I've been known to promote a particular picture of what *knowledge* is, which I'll say a word about today, but I've now come to a view about what it is to *understand*. This came about unintentionally by thinking about a classic epistemology problem, the Gettier problem. So here's the story of that.

SLIDE 2

For as long as anyone in the west could remember, knowledge was taken to be justified, true belief. The belief had to be true because you had to get it right, *and* you had to have a good reason to believe – a justification – to rule out the possibility of having your true belief ***accidentally***.

SLIDE 3

Things went very well for 2,500 years – and we reached the highest pinnacle of human evolution so far – until 1962 when the following kind of example was published:

SLIDE 4

Imagine an office with three workers, Smith, Jones, and Brown. Smith and Jones are friends but neither of them knows Brown. Smith was with Jones when he bought a Ford car, and since then Jones has given Smith a ride to work in it every day.

SLIDE 5

So Smith believes ***Jones*** owns a Ford, and has good reason to believe it. Smith also believes that ***someone*** in the office owns a Ford. After all, that follows logically.

SLIDE 6

And so, he has a justified belief in p, surely. But now, it turns out that a few years back Jones decided to change the status of his car, and now *leases* it instead of owning it.

SLIDE 7

So q is false, though Jones didn't tell Smith about this change because it wasn't an interesting life event, and he's still driving the same car. We *know* that q false doesn't imply p is false. It could be that p is true for a different reason.

SLIDE 8

And so it is. You see, although Smith doesn't know Brown, and so doesn't know what kind of car she has, it turns out that … *Brown* owns a Ford!

SLIDE 9

But now something is wrong here.

SLIDE 10

Smith's belief in p was justified. **PAUSE** He had good reason to believe it – we can't expect him to rule out all logically possible ways of being wrong or we'd never count anyone as justified.

SLIDE 11

But a lot of us have the feeling he doesn't have knowledge. Despite the fact that he's justified he still gets a true belief only accidently by double luck. He had *bad* luck with his justification, and then luckily had perfectly compensating *good* luck with p. So, this and many like it look like counterexamples to the ancient theory of knowledge.

SLIDE 12

And now we're off to the races. Philosophers began to add more conditions beyond justification to rule out these cases from counting as knowledge. And the problem opened up space for *entirely new* conceptions of what knowledge is involving the dimensions of reliability, virtue, tracking. /// And of course

SLIDE 13

distinguished people have been involved in the theorizing.

But all the theories so far fall to some Gettier case or other.

SLIDE 14
SLIDE 15
SLIDE 16

So fatigue has set in, along with a suspicion that the enterprise of giving a theory of knowledge is pointless. Besides the failures, one way of motivating this is to notice that though the Gettierized subject, Smith, is merely lucky that his belief is true, his luck is

good. So why do we care? The only reason to care is that he doesn't have *knowledge*; but what is the added value of ***that***? Maybe the problem exists only because we're trying to define a property that has no added value.

Jon Kvanvig has developed these questions in a sophisticated way, arguing that the two goals of defining away the Gettier cases, and explaining the value of knowledge are in a trade-off situation and cannot be jointly satisfied,

and that any attempt to explain the value of avoiding **Gettier**-style luck can only succeed by **presupposing** the value of knowledge.[1]

SLIDE 17

So, whether from fatigue or sophisticated arguments, many people think the hamster wheel of Gettier cases should be telling us something.

SLIDE 18

Maybe we should step off.

SLIDE 18

Try some *different* theoretical goals than defining knowledge.

SLIDE 19
SLIDE 20

I do see the hamster wheel this all points to but I don't draw the same conclusion that Jon and many others do.

SLIDE 21

What looks like a wheel from the **inside** can also be a spiral, that broadens our vision or takes us deeper in. And not only that.

SLIDE 22

[1] In other words, the Gettier problem is only worth worrying about if knowledge is worth worrying about.

Though the discussion repeats the question and repeatedly fails to get necessary and sufficient conditions for knowledge, it also covers and *dis*covers **new** ground all the time.[2]

And Jon's challenge about the added value of knowledge in relation to its nature is a great question, but I think there's at least one very good answer,

SLIDE 23

and the first thing I'm going to do today is explain the value part of that answer for the probabilistic tracking theory of knowledge.

Secondly, I'm going to give a probabilistic characterization of what the gettierized subject is missing and I'm going to call it relevance matching. Then I'll show that tracking promotes this and so, the same thing that explains the value of knowledge and was there as the definition of its nature ***also*** directly addresses the Gettier Problem. We'll see that steps of improving relevance matching (and so Gettierization-avoidance) do not necessarily improve tracking, but the matching by itself serves something precise that seems to me best interpreted as ***understanding***. So we'll see that though tracking directly promotes G-avoidance, you can explain the **value** of Gettierization avoidance without **assuming** the value of knowledge. As a side effect of this we'll get a picture of what it is to understand why p is true and it will say that understanding is simulation.

So, my first point today is that if knowledge is tracking, then it has a high value that mere true belief does not. To explain that value, I'll go through a few steps of how it's derived.

SLIDE 24

So, value of knowledge. We start with a game that a subject plays with Nature. In any given round of the game Nature has two options, to make proposition p true or to make p false. In a given round the subject's options are to believe p or not believe p. To have a game there must be payoffs associated with the four possible outcomes. Nature's payoffs are in the first position in each ordered pair. (She's indifferent [to the consequences of the play] so they're all zero.) Your payoffs are in the second spot. The

[2] I also just think it's decadent for an epistemologist to say the concept of knowledge isn't important.

particular numbers here don't matter but some of their relationships do. I'm going to assume that we're talking about p for which when p is true it's better for you to believe it than not to believe it, and when p is false it's better not to believe it than to believe it. It will only be for those p that ***fulfill*** these criteria that we can prove the added value of tracking-type knowledge in this way. But there are a lot of such p.

SLIDE 25

These assumptions correspond to some of the relative values of the different possible types of belief **states**. Some of those states are typically good and some bad.

SLIDE 26

For the derivation of the **value** of knowledge, it's sufficient to make the two assumptions I listed as defining the payoffs, that the top blue state is better than the bottom blue state, and the top green one is better than the bottom green one.[3]

When you see what comes out of this derivation you may suspect that the result was rigged by these payoff assumptions, but this is definitely not so, for several reasons.

First, the challenge to explain the value of knowledge is that of explaining why it is *more valuable than mere true belief*. No one expects us to explain why knowledge of p is valuable for p where true belief isn't valuable, or where avoiding false belief has no value. So, the restrictive assumptions about which states are more valuable than which are not crutches smuggled in to the discussion by this representation. **PAUSE**

The payoff assumptions also do not trivialize the result I will describe by being only ***trivially different*** from the result.

SLIDE 27

And this is because the payoffs concern belief *states*, and the result will be about *strategies*.

[3] [True belief > false belief doesn't follow because of scale, e.g. if false belief still has positive value, but you could add it and it wouldn't undermine the result because the other conditions still hold.]

Strategies are *dispositions* or *regularities* in your responses to the plays of the other player. So, they're conditional, and modal, and general. They're rules. (I don't say the particular rules depicted are good, of course.)

SLIDE 28

States pertain to a single round of the game. In a given round, you may land in a good state (square) or a bad state (square). In our game the peachy outcomes are good for you because they have the best payoffs. 10 > -20 and 5 > -7. The yellow outcomes are bad. [Yellow for caution.] That's good and bad outcomes in a given round. It is a co-incidence of two things, p, and belief in p on a given occasion.

SLIDE 29

But now **strategies** are a different matter. A *strategy* is a general rule for responding to all the possible plays of the other player – so it's a ***set*** of conditionals – and it will have consequences for where you land how frequently in *all* or a large number of trials of the game.

There are some important things to know about strategies *as discussed here*: first, as in all game theory, no assumption is made that they involve conscious or ***deliberate*** actions. A strategy is only a regularity of [belief] response.[4] Also, here we don't assume in the *definition of the game* that the player of belief states has *knowledge* of which play Nature has made. We're going to judge whether he has knowledge or not by looking at his strategy and asking about its game-theoretic properties. [Note also that all possible response rules are included in the set of possible "strategies". Even responding *randomly* to nature's plays is in there as the degenerate strategy of using a random generator.]

SLIDE 30

Okay, so **here's** a strategy. The ***tracking conditions*** are dispositional, and the pair of them form an example of a strategy in the true belief game because they cover all the possible plays of nature, p and not-p: You satisfy them if the probability you **don't** believe p given that p is false is high, and the probability you **do** believe p given that p is

[4] Pancreatic cells have strategies in this sense. They respond to the presence of sugar by producing insulin and the non-presence by not producing insulin.

true is high. That is, in response to nature's play of not-p, play no belief and in response to nature's play of p play belief, say, in all scenarios with greater than 95% probability. (so s = t = .95)

SLIDE 31

In other work I've defended these conditions as a definition of knowledge. The first, not believing p when it's false, is called the *variation* condition, or more recently "sensitivity," and the second, believing it when it's true regardless of variation in **other,** irrelevant**,** things, is called *adherence*.

Having the tracking strategy is an achievement not a choice or a mere choice. It's possible to exhibit the tracking strategy without a decision or deliberate effort, and often a mere decision to have it does nothing to help you fulfill it. An example of the first occurs in perception, where we get the equipment to fulfill this strategy for free from having normal organs. We don't choose to believe there's a tiger if and only if there is one. An example of the second, where a mere choice doesn't get us the status of having the strategy, is a proposition like the existence of the Higgs boson. For human beings to have the tracking dispositions for this p is going to require building something like a ***Large Hadron Collider***. However, that doesn't mean the tracking conditions are too strong: I think we can all agree that the existence of the Higgs Boson is a *very hard thing to know*.[5]

Now, there are an infinite number of other possible response **strategies** in the true belief game, and they correspond to all possible general conditions on the relation between p and your belief in p, and so *all possible definitions of knowledge of the form true belief + X, that do **not** have tracking among their requirements*.

[As long as those conditions are general, as they always are, they can be written as strategies. The X of a given theory of knowledge (e.g., good reason to believe, reliable process of belief formation) is extensionally equivalent to some strategy because it provides an answer for every situation where p is true and p is false as to whether the subject who has what's counted as knowledge of it believes it. Decomposed this way

[5] [What a particular kind of being has to do to achieve the tracking conditions is important, but it is a contingent matter.]

those X conditions that aren't tracking will typically be disjunctive, but that just makes them what are called "mixed" strategies.]

Now let's see how all these strategies serve us in the true belief game.

SLIDE 32

Obviously the tracking strategy leads us to land in those good peachy squares almost all of the time. I.e., 95% of the time we win.

Now – here's where I'm skipping over technical details – this game we play with Nature can be converted into a symmetric signaling game with different strategies compared to ***each other*** in trying to win the game with Nature. If we call the tracking strategy R then all the other possible strategies (i.e., other theories of knowledge) are members of –R, and we get an interesting implication:

SLIDE 33

The subject who is a tracker of p has what's called an ***evolutionarily stable strategy***, an ESS.

SLIDE 34

What this means is that

> The tracking type strictly dominates any type following *any* other conditions beyond true belief , in the struggle for survival and utiles. [Any given round can be lost by any individual, but this type wins over time.]

This is significant for the triviality worry you might be having again: far from being trivial it is not just that tracking *has* an added value, but that it has a unique added value: no other logically possible added condition on knowledge will swamp or even match the added value of tracking. (For example, justifiedness, virtue, reliable process, etc., etc.).

> Secondly, what it means for it to dominate is that once this strategy is achieved by some level of majority of the population, no small population with an alternative strategy can "invade" and drive it out.
>
> These properties hold no matter what the dynamics of interaction are. That makes it an extremely strong property game-theoretically.

SLIDE 35 *It follows that if we think intuitively that knowledge can be of evolutionary or utilitarian value, then this is a unique* ***explanatory*** *advantage of the tracking theory.*

Secondly, the ESS property shows that (tracking) knowledge is more valuable than mere true belief, *and,* ***contra Jon,*** *shows it in terms of the value of its key component, namely the tracking*. I didn't do any ad hoc tinkering.

But what does this added value I've derived amount to *intuitively*, for a human being in real life?

SLIDE 36 Well, suppose you want to go to Larissa.

SLIDE 37 You believe route A will get you there by 12, and suppose this is actually true. Suppose that when p is true it's more valuable to you to be in a state believing p than not believing p and when p is false it is more valuable to you to not believe p than to believe p. To satisfy these conditions it's sufficient that you ***do want*** to go to Larissa by 12.

Then you're better off using real-time GPS like S' does than a paper map like S does. S is right when he starts, it turns out, and may continue to be right as he wends his way down the road, but S' is prepared for the possibilities of road work and impassible traffic. In these scenarios S will detect the obstruction when he gets to it. But S' will detect it ahead of time, and have more time to adjust. S' is more likely to get to his appointment on time in this scenario. This scenario that might not end up being actual – in which case S and S' will both turn out to actually have true beliefs. But impassable traffic is possible at the beginning and remains a realistic possibility as they continue down the road.

SLIDE 38

What does S' have in that GPS machine that S doesn't? She has the means to **track**, in the specific sense of **a strong disposition to believe p when it's true and not believe p when it's false.** This disposition that she has now nevertheless ***applies*** to all the different times and stretches of road along the way. [It's present even for cases where it's not activated.] // S tracks a little but at a much lower level of resolution.

SLIDE 39

The general upshot is that S' has a true belief plus a contingency detector, one value of which is preparedness, a value that S' possesses *now* because of the GPS and hence tracking she has now. If your utilities concerning belief states are going to remain the same, then a tracking device is what you want to have now. This applies to any contingent proposition whose truth value matters to you, because its truth value may change. [I believe necessary truths are *known* in a different way. And I don't yet have a view about what the value of that knowledge is.]

Importantly, the value may not be practical or consequential. You may not want to go to Larissa but may still want to have appropriate beliefs states, perhaps because you are writing a dissertation on traffic flow. [And I think this is true: don't be misled by the connotations of the word "utility" – having the appropriate belief states may not be of consequentialist value to a person; they may affect what one is willing to do just because. Their value to a person could be intrinsic as long as they will continue to have that value.]

SLIDE 40 I discuss this argument about tracking in more detail in the paper "The Value of Knowledge and the Pursuit of Survival."

Now I want to turn to Gettier cases, the avoidance of which has I think a recognizably different but related value. [In other words, you don't have to care about knowledge in order to care about Gettier cases.]

SLIDE 41

There is an impressive menu of interesting, intuitive, and fruitful ways of thinking of these cases. I don't claim any of these accounts is wrong, but I'm going to add another to the list. I will characterize Gettier cases in terms of probabilistic relevance.

SLIDE 42

To begin with, a bare bones intuitive way of describing what is wrong in these cases is that what makes the subject believe p is not what makes p true.

I'm going to gloss this by saying that what is *relevant* to S's believing p is not *relevant*, to the same degree, to p's being true, or vice versa. In our case, whether Brown owns a Ford is not relevant to whether Smith believes someone in the office owns a Ford. Only Jones is relevant to that. But whether Brown owns a Ford *is* relevant to whether *in fact* someone in the office owns a Ford.

SLIDE 43

More precisely, as the first equation says, the probability that Smith ***believes*** p given that Jones doesn't own a Ford is the same whether Brown owns a Ford or not.

And the problem is that Brown DOES make a difference to whether p is ***true*** when Jones doesn't own a Ford.

$P(b(p)/\text{-}q.r) = P(b(p)/\text{-}q.\text{-}r)$ but $P(p/\text{-}q.r) \neq P(p/\text{-}q.\text{-}r)$

The probability of **p** given not-q is ***different*** depending on whether Brown owns a Ford or not. (And because it makes a difference when q is false it makes a difference overall.)

Now when a Smith so described who believes p when he believes q makes an unusual mistake of believing q when q is false, this difference between his belief dispositions and the dependences between things in the world will be exposed. His belief in p will be either false or Gettiered, depending on whether **Brown** doesn't or does own a Ford respectively.

We can get a better feel for the relevance point by writing it out more fully:

SLIDE 44

First, here's what it is for q to be highly positively relevant to whether you ***believe*** p:

$P(b(p)/q)/P(b(p)/\text{-}q) >> 1$

you are much more likely to believe p given that q is true than you are given that q is false. q's truth value is something you rely on in believing or not believing p. You can also write that as the ratio of those two terms being much greater than 1.

SLIDE 45

But there's another, distinct, question: whether q is positively relevant to p's being ***true***: This can be represented as

p is more likely when q is true than when q is false. q makes a positive difference to whether p is true. And that can also be written as the ratio being much greater than 1.

SLIDE 46

Now these two things, the relation between q and p and between q and your ***belief*** in p, may be in line with each other or not. That is, those relevance relations can be *matched* or *mismatched*.

Comparing the two relevances by the ratios, they are ***matched*** when the difference that q makes to your ***belief in p*** (the term on the left hand side) is the same as the difference that q makes to p (the right hand side). That's the first equation.

They're **mismatched** when q's truth value makes more of a difference or less of a difference to ***your belief*** in p – left hand side – than it does to ***p's being true*** – right hand side. That is, the inequality on the bottom.

[It's not important to parse these equations here. If you want to I can give you the slides later.] [P(b(p)/q)/P(b(p)/-q) not equal to P(p/q)/P(p/-q)]

SLIDE 47

In our Gettier case we had a relevance mismatch because q – Jones owning a Ford -- makes more difference to whether Smith **believes** p – someone owns a Ford – than it does to whether p is **true**. There's also a mismatch on r because r – Brown owning a Ford – makes more difference *to p* than it does to whether Smith *believes* p.

SLIDE 48

So my hypothesis about gettierization is this: when you have such a case you'll always have a relevance mismatch on a q the subject is depending on for his belief in p,

Or a mismatch on some r that p's truth value depends on. (And either implies the other.)

SLIDE 49

The arrow doesn't go in the opposite direction because not all cases of mismatch are ACTUAL gettierizations. For that the mismatch has to be exposed. The good subject's grip on q is often enough to avoid exposure of the mismatch: as long as our Smith doesn't believe q when q is false he won't be believing **p** when q is false either (and so, won't be Gettiered) because he's a guy who's only using q to decide whether to believe p. So the question what is happening with that relevant proposition he isn't paying attention to doesn't often arise for a good subject.

SLIDE 50

To summarize what I think a Gettier case is: we need three things, that the person bases belief in p on q when q is false, that there is a relevance mismatch on q for the first condition to exploit, and that p is true.

Improving tracking of p (by being more sensitive) decreases the probability of the first feature directly. But 2 is the distinctive element in a gettier case. Obviously, increasing ***relevance matching*** decreases the probability of 2 directly, So …. we can ask whether improving **tracking** addresses the distinctively gettier problem by asking whether improving **tracking** improves ***relevance matching***.

And it turns out you can just write down the relationship between tracking and relevance matching very neatly:

SLIDE 51

$$P(b(p)/q) = \frac{P(b(p)/p)P(q/b(p).p)P(p/q)}{P(q/p)} + \frac{P(b(p)/-p)P(q/b(p).-p)P(-p/q)}{P(q/-p)}$$

$$P(b(p)/-q) = \frac{P(b(p)/p)P(-q/b(p).p)P(p/-q)}{P(-q/p)} + \frac{P(b(p)/-p)P(-q/b(p).-p)P(-p/-q)}{P(-q/-p)}$$

The colors here are the important part. (Same color doesn't necessarily mean same term, but term with the same interest to us.) The ratio of the purple terms on the left-hand sides expresses the relevance of q to my **belief in p**. The probability that I believe p given q and the probability that I believe p given that q is not the case. The ratio of the red terms on the right-hand sides expresses the relevance of q to **p**. The probability that p is *true* given q is **true** and the probability p is true given q is **false**. For relevance matching we want the relevance level of q to p (the red) to translate through these

other terms into the same relevance level of q to my *belief* in p (the purple); *we want the ratio of the purples to equal the ratio of the reds*.

SLIDE 52

SLIDE 53

Same equations, more colors.

In this slide we see that level of relevance matching for p on a given q depends on how well I fulfill the *tracking* conditions **for p**, the blue and green terms,

and on a couple of other things.

The dependence looks complicated but it's simple in extreme cases.

If I'm perfectly sensitive – the first tracking condition -- then I never believe p when it's false, and the blue terms, and the whole second summands with them, go to zero.

SLIDE 54

If in addition I'm a perfect adherer – I always believe p when it's true, the second tracking condition -- then the green term is 1 and that forces the tan terms to be equal, so what is multiplied by the top red is equal to what is multiplied by the bottom red.

[P(q/b(p).p)/P(q/p) = P(b(p)/q.p)/P(b(p)/p). Now P(b(p)/p) = 1, so that whole green and tan term becomes: P(b(p)/q.p). But since again adherence is 1, this is 1 too.]

SLIDE 55

That means the factor that lies between the red and the purple is the same in both equations, so

SLIDE 56

the ratio of the purples equals the ratio of the reds.

SLIDE 57

And that means that perfect *tracking* of p implies perfect *relevance matching* for p on q.

We can do this argument for p on any q,

SLIDE 58

so we have that perfect tracking gives perfect relevance matching on all q.[6]

SO if I fulfill the extra condition for knowledge ***infallibly***, then I can't be Gettiered. But of course any theory with an infallibilist requirement can disqualify gettier cases because it allows no room for luck. What matters most is what happens to a subject who tracks ***fallibly***. The (fallibilist) tracking theory actually disqualifies the Gettier case I went over, and many standard cases, though maybe not all.[7]

However, we know a priori that there is no fallibilist theory that disqualifies *all* such cases, so to get anywhere we have to ask comparative questions of better and worse.

SLIDE 59

Fallible tracking addresses the Gettier problem *directly* in the following way: **NUMBER 2** the ***better*** you track p the ***better*** you relevance match for p on ***every*** q. Fallible tracking won't make gettierization impossible, but improving your tracking will always make it less likely. To put that into context, there ***are*** fallibilist theories for which this improvement relation doesn't hold: for example, increasing internalist justification of belief in p does not necessarily reduce the chances of mismatch because internalist theories have no condition at all on your relation to the world other than the obvious one that your belief actually be true.

But given that fallible tracking allows Gettier cases as possible one might wonder how I can take myself to have addressed Kvanvig's argument that you can't understand both the nature and the value of knowledge simultaneously, because of the gettier cases.

Jon's argument wasn't merely that no criterion could rule out all gettier cases – we knew that no fallibilist theory can do that. His point was that when you come up with conditions that seem to explain the value of knowledge they face the gettier problem,

[6] [This is because *perfect* tracking requires you to be someone who does the right thing with your belief in p not just in the probable scenarios but in all possible scenarios.]

[7] [One recipe is: make P(r/-q) superlow, so it falls below what you can reasonably expect a subject to relevance match to, but the occurrence of r when you believe q falsely is still possible.]

and the fixing that you have to do to even begin to rule out gettier cases is ad hoc from the point of view of the condition that insured the value of knowledge.//

So for *that* claim what's important in what I've found is that the tracking conditions got no ad hoc tweaking, no adjustments at all, when they went from explaining the value of knowledge to rooting out gettier cases where they live. The same conditions did both. Secondly, the tracking conditions addressed what is distinctive about the gettier cases, rather than just generically decreasing luck: not only does tracking make it less likely that a subject will believe q when it's false – the first thing we need for a gettier case – but also better tracking makes it less likely that the subject will get gettiered if he *does* have that false belief in q – the second, relevance mismatch, aspect.

SLIDE 60

Okay, now looking in the other direction, from relevance matching to tracking, actually, that first claim I made doesn't ***only*** go left to right. It also goes right to left. (green double arrow) **Perfect** relevance matching for p on every q gives perfect tracking.

But while better tracking of p improves relevance matching on every q,

We have number 3: increasing your relevance matching for p on **some q** doesn't necessarily increase your tracking of p.

SLIDE 61

Here's a diagram that can help explain these claims. If you track p *perfectly* then whether you use an indicator q or not, your belief in p co-varies perfectly with p. So the proposition *that you believe p* is substitutable for p in every relevance relation p has to any q. So, your belief in p relevance matches for p on every q.

Other direction: If your belief in p perfectly relevance matches for p on every q, that means that it stands in the same relation to each q as p stands to that q. So your believing p is substitutable for p in every relation that could affect the truth value of p. That is, you track p.

SLIDE 62

Okay, that's perfection. What about imperfection? What if your belief relates to p via an intermediary? Your belief in p depends on q strongly as a proxy – bottom picture. Whether you track p depends on how good a proxy q is.

If it's a perfect proxy for p then there's a solid double-sided arrow between q and p and the fidelity of your tracking of p is limited only by how closely your belief in p depends on that perfect proxy.

SLIDE 63

If q is not a perfect proxy for p then the tracking has two limitations so the relation between p and your belief in p is *weaker*. About relevance matching,

SLIDE 64

Let's call the proposition your belief uses as a proxy q_1, to make room for some more factors. When q_1 is a better proxy for p, your tracking of p is better. But also, the substitutability between q_1 and p is tighter, and that makes q_1 vary with the other factors in p's web of relevance in a way closer and closer to exactly the way p does. Because your belief in p is following q_1, your belief in p automatically becomes more relevance matched to those other q's as q_1 becomes substitutable for p in all of p's relations to other propositions.

SLIDE 65

In the opposite direction, you may improve your relevance matching for p on q_2 by getting your belief in p into a relation to q_2 that is more similar to p's relation to q_2, but you might do it at the expense of your belief's nice relation to q_1. (Maybe you outsource **detection** of q_1 so that you can investigate q_2, and your third party isn't as good as you are.) So you improve your relevance matching to one q but because you reduce it wrt another q you don't track p any better. Improving relevance matching on one q doesn't ***imply*** improved tracking.

OKAY, and so on and on. What is relevance matching intuitively and what is it for?

We know that tracking, consequentially speaking, serves the bottom line, dominance in the struggle for utiles. Now we know that piecemeal relevance matching and hence

piecemeal Gettierization avoidance doesn't necessarily achieve that. So what ***are*** Gettierization avoidance and hence relevance matching, for?

SLIDE 66

I think it's unmistakeable in the case I went over that the Gettierized subject, the one who fails to relevance match on a matter that is relevant to his believing p, *doesn't understand why p is true*. ***Understanding*** is what I would say relevance matching and Gettierization avoidance are *for*. // [And this could be a reason not to fret that tracking allows a few gettier cases, because tracking is for knowledge, which is a different thing.]

Notice that this is an answer to Kvanvig's challenge to explain the value of *gettierization avoidance* without assuming the value of *knowledge*. // We don't have to assume that I want to be able to go to Larissa by noon in order for it to be valuable to me to *understand* something about how to get there by noon. Those roads and traffic patterns may be interesting in themselves, or important for my dissertation.

SLIDE 67

I like this interpretation of relevance matching so much I'm trying out ***defining*** understanding in these terms. So, as a first pass, if p is true then for a subject who believes p

Understanding of why p is true *improves* iff there is an increase in relevance matching for p on some q, and no decrease in relevance matching on any other q.

So, your level of understanding goes as your level of ***net*** relevance matching.

SLIDE 68

So to re-state our previous implications in the terms of understanding, we have:

Increase in level of tracking p will increase relevance matching for p on every q. In other words, tracking brings relevance matching and so Gettierization-avoidance, and – now we know – understanding.

And we saw also that improved relevance matching for p on a *given* q doesn't necessarily improve your tracking of p.

HOWEVER, recall another thing I cited earlier:

Increasing your relevance matching for p on ***every*** q *does* increase your tracking of p. And overall increase in relevance matching is what we required for improving understanding. So,

SLIDE 69

Improved *understanding* of p always improves tracking level (i.e., promotes knowledge) of p.

So, knowledge improves understanding (as so far defined), and understanding gets us closer to knowledge.

You might also have noticed that perfect understanding as I've defined it so far is equivalent to perfect tracking. So, what's really the difference? There's going to be a difference even in the perfect cases when I refine the definition of understanding in a minute, but for now there is at least a very important pragmatic difference. For contingent reasons, unlike god we have to use intermediaries in order even to track p. At any given point we only have enough resources to address some factors relevant to p and not others. If we choose to hone our grasp on some q that is a very good indicator of p we will have something that in fact makes our belief in p relevance match for p on every q, but we won't be aware of what those other q are. Nevertheless this is what we should do if we want simply to be able to predict p's disposition in the future. (Of course, for contingent reasons we may have to look for more than one q in order to get a set that constitutes an indicator but the point is that we *may* get by with only one. I'll show you an example in a moment.)

We could choose instead to spread our resources and investigate multiple q's, trying to get their relation to p. Since we wouldn't necessarily increase ***net*** relevance matching we wouldn't necessarily improve tracking of p, or understanding strictly speaking, but we would have made progress toward understanding what makes p true, and as we went about this we'd be making ourselves ***aware*** of more factors relevant to whether p is true, whether they suffice together to be a highly tracking indicator or not. So whether you take a knowledge approach or an understanding approach to p, your

efforts will also take you along the road to the other, but you'll have different results in the short term, so which approach you choose matters.[8]

There could be an insight here about a difference that has frequently been commented on between natural science and philosophy. Even when philosophers can't decide or agree on whether p, say justice is fairness, is true, we increase our knowledge of what its truth does and doesn't depend on, and even when we don't know whether those things are true either. The framework here gives a way seeing us as improving understanding of why p would be true or false, even if we don't settle whether it is or isn't. **PAUSE**

SLIDE 70

I made a choice in defining relevance matching that has an implication for the definition of understanding. I didn't say your belief in p must relevance match on the q's ***via beliefs*** in the q's. You can relevance match your belief in p to the web of q's around p by your belief in p co-varying directly with whether or not the q's are ***true*** as long as your belief in p does that in the same way that p does that.

This means that in order to understand why p is true you don't need to be able to *cite* those q that are in p's web of relevance, and this allows me to distinguish between understanding why something is true and the ability to give an explanation of it.

We could ***add*** a citation requirement, and then we'd get a definition of explanation that is more or less equivalent to Wesley Salmon's statistical relevance view of explanation.

[8] ~~Second, I conjecture that if you improve relevance matching for some q that are positively relevant to p, that are in p's web of relevance, so to speak, then you improve *tracking of* ***those q***. (But wait, that has to assume you have a belief or other mental state to do the tracking. Hmm.) This is a kind of robustness that I think could explain our sense that in risking gettierization you are risking a mistake that goes beyond that of being wrong about p. You are risking not tracking the q that are relevant to p. You have a delicate hold on the area. I think this could be practically valuable in another way, These are ideas.~~

And this distinction between understanding and being able to give an explanation is a good thing, I think, because I think there *are* things we understand but for which could not list the states of affairs that make it so.

SLIDE 71

Here's one kind of familiar example: **READ**

S is probably able to say *something* in support of the prediction. But she doesn't need to be able to become conscious of a set of factors sufficient for that prediction in order to understand. Her belief about what the person will do has a "feel" for what that person will do and I'd say that feel comes from the co-variation with the web of relevant factors that I described.

So, here's a set of pictures to illustrate what I'm saying about understanding.

SLIDE 72

This is p and its web of relevance, that is, the q that are relevant to p and the degrees of their relevance. The triangles are indicating dispositional or dynamical relationships, not just co-incidences.

Now THIS

SLIDE 73

Is going to be what it is for someone to have merely a true belief, a belief in p that is co-incident with p in the actual world. There's no triangle between p and belief in p, no dispositional or dynamical relationship between them.

And THIS

SLIDE 74

Is what it's going to be for that belief to relevance match for p on every q. The belief in p has the same relation to every q that p has to that q. In contrast,

SLIDE 75

THIS is something I didn't do, and that spared us a citation requirement: I *didn't* require that in order to relevance match for p on q the subject has to have a *belief* in q that

coincides with q and that co-varies with the belief in p in the same way as q co-varies with p, so beliefs in the q are not required for understanding.

Okay, so this is a good consequence of the way I defined relevance matching for understanding. We're not requiring for understanding something that people often don't have. Here are a couple *awkward* consequences of my choice.

SLIDE 76

Concerning understanding, there is an obvious distinction between understanding why you should believe p and understanding why p is true,

SLIDE 77

which corresponds to a distinction between ***indicators*** of p and things that ***make*** p true. The current definition of understanding counts you as understanding why Jefferson is dead merely by having your belief that he's dead co-vary with the lack of pulse. That's because the pulse will give you tracking and tracking promotes relevance matching for all q. You'll probably be aware of the dependence between p and the lack of pulse, but on the current definition of understanding you don't need to have knowledge of or even ***ANY AWARENESS AT ALL*** OF those *other q* in order to count as understanding why Jefferson is dead.

Now I can make that distinction between indicators and what makes p true in a principled way but that won't solve this problem. Because …

SLIDE 78

Any belief that co-varies with a tracking indicator DOES co-vary appropriately with the other, substantive, q. That follows logically from the tracking. So, the definition of understanding is just too weak. *And there's another problem in the same direction.*

SLIDE 79

Suppose your believing p (that there will be a hurricane tomorrow) co-varies with the output of a great computer simulation programmed by someone else. Then you track p, from which it follows that you relevance match for p on all q. Hence you count as understanding why p is true. But surely this isn't a case where you understand. (Maybe the ***computer*** understands. But ***you*** don't.)

So, we need two things. We need a requirement that you be in closer touch with the ***causes*** and not merely the indicators of p, and we need you to **"own"** the co-variation with those factors in a way that you don't when what guarantees your co-variation is only housed in the computer.

SLIDE 80

Both of these would be satisfied if we required you to have beliefs for every q. But that was too strong.[9]

SLIDE 81

This is too weak, as we just saw. But there is an in-between option:

SLIDE 82

In this option you're required to have **mental states** for the q in p's web of relevance, but *not necessarily* ones that are consciously accessible to you, so not necessarily beliefs.[10] Having mental states means you'll have things in your mind that correspond to the factors that cause p, and thereby your mind is ***following*** these factors, if only implicitly. And it means that ***you*** and not just the computer own something supporting the relevance matching your belief in p is doing. So I will be counting you as understanding only if you have the ability to walk away from the computer and still relevance match to the q.

Notice that this change makes understanding different from knowledge, and strictly ***stronger***. For knowledge no intermediaries of any sort are required because tracking could be fulfilled directly (although we never do that). Since for tracking there don't need to be any indicator facts in the world that help the belief to co-vary with p, there definitely don't need to be mental states corresponding to them. But if you have full understanding of why p is true, you *will* have mental states for all q's that are relevant

[9] because that would mean that in order to understand you must have the ability to become conscious of and report all the relevant factors.

[10] You will have to be able to recognize their propositional content as propositional content, if prompted, but you do you even have to be right about whether your belief depends on that? Well, you might not know, but it doesn't sound okay to actually have a belief about that matter that is wrong. Not clear computers have mental states in our sense. But when they have the semantic skills for attributing compositionality then I think we'll say they do. So I think there's a good chance that if we saw a computer doing what Searle imagines then we WOULD attribute understanding (even though the human being in question doesn't have (because he's not doing the computations). More importantly, I think we do not now know how we would react.

to p. What follows from this is that it is possible to *know* that p is true, without knowing *why* it is true. And this matches our commonplace experience of using a mere indicator.

SLIDE 83

With this strengthening of the definition of understanding we maintain that good thing we had where understanding another person doesn't require the ability to cite all of the things that make you able to predict her behavior. Mental states aren't necessarily things you can become conscious of. Nevertheless if you understand in the sense I defined there ***will*** be something going on in your mind that you're carrying with you and that is making you able to have the right belief about what she will do. And that thing is …

SLIDE 84

… a ***simulation***. So, the definition I'm offering is a way of making *out* the simulation view of what it is to understand other minds. It's a simulation in the technical sense because those triangles are dynamical relationships, rules that tell you how the bubbles behave in all possible situations.

SLIDE 85

There's a lot more to say and decide here, and another metaphor has begun to pop out for me, with new pictures. But, I'll stop there for now to summarize.

1. Knowledge (tracking) is more valuable than mere true belief; it is an ESS.
2. What explains that value (tracking) also directly opposes gettierization.
3. Gettierization avoidance for p has a value – contributing to understanding p – even if we don't assume knowledge of p has value.
4. Understanding goes as relevance matching for p via mental states tracking the various q's relevant to p, and understanding why p is true is *simulation of p's web of relevance* rather than having beliefs and inferences *about* why p is true.