1 The safety principle, modal epistemology and epistemic luck

Some epistemologists, dissatisfied with simple reliabilism but akin to the view that cognitive processes tend to produce true beliefs in a range of not only actual or past but also possible cases, proposed a family of theories which have as core the metaphor that knowledge is a matter of tracking the truth. Two principles in the literature shed light on what it exactly takes to track the truth. The sensitivity principle, originally attributed to Nozick (1981), was the first to appear:

*Sensitivity*: If S knows that p, then if p had not been true, S would not have believed that p.¹

Sensitivity-based theories were regarded at the time as very developed analyses of knowledge, since they could deal with all sorts of chief counterexamples and problematic cases like usual Gettier-type cases, relevant alternatives cases or lottery cases. But many counterexamples and objections were posed against sensitivity so that the principle was dismissed.

The safety principle, originally attributed to Sosa (e.g. 1999), kept alive the project of modal epistemology, for it seems to deal with all the problematic cases covered by sensitivity but also with its counterexamples:

*Safety*: If S knows that p, then S’s true belief that p could not have been easily false.²

Pritchard (2009; 2010) argues that sensitivity and safety-based accounts are basically motivated by the intuition that knowledge is incompatible with (dangerous) epistemic luck, in the sense that one’s cognitive success is not a matter of luck, what he calls the anti-luck intuition. As a consequence, modal epistemology can be regarded as a type of anti-luck epistemology.

Consider now the definition of luckily true belief that Pritchard (2007) provides:

*(LTB)* S’s true belief is lucky iff there is a wide class of near-by possible worlds in which S continues to believe the target proposition, and the relevant initial conditions for the formation of that belief are the same as in the actual world, and yet the belief is false.

And the definition of non-lucky true belief:

¹ Formally: \( Kp \rightarrow (\neg p > \neg Bp) \). Standard possible-worlds interpretation: if S knows that p, then in most nearby worlds where p is false, S does not believe that p.

² Formally: \( Kp \rightarrow (Bp > p) \). Standard possible-worlds interpretation: if S knows that p, then in most nearby worlds where S believes that p, p is true.
S’s true belief is non-lucky iff there is no wide class of near-by possible worlds in which S continues to believe the target proposition, and the relevant initial conditions for the formation of that belief are the same as in the actual world, and yet the belief is false.

A quick look to AL suffices to realize that it is tantamount to the definition of safe belief (in its possible-worlds interpretation). As a result, a condition which holds that only safe beliefs are known amounts to the naive view that beliefs must not be luckily true in order to be known, a requirement which does not throw much light on the nature of knowledge. In any case, the criticism against the safety principle will not be that it is uninformative but that it is blind to what causes epistemic luck. That criticism will serve to argue that there is a sense in which epistemic circumstances themselves can be regarded as safe. But first, in order to develop such an argument, it will be necessary to assume the view that the concept of knowledge has a bipartite structure, a view originally advanced by Pritchard (2010) that will be nevertheless qualified.

2 The two master intuitions and the independence claim

Pritchard (2010) argues that there are two overarching intuitions which govern how epistemologists think about knowledge, and particularly their thinking about what turns true belief into knowledge: the already seen anti-luck intuition and the ability intuition, the intuition that knowledge requires cognitive ability, in the sense that when one knows one’s cognitive success is, in substantial part at least, creditable to one’s cognitive abilities. Each of them motivates a type of epistemic condition. On the one hand, anti-luck conditions like the sensitivity or the safety principles stem from the anti-luck intuition. On the other hand, the ability intuition originates ability conditions like those of simple reliabilism and of standard epistemic internalism, but more importantly, conditions of virtue epistemology like these two generic ones:

WVE (Weak Virtue Epistemology): If S knows that p, then S’s cognitive success is to a significant degree creditable to the reliable operation of S’s cognitive abilities.

SVE (Strong Virtue Epistemology): If S knows that p, then S’s cognitive success is primarily creditable to the reliable operation of S’s cognitive abilities.

One of the virtues of Pritchard’s classification of epistemic conditions is that it makes very easy to comprehend contemporary theorizing about knowledge. The most interesting point, however, is the purpose to which he introduces it: to argue that no theory of knowledge hitherto is sufficient for knowledge. In a nutshell, his argument can be reconstructed as follows:

i. In order to be sufficient the conditions of one’s definition of knowledge must satisfy both the anti-luck and the ability intuitions.

ii. No formulation of an anti-luck condition can fully accommodate the ability intuition and no formulation of an ability condition can fully accommodate the anti-luck intuition.

iii. Theories in the market either include anti-luck or ability conditions.

Conclusion: No theory hitherto is sufficient for knowledge.

It is premise (ii) that carries the burden of proof of the argument. Its justification involves showing case by case how theories that incorporate anti-luck conditions face counterexamples which are solved by ability conditions, and the other way around, how theories with ability
conditions are in trouble in cases where anti-luck conditions have no problems. Pritchard, by appealing to the long-record of counterexamples in the literature, compellingly provides that justification so it will be just assumed that the conclusion of his argument obtains.

The focus of the discussion will be rather on Pritchard’s intriguing diagnosis of why there is such state of theoretical insufficiency. According to him, epistemologists have obviated the fact that the two intuitions actually impose independent epistemic constraints on our theory of knowledge (what henceforth will be called the independence claim). That claim, Pritchard presumes, amounts to saying that the concept of knowledge has a bipartite structure.

To be more precise, the independence claim can be construed as the claim that the tasks of excluding luck and solving epistemic problems concerning the agent’s cognitive character are two tasks that cannot be fully fulfilled by one single condition (i.e. either just by an anti-luck condition or just by an ability condition).

One strategy that one may adopt so that one’s definition satisfy the independence claim is to opt for a mixed definition of knowledge, one that combines an anti-luck condition with an ability condition in such a way that the resulting condition can fulfill both the tasks of excluding luck and ruling out cases in which something is amiss with the agent’s cognitive character.

For example, accounts based on sensitivity or safety commonly relativize the principles to reliable methods of belief formation. In general, the resulting package (modal epistemology plus simple reliabilism), though it excludes luck, is unsuccessful in ruling out certain cases in which something is wrong with the agent’s cognitive character like, for instance, Platinga’s case of the ‘strange and fleeting’ (albeit reliable) process.

Another example of a mixed definition is Pritchard’s new theory of knowledge which neatly combines an anti-luck condition (safety) with a condition of virtue epistemology (WVE), a move that allows Pritchard to avoid the problems that threaten simple reliabilism:

*Anti-luck Virtue Epistemology* (ALVE): Knowledge is safe belief that arises out of the reliable cognitive traits that make up one’s cognitive character, such that one’s cognitive success is to a significant degree creditable to one’s cognitive character (Pritchard et al. 2010, 55).

The theory, reconstructed in a more clear way:

ALVE: S knows that p iff:

i. p is true,
ii. S believes that p,
iii. S’s safe cognitive success is to a significant degree creditable to S’s cognitive character.

### 3 Overlapping conditions

Possibly, it is true that there are two ‘master’ intuitions which play a fundamental role in epistemology. This paper will not deny that. Rather, it will just show that, if the concept of knowledge really is (in some sense) a twofold concept, its bipartite structure is not elucidated in an intelligible way by means of the independence claim, that is, by requiring one’s definition of knowledge to have, on the one hand, a condition which aims at luck exclusion and, on the other, a condition which aims to incorporate into the concept of knowledge the contribution
made by our cognitive abilities. In addition, an alternative way of conceiving knowledge as having a bipartite structure will be proposed. That is, it will be given a different view of the nature of knowledge, but no necessary and sufficient conditions will be offered.

The general line of reasoning is the following: incorporating to one’s definition an epistemic condition which is basically aimed at excluding luck is problematic insofar as it overlaps in certain cases with the relevant ability condition. The two epistemic conditions should be completely independent in their roles. For otherwise it is simply not clear how to make sense of the metaphor that the concept of knowledge has a “bipartite structure”.

To be more specific, the crucial qualification that it will be made is that the safety principle obviates what causes veritic epistemic luck, and since it can be caused by unreliable cognitive traits, safety may well be doing a job ascribed to ability conditions. But let us proceed step by step.

First, let us just assume (uncontroversially) that knowledge is a kind of success from cognitive ability, that is, that by displaying certain reliable cognitive abilities (e.g. sense perception, memory, reasoning, etc.) agents can gain true beliefs. As a consequence, let us assume that it is to a certain degree creditable to the reliable operation of those abilities that agents cognitively succeed.\(^3\)

Now, the question is: which types of epistemic problems are to be solved by the safety principle and which ones by ability conditions so that they do not overlap? Note, in the first place, that two factors are involved in an agent’s cognitive success:

CA: The deployment of certain cognitive abilities.

EC: The epistemic circumstances in which those cognitive abilities are displayed.

1. On the one hand, it is clear that ability conditions (i.e. conditions ranging over one’s cognitive character) are the best suited to solve problematic cases concerning CA. One example: non-knowledge cases in which certain cognitive abilities are unreliably displayed (think of a visually impaired person trying to spot an object at long distance). Another: cases in which the relevant cognitive abilities are reliable but defective in some sense, as when an agent reliably forms a belief but that fact does not explain to some relevant degree why the agent cognitively succeeds. Consider the following case by Lackey (2007, 347): imagine that S comes to truly believe that drinking Bacardi rum enhances one’s sex appeal, but only because of the subliminal suggestions contained in their ads. What is amiss with S’s cognitive success, following Lackey’s diagnosis, is that, while S does form her belief by means of reliable cognitive abilities (sense perception and memory), it is not significantly creditable to their reliable operation that she holds her belief. Instead, it is creditable to the subliminal suggestions of the Bacardi’s ads.

On the other hand, it does not seem that ability conditions are suitable for dealing with problems which stem from EC. After all, they are conditions on one’s cognitive character and not on one’s epistemic circumstances.

2. Safety is aimed at excluding, at least, cases of veritic epistemic luck (consider again LTB). Safety, then, is supposed to deal with all cases in which veritic epistemic luck affects the way that one’s belief hits the truth. The question is: is that compatible with a full independence

\(^3\) What is being implicitly assumed is that some virtue-theoretic condition is necessary for knowledge.
between types of conditions, that is, with the constraint that the epistemic conditions of one’s definition should not overlap in their roles?

Consider the following piece of reasoning of why that is not the case. Becker (2008) rightly argues that veritic luck has two types of causes: faulty environmental conditions and unreliable belief-forming abilities. In accordance with that, he distinguishes between world luck (when luck is caused by the first) and process luck (when luck is caused by the latter). In addition, he explains that two different kinds of mechanisms are needed for fixing the two varieties of luck. Thus, in order to eliminate process luck we need an ability condition, and in order to get rid of world luck we need an anti-luck condition, preferably safety. As a consequence, it seems that safety should take care of problems concerning EC, and namely of world luck, while ability conditions should take care of problems concerning CA like process luck and other problems stemming from faulty cognitive ability.

The problem is that it is ambiguous whether the safety principle, as it stands in the literature (as a condition on doxastic states), is aimed at eliminating world luck or process luck. Ideally, it should be a more informative and specific principle on this regard.

With respect to Pritchard’s theory, the point is that the allegedly independent conditions in ALVE play no independent roles, for safety and WVE (the ability condition) are both aimed at excluding world and process luck. If one’s ‘independent’ conditions apply to the same field, how can those conditions make intelligible the metaphor that the concept of knowledge has a bipartite structure?

4 Circumstantial safety and the bipartite structure of knowledge

The bipartite structure of knowledge

Nevertheless, there is a way in which our theory of knowledge could incorporate fully independent epistemic conditions. Consider the following strategy. One could include in one’s definition of knowledge at least a condition which only applies to the agent’s epistemic circumstances (call it a circumstantial or environmental condition) and at least a condition which only applies to the agent’s belief-forming abilities (call it a belief-forming condition). As a consequence, a definition including both would shape knowledge as having a bipartite structure consisting of two interrelated dimensions:

- What the agent is able to know by displaying her cognitive abilities in certain epistemic circumstances.
- What the epistemic circumstances allows the agent to know, in the sense that the circumstances may make more or less difficult (or easy) for the agent the task of achieving knowledge.

Thus, knowledge is true belief that arises out of reliable cognitive abilities displayed in certain epistemic circumstances, such that one’s cognitive success is creditable to (i) and (ii):

i. One’s cognitive character.

ii. Certain properties of the circumstances.⁴

⁴ How much creditable is to (i) and (ii) may vary depending on the case in such a way that the more creditable is to (i), the less creditable is to (ii), and vice versa.
With respect to (i) some kind of belief-forming condition is needed, while with respect to (ii) it is required an environmental one. We are already familiar with the former type of conditions (e.g. simple reliabilist conditions, WVE, SVE) but what kind of epistemic condition could be the latter? The variability of the agent’s epistemic circumstances suggests that an environmental condition should be modal in nature. In particular, note, on the one hand, that the key problem caused by faulty environmental conditions is world luck (a modal notion) and, on the other hand, that safety (a modal condition) is the most appropriate luck-excluding condition. Then, what we just have to do is to consider that environmental conditions are faulty when unsafe. World luck could be thus eliminated with a properly formulated safety condition ranging only over epistemic circumstances while a properly formulated belief-forming condition (like WVE or SVE) could eliminate process luck and other problems concerning the display of the agent’s cognitive abilities.

If one feels that it is wrong to regard the epistemic circumstances themselves as safe or unsafe, one just have notice that in ordinary language, when generally talking about abilities the circumstances in which an agent does \( \phi \) are regarded as safe or unsafe for \( \phi \)-ing independently, in a way, from her more or less reliable abilities to do \( \phi \). For example, drivers are commonly said to be reliable or unreliable, but their cars, roads or climatic conditions, i.e. the driving circumstances, are commonly regarded, independently of how well they drive, as safe or unsafe.

Consider the following analogy that Williamson (2000, 123) uses to explain the notion of safety:

Imagine a ball at the bottom of a hole, and another balanced on the tip of a cone. Both are in equilibrium, but the equilibrium is stable in the former case, unstable in the latter. A slight breath of wind would blow the second ball off; the first ball is harder to shift. The second ball is in danger of falling; the first ball is safe. Although neither ball did in fact fall, the second could easily have fallen; the first could not.

The right intuition does not seem that the ball itself is safe. What it is rather safe is the state of equilibrium of the ball, that is to say: its circumstances.

**Epistemic circumstances**

The relevant circumstances at issue are ‘epistemic’ because they affect whether agents form true beliefs or not, that is, they have an effect on the way that agents cognitively succeed. To see how they affect cognitive abilities notice how cognitive abilities are relativized to environments. Lepock (MS) observes that the reliability of one’s cognitive abilities may vary depending on:

- Conditions of the circumstances.
  For example: vision is more reliable in good light conditions than at twilight in a dark forest; testimony is more reliable in a secondary school (where teachers usually tell the truth) than in a poker game (in which players typically bluff).

- The content of the information acquired.
  For example: smell is reliable for perceiving sulphur but unreliable for perceiving carbon monoxide.

In this way, what counts as epistemic circumstances are, according to Lepock (2006; MS), the conditions under which an ability is displayed and the type of believed propositions that it is being used to produce. For example, by using visual abilities one comes to believe propositions about shapes, colours or locations of objects but not about their smell. On the other hand, what
counts as ‘conditions of the circumstances’? Imagine that S visually perceives a glass that is on the table which is in front of him and forms the belief that there is a glass on the table. Conditions of her epistemic circumstances are not only those conditions that affect the reliability of her visual system (e.g. the light conditions of the room), but whatever affects the reliability of her visual ability; for instance, misleading evidence (e.g. someone telling S that the glass in front of her is a holographic image of a glass).

**Reliability**

How should we understand the notion of reliability at stake? Reliability is typically defined as a tendency of certain cognitive abilities to form true beliefs. ‘Tendency’ can be interpreted either statistically or modally. On the statistical interpretation ‘tendency’ just means a frequency regarding actual and past cases. On the modal interpretation it means a propensity or regularity throughout actual and possible circumstances. Henceforth the term ‘reliable’ is to be understood in terms of statistical reliability. Thus, a cognitive ability is reliable when (considering a representative number of cases) it yields to true beliefs at least more than half of times. Note that this recipe fixes when a cognitive ability is reliable. A different issue, though nevertheless related, is that of fixing when a cognitive ability will be reliable, that is, the issue of determining the probability that certain cognitive ability will yield to true beliefs in future cases.

As we have seen, the reliability of cognitive abilities is relativized to conditions of the circumstances. In general, for each type of cognitive ability A there is a set of conditions which are typically appropriate for the reliable operation of A and by ‘appropriate’ it is just meant those conditions under which A generally yields (at least more than half of the times) to true beliefs. Good lighting and a properly integrated and functioning visual system are examples of conditions which are typically appropriate for forming true beliefs via visual ability. In the case of testimony, some people argue that what makes testimony possible is the fact that the informed person establishes an asymmetric relation of epistemic dependence with the informant and that both the informant and the informed person recognize (more or less implicitly) that situation. The point is that certain conditions of that kind of situation may affect the reliability of the informed person’s ability to select good informants.

Anyhow, individuating the conditions which are typically appropriate for each source of knowledge is something that goes beyond the scope of this paper and sometimes, of epistemology (for example, it is a task of cognitive science to determine the parameters that make perception reliable). It is enough for the present purposes to say that certain conditions are typically appropriate for the reliable operation of cognitive abilities while remaining as neutral as possible concerning which are the conditions that typically count as appropriate.

**Circumstantial safety and risk**

We are now in position to come up with tentative definitions of circumstantial safety and circumstantial risk:

**Circumstantial safety**: Certain epistemic circumstances EP are (to a certain degree) safe for forming true beliefs by means of certain cognitive ability A iff:

---

5 They are not necessary or sufficient epistemic conditions but definitions of two concepts that may help to provide a proper definition of knowledge.
i. The conditions of EP to which \( A \) are relativized are typically appropriate for the reliable operation of the type of cognitive ability which \( A \) is.

ii. The believed propositions are the type of propositions which cognitive abilities like \( A \) can reliably generate.

iii. In most nearby worlds in which the agent continues to form beliefs via \( A \) the conditions are still appropriate.

*Circumstantial risk:* Certain epistemic circumstances EP are (to a certain degree) risky for forming true beliefs by means of certain cognitive ability \( A \) iff:

i. The conditions of EP to which \( A \) are relativized are not typically appropriate for the reliable operation of the type of cognitive ability which \( A \) is.

ii. OR: the believed propositions are not the type of propositions which cognitive abilities like \( A \) can reliably generate.

iii. OR: in most nearby worlds in which the agent continues to form beliefs via \( A \) the conditions are not appropriate.

5 Cases

The picture we get is that while epistemic circumstances are safe or risky, cognitive abilities are reliable or unreliable. Now, by combining these two factors we can predict that there will be three types of cases of knowledge:

- Type 1: Circumstances are safe and cognitive abilities are reliably deployed.
- Type 2: Circumstances are risky and cognitive abilities are reliably deployed.
- Type 3: Circumstances are safe and cognitive abilities are unreliably deployed.

(There is no type 4 since it does not seem that one can know in risky circumstances by displaying unreliable cognitive abilities)

The basic intuition underlying all these cases is the following: the safer the epistemic circumstances are, the less cognitive ability is required in order to know, and the other way around, the more cognitive ability is displayed, the less safety of the circumstances is required. The challenge for any theory of knowledge which incorporates ability conditions is to include these three types of cases. That is, any successful ability condition must be flexible enough to cover all of them. In what follows, I will provide some examples.

I. Type 1 (Normal cognitive ability cases):

*Good vision*

In good light conditions \( S \) sees a glass on the table and consequently forms the belief that there is a glass on the table.

Not only the epistemic circumstances are safe but also \( S \)’s cognitive success is creditable, at least in a substantial part, to her reliable cognitive abilities. A condition like WVE suffices to cover this type of knowledge case.

II. Type 2 (High cognitive ability cases):

*Chicken-sexer*

Farmers often want to know the sex of newly hatched chicks so they can keep the females, and “sacrifice” the males. To do so, they hire a chicken-sexer, an expert in distinguishing male and female chicks. This is not an easy task. It is an art developed in Japan. The chick has an external opening called the cloaca, which serves digestive,
urinary, and reproductive purposes. This opening is closely examined for a degenerate penis, which is found in all males but also 15% of females. The developed skill of a professional chicken-sexer is in determining the sex of this 15%. It requires effort and many hours of training and practice. Novices are not much better than chance at determining a chick’s sex. But experts get faster and faster. The best sexers in the world are able to sex about 800 chicks an hour with 99% accuracy. They have so developed their ability to distinguish the male from female pattern that they can do it instantly (Sloman 2005).

The intuition in this type of cases is that, after hard training, experts are able to acquire a wide range of true beliefs in circumstances that are commonly regarded as epistemically risky for knowing the propositions at issue. Since in those cases the expert’s cognitive success is primarily creditable to her reliable cognitive abilities, a so demanding condition like SVE seems to be required.

III. Type 3 (Low cognitive ability cases):

_Gully_

Gully, a teenager, lacks the ability of picking out reliable informants. Whenever someone tells him a false story, he believes it. His history teacher, knowledgeable of this fact, makes sure that everything she says to Gully is true (e.g. she takes care of not telling obviously false historic events that Gully would easily believe). Gully knows what his teacher teaches him.

This is a knowledge case in which very safe ‘assist’ an agent displaying poor cognitive ability. Gully’s circumstances are very safe for knowing propositions about history: not easily would have Gully believed a false proposition while having lessons. However, Gully’s cognitive success is not creditable (at least in a substantial part) to Gully’s reliable cognitive abilities but to his teacher’s. Therefore, while a safety condition on the epistemic circumstances would obviously obtain, an ability condition like WVE would not. The ability condition which obtains in this case is instead:

_VWVE (Very Weak Virtue Epistemology): If S knows that p, then S’s cognitive success is to a low degree creditable to the reliable operation of S’s cognitive abilities._

[Alternatively, in explanatory terms: If S knows that p, S’s reliable cognitive abilities are a relevant important part of the total set of factors which explain S’s cognitive success (neither a very important, nor the most important part)]

Note that since neither SVE, nor WVE obtain the case is a counterexample to ALVE.

6 Conclusion

Pritchard argues that one good reason why theories of knowledge in the market are insufficient is that they do not take into account that two different intuitions impose very different constraints on our theory of knowledge: that knowledge is incompatible with luck and that it requires cognitive ability, so that they just incorporate either anti-luck or ability intuitions. His solution is to propose a theory, ALVE, which mixes an anti-luck condition (safety) and an ability condition (WVE), what allegedly clarifies the sense in which the structure of knowledge is bipartite.

It has been argued that the safety principle, as it stands in the literature, is blind to what causes veritic epistemic luck and so it is not clear whether it is aimed at excluding world luck (when luck is caused by faulty circumstantial conditions) or process luck (when luck is

---

6 Though better placed than many other theories, ALVE nevertheless faces more counterexamples.
caused by unreliable cognitive abilities). The solution has been to propose that safety should be better understood as a condition ranging only over epistemic circumstances. In this way, it becomes clear in which sense the structure of the concept knowledge is bipartite. It simply incorporates two different epistemic dimensions: what the agent is able to know by displaying her cognitive abilities in certain epistemic circumstances; what the epistemic circumstances allows the agent to know, in the sense that the circumstances may make more or less difficult (or easy) for the agent the task of achieving knowledge. In the end, it has been challenged any prospectively successful theory of knowledge to incorporate three kinds of knowledge cases which seem to pull us in opposite directions: while certain type of cases compels us to strengthen our ability conditions, other cases compel us to weaken them. Any prospective definition which aims to be sufficient for knowledge should take these arguments into account.

References

- Lepock, Chris. (manuscript). Metacognition and Epistemic Virtue. URL: <http://www.ualberta.ca/~clepock/home.htm>