Improbable Knowing [notes]

Timothy Williamson
Timothy.williamson@philosophy.ox.ac.uk

Can we turn the screw on counter-examples to the KK principle (that if one knows that P, one knows that one knows that P)? The idea is to construct cases in which one knows that P, but the epistemic status for one of the proposition that one knows that P is much worse than just one's not knowing it. Of course, since knowledge is factive, there can't be cases in which one knows that P and knows that one doesn't know that P (we can't strengthen ¬KKp to K¬Kp)! If we can construct such cases, we may be able to use them to understand some puzzling epistemic phenomena.

We begin by recalling the formal structure of counter-examples to the KK principle within the framework of the standard possible worlds semantics for epistemic logic.

Non-transitivity and the failure of the KK principle

For any formula A and world w, write w |= A just in case A is true at w.
R is the relation of epistemic accessibility between worlds (Rwu just in case for all that is known at w, one is at u).
As usual, w |= KA just in case for all u such that Rwu, u |= A.

Example: R is reflexive and symmetric on the domain {x, y, z}; Rxy, Ryz, not Rxz.

Thus: {u: Rxu} = {x, y}
{u: Ryu} = {x, y, z}
{u: Rzu} = {y, z}

x ———————— y ———————— z

x	y	z
p	p	¬p
Kp	¬Kp	¬Kp
¬KKp	¬KKp	¬KKp

More generally, consider any non-transitive relation R. Thus for some x, y, z, Rxy, Ryz, not Rxz. For all w in the domain, set w |= p just in case Rxw; hence y |= p but not z |= p. Then x |= Kp by the truth-conditions for p but not y |= Kp (because Ryz); consequently not x |= KKp (because Rxy).

Conversely, consider any model with a transitive accessibility relation R. Suppose that x |= Kp for a world x. Thus z |= p for any z such that Rxz. Suppose that Rxy for a world y. Thus for any z such that Ryz, by transitivity Rxz, so z |= p; consequently y |= Kp. Thus whenever Rxy, y |= Kp. Consequently, x |= KKp.

<u>Adding evidential probabilities</u>

Start with a prior distribution $\text{Prob}_{\text{prior}}$ over sets of worlds, interpreted as propositions. Define conditional probabilities by ratios in the usual way:
$\text{Prob}_{\text{prior}}(X \mid Y) = \text{Prob}_{\text{prior}}(X \cap Y)/\text{Prob}_{\text{prior}}(Y)$ (for $\text{Prob}_{\text{prior}}(Y) > 0$).

Identify the evidential probability of a proposition with its probability conditional on the evidence, and the evidence with what is known (see *Knowledge and its Limits*, ch. 10). In the model, what is known at a world w is just {u: Rwu}. Thus evidential probabilities at w result from conditionalizing $\text{Prob}_{\text{prior}}$ on {x: Rwx}. More formally, let a singular term c (rigidly) denote the real number <u>c</u> in the model. Let Pr express evidential (posterior) probability. Then:

$$w \models \Pr(A) = c \text{ just in case } \text{Prob}_{\text{prior}}(\{x: x \models A\} \mid \{x: Rwx\}) = \underline{c}$$

Uniform prior distribution:
$\text{Prob}_{\text{prior}}(\{x\}) = \text{Prob}_{\text{prior}}(\{y\}) = \text{Prob}_{\text{prior}}(\{z\}) = 1/3$.

x |= Pr(p) = 1	y |= Pr(p) = 2/3	z |= Pr(p) = ½
x |= Pr(Kp) = ½	y |= Pr(Kp) = 1/3	z |= Pr(Kp) = 0

For models of this simple kind, any failure of the KK principle involves a world at which Kp is true but its evidential probability is less than 1 (where p expresses the proposition at issue and every world has non-zero prior probability). By contrast, given the KK principle, w |= Pr(Kp) = 1 whenever w |= Kp (for then w |= KKp, and w |= Pr(A) = 1 whenever w |= KA).

For any rational number m/n such that 0 < m/n < 1 (so 0 < m < n), we construct a finite model with a uniform prior probability distribution and a world at which Kp is true while the evidential probability of Kp is m/n.

The set of worlds is $\{x_1, \ldots, x_m, y_{m+1}, \ldots, y_n, z\}$.
R is reflexive and symmetric; for any i, j Rx_ix_j, Rx_iy_j, Ry_iy_j; Ry_iz, $\neg Rx_iz$

$\{x_1,\ldots,x_m\}$ ———— $\{y_{m+1},\ldots,y_n\}$ ———————— z

p	p	¬p
Kp	¬Kp	¬Kp
¬KKp	¬KKp	¬KKp

Since $\text{Prob}_{\text{prior}}$ is uniform, $\text{Prob}_{\text{prior}}(\{w\}) = 1/(n+1)$ for any world w.

For any i, j ($1 \le i \le m$, $m+1 \le j \le n$):

x_i |= Pr(p) = 1 y_j |= Pr(p) = n/(n+1)   z |= Pr(p) = (n-m)/(n-m+1)
x_i |= Pr(Kp) = m/n y_j |= Pr(Kp) = m/(n+1)   z |= Pr(p) = 0

Thus any x_i will serve as the world of the required type.

Moreover, for every world w in the model, w |= Pr(Kp) ≤ m/n. Consequently, at every world w, and so in particular at x_i, there are any number k of iterations of knowledge that the probability that p is known is at most m/n; formally, w |= K^k(Pr(Kp) ≤ m/n).

Incidentally, such cases also make trouble for accounts of assertion on which one has sufficient warrant to assert p when p has high evidential probability for one, even though one does not know p. For in such models the Moorean conjunction p & ¬Kp can have very probability; specifically, x_i |= Pr(p & ¬Kp) ≥ (n – m)/n. Thus by setting m = 1 and letting n go high enough, we can find a case in which the Moorean conjunction is probable enough for its assertion to be warranted.

Note that the strong assumptions built into the structure of these models (conditionalizing on what is known, uniform prior distribution) should make it *harder* to construct models with the target phenomenon, since it involves evidential probabilities that are misleading as to the epistemic facts — loosening the tie between evidential probabilities and knowledge and allowing special weight to be given to ‘bad’ worlds should make it easier to construct such models. In particular, the models already constructed could hardly be ruled our.

More realistic examples

Consider an unmarked circular dial with a pointer that rotates in discrete steps. It can point at any one of n equally spaced positions on the perimeter of the dial. We measure distances between positions by the minimum number of steps needed to go from one to the other (clockwise or anti-clockwise). Let the positions be numbered 0, .., n-1, where the pointer is pointing vertically up if it is pointing at position 0 (of course, these numberings are not marked on the dial). For simplicity, we consider just n mutually exclusive and jointly exhaustive circumstances (‘world’) x_0, .., x_{n-1}, where in x_i the pointer is pointing at position i. We measure distances between worlds by the corresponding distances between positions. An *interval* of worlds is a nonempty proper subset of the set of worlds such that any world in the set can be reached from any other world in the set by a sequence of steps from one world to another (corresponding to steps between neighbouring positions), such that all worlds in the sequence belong to the set. Any interval has at least one endpoint (a member next to a non-member). An interval with an odd number of members has a unique midpoint, equidistant from its endpoints.

Your only source of knowledge as to which position is current being pointed at is naked eye visual perception from a fixed point of view equidistant from all points on the perimeter. You can make some discriminations between positions, but the difference between neighbouring positions is well below your threshold of discrimination. Assume that your capacity to discriminate between positions depends only on their relative distance; thus if world w is at least as close to world x as world y is to world z and you can discriminate w from x then you can discriminate y from z. Consequently, if R is the epistemic accessibility relation, for some positive h and all worlds w, x, Rwx just in case the distance between w and x is at most h (so R has both reflective and rotational symmetry). Thus for any world w, {x: Rwx} is an interval with an odd number of members and w as its midpoint. Consequently, for any worlds v, w, {x: Rvx} = {x: Rwx} only if v = w. Furthermore, for any worlds v, w, {x: Rvx} = {x: Rwx} have the same number of members, i.e. 2h + 1.

Consider any world w. For all worlds x, let x |= p_w if and only if Rwx. Clearly, w |= Kp_w (in fact, p_w expresses the strongest proposition known at w). Conversely, for any world v, v |= Kp_w only if {x: Rvx} $\subseteq$ {x: x |= p_w} = {x: Rwx}. Since {x: Rvx} and {x: Rwx} have the same number of members, {x: Rvx} $\subseteq$ {x: Rwx} only if {x: Rvx} = {x: Rwx}. As noted above, {x: Rvx} = {x: Rwx} only if v = w. Consequently, v |= Kp_w only if v = w. Thus p_w expresses a proposition known at w and at no other world.

If we add evidential probabilities to the model as explained above, with a uniform prior distribution, $Prob_{prior}$({x: x |= Kp_w} | {x: Rwx}) = 1/(2h+1). Consequently, w |= Pr(Kp_w) = 1/(2h+1). By increasing the number of positions while keeping your discriminatory capacities fixed, we can increase h without limit, and thereby make the probability of Kp_w at w as small as we like, even though Kp_w is in fact true at w.

For the same reason, w |= Pr(p_w & ¬Kp_w) = 2h/(2h+1). Thus the probability of a Moorean proposition can be arbitrarily high, short of 1. As already noted, this makes trouble for ‘justificationist’ accounts of assertion.

Moreover, for every world x in the model, x |= Pr(Kp_w) ≤ 1/(2h+1), for {u: Rxu} has 2h + 1 members, of which at most one (w) verifies Kp_w. Consequently, at every world x, and so in particular at w, there are any number k of iterations of knowledge that the probability that p is known is at most 1/(2h + 1); formally, w |= K^k(Pr(Kp) ≤ 1/(2h + 1)).

Even if we consider other types of probability distribution, when the number of positions is very large the evidential probability at w that one is at w will still be very small, so the example will stand.

Similarly, even when one considers possibility spaces with different structures (e.g. linear or multi-dimensional, perhaps continuous rather than discrete, perhaps with dimensions of variation other than the immediately relevant one, e.g. the position pointed at) similar phenomena will arise. The phenomenon also seems robust to more psychologically realistic descriptions of knowledge (they don’t increase the probability on one’s evidence that one knows the strongest relevant proposition one can know, even though one may in fact know it).

A structural generalization

A key structural feature of the example is:

(*) For all worlds v, w: $\{x: Rvx\} \subseteq \{x: Rwx\}$ only if v = w.

That is, shifting from one world to another (e.g. from w to v) always opens up new epistemic possibilities as well as closing down old ones. This is a plausible feature of real-life examples of inexact knowledge. In fact, this more limited feature suffices:

(**) Some world w is such that for all worlds v, $\{x: Rvx\} \subseteq \{x: Rwx\}$ only if v = w.

Note that (**) implies that R is non-transitive if at least one of the worlds w described in (**) has R to another world. For suppose that Rwv, $v \neq w$ and $\{x: Rvx\} \subseteq \{x: Rwx\}$ only if v = w. Then for some world x, Rvx but not Rwx. Thus transitivity fails.

Variable margins for error

One restrictive feature of the original model is that the width of the margin for error is in effect treated as beyond doubt, since it is built into the structure of the model. More specifically, since the model has only one world at which the pointer occupies a given position, worlds can differ over what positions are epistemically possible for the pointer only by differing over which position it in fact has. Yet it is plausible that there is inexactness in our knowledge of the width of the margin for error in addition to the inexactness in our knowledge of the position of the pointer. If so, then in more realistic models the world accessible from a given world w will include some at which the margin for error differs slightly from that at w, while the position of the pointer is the same. In particular, Rwv for some world v at which the margin for error is slightly less than at w. In such cases we may have $\{x: Rvx\} \subseteq \{x: Rwx\}$ when $v \neq w$. Pictorially: a sphere may contain a sphere of slightly smaller radius whose centre is a slight distance from the centre of the first sphere (Anna Mahtani makes a related point in recent work on higher-order vagueness). Then whatever is known at w is also known at v. In such cases, (**) may fail.

To construct models with a variable margin for error is easy. But doing so without making *ad hoc* choices is harder. In effect, we will need higher-order margins for error distinct from the first-order margins for error. There is no obvious non-arbitrary way of determining the relation between the width of the margins at different orders.

Nevertheless, we can still give an informal argument argue for a conclusion similar to that already reached in the constant margin case. Let p_w in the new setting be the strongest proposition known at w about the position of the pointer (or whatever other non-epistemic fact is relevant). Thus p_w may be true at worlds other than w; its truth-value remains constant across worlds where the position of the pointer is the same. Let c be the margin for error at w for knowledge of the position of the pointer. Thus p_w is true

at exactly those worlds where the distance of the pointer position from that at w is at most c. Let $ME_{<w}$ be true at just those worlds at which the margin of error for knowledge of the position of the pointer is less than c, $ME_{>w}$ true at just those worlds at which the said margin for error is greater than c, and $ME_{=w}$ true at just those worlds at which the margin for error is equal to c. On the margin for error conception, these three possibilities are mutually exclusive and jointly exhaustive. Therefore, by the definition of conditional probability, for any world x, x |= $Pr(Kp_w)$ =
$Pr(Kp_w \mid ME_{<w}).Pr(ME_{<w}) + Pr(Kp_w \mid ME_{=w}).Pr(ME_{=w}) + Pr(Kp_w \mid ME_{>w}).Pr(ME_{>w})$. From any world v in $ME_{>w}$ some world is accessible at which p_w is false, because the margin for error at v is some d > c, and a sphere of radius d cannot be contained in a sphere of radius c. Thus $ME_{>w}$ is incompatible with Kp_w, so x |= $Pr(Kp_w \mid ME_{>w}) = 0$. Consequently, x |= $Pr(Kp_w) = Pr(Kp_w \mid ME_{<w}).Pr(ME_{<w}) + Pr(Kp_w \mid ME_{=w}).Pr(ME_{=w})$. Since x |= $Pr(Kp_w \mid ME_{<w}) \leq 1$, x |= $Pr(Kp_w) \leq Pr(ME_{<w}) + Pr(Kp_w \mid ME_{=w}).Pr(ME_{=w})$. For simplicity, it is reasonable to assume that w |= $Pr(ME_{<w}) = Pr(ME_{>w})$, that is, the margin for error is equally likely to be less or greater than what is in fact its actual value. Since for any x x |= $Pr(ME_{<w}) + Pr(ME_{=w}) + Pr(ME_{>w}) = 1$, w |= $Pr(ME_{<w}) = (1-Pr(ME_{=w}))/2$. Thus w |= $Pr(Kp_w) \leq (1-Pr(ME_{=w}))/2 + Pr(Kp_w \mid ME_{=w}).Pr(ME_{=w})$. But $Pr(Kp_w \mid ME_{=w})$ is in effect the probability of Kp_w in the case considered previously of a constant margin for error (c). From that case we have at least w |= $Pr(Kp_w \mid ME_{=w}) < ½$. Consequently: w |= $Pr(Kp_w) \leq (1-Pr(ME_{=w}))/2 + Pr(ME_{=w})/2$. By simple calculation, therefore, w |= $Pr(Kp_w) \leq ½$. In other words, although p_w is in fact known at w, it is no more probable than not on the evidence at w that p_w is known. Even if we slightly relax the simplifying assumption that w |= $Pr(ME_{<w}) = Pr(ME_{>w})$, the probability on the evidence at w that p_w is known will not rise significantly above evens. Indeed, the probability may well be close to zero. For if the width of the margin for error varies only slightly (as a proportion of c) over the worlds accessible from w, then $Pr(Kp_w \mid ME_{<w})$ will be close to $Pr(Kp_w \mid ME_{=w})$ at w. Thus $Pr(Kp_w)$ at w will be only slightly greater than $Pr(Kp_w \mid ME_{=w}).Pr(ME_{<w}) + Pr(Kp_w \mid ME_{=w}).Pr(ME_{=w})$, in other words $Pr(Kp_w \mid ME_{=w}).Pr(ME_{\leq w})$. But, as just noted, $Pr(Kp_w \mid ME_{=w})$ is in effect the probability of Kp_w in the case of a constant margin for error, which goes to zero as the number of pointer positions increases. Hence $Pr(Kp_w)$ may well be close to zero even when the width of the margin for error varies. But even without that stronger conclusion, the result of the informal argument is enough for present purposes.. Uncertainty about the width of the margin for error does not undermine the possibility of knowing something without its being probable on one's evidence that one knows it.

Epistemic binds

Suppose that one knows that P but does not know that one knows that P. Given the knowledge account of assertion (*Knowledge and its Limits*, ch. 11), one is therefore in a position to answer 'Yes' to the question 'P?'. However, one is not in a position to answer 'Yes' to the question 'Are you in a position to answer "Yes" to the question "P?"?'. Thus reflection can appear to undermine one's epistemic position. The problem is exacerbated when one knows that P but it is very improbable on one's evidence that one knows that P, for then it is very probable on one's evidence that the true answer to the question 'Are you in a position to answer "Yes" to the question "P?"?' is 'No'. The same phenomenon occurs at higher iterations of knowledge.

<u>Closure cases</u>

Consider cases in which it is highly plausible that one knows each of many propositions (denial that one knows looks like a form of scepticism). From those propositions one competently deduces a further conclusion and believes it on that basis, while still knowing the premises. According to any standard version of multi-premise closure for knowledge, one thereby comes to know the conclusion. However, it may seem utterly implausible that one knows the conclusion, given the multiplication of error possibilities (the preface paradox; knowledge of the future and other cases whose importance John Hawthorne has recently emphasized in *Knowledge of Lotteries*). A similar problem arises for single-premise closure principles when one competently carries out a long chain of deductive steps, each with a small epistemic probability of error (in the multi-premise case, for simplicity, one's deductive competence is treated as beyond doubt); see Maria Lasonen-Aarnio, 'Single Premise Deduction and Risk'. Such examples have been used to motivate various non-traditional proposals about the truth-conditions of 'knowledge'-ascriptions (subject-sensitive invariantism, perhaps contextualism).

Here is a possible account of the multi-premise case, consistent with a traditional view of the truth-conditions of 'knowledge'-ascriptions. One does indeed know each premise, without knowing that one knows it. By multi-premise closure, one also knows the conclusion (without knowing that one knows it). For each premise, it is very probable on one's evidence that one knows it. However, it is very improbable on one's evidence that one knows every premise (compare the lottery paradox). Since it is granted that one knows the conclusion only if one knows every premise, it is very improbable on one's evidence that one knows the conclusion.

Here is a possible account of the single-premise case, consistent with a traditional view of the truth-conditions of 'knowledge'-ascription. One does indeed know the premise, without knowing that one knows it. For each deductive step, one carried it out competently but one does not know that one did. By single-premise closure, one also knows the conclusion, without knowing that one knows it. For each deductive step, it is very probable on one's evidence that one carried it out competently. However, it is very improbable one one's evidence that one carried out every deductive step competently. Since it is granted that one knows the conclusion only if one carried out every deductive step competently, it is very improbable on one's evidence that one knows the conclusion.

In some multi-premise cases, it is plausible that for each premise one knows that one knows it. Similarly, in some single-premise cases, it is plausible that for each deductive step one knows that one carried it out competently. Given those plausible claims and plausible background assumptions (e.g. that one knows the relevant closure principle), one knows that one knows the conclusion. Consequently, on the type of model explained above, the probability on one's evidence that one knows is one. In some cases, even more iterations of knowledge may be claimed. However, higher-order versions of the above accounts can be given. Here are those for n iterations of knowledge:

Multi-premise case: One has n but not n+1 iterations of knowledge of each premise. By multi-premise closure and plausible background assumptions, one also has n but not n+1 iterations of knowledge of the conclusion. For each premise, it is very probable on one's evidence that one has at least n iterations of knowledge of it. However, it is very improbable on one's evidence that one has at least n iterations of knowledge of every premise. Since it is granted that one has at least n iterations of knowledge of the conclusion only if one has at least n iterations of knowledge of every premise, it is very improbable on one's evidence that one has at least n iterations of knowledge of the conclusion.

Single-premise case: One has at least n iterations of knowledge of the premise. For each deductive step, one has n-1 but not n iterations of knowledge that one carried it out competently. By single-premise closure and plausible background assumptions, one has n but not n+1 iterations of knowledge of the conclusion. For each deductive step, it is very probable on one's evidence that one has at least n-1 iterations of knowledge that one carried it out competently [NB one has at least 0 iterations of knowledge of a proposition just in case it is true]. However, it is very improbable one one's evidence that one has at least n-1 iterations of knowledge that one carried out every deductive step competently. Since it is granted that one has at least n iterations of knowledge of the conclusion only if one has at least n-1 iterations of knowledge that one carried out every deductive step competently, it is very improbable on one's evidence that one has at least n iterations of knowledge of the conclusion.

These accounts can be modelled formally. In particular, it is easy to show that the standard possible worlds semantic framework builds in the deductive closure of any given number of iterations of knowledge (irrespective of whether the subject carries out the deductions). That is not to say that such a strong closure principle is realistic, just that the phenomena just described do not depend on any failure of closure.

In the accounts above, one's number of iterations of knowledge of the conclusion for one does not dwindle as the number of premises (multi-premise case) or deductive steps (single premise case) increases. What dwindles is the epistemic probability for one that one has that number of iterations of knowledge: hence the impression that knowledge is being lost as the number of premises or deductive steps increases. All that is really being lost is probability on one's evidence, but it is the probability on one's evidence of self-ascriptions of (iterations of) knowledge.

A large topic for further investigation: How far can this style of treatment be generalized to other cases of the apparent undermining of knowledge by higher-level considerations? For example, does it provide a credible defence of knowledge of conclusions about the future with a very low objective chance in Hawthorne-type examples involving multiple independent lotteries. Such results may be hard to avoid if one combines attractive general principles about the preservation and extension of knowledge (closure, memory) with a denial of scepticism.

28.vi.2007