KNOWLEDGE, SAFETY, AND GETTIERIZED LOTTERY CASES: WHY MERE STATISTICAL EVIDENCE IS NOT A (SAFE) SOURCE OF KNOWLEDGE

Fernando Broncano-Berrocal
Autonomous University of Madrid

Abstract: The lottery problem is the problem of explaining why mere reflection on the long odds that one will lose the lottery does not yield knowledge that one will lose. More generally, it is the problem of explaining why true beliefs merely formed on the basis of statistical evidence do not amount to knowledge. Some have thought that the lottery problem can be solved by appeal to a violation of the safety principle for knowledge, i.e., the principle that if *S* knows that *p*, not easily would *S* have believed that *p* without *p* being the case. Against the standard safety-based solution, I argue that understanding safe belief as belief that directly covaries with the truth of what is believed in a suitably defined set of possible worlds forces safety theorists to make a series of theoretical choices that ultimately prevent a satisfactory solution to the problem. In this way, I analyze several safety principles that result from such choices—the paper thus gives valuable insights into the nature of safety—and explain why none solves the lottery problem, including their inability to explain away Gettierized lottery cases. On a more positive note, I show that there is a viable solution in terms of safety if we get rid of the unquestioned assumption that safe beliefs directly track the truth. The alternative is a conception of safe belief according to which what safe beliefs directly track is the appropriateness of the circumstances and, indirectly, the truth. The resulting safety principle, I argue, explains why mere statistical evidence is not a safe source of knowledge.

1. Introduction: safety and the lottery problem

It is fair to say that views based on the *safety principle* (if an agent *S* knows a proposition *p*, not easily would *S* have believed that *p* without *p* being the case) stand among the most popular in the epistemological landscape.[1] Several good reasons explain this popularity. According to safety theorists, the principle helps rule out *Gettier-style cases* as cases of knowledge in a way that reveals how knowledge is incompatible with *epistemic luck*. Second, unlike other modal principles such as the Nozick's (1981) *sensitivity principle* (if *S* knows that *p*, had *p* been false, *S* would not have believed that *p*), it helps explain why we can know the *denial of skeptical*

[1] The safety principle was influentially introduced by Sosa (1999), but can be traced back to Luper (1984).

hypotheses. Third, epistemologists tend to prefer safety because, unlike sensitivity, it doesn't seem to violate *knowledge closure* (roughly, the principle that if *S* knows that *p* and deduces *q* from *p* while retaining knowledge of this entailment, then *S* knows that *q*). Finally, unlike sensitivity, safety is said to be able to account for knowledge that results from *inductive reasoning*. In sum, if safety has all the virtues its supporters claim it has, it is a principle worth having in one's epistemological toolbox.

In spite of these virtues, safety theorists have a pebble in their shoes: the so-called *lottery problem*. The lottery problem is the problem of explaining why mere reflection on the long odds that one will lose the lottery does not yield knowledge that one will lose.[2] More generally, it is the problem of explaining why true beliefs *merely* formed on the basis of statistical evidence do not amount to knowledge.

Unless more is said, the foregoing formulation of safety does little to solve the problem. After all, if you buy a lottery ticket, you wouldn't lose very easily, so your belief that you will lose is safe. The way safety theorists typically try to get around this problem—call it the *standard safety-based solution*—is by tailoring the notion of *easy possibility*. In particular, their preferred *modal conception* has it that the occurrence of an event is easily possible if and only if the event would occur in possible worlds that are close to the actual world.[3] On the resulting notion of safe belief, safe beliefs directly covary with the truth of what is believed in close possible worlds. Drawing on this, the standard solution argues that there is a structural mismatch between knowing that certain empirical facts will occur and ignoring lottery results. For instance, it's a modally robust fact that you won't become a millionaire by writing philosophical papers: given how things stand in the actual world, too much would need to change for philosophical papers start making profit (for philosophers) in some close possible world. Thus, if you believe that you won't become a millionaire by writing philosophy, your belief is safe and knowledgeable—you would continue to get things right in close possible worlds in which you believe that. Now, supposing that the likelihood of becoming filthy rich by writing philosophy is the same as by playing the lottery, shouldn't we draw the same conclusion? Contrary to what one might think, the proponent of the standard solution argues, the lottery case is special, because all it takes for you to win is a slightly different configuration of the balls in the lottery machine, which means that the *unlikely* event of winning would still occur in close possible worlds. Presumably, in some such possible worlds you would continue to believe, falsely, that you will win, so your belief in the lottery proposition is unsafe and, therefore, unknowledgeable.

Against the standard safety-based solution to the lottery problem, I will argue that understanding safe belief as belief that *directly* covaries with the truth of what is believed in a

[2] The orthodox position in epistemology is that one *ignores* that one will lose a fair lottery if this judgment is only based on the statistical evidence available, namely on the fact that it is overwhelmingly likely that one will lose the lottery. For some notable exceptions, see Reed (2010), Gibbons (2013), and Sosa (2015).

[3] A competing conception of 'easy possibility' is *probabilistic*: the occurrence of an event is easily possible if and only if it is very likely that it will occur. For the most part, both conceptions coincide in their predictions, because what is likely in the actual world *typically* occurs in close possible worlds, and the other way around.

suitably defined set of possible worlds forces safety theorists to make a series of theoretical choices that ultimately prevent a satisfactory solution to the problem. To illustrate the point, I will analyze different safety principles that result from such choices and explain why none solves the lottery problem, including their inability to explain away *Gettierized lottery cases*. Although this may lead to the pessimistic conclusion that safety theorists are unable to explain why mere statistical evidence is not a source of knowledge, I will show that there is a viable solution in terms of safety if we get rid of the unquestioned assumption that safe beliefs *directly track the truth*. The proposed alternative is a conception of safe belief according to which what safe beliefs directly track is the appropriateness of the circumstances and, indirectly, the truth. The resulting safety principle, I will argue, explains why mere statistical evidence is not a safe source of knowledge.

2. Two theoretical junctures

As we have just seen, key to the standard safety-based solution to the lottery problem is the idea that, however unlikely it is, winning the lottery is a possibility that would materialize in close possible worlds, so believing that you will lose the lottery is unsafe because you could *easily* have believed a false proposition. Of course, the fact that you would *lose* and therefore continue to get things right in many close possible worlds (presumably, in most of them) immediately gives rise to the following question: *how much risk of error* does knowledge tolerate and, accordingly, how much risk of error should the safety principle prescribe? In other words, what is the threshold of close possible worlds beyond which messing up epistemically makes a belief fall short of safety and hence of knowledge?

There are at least two options available to the safety theorist, as Williamson (2009) helpfully explains. One is the view that a belief that *p* can be safe when there is *at most* a small risk of error. Williamson calls this the '*small risk of error*' conception of safety. An alternative view is the 'no risk of error' conception, which tolerates no risk of being mistaken about *p* *whatsoever*. However, such an infallibilist requirement on knowledge is too demanding and not in the spirit of the neo-Moorean intuitions that motivate the safety principle in the first place. So a better way to understand it is what Williamson calls the '*no close risk of error*' conception. Put in terms of possible worlds, this conception holds that *S*'s true belief in *p* is safe only if *S* doesn't erroneously believe that *p* in *any* close possible world. By contrast, the 'small risk of error' conception permits that *S*'s true belief in *p* counts as safe even when *S* makes a mistake about *p* in some close possible world.[4]

The second choice that safety theorists must make—especially when adopting a possible worlds framework—concerns *degrees of safety*. For example, is your belief that you will win the lottery *as* safe (or unsafe) as the belief that you will lose, i.e., are these two beliefs safe to the same degree? Whether you think that the former belief is safer (e.g., because you would lose in

[4] I avoid the complication of whether a belief can be unsafe when an agent would easily believe false propositions that are distinct but closely related to the proposition actually believed. This complication is irrelevant for the purposes of the paper. For discussion of this issue, see Broncano-Berrocal (2016).

more close possible worlds) or that both beliefs are safe to the same degree (e.g., because possible worlds in which you lose are *as close* to the actual world as possible worlds in which you win), it is something that requires a principled explanation of the gradualness of safety.

In this sense, following again helpful conceptual work by Williamson (2009), there are two theoretical options available. One is that the degree to which an agent *S* is safe from error with respect to *p* is a function of *how large* the proportion of close possible worlds in which *S* falsely believes *p* is. This is the *proportion view* of safety. An alternative is that how much *S* is safe from error vis-à-vis *p* is a function of *how distant* (i.e., dissimilar) from the actual world possible worlds in which *S* falsely believes *p* are. This is the *proximity view* of safety.

3. Strong safety

By combining the 'no close risk of error' conception of safety with the proportion view, we can come up with the following formulation of the principle, which has been dubbed 'strong safety' elsewhere:[5]

> *Strong Safety*: If *S* knows that *p*, then *S*'s belief that *p* is true in the actual world and in *no* close possible world in which *S* forms the belief that *p* via the same type of method *S* uses in the actual world, *p* is false.[6]

Strong Safety apparently gets the lottery intuition right. After all, while the proposition that you will lose the lottery is true in a majority of close possible worlds (on the assumption that the proportion view is correct), there are still some close possible worlds in which you win the lottery (by stipulation, the lottery always has a winner), which means that there are some close possible worlds in which your belief is false. Since *Strong Safety* has zero tolerance for error in close possible worlds, your belief is unsafe.

However, explaining the lottery intuition comes at a cost: not being able to account for knowledge acquired by inductive methods. Consider this often-discussed case of *inductive knowledge* by Sosa (1999):

> *Trash Bag*
> On his way to the elevator, Ernie releases a trash bag down the garbage chute from his high rise condo. He does this every morning and has good reason to think that his trash bags have reached the basement each time—e.g., he has checked several times in the past that the bags have, in effect, reached the basement; so far, the waste collection service has never reported any anomaly; quite often, he has seen garbage trucks filled with trash bags leave the condo's basement, and so on. Presumably, this time Ernie also knows that the trash bag he has just released will soon be in the basement. (adapted from Sosa 1999)

[5] See Greco (2007).
[6] See Pritchard (2005) and Williamson (2009).

Strong Safety rules out *Trash Bag* and similar cases of inductive knowledge as cases of knowledge.[7] The culprit is the standard way to cash out the notion of closeness in terms of similarity between possible worlds that *Strong Safety* assumes: although possible worlds in which the bag is snagged in the chute are less similar to the actual world than possible worlds in which this doesn't occur, the possibility that the bag gets snagged is not so bizarre not to figure in the range of close (i.e., similar) possible worlds, unlike skeptical possibilities such as Matrix-like scenarios, which involve too great a departure from the actual world. In other words, we may well presume that there is at least one similar enough possible world in which Ernie forms a belief about the trash bag being in the basement when it is not, which suffices to violate *Strong Safety*.

4. Weak safety

Not being able to account for inductive knowledge is a price too high to pay. After all, a significant subset of the propositions we know are propositions known by inductive methods. In comparison, accounting for the lack of knowledge of lottery results does not seem so important from an epistemological perspective, not least from a practical viewpoint.

The safety theorist might accordingly come to the conclusion that *Strong Safety* needs to be weakened. For instance, the 'no close risk of error' conception of safety could be replaced with a 'small risk of error' conception and be combined with the proportion view in the following way:

> *Weak Safety*: If *S* knows that *p*, then *S*'s belief that *p* is true in the actual world and in *most* close possible worlds in which *S* forms the belief that *p* via the same type of method *S* uses in the actual world, *p* is true.

Weak Safety gets cases of inductive knowledge right. For example, it considers Ernie's belief in *Trash Bag* safe, because in most close possible worlds the bag reaches the basement, even if in some this is not the case. However, weakening *Strong Safety* comes at the cost of getting lottery cases wrong. After all, in most close possible worlds in which you believe that you will lose the lottery, you lose, which means that, as per *Weak Safety*'s standards, your true belief that you will lose is safe.

One might think that this is not a price too high to pay if one can account for inductive knowledge in exchange. Be that as it may, this bolsters the point that, despite all the virtues safety-based epistemology might have, safety theorists have a pebble in their shoes: the lottery problem. In fact, Greco (2007) presents the delicate dialectical situation of safety theorists as a *dilemma*: if one weakens *Strong Safety* in order to account for inductive knowledge, one fails to solve the lottery problem (*first horn*); if one strengthens *Weak Safety* in order to solve the lottery

[7] See Vogel (2007) for other cases.

problem, one fails to account for inductive knowledge (*second horn*). Either way, safety-based epistemology falls short of theoretical adequacy.[8]

5. Hybrid safety

Pritchard (2007a; 2007b) agrees with the second horn of the dilemma: *Weak Safety* is needed to account for inductive knowledge. However, he thinks that one can steer clear of the first horn by endorsing a qualified version of *Strong Safety*, which is motivated by the idea that if one doesn't know that one will lose the lottery, it is not because there are *a few* close possible worlds in which one's belief is false, but by the fact that such worlds are actually *very close*. The resulting safety principle conjoins a 'small risk of error' requirement ranging over close possible worlds with a 'no risk of error' requirement ranging over *very* close possible worlds:

> *Hybrid Safety*: If *S* knows that *p*, then *S*'s belief that *p* is true in the actual world and (i) in *most* close possible worlds in which *S* forms the belief that *p* via the same type of method *S* uses in the actual world, *p* is true, and (ii) in *no* very close possible world in which *S* forms the belief that *p* via the same type of method *S* uses in the actual world, *p* is false.

While *Hybrid Safety* has small tolerance for error in close possible worlds, it tolerates no error in very close possible worlds. The former is meant to accommodate inductive knowledge. The latter is introduced to handle the lottery problem. Suppose that you come to believe you will lose the lottery by reflecting on the odds. The proportion of nearby worlds in which you form a false belief in that proposition via this method is very small, indeed. Crucially, however, some of those possible worlds are very close to the actual world (at least this is what Pritchard thinks). Since *Hybrid Safety* tolerates no error in very close possible worlds, your belief is unsafe, in keeping with the lottery intuition.

However, *Hybrid Safety* is, to a large extent, *ad hoc*. Clauses (i) and (ii) are independently introduced to handle inductive knowledge and lottery cases respectively. That being so, it is unclear why they should be part of a single safety principle instead of two independent principles specifically tailored to handle counterexamples; but, of course, that a principle handles counterexamples cannot be the only theoretical motivation for it. Further justification is required.

If this is not enough, the main reason for abandoning *Hybrid Safety* as a safety-based solution to the lottery problem is that, although it can get the intuitions about some cases right, it cannot accommodate the intuition of ignorance in *all* lottery cases. Consider the following examples:

> *Eaten Ticket*
>
> Lisa bought a ticket for her twin sister Lottie and put it in her drawer to keep it there until the lottery draw. Lottie comes to know that Lisa bought a ticket for her and, based on the

[8] See Dodd (2012) for the same kind of dilemma.

odds, comes to believe that she won't win. As a matter of fact, the winning number is Lottie's ticket number. Unbeknownst to both, when cleaning the drawer, the household helped drop the ticket on the floor where it was swiftly eaten by the dog (weeks ago), so Lottie no longer owns the ticket.

Address Lottery
In order to stimulate the housing market and incentivize home sales, the government creates *Address Lottery. A Lottery for Property Owners*. To participate in Address Lottery, one needs to own a home. Lottery tickets are not arbitrarily purchased but the addresses of the participants become their ticket numbers after paying an annual subscription fee, which entitles them to participate in several lottery draws per year. On each of those occasions, a random system chooses one of the registered addresses as the winner, whose owner needs to prove that she effectively owns the property so as to become the official winner (e.g., tenants cannot collect the prize despite living in the winning home). As always, Lottie, an Address Lottery subscriber and flat owner, reflects on the odds and concludes that this time, she will lose the lottery again. Unbeknownst to Lottie, some time ago a hedge fund decided to purchase all flats in her building for reasons that have nothing to do with the lottery (an unjust law sanctions such purchases but never after lottery draws). Coincidentally, the purchase order has been scheduled and made effective (again, for completely independent reasons) *just before* Lottie is about to win the lottery. In this way, while her address is finally chosen as the winner, she doesn't win the lottery. To the surprise of the investors, the hedge fund does.

Pritchard's *Hybrid Safety* fails to explain why Lottie doesn't know that she won't win in *Eaten Ticket* and *Address Lottery*. To see this, let's consider the structure of the cases. Both involve *coincidences*, where a coincidence is composed of two events occurring simultaneously without a common close nomological or explanatory antecedent connecting them.[9] In *Eaten Ticket*, the two events composing the relevant coincidence are (1) the result of the lottery corresponding to Lottie's ticket number and (2) Lottie's ticket being eaten by the dog. In *Address Lottery*, the two events composing the relevant coincidence are (1) the result of the lottery corresponding to Lottie's address and (2) the purchase of Lottie's flat by the hedge fund.

Now, if we fill in the details of the cases so that events of type (2) are bound to occur, in the sense that there is no chance whatsoever that they would not happen in *all* close possible worlds, then in *no* close possible world events of type (1) make the relevant lottery propositions false. Furthermore, if we fill in the details of the cases in this way, not even *Strong Safety* is able to explain the intuition that Lottie doesn't know that she will lose, because, once again, given the

[9] See Owens (1992) for this account of coincidence.

occurrence of type-(2) events in all close possible worlds, in no close possible world Lottie's belief is falsified by a favorable lottery result.[10]

Eaten Ticket and *Address Lottery* are *Gettierized lottery cases*. They are *lottery cases* because the target beliefs are (1) about the outcomes of a random lottery and (2) formed merely on the basis of evidence about lottery odds (i.e., on the basis of mere statistical evidence rather than, say, testimonial evidence). They are *Gettierized* because they involve the kind of coincidences that are typical of Gettier cases or, as Zagzebski (1994) puts it, the kind of double structure of luck that gives rise to Gettierized beliefs: an agent with good justification to believe something (e.g., solid statistical evidence that she will lose the lottery by holding ticket *n*) has the bad luck that, in spite of her good evidence, her belief is a candidate false belief (because ticket *n* will become the winner). Yet, due to a stroke of good luck (the intervention of an external factor, such as the dog or of the hedge fund), her belief turns out true. Interestingly, many epistemologists (especially many safety-theorists) assume without argument that whatever factor produces the second stroke of good luck must be a factor that occurs *but could easily have failed to occur*. This is an obvious mistake. Nothing in the description nor the structure of standard Gettier cases hangs on this: the factor that produces the second stroke of good luck (or better said, the second component of the relevant Gettier coincidence) may well be, as *Eaten Ticket* and *Address Lottery* exemplify (and many other Gettier cases for that matter), a modally robust, or even an inexorable factor.[11] In other words, most Gettier cases, including Gettierized lottery cases, are generated by coincidences, not by the fact that one of their components could easily have failed to obtain.[12]

[10] Pritchard could still try to argue that in some *very* close possible world Lottie wins the lottery. Unfortunately, if we fill in the details of the cases so that more significant changes to actuality are needed than a mere change in the configuration of the balls to prevent type-(2) events from occurring, then *very* close possible worlds will be worlds in which such events *occur* (e.g., in which the hedge fund still purchases Lottie's flat), which means that Lottie's belief remains safe.

[11] Consider Gettier's original cases (Gettier 1963). What if the job was *never* meant to be for Jones, but for Smith only, and what if Smith would *never* leave his home without ten coins in his pocket? What if there is *no chance whatsoever* that Brown is in a place different from Barcelona? Consider Russell's stopped clock case (Russell 1948: 170-171). What if the person in question would *never* have looked at the clock at a different time? Consider Chisholm sheep-in-the-field case (Chisholm: 1997). What if the sheep was, is and will *always* be chained to the rock? In such cases, the relevant coincidences (e.g., forming the true belief that there is a sheep in the field while there is a sheep that would *never* leave the field) make the target beliefs lucky, but *such coincidences are modally robust*. For related discussion, see Miracchi (2015) for a procedure for generating what she calls 'systematic Gettier cases', i.e., scenarios in which the elements of bad and good luck that are typical of Gettier cases are systematically related in the following way: *whenever S* has the bad luck that her belief is a candidate false belief (despite having good epistemic justification for it), *S* has the good luck that such a belief ends up being true.

[12] In Gettier cases with the structure of Goldman's fake barn case (Goldman 1976), the fact that the agent could have easily been mistaken definitely plays a role in giving rise to the ignorance intuition, but because they have a different structure than standard Gettier cases, they are treated differently (see, e.g., Pritchard 2012). My claim is that in *standard* Gettier cases the kind of knowledge-undermining luck at issue (whether one wants to call it 'intervening' or otherwise) always arises from a *coincidence*, and a feature shared by all coincidences is that they involve luck independently of the modal robustness or fragility of their components. For discussion of the notion of

6. Proximity safety

That, by *Hybrid Safety*'s standards, Gettierized lottery cases are cases of safe belief is bad news for Pritchard. That, by *Strong Safety*'s standards, they also qualify as safe is bad news for the prospects of a safety-based solution to the lottery problem, as it does not matter whether one adopts a 'no close risk of error' or a 'small risk of error' conception of safety: the lottery problem cannot be solved that way.

One plausible diagnosis is that what prevents the foregoing safety principles from giving a successful solution to the problem is the very *proportion view* of safety, i.e., the idea that degrees of safety are to be modeled in terms of how large the proportion of close possible worlds in which relevant errors occur is. For one might be perfectly safe from error in this sense while still lacking knowledge of the relevant proposition—as we have seen, this happens in cases in which the relevant proposition is inexorably true in all close possible worlds due to a modally robust coincidence.

At this point, the safety theorist might think that the best she can do is to give up the proportion view altogether and opt for a safety principle entirely based on a *proximity view* instead. Here is one possible formulation of that kind of safety principle:

> *Proximity Safety*: If *S* knows *p*, then *S*'s belief that *p* is true in the actual world and close possible worlds in which *S* believes *p* truly via the same method of belief formation are *closer* to the actual world than close possible worlds in which *S* believes *p* falsely via that method.

Proximity Safety is superior to the other versions of safety in several respects. First, it is better than *Strong Safety* in that, like *Weak Safety*, is compatible with inductive knowledge. Consider *Trash Bag*: close possible worlds in which Ernie correctly believes that the bag is in the basement by reasoning inductively are closer to the actual world than close possible worlds in which he incorrectly believes the same proposition by the same method. To see this, notice that close possible worlds in which the bag is snagged in the chute—and Ernie forms the false belief that it has reached the basement—are more dissimilar to the actual world than close possible worlds in which it is in the basement and Ernie forms a true belief, because more changes need to be made to the actual world to obtain the former worlds than to obtain the latter. In sum, Ernie's belief counts as safe.

Second, *Proximity Safety* is better than *Weak Safety* in that, like *Strong* and *Hybrid Safety*, seems to account for the lottery intuition. For example, consider standard lottery cases: possible worlds in which you falsely believe that you won't win the lottery solely on the basis of statistical evidence are presumably *as close* to the actual world as close possible worlds in which you correctly believe the same proposition by the same method. After all, as Pritchard (2007a) points out, all it takes to obtain a different result in a fair lottery is that a few colored balls fall in

---

coincidence and how it affects the modal accounts of luck and epistemic luck, see Broncano-Berrocal (2018a; *forthcoming*).

a slightly different configuration. Since true-belief close possible worlds are *not closer* to actuality than false-belief close possible worlds, the belief that you will lose counts as unsafe, according to *Proximity Safety*.

In addition, it is worth noting an advantage of *Proximity Safety* over *Hybrid Safety*: the former explains inductive knowledge and the lottery intuition in standard lottery cases *uniformly*—i.e., by means of only one and the same requirement—, so the solution it provides to Greco's dilemma is not *ad hoc*.

However, while *Proximity Safety* is certainly better than its rivals, it still doesn't solve the lottery problem, and the reason is that whether or not false-belief close possible worlds are closer to actuality than true-belief close possible worlds depends on how things stand in the actual world, which means that *Proximity Safety* will encounter lottery cases in which the relevant beliefs turn out undesirably safe.

Consider the second Gettierized lottery case, *Address Lottery*. According to *Proximity Safety*, Lottie's belief that she won't win counts as safe, because close possible worlds in which her belief is true (i.e., worlds in which she loses the lottery because the hedge fund buys her flat) are *closer* to the actual world than close possible worlds in which her belief is false (i.e., worlds in which the hedge fund does not buy her home and Lottie wins the lottery). After all, more significant changes need to be made to the actual world in order to change the hedge fund's *firm* decision to buy Lottie's flat than to leave things as they stand.

At this point, the dialectical position of the safety theorist is quite delicate. For one thing, neither the 'no close risk of error', nor the 'small risk of error' conceptions of safety are valid options. For another, we have just discovered that replacing the proportion view with the distance view is of no help either. Perhaps it's time to give up one of the key assumptions of mainstream safety theorizing: that safe beliefs are beliefs that covary with the truth in *close* possible worlds.

7. Normalized (or normic) safety

Safe beliefs are supposed to preserve conformity between what is believed and what is the case in *close* or *nearby* possible worlds. One alternative is to understand safe beliefs as beliefs that covary with the truth in *normal* possible worlds. For example, the following is a safety principle formulated in terms of a 'no *normal* risk of error' conception of safety combined with the proportion view—Dutant (2010) and, more recently, Littlejohn & Dutant (*forthcoming*) defend this kind view; a similar view about justification is defended by Smith (2016):

> *Normalized Safety*: If *S* knows that *p*, then *S*'s belief that *p* is true in the actual world and in *no* possible world which is *at least as normal as the actual world* and in which *S* forms the belief that *p* via the same type of method *S* uses in the actual world, *p* is false.

The devil, of course, is in the details: *Normalized Safety* makes different predictions depending on what is considered a 'normal' possible world. In this sense, there are three candidate conceptions of the notion of normalcy that are either endorsed or discussed in the literature: (1) in statistical terms (Smith 2016); (2) in explanatory terms (Smith 2016); (3) in terms of different respects of normalcy akin to David Lewis's (1979) proposed rules for weighting respects of similarity (Dutant 2010).

Understanding normalcy in statistical terms and, specifically, in terms of frequency of occurrence—so that the more frequent the occurrence of an event *E* is, the more normal *E*'s occurrence is—makes *Normalized Safety* deliver the wrong results in the lottery case. After all, because losing a fair lottery is frequent and therefore normal (while winning is infrequent and therefore abnormal), normal possible worlds are worlds in which one loses the lottery. If so, one's belief that one will lose is true in all normal possible worlds, and hence counts as safe.

Dutant's view (2010: 12-13) is that normalized alternatives to the actual world *w* (i) "are furnished with the same kind of things at broadly the same locations" (e.g., no dinosaurs in Central Park); (ii) "satisfy the basic laws of *w*"; (iii) "do not breach *ceteris paribus* laws of *w* more than *w* does, especially within the area of interest. If the woman does not normally lie, do not add a world where she does"; and (iv) "do not have less natural properties that are highly likely (by the lights of the laws of *w*) than *w* does, especially within the area of interest. If the world contains just 100 fair coin tosses, do not add a world with a series of 100 heads; but if it contains 2^{100} tosses, do so".

Rule (iv) is the one of interest to our purposes. According to Dutant, (iv) predicts that if you will *de facto* lose in *each* of nine different small ten-ticket lotteries, the possibility that you will win in each is *less normal* than the future result, whereas if you will *de facto* lose in a big one-million-ticket lottery, that you eventually win is *as normal* as the actual result. In this way, in possible worlds that are as normal as the actual one, your true belief that you will lose *every* small lottery will continue to be true, whereas your true belief that you will lose the big lottery would be falsified in some of those as-normal-worlds. In the former case, your belief is safe (and amounts to knowledge); in the latter, your belief is unsafe (and not knowledge), which, according to Dutant (and in line with what *Normalized Safety* would predict) is the right result.

However, *Normalized Safety*, when understood in terms of Dutant's metric for normalcy, fails to account for Gettierized lottery cases. Consider *Address Lottery*. Recall that it involves a coincidence with two components: (1) the result of the lottery corresponding to Lottie's address and (2) the purchase of Lottie's flat by the hedge fund (for completely independent reasons). Although possible worlds in which (1) is not the case are *as normal* as the actual world by rule (iv), possible worlds in which (2) does not happen are *less normal*. To see this, we can further fill in the details of the case as follows: the investors never change their decisions once made and that they always get what they want, because the hedge fund is the most powerful organization in the world. If this is how things normally stand in the actual world, then, by rule (iii) (don't breach *ceteris paribus* laws), the investors would continue to purchase Lottie's flat in normal possible worlds; and if in normal possible worlds Lottie loses her flat, her belief that she will lose the lottery continues to be true in them, which means that her belief is safe, according to *Normalized Safety*.

Finally, understanding normalcy in explanatory terms won't do either. Smith's proposal (2016) is that abnormal events call for special explanations (or more explanation) than normal events. If someone claims to have seen dinosaurs in Central Park, we would surely demand an explanation of such a bizarre event, whereas if someone said the opposite, we wouldn't. Smith's idea, then, is that the outcomes of fair lotteries, especially when favorable, demand *no* special explanation, or at least no more explanation than when disfavorable: in order to explain the fact that someone has won *or* lost the lottery you simply need to point to the fact that certain balls got out of the lottery machine first (i.e., the explanation is the same for both kinds of outcomes).

However, Gettierized lottery cases are problematic for this proposal too. As we have seen, in Gettierized lottery cases subjects form their beliefs in exactly the same way as standard Lottery cases, namely by reflecting on the odds of a fair lottery. However, although both types of cases involve fair lotteries and the same belief-forming methods, Gettierized lottery cases feature, in addition, certain independent factors such as the intervention of the dog (*Eaten Ticket*) or of the hedge fund (*Address Lottery*). These factors make the relevant lottery beliefs become true in a way that neither affects the fairness of the lottery, nor how such beliefs are formed—precisely because such factors are independent of the inner workings of the lottery as well as of the relevant methods of belief formation, we should expect that the standard lottery intuitions stay stable in the Gettierized cases. Crucially, however, these independent factors make a difference in terms of normalcy between both types of cases. In standard lottery cases, a change in the result of the lottery demands no more explanation than leaving things as they stand. After all, as Smith thinks, all it takes to obtain a different outcome is a slightly different configuration of the balls in the lottery machine. By contrast, in Gettierized lottery cases a change in the result of the lottery implies a change in the aforementioned independent factors; but modifying such factors *does* require a special explanation or, at any rate, *more* explanation than a slight change in the configuration of the balls.

To see this, consider *Address Lottery* again. Let’s suppose, as before, that it is widely acknowledged that the investors never change their decisions once made and that they always get what they want because the hedge fund is the most powerful organization in the world. Given that this is how things normally stand in the actual world, it is fairly obvious that reversing the investors’ firm decision to buy Lottie’s flat or thwarting the purchase requires more explanation than leaving things as they stand. If so, normal possible worlds, as per Smith's conception of normalcy, are possible worlds in which Lottie would still lose her flat *and* the lottery, and in which she would accordingly continue to get things right about the fact that she will lose the lottery. This makes her belief count as safe, as per *Normalized Safety*.

9. Moving forward: the indirect truth-tracking approach to knowledge

Let’s take stock. The basic idea underlying safety-based theories (and any theory featuring modal principles) is that knowledge is a matter of preserving conformity between what is believed and what is the case in certain possible worlds. From this core assumption, safety theorists make certain theoretical choices to delimit the relevant set of possible worlds to the one that befits knowledge. As we have seen, one can delimit the range of covariation between belief and truth over *normal*, or else *close* possible worlds. One can also partition such worlds in terms of *proportions* (the proportion view) or in terms of *distance* to the actual world (the proximity view). Finally, one can opt for a *no risk of error* or else for a *small risk of error* policy. The resulting combinations can be represented as follows:[13]

Safety Principle	Relativization to possible worlds (PW)	No risk of error in the relevant set of possible worlds	Small risk of error in the relevant set of possible worlds	Proportion view	Proximity view
Strong Safety	Close PW	X		X	
Weak Safety	Close PW		X	X	
Hybrid Safety	Close PW	X	X	X	X
Proximity Safety	Close PW				X
Normalized Safety	Normal PW	X		X	

As we have seen, none of these safety principles solves the lottery problem, at least not without unwelcome consequences. Fortunately, there is a solution available to the safety theorist, one that involves *revising the key assumption* that believing safely is a matter of preserving conformity between what is believed and what is the case in a suitably selected set of possible worlds.

9.1 The indirect-truth tacking approach to knowledge explained

Independently of how the space of possible worlds is delimited in the formulation of one's preferred modal condition on knowledge, keeping conformity between belief and truth can be understood in two ways: (1) *direct covariation* between one's doxastic state about *p* and *p*—this is the orthodox *direct truth-tracking approach* to knowledge that extant modal principles presuppose; (2) *indirectly*, by requiring one's doxastic states about *p* covary with the kind of circumstances that are appropriate for knowledge—this is the alternative *indirect truth-tracking approach* to knowledge that I will endorse in what follows.

The conception of safe belief that ensues from (2) is one according to which what safe beliefs track is the appropriateness of the circumstances rather than, directly, the truth. In Broncano-Berrocal (2018b), I have explored this proposal in some detail and argued that a safety principle formulated in those terms (see below) not only handles Gettier cases and counterexamples to the necessity of safety, but also nicely fits Nozick's original (and often

[13] These are not all the possible combinations, but at least some plausible ones, and certainly some of the most relevant when it comes to the lottery problem.

overlooked) motivation for the truth-tracking approach to knowledge. Here, I will argue that such a safety principle also gives a solution to the lottery problem.

Let's give some theoretical motivation first. For Nozick (1981), the most plausible hypothesis for explaining why we came to have *knowledge in a changing world* (the kind of hypothesis that should be expected, on theoretical grounds, from an evolutionary perspective) is that we developed belief capabilities for detecting changes in facts that would change beliefs accordingly—this is certainly more plausible than postulating a pre-established harmony between facts and beliefs from birth until death, or a god constantly intervening in the world to make facts match beliefs. In particular, Nozick thinks that whatever evolutionary process might have transformed our antecessors did "not bestow upon them a capability for true beliefs so powerful that in no logically possible situation would their beliefs be mistaken", because "there would not be strong selection for it; there would be no selection against those other organisms whose lesser capacities function just as well in the actual range of situations." (Nozick 1981: 285). Accordingly, Nozick thinks that it would be a too demanding condition on knowledge to require covariation with the truth in *every* or *any* possible circumstances, hence the narrower sensitivity, adherence, or safety requirements.

As I will suggest next, Nozick's evolutionary explanation can be plausibly reinterpreted as follows: knowledge in a changing world is possible, not because the relevant belief capabilities track the facts, but because they *track appropriate circumstances for knowledge*. On this new approach, just as it would be evolutionarily suboptimal and certainly too epistemically demanding to require knowers to track the truth in every or any possible circumstances, it would be suboptimal and too demanding to require them to track every or any appropriate circumstances.

By way of illustration, the presence of oxygen or food surely makes the environment appropriate for knowing—because dead bodies can't form beliefs—, but such conditions don't *determine* whether or not a given doxastic state ends up matching the facts: they *merely enable belief formation*. Understandably, tracking such *enabling conditions* is evolutionarily suboptimal, a waste of cognitive resources vis-à-vis the purpose of knowledge acquisition. By contrast, there are conditions (environmental or internal) that do *determine whether an agent's doxastic state ends up matching the facts*. For example, the light conditions, the distance and the size of an object, the degree of attention paid, and so on, have a *direct* bearing on whether one ends up forming a true or a false visual belief of that object. In this way, the right (viz., not too demanding) requirement is that the kind of appropriate conditions that safe beliefs track are not enabling, but *determining conditions*. One possible formulation of a corresponding safety principle along these lines is the following:

> *Indirect Safety*: If *S* knows that *p* via a method of belief formation *M*, then the determining conditions in the actual world are appropriate and in nearly all (if not all)

> close possible worlds in which *S* continues to believe that *p* via *M*, the determining conditions for *S*'s belief that *p* continue to be appropriate.[14]

A remark on appropriate determining conditions is in order. Circumstances in which the determining conditions are appropriate for knowing *p* are typically circumstances in which it is *likely* that, if one believes that *p*, *p* is true. However, *not all* circumstances that make it likely that one forms true beliefs are appropriate for knowing. To see this, we need to make a further distinction between (a) conditions that make success likely and (b) conditions that make success *in the right way* likely.

The former are appropriate conditions for a performance to be considered *successful*, whereas the latter are appropriate conditions for a performance to be considered an *achievement*. Normally, these nuances don't matter, because the two types of conditions coincide often times. Consider archery. Normally, circumstances in which the light conditions are good, the wind is calm, the target is at a reasonable distance, and so on (i.e., conditions under which one is *likely* to hit the target) are also circumstances in which one is likely to hit the target *in the right way*, namely in a way that the exercise of one's shooting abilities is a salient factor in the explanation of why the arrow hits the target. Thus, the circumstances are both appropriate for one's shot being successful *and* an achievement. However, in some cases the two types of conditions come apart. Here is one such case:

> *Electromagnets*
> Unbeknownst to everyone, all targets in the archery field have been replaced with indistinguishable powerful electromagnets, so when *anyone* releases an arrow (all arrows are metallic), it hits the bull's eye. Success in this field is as likely for a four-year-old as it is for a professional archer. In fact, if a professional were to shoot without the presence of the electromagnets, her ratio of successful-to-unsuccessful shots would be lower than when they are in place.

The circumstances in *Electromagnets* are certainly appropriate—in fact, very appropriate—*for successful shooting*. However, they are utterly inappropriate for such shots being *successful in the right way* and hence for being considered achievements. In a sense, despite the high probability of success, the circumstances are achievement-precluding because they are *unduly helpful*: they make it likely that one hits the target, but they do so in a way that one is *not* thereby likely to hit the target while preserving the right kind of causal or explanatory connection between one's competent shot and the fact that the arrow hits the bull's eye.

[14] The safety principle I propose in Broncano-Berrocal (2018b) does not entail that one knows in circumstances *C* only if the determining conditions in *C* are appropriate, but rather says that if one knows in circumstances where the determining conditions are appropriate, one's beliefs must be indirectly safe. As a result, there will be cases in which the principle I propose here applies but in which the latter doesn't. For discussion of the assumption that knowledge requires appropriate circumstances, see Broncano-Berrocal (*forthcoming*).

9.2 A different safety-based solution to the lottery problem

Back to epistemology. While, often enough, circumstances that are appropriate for cognitive success are also appropriate for cognitive achievement (i.e., knowledge), we can also expect cases in which they come apart. My claim is that *lottery cases*, in particular, and cases in which one forms beliefs *solely* on the basis of statistical evidence, in general, *are* such cases.

More specifically, the kind of circumstances in which one merely bases one's beliefs on (strong) statistical evidence are analogous to the kind of circumstances in which one shoots metallic arrows in a field with powerful electromagnets. In both cases, the circumstances make success (shooting an arrow accurately) and cognitive success (believing correctly that one will lose the lottery) very likely, almost guaranteed. In other words, in both cases the circumstances are appropriate for *succeeding simpliciter*, cognitively or non-cognitively. However, they are *in*appropriate for *succeeding in the right way*, i.e., for achievement or cognitive achievement (knowledge).

As before, the reason is that, like the effect of the electromagnets, strong statistical evidence is *unduly helpful*. It is *helpful* because it makes one get things right in a great number of instances of belief formation, just as the electromagnets make one hit the bull's eye in many instances of shooting. It is *undue* (in the sense of improper) because, just as the electromagnets prevent each particular successful shot from being explained in the right way by the competent exercise of the archer's shooting abilities, *mere* statistical evidence makes the exercise of our cognitive abilities irrelevant in the explanation of why we come to get things right *on each particular occasion*.

To see this more clearly, consider an example by Adler (2005): merely using the knowledge that 999 of the 1000 apples of a barrel are good in order to judge *each one* as good is epistemically improper, insofar as the evaluation of *each* particular apple as being good is based on assigning the apple in question to the category 'good' just because, *in general*, this is the case in a high proportion of cases. By contrast, if a careful and skilled, yet fallible, evaluator examines each, it doesn't matter whether that person makes less correct judgments (in general) than the mere user of statistical evidence (i.e., it doesn't matter whether she is less reliable): for *each* correct assessment we can say that the fact that she comes to have a true belief about *that particular apple* is saliently explained by her apple-sorting abilities.

We cannot say the same of the mere user of statistical evidence. *Only* using statistical evidence to determine whether a certain apple is good (or whether one will lose the lottery) gives rise to the kind of inappropriate epistemic circumstances in which cognitive success *simpliciter* is likely, but in which cognitive success *in the right way* is unlikely.

Of course, if this is the case, the relevant beliefs (about apples or lotteries) cannot be safe, according to *Indirect Safety*, not just because of what happens in possible worlds, but mainly because of what happens *in the actual world*: we ignore that we will lose the lottery merely on the basis of evidence about the odds because the *actual* circumstances of belief formation are inappropriate for knowledge or cognitive achievement.[15] By contrast, the kind of circumstances in which one uses one's apple-sorting abilities, testimonial evidence from a reliable newspaper, and so on, make cognitive success in the right way likely. That's why we can form safe beliefs in them.[16]

This is not to say that we should avoid using statistical evidence altogether. The quarrel is with *merely* using statistical evidence. Evidence about the likelihood of a certain type of event can be a powerful ally when one's cognitive abilities are already in use: it can play an *auxiliary role*. For example, if a lottery player doesn't know anything about statistics but happens to glance at the winning number, availing this person with the relevant statistical knowledge would boost her confidence in her visual belief.

Analogously, in cases of *inductive knowledge* statistical evidence plays, if not an auxiliary, at least not a determining role. To give an example, by exercising one's cognitive abilities one can figure out that the waste collection service has never reported any anomaly in the garbage chute. One can also see that one's trash bags have reached the basement many times in the past, or that garbage trucks often leave the condo's basement filled with those bags. This kind of 'precautionary' background knowledge *rules out* the possibility that the bag is snagged in the chute *in the first place*, i.e., when it comes to forming the belief that one's trash bag will soon reach the basement—at least it ensures that such a possibility wouldn't materialize in most close possible worlds. This is still true if, when forming this belief, one *also* uses auxiliary evidence that, in a high proportion of cases, trash bags don't get snagged in garbage chutes. Such general statistical evidence helps reassure one that one's background assumptions are likely to be correct, and perhaps it thereby endows one with epistemic justification to believe that they are, in fact, correct.

By contrast, in the lottery case, the relevant statistical evidence doesn't play this auxiliary role, because *there is nothing to be auxiliary to*; unless one has conclusive reasons for thinking that the lottery is rigged, the kind of precautionary background knowledge that is in place in cases of inductive knowledge is not available to lottery players, either because of the presence of a random mechanism, so that one cannot avail oneself of knowledge of some causal pattern (*standard lottery cases*) or (also) because there are Gettierizing events lurking around, events

---

[15] In Gettierized lottery cases, the source of inappropriateness is double: it not only comes from the use of mere statistical evidence, but also from the bizarre factors that are typical of Gettier cases.

[16] One possible methodological objection to my solution could be that, first, I have appealed to the concept of knowledge to characterize the notion of appropriateness ("appropriate conditions *for knowledge*", I have said), and then I have used this notion to formulate *Indirect Safety*. This would make an eventual definition of knowledge circular if *Indirect Safety* were to figure among its necessary conditions. My reply is that we can leave things as they stand and abandon hope for a reductive analysis of knowledge. However, if one dislikes knowledge-first theorizing, one might have noticed that I have also characterized the notion of appropriateness in terms of the notion of cognitive achievement and assumed that cognitive achievement and knowledge are the same thing. This assumption can be dropped and appropriateness be characterized solely in terms of cognitive achievement (i.e., without appeal to the concept of knowledge). In this way, we can formulate *Indirect Safety* in those terms, and *then* argue for the thesis that knowledge is a cognitive achievement on independent grounds. This should please the traditionalist. For relevant discussion of the cognitive achievement thesis, see Broncano-Berrocal (2017).

that are, by all lights, beyond their ken (*Gettierized lottery cases*). This explains why we often attribute knowledge to people who employ inductive methods, while we refrain from attributing knowledge to lottery players.[17]

Acknowledgments

Many thanks to Cameron Boult, Chris Kelp, Lisa Miracchi, Harmen Ghijsen, Francesco Praolini and audiences at Lisbon, Copenhagen, Bled, and Madrid for helpful comments. This work was supported by the research grant *New Perspectives on Epistemic Risk* (Ref. PGC2018-098805-B-I00) from the Spanish Ministry for Science, Innovation, and Universities.

References

Adler, J. E. (2005). Reliabilist justification (or knowledge) as a good truth-ratio. *Pacific Philosophical Quarterly* 86: 445–58.

Broncano-Berrocal, F. (*forthcoming*). Difficulty and knowledge. In C. Kelp & J. Greco (eds.) *Virtue-theoretic epistemology: New methods and approaches*. Cambridge University Press.

————(2018a). Purifying impure virtue epistemology. *Philosophical Studies* 175: 385-410.

————(2018b). Knowledge and tracking revisited. *Analysis* 78: 396–405.

————(2017). A robust enough virtue epistemology. *Synthese* 194: 2147–174.

————(2016). No luck in the distance: A reply to Freitag. *Theoria: A Swedish Journal of Philosophy* 82: 89–100.

Chisholm, R. M. (1977). *Theory of knowledge*. Prentice-Hall.

Dodd, D. (2012). Safety, skepticism, and lotteries. *Erkenntnis* 77: 95–120.

Dutant, J. (2010). Two notions of safety. *Swiss Philosophical Preprints* 87.

Greco, J. (2007). Worries about Pritchard's safety. *Synthese* 158: 299–302.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy* 73: 771–91.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis* 23: 121–23.

Gibbons, J. (2013). *The norm of belief*. Oxford University Press.

Hawthorne, J. (2003). *Knowledge and lotteries*. Oxford University Press.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs* 13: 455–76.

Littlejohn, C. & Dutant, J. (*forthcoming*). Justification, knowledge, and normality. *Philosophical Studies*.

Luper, S. (1984). The epistemic predicament: knowledge, Nozickian tracking, and skepticism. *Australasian Journal of Philosophy* 62: 26–50.

Miracchi, L. (2015). Competence to know. *Philosophical Studies* 172: 29–56.

Nozick, R. (1981). *Philosophical explanations*. Oxford University Press.

Owens, D. (1992). *Causes and coincidences*. Cambridge University Press.

[17] See Adler (2005) for this kind of explanation of how to tell apart cases of inductive knowledge from lottery cases and especially for his explanation of the "illusion" of thinking that they are the same kind of case (cf. Hawthorne 2003).

Pritchard, D. (2012). Anti-luck virtue epistemology. *The Journal of Philosophy* 109: 247–79.
————(2007a). Anti-luck epistemology. *Synthese* 158: 277–97.
————(2007b). Knowledge, luck, and lotteries. In V. F. Hendricks & D. Pritchard (eds.), *New waves in epistemology*. Palgrave Macmillan.
————(2005). *Epistemic Luck*. Oxford University Press.
Reed, B. (2010). A defense of stable invariantism. *Noûs* 44: 224–244.
Russell, B. (1948). *Human knowledge: Its scope and its limits*. Allen & Unwin.
Smith, M. (2016). *Between probability and certainty: What justifies belief*. Oxford University Press.
Sosa, E. (2015). *Judgment and agency*. Oxford University Press.
————(1999). How to defeat opposition to Moore. *Philosophical Perspectives* 13: 137–49.
Vogel, J. (2007). Subjunctivitis. *Philosophical Studies* 134: 73–88.
————(1999). The new relevant alternatives theory. *Philosophical Perspectives* 13: 155–80.
Williamson, T. (2009). Probability and Danger. *The Amherst Lecture in Philosophy* 4: 1–35.
Zagzebski, L. (1994). The inescapability of Gettier problems. *Philosophical Quarterly* 44: 65–73.