (1998). Individual differences in framing and conjunction effects. *Thinking and Reasoning*, 4, 289-317. doi:10.1080/135467898394094
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 22, 645-665. doi:10.1017/S0140525X00003435
- Stich, S. (1985). Could a man be an irrational animal? *Synthese*, 64, 115-135. doi:10.1007/BF00485714
- Stich, S. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge: MIT Press.
- Tanaka, K. (1996). *An introduction to fuzzy logic for practical applications*. New York: Springer.
- Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction. *Cognitive Science*, 28, 467-477. doi:10.1207/s15516709cog2803_8
- Tversky, A., & Kahneman, D. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315. doi:10.1037/0033-295X.90.4.293
- Zadeh, L. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23, 421-427. doi:10.1016/0022-247X(68)90078-4
- Zadeh, L. (1984). Fuzzy probabilities. *Information Processing & Management*, 20, 363-372. doi:10.1016/0306-4573(84)90067-0