A Note On Comparative Probability

Nick Haverkamp and Moritz Schulz

Penultimate draft. Please quote from the published version (Erkenntnis 2012).

Abstract

A possible event always seems to be more probable than an impossible event. Although this constraint, usually alluded to as *regularity*, is *prima facie* very attractive, it cannot hold for standard probabilities. Moreover, in a recent paper Timothy Williamson has challenged even the idea that regularity can be integrated into a comparative conception of probability by showing that the standard comparative axioms conflict with certain cases if regularity is assumed. In this note, we suggest that there is a natural weakening of the standard comparative axioms. It is shown that these axioms are consistent both with the regularity condition and with the essential feature of Williamson's example.

Comparative probabilities are studied for various reasons. As a starting point, one may be motivated by the observation that comparative probabilities are psychologically more fundamental than numerically measured probabilities. Usually, we compare events by thinking one to be more probable than the other or by taking them to be equiprobable without assigning numerical values to them. Thus, it seems that a comparative conception of probability provides a more realistic model of our uncertain beliefs.

An integral part of our pre-theoretic conception of probability is the so-called *regularity* constraint: An event which is possible is always taken to be more likely than an event which is impossible.[1] As a consequence, one would infer that an event which is not necessary is always thought to be less likely than an event which is necessary.[2] More formally, representing events as sets of possibilities, the regularity constraint can be expressed as follows:

(1) $A > \emptyset$, for every non-empty set of possibilities A.

The regularity constraint cannot always be satisfied in a plausible way by standard numerical probabilities if there are infinitely many events to be considered. This can be illustrated most easily by the famous dart board problem.

[1] The regularity condition has already been proposed by Finetti (1964: 100).

[2] Generally, we will presuppose a subjective conception of probability and possibility. But most of what is said will apply *mutatis mutandis* to objective kinds of probability and possibility.

There is a dart board with infinitely many points on it. A point-sized dart will be thrown randomly at the board. So, for any point i, it is possible that the dart hits i, and that i is hit is as likely as that any other point is hit. By regularity, the event that point i will be hit is more probable than the impossible event that i is hit and not hit. But since standard probabilities are additive and do not sum up to more than 1, infinitely many incompatible events cannot be assigned the same positive probability.[3]

Although one cannot always validate regularity in the standard real-valued framework, it seems possible to satisfy the regularity constraint within a comparative setting. One will look for plausible sets of comparative axioms which either contain or imply regularity. For instance, Finetti suggests a set of axioms for comparative probability which simply contains the regularity constraint as one of the axioms. Indirectly, the possibility of satisfying the regularity condition can thus be seen as an argument in favour of comparative probabilities.[4]

However, Williamson argues in a recent paper that the problem of regularity is deeper than it has been assumed.[5] It is shown that there are cases which indicate that the regularity constraint clashes with one of the major axioms of comparative probability. Hence, Williamson's argument is a challenge for the idea that regularity can be satisfied in a comparative framework which is strong enough to be a plausible candidate for modeling the structure of rational epistemic states.

In this note, we will take Williamson's example seriously and discuss some of its consequences for the prospects of integrating the regularity constraint into a comparative framework. Before we come to this, we will have a quick look at a recent paper by Weintraub (2008) which calls Williamson's example into question. In the main part of this paper, we will then show that there are natural comparative axioms which are consistent both with the regularity constraint and with the essential feature of Williamson's sample case.

Let us start by looking at Williamson's example. There is a chance device which tosses a fair coin infinitely many times. The tosses are probabilistically independent of each other. Let B be the event that the coin always comes

[3] It should be noted that the regularity constraint can be satisfied in a setting with countably many possible outcomes if the outcomes are *not* equiprobable. Yet as soon as there are uncountably many possible outcomes, the regularity constraint becomes unsatisfiable.

[4] Of course, standard probabilities will remain a valuable tool for modeling various other important phenomena.

[5] Cf. Williamson (2007).

up heads, let C be the event that the first toss comes up tails and every toss afterwards comes up heads, and let D be the event that the coin either comes up heads or tails on the first toss and always comes up heads afterwards, i.e. $D = B \cup C$. Williamson argues that B is as likely as D (formally, $B \approx D$). To this effect, he introduces a second chance device which is physically isomorphic to the first chance device but which tosses a fair coin for the first time when the first chance device tosses the coin for the second time. The two chance devices then proceed to work in a completely synchronical way. Now, let B^* be the event that the coin always comes up heads within the second chance device. Since the two chance devices are physically isomorphic and it does not make a difference when one starts to toss the coins, B is as probable as B^*. We should think it equally likely that the coin always comes up heads within the first chance device and that the coin always come up heads within the second chance device if these two devices agree in all relevant physical respects. But since the coins are tossed synchronically after the first toss of the first device, the chance that the first device produces always heads after the first toss equals the chance that the second device produces always heads, i.e. D is as probable as B^*. Finally, by the transitivity of equiprobability, one finds that B is as probable as D, i.e. $B \approx B \cup C$.

In a reply to Williamson (2007), Weintraub (2008) doubts the cogency of Williamson's example. She objects:

> The *set of temporal points occupied by one sequence* is a proper subset of those occupied by the second sequence. So the two sequences do not share *all* their physical properties. And this serves – yet again – to show that isomorphism, despite its seeming stringency, isn't sufficient for equiprobability, thus blocking the paradoxical reasoning. (Weintraub 2008: 249)

Weintraub points to the fact that the two chance devices start at different times. The second chance device begins to work an instance later: it tosses a fair coin for the first time when the first chance device tosses a fair coin for the second time. As a consequence, the set of temporal points at which a coin is tossed by the second chance device is a proper subset of the set of temporal points at which a coin is tossed by the first chance device.

The important question is, of course, how the different starting time is supposed to effect the corresponding probabilities. Williamson's example involves the following assumption:

(2) The event B that the first chance device produces always heads is as likely as the event B^* that the second chance device produces always heads.

This assumption seems plausible enough. After all, the two chance devices agree completely in their physical make-up, so how could the probabilities that they produce a sequence of always heads be different?

Weintraub seems to think that the different starting time can make for a difference in probability despite the complete congruence of the physical make-up. As far as we can see, her idea must be that the first chance device has to produce heads at one time more, namely when the coin is tossed for the first time, than the second chance device which only starts one instance later. For this reason, the first chance device producing a sequence of always heads would be a little less likely than the second chance device producing a sequence of always heads.

It is not clear to us whether one should follow Weintraub in this argument. Weintraub's approach has a fairly extreme and in our eyes unattractive consequence which becomes visible when we consider multiple runnings of the two chance devices.[6] Suppose that we repeat the experiment at some time in the future, but this time the first chance device starts an instance later than the second chance device. By the same line of argument, Weintraub will now need to hold that it is a little more likely that the first chance device produces always heads than that the second chance device, which now starts one instance earlier, delivers always heads. But on this assumption, we would need to accept cases in which *two runnings of the same chance device* are not governed by the same probabilities, e.g. the probability that the first device produces always heads at its first running would not be the same as the probability of it producing always heads at its second running. To see why this would need to be accepted, assume that the latter were not the case, i.e. the two runnings of the first chance device were governed by the same probability. Now, the first chance device producing always heads at the first running is, by Weintraub's argument, a little less likely than the second chance device producing always heads at its first running. Similarly, the first chance device producing always heads at the second running is a little more likely than the second chance device producing always heads at its second running. Jointly, it follows that the probability that the second chance device produces always heads at its first running does not equal the probability that it produces always heads at its second running, in fact the former will

[6]For such multiple runnings of an experiment involving infinitely many tasks to be possible, the tasks need to be performed at an accelerating and converging time scale so that a coin is tossed infinitely many times within a finite amount of time. For instance, the coin may be tossed for the first time at point 0, for the second time at $1/2$, for the third time at $2/3$, and so on and so forth; the n-th toss of the coin occurring at $1 - 1/n$.

be larger than the latter. So, not only would we need to accept cases where two chance devices with the same physical make-up are governed by different probabilities, we would also have to swallow that running the same experiment on the same chance device twice may not produce a certain outcome with the same probability.

Since Weintraub's suggestion comes with considerable costs, we think it worthwhile to explore the possibility that Williamson's example is genuine. This is not to say that it cannot be rejected. However, at the present stage of the debate, rejecting it will be fairly costly in allowing a surprising dependence of the probability of an event on its time of occurrence, which would dramatically change our picture of the basis of objective chances. But only following up the possibility that Williamson's example might be genuine will put us into a position to compare the costs of accepting it with the costs of rejecting it.

The essential feature of Williamson's example is that there are two events B and C with the following property:

(3) $B \approx B \cup C$, where B and C are non-empty and disjoint sets and $B \cup C$ is not the universal set of all possibilities.

Given the regularity constraint, the existence of events which satisfy (3) conflicts with the principle of additivity usually assumed to govern comparative probabilities. The principle of additivity states that for all events A, B and C, the following is true:

(!) $A \geq C$ iff $A \cup B \geq C \cup B$, if A and C are disjoint from B.

This principle, which can be seen as a comparative variant of the principle of additivity for standard probabilities, states that the comparative relation is preserved under the addition of a set of possibilities which is incompatible with each of the two original sets of possibilities.[7]

To see that (!) is violated by the existence of events satisfying (3) if regularity is assumed, take two sets B and C as in (3). Since C is non-empty, the regularity constraint implies $\emptyset \not\geq C$. Since B is non-empty and disjoint from C, principle (!) would allow us to infer $\emptyset \cup B \not\geq C \cup B$ contradicting $B \approx B \cup C$. Thus, if regularity is assumed and Williamson's example is granted, one cannot stick to the additive principle (!). To reiterate, the regularity constraint

[7] This principle is introduced by Finetti (1964: 100f.). For a discussion of this and related principles, see Fishburn (1986).

requires a possible but "infinitely improbable" event to be more probable than the impossible event, but in cases like Williamson's example such an event is not supposed to raise the probability if added to certain non-empty sets of possibilities disjoint from it. This, of course, conflicts with the idea that the comparative relation is always preserved under the addition of disjoint sets of possibilities.

In the light of Williamson's example it is an interesting question whether one can satisfy regularity in a plausible comparative framework which will then, of course, not imply the additive principle (!). In what follows, we start by focussing on a natural set of comparative axioms. By introducing *regularizations* of comparative probability relations we then show that these axioms are consistent both with regularity and the existence of events satisfying (3).

To fix ideas, let Ω be a non-empty set (the *universal* set of possibilities), and let $\mathcal{A}$ be a Boolean algebra on Ω.[8] We set $\overline{A} := \Omega \setminus A$. Now, let us call a relation $\leq$ on $\mathcal{A}$ a (*comparative*) *probability relation* if it satisfies the following axioms:[9]

(i) $\leq$ is transitive (*Transitivity*).

(ii) $A \leq A \cup B$, for all $A, B \in \mathcal{A}$ (*Monotonicity*).

(iii) If $A \leq B$, then $\overline{B} \leq \overline{A}$, for all $A, B \in \mathcal{A}$ (*Complementarity*).

By the first axiom, if A is at least as likely as B and B is at least as likely as C, then A is at least as likely as C. The second axiom requires that it is always at least as likely that A or B occurs as it is that A alone occurs. Finally, the third axiom states that if B is at least as likely as A, then it is at least as likely as that A does not occur as it is that B does not occur.

Let us note a few consequences of these axioms. Monotonicity implies that a probability relation is reflexive, for $A \leq A \cup \emptyset$. Hence, a probability relation is a quasi-ordering.[10] The axioms do not require that two events are always comparable and they do not exclude the trivial ordering according to which any two events are equiprobable. These constraints could be added if one is inclined to do so; in the discussion to follow, nothing hangs on this choice.

---

[8] Recall that $\mathcal{A}$ is called a *Boolean algebra* on Ω if $\mathcal{A}$ is a set of subsets of Ω which is closed under union and complementation and which contains Ω.

[9] Adopting the nonstrict relation $\leq$ as basic, the relations $\approx$ ('equiprobable') and $<$ ('less probable than') can be defined in the usual way:

$$A \approx B \quad :\Leftrightarrow \quad A \leq B \,\wedge\, B \leq A \qquad \text{and} \qquad A < B \quad :\Leftrightarrow \quad A \leq B \,\wedge\, B \not\leq A.$$

[10] Sometimes transitivity is doubted, but we would like to think of the axioms as stating epistemic norms rather than psychological facts.

Note that the second axiom implies that nothing is more likely than the necessary event and that everything is at least as likely as the impossible event: $\emptyset \leq A \leq \Omega$ for every $A \in \mathcal{A}$. The second axiom is equivalent to the requirement that if $A \subseteq B$, then $A \leq B$. So, if A cannot occur without B, i.e. if A implies B, then B is at least as likely as A. Generally, we were looking for a mathematically natural set of axioms which is strong enough to be a plausible weakening of stronger systems comprising additive principles such as (!).

How do these axioms square with regularity? To answer this question, let us call a probability relation $\leq$ *regular* if $A \not\leq \emptyset$ for every non-empty set of possibilities A. By complementarity, if $\leq$ is regular, $\Omega \not\leq A$ for every set of possibilities A distinct from Ω. Thus, by defining $\mathcal{A}^* := \mathcal{A} \backslash \{\emptyset, \Omega\}$, a probability relation is regular iff $\emptyset < A < \Omega$ for every $A \in \mathcal{A}^*$. Now, there is a surprisingly simple method of turning a non-regular probability relation into a regular one which differs only minimally from the original one. The idea is to refine the original ordering by making the empty set less likely and the universal set more likely than any other set of possibilities. Apart from that, nothing changes.

More precisely, given a comparative probability relation $\leq$ on $\mathcal{A}$, we define the *regularization* $\leq^*$ of $\leq$ by setting $A \leq^* B$ iff

(i) $A \leq B$ and $A, B \in \mathcal{A}^*$, or

(ii) $A = \emptyset$ or $B = \Omega$.[11]

What is the effect of regularizing a probability relation? The probability of an impossible event on the one hand and an infinitely improbable event on the other hand can now be distinguished.[12] Suppose that in the original ordering the empty set of possibilities and a certain non-empty set A were equiprobable. Then the regularization will remove this feature. To see this, suppose $A \leq^* \emptyset$. Since $\emptyset \notin \mathcal{A}^*$, clause (i) is not satisfied. Hence, clause (ii) must be satisfied. This means that either $A = \emptyset$ or $\emptyset = \Omega$. But Ω is never empty, and so we can infer that $A = \emptyset$. What we have shown is that if $A \leq^* \emptyset$, then A is the empty set. This shows that no possible event is equiprobable with the impossible event in the regularization of the original ordering. In the same way, the necessary event becomes more likely than any other event.

It is not hard to verify that the resulting relation $\leq^*$ is again a comparative

[11] In set-theoretical terms: $\leq^* := \leq \backslash \big(\{(A, \emptyset) : A \in \mathcal{A} \wedge A \neq \emptyset\} \cup \{(\Omega, A) : A \in \mathcal{A} \wedge A \neq \Omega\}\big)$.

[12] If $\leq$ is already regular, we have $\leq^* = \leq$.

probability relation which is, as we have just seen, regular.[13] Hence, the axioms are consistent with regularity. Moreover, if there are events satisfying (3) with respect to the original relation, then these events satisfy (3) with respect to the regularization, for (3) concerns only events in $\mathcal{A}^*$. Taken together, this shows that the axioms are consistent with the conjunction of regularity and the assumption that there are sets of possibilities satisfying (3). Therefore, we can be sure that within the suggested comparative setting, regularity does not conflict with Williamson's example.

Of course, in general it will not be the case that if a probability relation satisfies (!) for all events, its regularization will do so as well. Its regularization will only have the slightly weaker property that (!) is true if all compared events are *contingent*, i.e. if the events A, B, C are such that $A, C, A \cup B, C \cup B \in \mathcal{A}^*$.[14] In a way, this observation can be seen as estimating the price one has to pay for regularity. In infinite settings, the assumption that all events satisfy (!) can only be maintained if no principled distinction is drawn between *contingent* events on the one hand and the *impossible* and the *necessary* event on the other hand. Once such a distinction is wanted, an additive principle like (!) can only be true with respect to contingent events.[15]

In conclusion, we hope to have shown that one can deal with regularity in a natural setting for comparative probability even in the light of Williamson's example.

Acknowledgements

We would like to thank the other members of the *Phlox* research group, Miguel Hoeltje, Benjamin Schnieder and Alexander Steinberg, for very helpful comments and support. Many thanks are also due to two anonymous referees of this journal and the participants of *Phloxshop*

[13] If one were to require a comparative probability relation to be linear and/or non-trivial, the class of comparative probability relations would still be closed under regularization.

[14] This is obvious, since the regularization agrees with the original relation on the relevant cases.

[15] The operation of *regularization* can be extended to the standard case, where probability distributions $P : \mathcal{A} \to [0, 1]$ are considered. One would expand the image of P by adding two new elements $s < 0$ and $1 < t$. Then a new function P^* could be defined which differs from P only in mapping $\emptyset$ to s and Ω to t. The induced quasi-odering of P^* would then be the regularization of the quasi-ordering induced by P. Similarly to the comparative case, and as one would expect anyway, the new function does not satisfy the numerical version of additivity: if $A \in \mathcal{A}^*$, then $t = P^*(A \cup \overline{A}) \neq P^*(A) + P^*(\overline{A}) = 1$. On the other hand, its restriction to $\mathcal{A}^*$ coincides with P's restriction to $\mathcal{A}^*$, and, as a standard probability distribution, it maps $\emptyset$ to the smallest and Ω to the largest value in the linearly ordered value space. To the extent that one takes additivity to be an essential feature of numerical probability, it might be debatable whether one is willing to call P^* a measure of probability.

I, Berlin 2008, at which an earlier version of the present paper was presented. Special thanks to Timothy Williamson whose comments on multiple drafts of this paper have led to significant improvements. Finally, research for this paper has profited from the generous support of the Deutsche Forschungsgemeinschaft (DFG).

References

[1] Finetti, B. de (1964): "Foresight: Its Logical Laws, Its Subjective Sources", in H. Kyburg and H. Smokler, eds., *Studies in Subjective Probability*, 93-158. Huntington, N.Y.: Krieger.

[2] Fishburn, P.C. (1986): "The Axioms of Subjective Probability", *Statistical Science* **1**, 335-345.

[3] Weintraub, R. (2008): "How probable is an infinite sequence of heads? A reply to Williamson", *Analysis* **68**, 247-50.

[4] Williamson, T. (2007): "How probable is an infinite sequence of heads?", *Analysis* **67**, 173-80.