

## INTENTION AS A MODEL FOR BELIEF<sup>1</sup>

RICHARD HOLTON, MIT

I start with a modern-day tale of two families, each happy in its own way. They have much in common: parents working in busy jobs, young children in various schools and childcare centers. Here though I focus on a difference. One family, the Gadgers, have long been enthusiasts for new communication technology. As a result they can, and do, contact each other many times during the day, confirming whereabouts, sharing information, asking questions, refining plans. In contrast the Luddeys have never owned so much as a mobile phone. They talk when they are together in the mornings and the evenings, but communication during the day, conducted by traditional landline telephone, is typically employed only in emergencies.

Thanks to their electronic resources, the Gadgers have capacities that the Luddeys lack. They can fine tune a last-minute rendezvous at a previously unvisited part of the city at any time of the day. But exactly because they have this capacity, their lives are constantly in flux. Try to make an appointment to see an adult Gadger, and they will tell you that they will need to get back to you. They will need to get back to you because they don't yet know quite what they will be doing at any particular time. Someone will need to pick up the children, but who will pick up whom will be decided at the last minute depending on who is where doing what. And from such uncertainties flow many others. The Gadgers' lives are governed by a complex of conditional and parallel plans, a constant to-ing and fro-ing between each other and those equally connected to the electronic highways of contemporary life.

Life with the Luddeys is rather different. They know where they will be during the day, because they agree on a plan in the morning (a plan which itself rarely changes from week to week), and then, barring accidents, they stick to it. Ask to see a Luddey and they will tell you when they are free, and will be ready to make a time unconditionally. But the offer is normally made on a take-it-or-leave-it basis. Ask them to move things around a little to accommodate your needs and they will typically refuse.

I shall not speculate on which family leads the better life. Neither would want to trade. Clearly though, despite the many things that they cannot do, the Luddeys' life has some advantages. It is simpler. The Luddeys think that that is a good in itself, but it also brings further benefits. It means that the Luddeys are much less likely to get muddled in their arrangements, less likely to forget which plans are still up in the air, less likely to fail to follow through on the plans that they have provisionally made. In contrast the Gadgers often find that they have lost track of what they are supposed to

---

<sup>1</sup> This paper develops ideas originally presented in my 'Partial Belief, Partial Intention' *Mind* 117 (2008), pp. 27-58. Versions of it have been presented at Harvard, Chapel Hill, Stanford and MIT. Thanks to the audiences for excellent discussion, and especially to Lynn Rudder Baker and Ruth Chang, my commentators on the first and second occasions, and to Selim Berker, Jamie Dreier, Dan Greco, and Rae Langton.

be doing. (Though of course, they have the resources, frequently invoked, to make rapid apologies for their failures.)

The Luddeys form their intentions ahead of time, and they benefit from it. This is the idea that lies behind Michael Bratman's hugely influential conception of intention.<sup>2</sup> Intentions stand as fixed points in our reasoning. There are very many things that, consistent with our wants, beliefs and values, we could do. We could keep all of these, or at least, many of them, as open possibilities. But if we foreclose our options we arrive at a simpler picture which enables us to achieve things that would otherwise be beyond us. Perhaps there are creatures who can coordinate any number of open possibilities. But for limited creatures like us, reducing the options to a number of fixed points is essential. It enables us to coordinate with our (equally limited) peers, and indeed to coordinate with ourselves: to keep track of what we need to do to implement our own plans. It also provides a way to keep the time spent on deliberation in check. Once we have deliberated long enough we form our intentions and then can turn our attention elsewhere. We need not go to the lengths of the Luddeys, but some balance between stability and flexibility has to be struck.

The idea I explore here is that much the same considerations govern belief. Just as we can imagine creatures who do not foreclose their practical possibilities, so we can imagine creatures who do not foreclose their epistemic possibilities. By this I mean that they do not resolve their uncertainties into all-out belief. In other words, they behave as many theorists hold that we do in fact behave: they form *credences*, attitudes that can stand anywhere between full belief and full disbelief. Such attitudes can then be represented, in a familiar way, as falling on a numerical scale, where belief of degree one represents full belief, belief of degree zero represents full disbelief, and the intervening degrees of belief are represented by the intervening real numbers, in proportion to their distance from the ends.

Let us be clear on what the credence picture involves. It doesn't simply involve the idea that we can entertain the thought that a certain outcome has a certain chance: that there is a 0.5 chance that a fair coin will come down heads. For that thought can be the content of an all-out belief. If that were all that was involved, we could simply add a few all-out beliefs with a probabilistic content to our stock of beliefs, and there can be no doubt that that would be a very useful thing to do. The credence picture is very different. It is that our beliefs are *essentially* probabilistic. This is because probability is not in the content of a few beliefs, but in the attitude of belief itself. It is not that we believe that something is, say, 0.5, or 0.7 probable. It is that we have a 0.5 or 0.7 credence in it happening. Every belief is thus a probabilistic belief (even if the credence happens to be 1 or 0).

A creature that had credences would benefit from the ability to keep many options open at once: such a creature would be able, for instance, to keep in play four mutually incompatible outcomes, regarding each with a credence of 0.25. But such abilities come at a price, parallel to that confronting the Gadgers. Unless their powers of memory and reasoning are very great, those who employ credences risk being overwhelmed by the huge mass of uncertainty that the approach generates. First they will have to store very

---

<sup>2</sup> Michael Bratman, *Intention, Plans and Practical Reasoning* (Cambridge: Harvard University Press, 1987).

much more information: rather than just discarding the propositions that aren't believed and focussing on those that are, they will have to keep track of all of them and their associated credences. Second, they will have to be able to deploy the complicated methods needed for probabilistic reasoning. The problem will be all the worse if, in Bayesian fashion, they update their credences by conditionalization. For then, in addition to keeping track of the probabilities of the various events, they would also need to keep track of the conditional probabilities of events upon each other: the probability of  $p$  given  $q$ .<sup>3</sup>

No one thinks that human beings can behave like perfect Bayesian agents. Typically, told that something is merely likely, we go on to use this information—i.e. to reason on the basis of it—in much the same way as we would if we had been told that it was certain. Told that a character in a story is an animal she glimpsed is probably a dog, subjects' assessment of the probability that it chews slippers is not significantly different from that of subjects who are told outright that it is a dog. And this is so even though subjects in the first condition remember that the information they got was only probable. It is not that we forget that things are less than sure; it is rather that we disregard the lack of certainty when we come to reason with the information.<sup>4</sup>

So if the Bayesian picture has a role, it is as an idealization. If it is to do that though, we had better *approximate* Bayesian agents. It had better be that we can form credences, and that we can conditionally update on them, even if we do not always do both perfectly. My contention in the first part of this paper is that even this minimal claim is false. I argue that we cannot form credences at all. The Bayesian approach is not an idealization of something we actually do. Instead, it is quite foreign to us. Just as our core native deliberative state is that of the simple intention, so our core native epistemic state is that of simple, all-out belief.<sup>5</sup> This is what makes it possible for creatures with the limited cognitive powers that we have to deliberate and act as effectively as we do.

There is a second way in which belief resembles intention. It is part of the nature of intention to be stable. If intentions are easily given up, they will not be able to play their coordinating role. If the Luddeys readily abandoned their intentions, their policy would be little different to that of the Gadgers (and they'd lack the technologies to implement it). Of course this does not mean that they will never revise an intention; that would be idiocy. Getting the right balance between stability and flexibility is a delicate business, one that is sensitive to many factors. Elsewhere I have argued that stability should come in at the point of resisting reconsideration, and that it therefore need involve no

---

<sup>3</sup> Gilbert Harman *Change in View* (Cambridge MA: MIT Press, 1986) p. 23.

<sup>4</sup> Barbara Malt, Brian Ross, & Gregory Murphy, 'Predicting features for members of natural categories when categorization is uncertain', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, (1995) pp. 646–61. See also Gregory Murphy & Brian Ross, 'Predictions from Uncertain Categorizations' *Cognitive Psychology* 27 (1994) pp. 148–93. An interesting issue, one that is addressed but not resolved in these papers, is to determine the rare circumstances under which probabilistic information is employed in reasoning.

<sup>5</sup> Actually, following Williamson, I suspect that our core epistemic state is knowledge, and that belief is best understood in terms of it: a belief is either a piece of knowledge or a failed attempt at knowledge (Timothy Williamson, *Knowledge and Its Limits* (Oxford: Clarendon Press, 2000)). But since knowledge is itself an all-out state, this only adds to the force of what is said here. I shan't make my discussion more contentious than it already is by phrasing it in terms of knowledge.

irrationality. Stability requires that the threshold at which new evidence will prompt the reconsideration of an intention is higher than the threshold it would have had to pass for it to be considered in the initial formation. Just how much higher will depend on many things.<sup>6</sup>

Again I want to say the same about belief. Once we think of an all-out belief as providing a stable coordination point, then it too should be resistant to reconsideration. If it were to be readily given up there would be little benefit from forming it in the first place. This is my concern in the second part of this paper.

My central argument then is that there are practical reasons for having stable all-out beliefs. Many other authors have suggested that all-out belief is intimately tied to practical considerations. For instance, in distinguishing belief from credence, Timothy Williamson has suggested that one all-out believes a proposition just in case one is willing to use it as a premise in practical deliberation.<sup>7</sup> Similarly Bratman writes:

To believe something is not merely to assign a high probability to its occurrence. I might assign a high probability to my failing to move the log without believing that I will fail ... what seems distinctive about believing that I will fail is that it puts me in a position to plan on the assumption of failure.<sup>8</sup>

In much the same mode, Keith Frankish thinks of all-out belief as involving a *premissing policy*: to all-out believe a proposition is to commit oneself to using it as a premise in one's reasoning.<sup>9</sup>

However, we can push that point too far. Whilst I agree that all-out belief is underpinned by practical considerations, I'm not suggesting that practical considerations come into the mind of the agent when they come to all-out believe a proposition. Still less I am suggesting that belief formation is voluntary. The point is rather that practical considerations have caused us, presumably by some evolutionary mechanism, to have all-out beliefs. That is why we can all-out believe propositions that can have no possible implications for practical reasoning.

---

<sup>6</sup> See my *Willing, Wanting, Waiting* (Oxford: Clarendon Press, 2009).

<sup>7</sup> Williamson, *op. cit.*, p. 99.

<sup>8</sup> Bratman, *Intention, Plans and Practical Reasoning* (Cambridge MA: Harvard University Press, 1987). Elsewhere Bratman distinguishes belief from acceptance. See 'Practical Reasoning and Acceptance in a Context' in *Faces of Intention* (Cambridge: Cambridge University Press 1999) pp. 15–34. I doubt that there is really such a state as the latter, apart from the highly sophisticated one of supposing something true for the sake of an argument (as one does, for instance, in a *reductio*). But that is a rather different idea.

<sup>9</sup> 'Partial Belief and Flat-Out Belief' in F. Huber and C. Schmidt-Petri (eds.), *Degrees of Belief* (Dordrecht: Springer, 2009) pp. 75–93. Frankish gives a very useful summary of other proposals in the area.

## I. BELIEFS AS ALL-OUT STATES

### DIFFERENT KINDS OF PROBABILISTIC BELIEF

In saying that all-out belief is the core notion, I do not mean that it will characterize all of our epistemic states. First, for beliefs as with intentions, there are cases in which one hasn't made up one's mind. So we need to distinguish the state of lacking the belief that *p* from the state of having the belief that not *p*. Second there surely are cases in which we believe that there is a certain chance that something is true. But here I will argue that the probability is in the content not in the attitude. Third, and distinct from this, I will argue there are cases in which, whilst we know that we lack knowledge, we do have some opinions on how things stand.

Let me introduce these ideas by pointing to some linguistic phenomena. Although the credence approach sees probability as residing in the attitude, our standard talk of probabilities that can take anything like a numerical value puts them instead in the content. Thus we say:

- I believe that there is a 50% chance that *p*.  
(Not: I believe to the 50% level that *p*.)
- I believe that *p* is twice as likely as *q*.  
(Not: I believe *p* twice as much as I believe *q*.)
- I believe that there is a one in six chance that *p*.  
(Not: I have a one in six belief that *p*.)

So far as I can see our ordinary talk of numerical probabilities is always of this form.<sup>10</sup> This suggests that we think of such beliefs as consisting of an all-out belief in a certain probabilistic proposition (where that proposition is itself formed from one or more propositions and a probabilistic operator: 'there is a 50% chance that \_\_\_').

Many other qualitative probabilistic judgements follow the same form:

- I believe that it is likely that *p*.
- I believe that there is a good chance that *p*.

---

<sup>10</sup> For instance I can find no instances of numerically modified belief ascriptions in the on-line *Corpus of Contemporary American English* (<http://corpus.byu.edu/coca/>).

I believe that it is not very probable that p.<sup>11</sup>

Here again it appears that it is the content that is being modified. But there are other constructions where we do appear to modify the attitude from that of simple belief to something stronger or weaker:

I am sure that p.  
I am certain that p.  
I am confident that p.

I doubt that p.  
I am sceptical that p.

I am more confident of p than of q.

I suspect that p, but I'm not sure.

How should we distinguish the two classes, those in which the modification goes into the content and those in which it goes into the attitude? An obvious first thought is that in the group in which the content is modified, the probability is in some way attributed to the world; whereas when the attitude is modified, some feature is self-attributed by the agent.

In much modern thinking about probability, these two ideas tend to get run together—or one gets subsumed to the other. Nevertheless, they are distinct. The point comes out very clearly if we look at the history of thought about probability. Ian Hacking makes a very credible case that our current ideas of probability only emerged in the late 17th century: it was only then that anything like a theory of probability was developed, even though dice, and before them animal knucklebones (the *talus*), have been in use for games of chance at least back to Ancient Egypt. Hacking's contention is

---

<sup>11</sup> A search on 'high \* of belief' (where '\*' is a wildcard) in the *Corpus of Contemporary American English* provides just one example of a qualitative modification of a belief state, which occurs in the abstract of a medical journal article written by two Swedes:

One-way analysis of variance, with eight possible groupings of high and/or low scores on the three subscales, indicated that those with a combination of strong belief in internal control of health, high degree of belief that health is determined by powerful others, and low degree of belief that health is a function of chance/luck had better self-esteem than those with a quite opposite pattern, involving high belief in chance and low belief in internal control.

I conclude that it is not standard English usage. Of course, such uses occur frequently enough in the writings of philosophers (see the first 50 entries one gets on a Google search); but we are a corrupt group. Talk of 'strongly believing' is much more common, but I suspect that this is just another way of saying that one is confident, an issue to be discussed later. As evidence that it does not really represent a point on a scale of belief, two considerations: (i) apart from one famous quotation from Thomas Jefferson, there is very little use of the idea of 'weakly believing'; (ii) none of the first 50 Google citations for 'believe \* more strongly than \*' brings a use in which the wildcards are both replaced by propositions; instead one gets usages like "women believe more strongly than men that..."

that when a theory of probability did emerge, it was ‘Janus-faced,’ bringing together ideas of aleatory (i.e. truly chancy) cases with those that involve uncertainty.

On the one side it is statistical, concerning itself with stochastic laws of chance processes. On the other side it is epistemological, dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background ... Pascal himself is representative. His famous correspondence with Fermat discusses the division problem, a question about dividing stake in a game of chance that has been interrupted. The problem is entirely aleatory in nature. His decision-theoretic argument for belief in the existence of God is not. It is no matter of chance whether or not God exists, but it is still a question of reasonable belief and action which the new probable reasoning can be applied.<sup>12</sup>

Of course, once a theory of probability that tends to run these ideas together was in place, people used it, and so their beliefs and actions need to be understood in the light of their grasp of that theory. But the underlying ideas are distinct.

So I want to defend the idea that there are two very different mental phenomena that need to be explained. First there are thoughts about aleatory probabilities widely construed. In their numerical form, these are the kinds of thoughts that we have only been able to have in any articulated form since the development of the theory of probability; and entertaining them involves grasping, perhaps in a very rudimentary form, something of that theory. These thoughts involve a range of cases—for example, thoughts about the chance of rolling two sixes with a pair of dice, about the chance of recovering from a disease, given a knowledge of others’ recovery rates, and about the chance of it raining tomorrow, given a certain evidential basis. This is a diverse collection. Philosophers, keen to theorize, will be tempted to insist that such beliefs must involve a commitment to objective probabilities, or else that they should be reconstrued as credences. But that is to move too fast. The notion of an objective probability, and the contrast with subjective probability, frequency and so on, are sophisticated ideas that have been refined over many years. Our pre-theoretic notion of something having a certain probability involves no such commitment.

Second, there are thoughts that we have when we realize that we lack knowledge, when we say things like ‘I think she’ll come, but I’m not sure.’ I take it that people were saying things like this long before the development of probability theory (Shakespeare is full of such states of doubt: consider Othello’s rising doubts about Desdemona). Thoughts like this might be accompanied by out-right probabilistic judgments, but they surely need not be.

Defenders of the credence approach will doubtless object that it is exactly these second states—the states of doubt and of certainty—that they have in mind when they posit credences: that by accepting them I have accepted credences. And, further, they will insist that they don’t deny that there are also beliefs about probabilities; indeed that much of their work has gone into determining the proper relation that such beliefs bear to credences. So am I really disagreeing with them?

---

<sup>12</sup> Ian Hacking, *The Emergence of Probability* (Cambridge: Cambridge University Press, 1984) p. 12.

A full response to this worry will take some time to spell out, but let me say a few things now. In his seminal discussion of the Principal Principle, linking credences to beliefs about probabilities, David Lewis writes:

A certain coin is scheduled to be tossed at noon today. You are sure that this chosen coin is fair: it has a 50% chance of falling heads and a 50% chance of falling tails. You have no other relevant information. Consider the proposition that the coin tossed at noon today falls heads. To what degree would you now believe that proposition?

Answer: 50% of course.<sup>13</sup>

That answer is indeed obvious *if* we can form beliefs that come in degrees. But it is far from obvious that we can do that: that in addition to having the belief that there is a 0.5 chance that the coin will fall heads, I can also have a 0.5 belief that it will fall heads. Why should we think that there are two such distinct states?

The need for the two states would be made out if the credence were understood just as a tendency to behave: a tendency, for instance, to accept bets at the relevantly favourable odds. But if the history of behaviourism and functionalism has taught us anything, it has taught us that that approach will not work.<sup>14</sup> Credences should be understood as the states that cause, or in some other way give rise to, such behaviour.

If we are to accept the existence of credences then, we need to think of them as a system of representational states that stand parallel to our beliefs about probability. They need to encode probabilities on a scale; they need to be manipulable in response to reasoning and to new information; and they need to have a tie to behaviour. Of course these last two features will not be perfectly realized: people will be more or less flawed in their abilities to reason, to update, and to behave rationally given their credences. But if there were no tendency whatsoever to do these things, then we would lack reason to think that the states represent probabilities at all.

Theorists who accept credences accept a picture along these lines. Probabilities are represented by the degree of belief; they to obey the rules of the probability calculus; they are updated by conditionalization; they are linked to beliefs about probabilities by the Principal Principle; and they give rise to behaviour such as the taking of bets at certain odds. Indeed, since proponents of credences are often sceptical of the role for beliefs about probabilities, credences are often seem as the primary way in which we deal with probabilities.

My argument here is that this gets things exactly the wrong way round. Insofar as we do register and manipulate probabilities, this is a difficult business done at the level of content. Much of the time these beliefs do not have any impact on our behaviour, which is done of the basis of simple all-out beliefs, with no probabilistic content.

---

<sup>13</sup> 'A Subjectivist's Guide to Objective Chance,' reprinted in his *Philosophical Papers*, Vol. II (New York: Oxford University Press, 1986) p.84

<sup>14</sup> See Lina Eriksson and Alan Hajek, 'What are Degrees of Belief?' *Studia Logica* 86 (2007) pp. 183-213, for a nice discussion of this. They go on to conclude that credences should be taken as ontologically primitive; but only since they think that beliefs about probabilities are too theoretically sophisticated to do the job. I will dispute this shortly.



When, sometimes, it becomes quite clear to us that we do not know how things will work out, we do behave accordingly: we make contingency plans. Not knowing whether you will meet me or not, I make plans both ways. But the mental states involved here are nothing like credences. They do not typically register anything like numerical degree of probability; we cannot manipulate them; they do not obey anything like the Principal Principle. You may call them 'credences' if you insist; I think that they do not deserve the name, and that we would do better saying that there are no such things as credences. When the time comes I will call them 'partial beliefs'.

One further clarification. I am talking here about explicit beliefs: the kinds of broadly propositional things that we can normally bring to consciousness and to express in words, and that are available for reasoning across different domains (which is not to say that they have to have any of those properties; science will tell us quite how the kind should be picked out). Plenty of recent research has shown that such beliefs do not constitute the whole of our mental lives; far from it. Indeed, there is evidence that our explicit beliefs are underpinned by many relatively modular unconscious systems that register aspects of the world and deliver the beliefs as outputs. Clearly that is so for perceptual beliefs, and the line between perception and inference is far from clear. So, for instance, the belief that a causal relation exists may well be based on the output of such a modular system. Such systems look to make use of probabilistic information. So, again, a causal relation may be identified on the basis of frequency data that are analyzed using something along Bayesian lines. Should such systems be thought of as deploying credences or as involving beliefs about probability? I doubt that either answer fits. Unless we have explicit beliefs of broadly propositional form I do not know how to understand the question. So here I confine the discussion to the explicit.<sup>15</sup>

#### CONTENT OR ATTITUDE?

So let me start with the issue of whether our explicitly probabilistic thoughts involve probabilistic attitudes, as the Bayesian theory suggests, or probabilistic contents, as I want to contend. It looks as though it should be an empirical issue. But our mental states do not come ready labelled. The issue is best seen as one of interpretation. Do we come up with a better picture of human probabilistic beliefs if we interpret them on the attitude model or the content model?

In resolving this, we might look to our prior convictions; or we might look to our behaviour, widely construed. I doubt that the former will be of much help. Do we have any pre-theoretic conviction about the matter? It seems unlikely. So when it comes to evaluating whether or not we have a certain attitude, behaviour is a more reliable guide than introspective conviction. Which behaviours should we look to? Many of the behaviours that are typically invoked in these discussions will not help us. Should we

---

<sup>15</sup> For the thought that the existence of such systems should not lead us to attribute probabilistic abilities to the agent, any more than the fact that some aspects of our visual systems are well modelled using Fourier series shows that we should attribute competence with them to the agent (or, indeed, to rodents); see Nick Chater, Joshua Tenenbaum and Alan Yuille 'Probabilistic Models of Cognition' *Trends in Cognitive Sciences* 10 (2006) pp. 287–91, at p. 288.

understand betting behaviour as stemming from our credences or our beliefs about probabilities? No obvious answer suggests itself, since both look able to do the job. We need to look elsewhere.

I have already mentioned our linguistic behaviour, which I think supports the idea that explicit probabilities are in the content rather than the attitude. But that is rather weak evidence; we cannot assume that our language accurately reflects how things are. I have also mentioned the history of probability beliefs: the idea that anything like numerical probabilities, together with any grasp on how to manipulate them, only arrived in the late-17th century. That too, I think, provides some evidence that explicit probabilities are in the contents. It is possible that people had been entertaining and manipulating credences for millennia before that without realizing what they were doing, but it strikes me as implausible. Equally, it is possible that people started having a distinctive new set of attitudes as recently as that; but it seems altogether more likely that they simply came to have beliefs with new contents. However, the history is far from established, so we should not put much weight on it either.

Instead I want to focus on our competence with probabilities. If we are to be good subjects for the ascription of probabilistic attitudes, then we should be able to make the kinds of transitions—whether in beliefs or in behaviour—that would be expected. Yet a host of now very familiar research shows that, in many cases, we are very bad at this. It might seem that this should count equally against an attitude account and a content account. But I think this is not so. Instead I will argue that the findings about our incompetences with probabilities suggest that these involve contents not attitudes.

My focus will be on the much discussed base rate fallacy. Since I shall be using this example to do quite a lot of work, let me spell it out in some detail. Consider the following question:

The probability that a woman has breast cancer is 1%. If she has breast cancer, the probability that a mammogram will show a positive result is 80%. If a woman does not have breast cancer, the probability of a positive result is 10%. What is the probability that a woman who has a positive mammogram result has breast cancer?

For most people, that is a very difficult question. Representing the conditional probability of a given b by 'P(a/b)', the answer, that is, the value of P(breast cancer/positive result), can be calculated using Bayes' Rule

$$\frac{P(\text{breast cancer}) \times P(\text{positive result}/\text{breast cancer})}{[P(\text{breast cancer}) \times P(\text{positive result}/\text{breast cancer})] + [P(\text{no breast cancer}) \times P(\text{positive result}/\text{no breast cancer})]}$$

which in this case is

$$\frac{0.01 \times 0.8}{[0.01 \times 0.8] + [0.99 \times 0.1]}$$

which, to two decimal places, is equal to 0.075. In other words, the chance that a woman who receives a positive result actually has breast cancer is somewhat less than 10%.

I doubt that there are many people who would arrive at that result without formal training; indeed, even formal training seems not to have much lasting impact. Worryingly, even those who we might expect to be able to do the calculation cannot. In a survey of experienced physicians who were presented with the data in this form, fewer than one in ten got the answer correct.<sup>16</sup> Two thirds estimated that the chance that a woman who received a negative result actually had breast cancer was 50% or more; and a full third thought that it was 80%. To arrive at this latter figure, the physicians had presumably ignored the base rate—the level of breast cancer in the population—altogether. The fact that breast cancer is rare in absolute terms greatly reduces the significance of the high reliability of the mammogram.

The fact that people are insensitive to base rates is now widely recognized. But the insensitivity is itself highly sensitive to how the data is presented. Suppose that, instead of giving the data in terms of conditional probabilities, it were given it in terms of natural frequencies, as follows:

10 out of every 1000 women have breast cancer. Of these 10 women with breast cancer, 8 will have a positive mammogram result. Of the remaining 990 women who do not have breast cancer, 99 will have a positive mammogram result. What is the probability that a woman who has a positive mammogram result has breast cancer?

Answering that question still requires some thought, but now the question is much more tractable. Of the 1000 women, how many will have a positive result? That is easily read off the data: it is 8 + 99, i.e. 107. Of these, 8 will have cancer. So the chance of having breast cancer if one has a positive result is 8/107, which is clearly somewhat less than 10%. When presented with the data in this form, around half of physicians got the answer exactly right, and over three quarters realized that the chance was less than 15%. Only one in twelve made the mistake of ignoring the base rate altogether.

In general people are far better at calculating probabilities if they are given natural frequencies than if they are given conditional probabilities. Indeed around half of twelve year olds are able to solve problems such as this when presented with natural frequencies, whereas virtually none can solve them when presented with conditional probabilities.<sup>17</sup>

What should we conclude from this for the question of whether the probabilistic elements in beliefs such as these are in the contents or the attitudes? I think it presents very good evidence that they are in the contents, for two reasons.

First, if the probabilities were really in the attitudes, then to do the calculation the subjects would need to have credences of the correct degree before they could apply Bayes' rule. But the relevant degrees of credence are exactly what they are given when

---

<sup>16</sup> Gigerenzer *Rationality for Mortals* (New York: Oxford University Press, 2008) ch. 9.

<sup>17</sup> Gigerenzer, *op. cit.* ch 12.

the problems are presented as conditional probabilities. In contrast, in the natural frequency presentation, the conditional probabilities are merely implicit in the data. The subjects would first have to calculate the relevant conditional probabilities and then go on to apply Bayes' rule. So you would expect the natural frequency presentation, requiring a further step, to be more difficult than the conditional probability presentation.

If, instead, the probabilities were represented in the content, one would expect exactly the results that we get. Since the presentation in terms of natural frequencies does not strip out the base rate data, the subjects are presented with a much easier calculation to do: instead of the complicated application of Bayes' rule, they have to calculate something of the form  $a/a+b$ . The subjects solved the problem just as I explained how to solve it, as an explicit bit of arithmetic that is the content of their belief.

The second reason for thinking that these findings support the content approach is rather more abstract. Suppose that someone were presented with both descriptions without being told that they were descriptions of the same situation. Mightn't they come to think that a positive mammogram in one case was far more significant than in the other?<sup>18</sup> It is easy to see how to explain this on the content account; or at least it presents an instance of the familiar problem that people can have different attitudes to logically equivalent propositions presented in different ways. That is a problem we will have to solve in some way, and there are familiar suggestions for how to do it. But how would we explain this on the attitude account? If the probability is built into the attitude, it looks as though we are going to have to say that there are different ways of having the same attitude; that is, for instance, different ways of having a confidence level of 0.3. Yet that is an idea that is not suggested by our normal account of attitudes, and I have no idea how it might be implemented.

I think then that this provides good reason for thinking that in cases like these, we reason with probabilities as contents, and not as attitudes. My conjecture, which admittedly could do with some more support, is that this is true for all of our explicit probabilistic reasoning. Certainly I know of no cases that go the other way.<sup>19</sup> Explicit probabilistic reasoning is a difficult business, something that we learn to do, much as we learn to do arithmetic. We do it best when we make it as simple as possible.

Are there any considerations against the content view? David Christensen argues that it would commit ordinary thinkers to a belief in objective probability, and that is to commit them to something too sophisticated.<sup>20</sup> But that doesn't sound right. It is us, the theorists, who might interpret the belief as requiring objective probabilities, understood in a sophisticated way. The ordinary thinker will have likely given the

---

<sup>18</sup> We don't need to rest with the hypothetical case here. Kahneman and Tversky have plenty of framing effects of just this form: the very same subjects take different attitudes to the same possibility presented in two different ways. See *Choices, Values, and Frames* (Cambridge: Cambridge University Press, 2000) pp. 4ff.

<sup>19</sup> Recall that I am accepting that we may well have sub-personal modules that make use of probabilistic reasoning. But here I doubt that the question of whether the probability is in the attitude or the content makes much sense.

<sup>20</sup> David Christensen *Putting Logic in its Place* (Oxford: Clarendon Press, 2004) pp. 18–20

matter no thought to the question of how to understand their intuitive talk of probabilities. Christensen might insist that their belief can only be legitimated if we can give it a plausible content. But there are many ways that we might try to do that, of which objective chance is only one. We might, for instance, say that the chances correspond to the credences that a Bayesian agent would have, even if we think that such a creature could not be us.

In a slightly different vein Keith Frankish objects that ascribing probabilities as contents involves over intellectualizing: “We speak of children and animals having more or less confidence in something and of preferring one thing to another, even though they do not possess the concepts of probability and desirability”<sup>21</sup> Frankish here speaks of degrees of desirability, an interesting topic, but not obviously one that should be treated in the same way as degree of belief (I’ll say a little about it shortly). But on degree of belief he is surely right: we do what to think of children and animals as having something like partial belief, without thinking that they have any real understanding of probability. I do not take this to show that we cannot treat explicitly probabilistic beliefs as having their probability in the content. What it shows, as I mentioned earlier, is that we need a separate treatment of the states that we can be in when we realize that we do not have knowledge . It is to this that I now turn.

#### PARTIAL BELIEFS

In an influential discussion, one that has a lot in common with the view developed so far, Gilbert Harman writes:

How should we account for the varying strengths of explicit beliefs? I am inclined to suppose that these varying strengths are implicit in a system of beliefs one accepts in a yes/no fashion. My guess is that they are to be explained as a kind of epiphenomenon resulting from the operation of rules of revision. For example, it may be that *P* is believed more strongly than *Q* if it would be harder to stop believing *P* than to stop believing *Q*, perhaps because it would require more of a revision of one’s view to stop believing *P* than to stop believing *Q*.<sup>22</sup>

I think that something along these lines might be right for desire. To say that one desires one thing more than another is to say, roughly, that one would choose that thing over the other if one had the choice of the two of them. (There are doubtless worries with any conditional analysis along these line, but I shan’t pursue them here.) This has the consequence that one may be have false beliefs about which thing it is that one desires more, but that seems to be to the good. Could one say much the same about belief? Could one say that one believes one proposition more strongly than a second if one would hang on to it, rather than the second, in circumstances in which one could not believe both? Doubtless there are again problems raised here by the conditional analysis (which circumstances?), but again I shall not pursue them, since it seems to me

---

<sup>21</sup> Frankish, *op. cit.*

<sup>22</sup> Harman *op cit*, p. 22

that there is a larger problem. Whilst this may give one criterion for strength of belief, one that is compatible with an all-out conception, it does not explain all that needs to be explained.

The point should be familiar. Suppose I believe something: for instance, that Brutus is an honorable man. I hear a few rumours to the contrary perhaps, but I discount them. Slowly though they mount up, and some other bits of evidence come in, not compelling, but enough to make me think. I do not altogether lose my belief that he is an honourable man, but my confidence is lessened. Now it may well be true that at this point I am closer to giving up my belief than I was. However, that is not all. My actual actions start to change. I certainly do not denounce Brutus; I probably do not even feel compelled to confront him. But I might stop confiding in him so freely, stop trusting him with the sensitive tasks.

Harman's proposal says nothing about this. My changed actions do not result just from a realization that I am more likely to change my belief. Even if I am not going to give up my belief in Brutus's honour altogether, and I know this—perhaps because I know I will never get more information—still I will treat him differently. So it is not the proximity to change of all-out belief that moves me, nor an apprehension of this; it is an actual change. How can we accommodate this?

We have seen why it makes sense to deliberate with all-out beliefs wherever possible. These examples show that we cannot always do so. Sometimes it is clear to us that we do not know what did or will happen. In such situations we have to make use of a more partial notion, and this involves a change in our attitude, and not in the content. Rather than working just with the attitude that Brutus is honorable, I have to work with two possibilities: that he may be, and that he may not.

Again there is a parallel here with intentions. Normally things are clear enough for us to make do with a simple plan. But sometimes we need contingency plans, that will depend on how things turn out. And again this is an old idea, long predating probability theory.<sup>23</sup> Keeping such complex plans in mind is costly, and the complexity quickly ramifies; so we only do so when the need is really pressing.<sup>24</sup>

But this does not take us to credences, for at least two reasons. First, this need not involve the assignment of anything like numerical probabilities. Indeed, it need not even involve an ordering. Suppose that I come to think that Brutus may be honourable, and that he may not; and that my cautious behaviour towards him reflects that. Must I thereby have some view of their relative probability? I think not. Both are simply worth taking seriously. Of course some will say that there is behaviour that will reveal it: betting behaviour. That however is a very particular type of behaviour, something to which I shall return shortly.

Second, suppose we have an agent who does understand the idea of probabilities, and who thus assigns probabilities (understood as contents) to a range of possibilities.

---

<sup>23</sup> Classical military history provides plenty of cases. See, for instance Caesar's *Commentaries on the Gallic War* 3; 3–6, for a plan to defend a city if possible, but to break out if this seems to be failing. Thanks to Anna Marmodoro for this and many other examples.

<sup>24</sup> Such plans can be understood either as conditional intentions, or as involving parallel 'partial intentions' that more closely mirror partial beliefs. I argue for the second interpretation in *Willing, Wanting, Waiting*, ch. 2, but do not assume it here.

What is the relation between the agent's assignments of probabilities and their probabilistic attitudes? If the attitudes really were credences, they would surely be governed by the Principal Principle: the attitudes would correspond to the probability assignments. In particular then, whenever an agent assigns a probability of greater than zero to an outcome, that should be reflected in their attitude.

In contrast, the kind of partial attitude that I am suggesting we have does not obey anything like the Principal Principle. Even when we do assign non-zero possibility to a range of propositions, we need not accept each of them as what we might call a *live possibility*. I assign a non-zero probability to the idea that MIT will be destroyed tomorrow by a rogue meteor, but I do not bring such a possibility into my practical reasoning in any way. I do not reason in the light of it, or make plans around it. And there are many things that are much more likely than that which I equally do not treat as live possibilities.

How high does one have to think that the possibility of something is before one treats it as a live possibility? We know for semi-formal reasons that there can be no simple threshold—if you do not treat  $p$  as a live possibility, nor treat  $q$  as a live possibility, then you will presumably not treat  $p$  or  $q$  as a live possibility; but if  $p$  and  $q$  were both close to the threshold, then  $p$  or  $q$  will likely be above it. More important than that though, we should expect that the factors that determine which possibilities will be live cannot be computed in a simple mechanical manner. For if belief is like intention, there will be all kinds of factors that govern its proper formation. Knowing when to conclude practical deliberation and form an intention, and when to open the deliberation up again, is an open-ended practical skill, one that is highly sensitive to the environment in which one finds oneself.<sup>25</sup> The same is true of belief formation.<sup>26</sup>

There will be times when the human limitations that constrain this practical skill are less pressing. At such times we may be able to entertain more possibilities as live. At the limit we might perhaps conform to something like the Principal Principle: we might treat every proposition to which we assign a non-zero probability as live. I think that this is what happens with the kind of idealized betting behaviour that is often—and, to my mind, misleadingly—seen as central to evaluating our attitudes. When you are asked how much you would be prepared to bet on a certain outcome, this does not require much from you. It does not require you to formulate complex plans contingent on that outcome, since it doesn't require further action on your part. The bet is purely imaginary, and what you are asked to imagine is itself perfectly undemanding—that if you win the winnings will be delivered to you. Things would be very different if this were real behavior—if, for instance, you were required to ascertain whether you had won, and then to claim the winnings yourself. For the extra cognitive load that this would impose would make it rational to avoid getting involved in betting behaviour where you estimate the chances to be small, or the gains to be slim. So we can concede to the orthodox approach that in certain imaginary circumstances we can form attitudes that are rather like credences, and that betting behaviour provides the

---

<sup>25</sup> I discuss this further in *Willing, Wanting, Waiting*, ch. 4, where I argue that weakness of will is best understood as the over-ready revision of an intention (or, more precisely, of a resolution).

<sup>26</sup> For a start at an empirical answer to the question of when we treat more than one option as live, see the papers mentioned in n.4 above.

touchstone for them, whilst insisting that this is really a very particular phenomenon, one that provides a bad model for our cognitive attitudes more generally.

I conclude then that there is a practical stance that one can take towards a proposition that is rather like all-out belief, but is partial. I will call it *partial belief*. Then we can summarize the discussion in a pair of definitions, taking the notion of a live possibility as primitive:

*All-out Belief*

One all-out believes  $p$  iff one takes  $p$  as a live possibility and does not take not- $p$  as a live possibility.

*Partial Belief*

One partially believes  $p$  iff one takes  $p$  as a live possibility and takes not- $p$  as a live possibility.<sup>27</sup>

Talk of live possibilities isn't very far from talk of belief; but it is far enough to give us some purchase.

## 2. THE STABILITY OF BELIEF

Now I turn to the second way in which I suggested that beliefs parallel intentions, namely their stability. I argued that intentions have stability for two reasons (I'll add a third shortly). First, stability provide a way to curtail deliberation. We deliberate, and then decide, and then move on. That way we can marshal our scarce cognitive resources, allocating reasoning time between the different deliberative demands upon us. Second, they allow for simpler coordination, both intra- and inter-personally. Once intentions are in place we can form other plans around them. The effective intender thus trades off flexibility for stability. By its very nature this trade off has to be something that is embedded in dispositions that cannot be considered afresh each time: I do not avoid deliberation if I have to deliberate, given some new evidence that bears on my intentions, whether to deliberate. Once I have formed an intention I simply have to be closed to new considerations, unless they pass some threshold that means that they cannot be properly ignored. If, after a lengthy perusal of the reviews, I decide to eat at a certain restaurant, then later coming across one further review shouldn't cause me to open the whole question up again; still less if I have coordinated to meet you there. But this is not to say that no information about the restaurant—a fire in the kitchen? a outbreak of salmonella?—should lead to reconsideration. A reasonable firmness of intention is distinct from pigheadedness.

Exactly the same considerations, I suggest, govern belief. We could go on indefinitely looking for evidence for a proposition, but we need to know when to curtail our search, and invest our conclusion with a reasonable degree of stability, so that we

---

<sup>27</sup> Here again I follow my earlier discussion in *Willing, Wanting, Waiting*, ch. 2.



can go on to reason and plan on the basis of it. Again here we need to strike a balance between sensitivity to the world and stability. And again it is part of the nature of that balance that it has to be embedded in brute dispositions. One cannot assess the new evidence each time to determine whether it should prompt a reconsideration, for that would already be to embark on a reconsideration. One has to simply ignore new evidence of a certain kind.

Stability is, at least in theory, quite independent of the idea that beliefs are all-out. One could imagine an agent that formed credences but whose credences were entrenched. Equally one could imagine an agent that formed all-out beliefs, but was ready to change them at the drop of a hat. But since both all-out beliefs and stability are underpinned by the same considerations of cognitive economy, it is not surprising that they typically go together. Indeed it seems very likely that, given the cost of maintaining them, partial beliefs will be less stable than all-out beliefs.

The need for stable beliefs will have many consequences for our picture of the ideal epistemic agent (or at least, of the ideal agent who has the kind of cognitive limitations that we have). Here I want to focus on just one topic, but one that has absorbed a deal of recent philosophical debate: How should we respond to epistemic disagreement? In particular, how should we revise our beliefs when we discover that an epistemic peer has different beliefs to ours? Much of the debate has assumed the credence picture that I have rejected: if my credence in a proposition is .3, and my peer's is .7, should I move mine to .5? In rejecting the credence picture I reject that way of putting the question, though of course it still arises in the (rather unusual) cases when peers have diverging beliefs with probabilities in the content. What I say will have some relevance for that, but my main concern is with what I take to be a rather more commonplace issue.

Let us start, once more, with intention. On the picture of intentions that I have given, there are two questions that confront an agent: whether to reconsider an intention; and whether to revise it once a reconsideration has taken place. It may well happen that it will not be rational to reconsider an intention even though it would be rational to revise it once it were reconsidered. So, for instance, it may be rational for us to persist with our intention to eat at the restaurant that we chose last week, even though, were we to reconsider in the light of this week's reviews, we would be rational to change our minds; the cost of the the reconsideration just isn't worth the likely benefit.<sup>28</sup> This is an instance of the familiar explore/exploit question: when is it rational to go on exploring for new possibilities, and when it is rational to exploit what one has already found? When exploration has costs, which in the real world it almost always does, there will be times when it is rational to exploit even though there are better options to be had. The benefits of exploration do not outweigh costs. Indeed, even if further exploration would have led to benefits that would have greatly outweighed the costs, exploitation may still be rational if the agent had no reason to expect that this would be so.

---

<sup>28</sup> I give a fuller account of this in *Willing, Wanting, Waiting*, ch. 7. There I argue that it can be rational to persist in an intention even if one thinks that, if one were to reconsider, one would probably revise. I think that the same can be true of belief. Things would be different, of course, if one antecedently knew *which* intention (or belief) one would arrive at after the reconsideration.

If all this is true for intention, and if intention provides a model for belief, we would expect it to be true of belief too. Once I have formed a belief about some matter, it will sometimes be rational for me not to reconsider it, even though, were I to reconsider it, it would be rational to revise the belief. In particular, it will sometimes be rational for me to disregard the opinions of a peer, even if, were I to reconsider my belief in the light of those opinions, it would be rational for me to revise my own.

In developing these ideas, let me start with a relatively straightforward case. The stability of intentions brings a further advantage in addition to those already mentioned. Intentions—more specifically *resolutions*—provide a crucial block against temptation. It is a feature of temptation that it tends to corrupt our judgments. Were we to reconsider what to do, we would judge it best to yield to the temptation. By maintaining a resolution without reconsideration, we can insulate ourselves from such corruption. Obviously this has immediate implications for belief. If I am going into a situation that will be epistemically corrupting, I should avoid reconsideration. And this will be equally true if the risk of corruption comes from another person: someone whose powers of persuasion will overwhelm what I would otherwise think.<sup>29</sup> Here again I should avoid reconsideration.

As I say, these cases should be uncontroversial, for here I do not take my interlocutor to be my epistemic peer: the problem is precisely that I think that they will corrupt me. But what of cases where they are my epistemic peer? Debate here has focussed on an opposition between two positions: the *conciliatory*, according to which one should move one's belief in the face of peer disagreement; and the *steadfast*, according to which one should persist in one's original belief.<sup>30</sup> It is widely thought that the debate comes down to the issue of whether one can use one's own belief as grounds for rejecting the views of others: Does the fact that I believe that p give me grounds in itself for thinking that a peer who does not believe that p is wrong?<sup>31</sup>

There looks to be something unacceptably bootstrapping about such an idea. Applied widely it would result in a horrible dogmatism, as the agent rejects all disagreement on the very grounds that it is disagreement, and hence misleading. But we can reject that idea and still have grounds for favouring a steadfast position. A reasonable steadfastness can result, not from the bootstrapping idea that my belief that p gives me reason to believe that p, but rather from the idea that my belief that p gives me reason to decline to reopen the question of whether or not p. In particular, it can sometimes give me reason to decline to reconsider my belief that p in the face of peer disagreement. Moreover, this isn't a practical reason to make oneself theoretically

---

<sup>29</sup> A familiar problem: 'You're so elegant, dear Badger, and so moving, and so convincing, and put all your points so frightfully well—you can do what you like with me in *there*, and you know it.' Kenneth Grahame, *The Wind in the Willows* ch. 6

<sup>30</sup> For a good introduction to the debate see the papers in R. Feldman and T. Warfield (eds.) *Disagreement* (New York: Oxford University Press, 2010).

<sup>31</sup> See, for instance, David Christensen 'Disagreement, Question-Begging and Epistemic Self-Criticism' *Philosophers' Imprint*, 11, no. 6 (2011) who terms a principle along these lines *Independence*, though he frames it rather differently. There are complexities here: as Christensen points out, the knowledge that I have reached a given conclusion after reasoning carefully can provide me with reason to trust myself over someone who I do not know to have been so careful. But in that case the grounds for my reasoning is more than just my belief.

irrational. In refusing to open a question in the face of disagreement, one doesn't come to believe anything inconsistent or anything like that; though if one did open the question, and then refused to be moved, one might.

To take a case close to home, consider what it is like to do philosophy; more specifically, what it is like for me to do philosophy. Over the years I have amassed a number of philosophical beliefs, some rather grand and overarching—a rejection of idealism, for instance; some very detailed—the manuscript I have just refereed is or is not worthy of publication in journal X; and many that fall somewhere in between. I also know that, for almost any of the philosophical beliefs I hold, I could find some peer who disagrees. And often not just a peer: often some philosophical great. It would be nice to say that in all these cases I have taken on board the various relevant arguments, and that I have come to my beliefs on the basis of a full assessment of each of them. But of course this is not true. It is not even approximately true. A glance at the new acquisitions shelves in the library is enough to show me that I couldn't even keep up with the new philosophical output. I may think that much of it is of poor quality, or in areas that don't concern me, but there is still far, far, too much good material, in interesting areas that have some bearing on what I am working on, for me to have a chance of reading and thinking about it all. Which is why a visit to the library will often be so depressing.

So do I live in a state of philosophical paralysis? Only when things are going very badly. Normally I get on with what I am thinking about, doing my best to keep up with the obviously relevant pieces, registering views that are inconsistent with mine, but not always fully engaging with the arguments. I'm surely not entirely self-serving in thinking that such a strategy need not be irrational. After all, it follows the advice we give to graduate students not to get bogged down in the literature. "The typical dissertation starts with two chapters," David Lewis used to say to those starting out. "The first summarizes the literature, the second makes the positive proposal. I suggest that you write the second first." Admittedly it is one thing not to try to read all of the literature; rather another to ignore an argument once it is put in front of you. But the line between the two is not clear-cut. I glance over the new journals in the common room, reading the abstracts. I skim some of the interesting pieces. One or two I might read carefully; rather more I vow to read carefully, though very often this vow is not kept. I know where much of the peer disagreement lies, but I cannot engage with it all.

Such an approach provides no legitimation of philosophical dogmatism. Instead it implies that, as with intentions, a balance must be struck, a balance between flexibility and stability. To be dogmatic is to get that balance wrong. Many factors will determine when a question is opened for reconsideration; and once it is opened, in a seminar room for instance, this approach provides absolutely no grounds for shutting it down.

My claim then is that it is practically rational to have a selectively closed philosophical mind; if one didn't, one would get nowhere. But if this is so in a discipline like philosophy, where a readiness to follow arguments to their conclusion is valued so

highly, then the lesson should be all the more appropriate elsewhere.<sup>32</sup> I suggest that it is very pervasive. My ability to operate in the work depends upon my keeping most questions closed: not for conceptual neurathian reasons (though those might hold too) but for simply practical ones. I can only focus on one question at a time. Steadfastness, I suggest, owes its main appeal to such considerations.

#### A ROLE FOR CREDENCES?

I have argued that most beliefs are like intentions: stable, all-or-nothing states that enable coordination and action. The picture that has become increasingly common in philosophy, according to which we form graded and readily changed credences, is not simply an idealization. It is quite misleading. Does that mean that there is no place for computations that make use of numerically graded states that are perfectly sensitive to the evidence? That would be to go too far. As theorists we can use such idealized models to determine what would be the best thing to do. But that does not mean that real agents actually instantiate that model, or anything like it. As in so many things, one does not achieve the second best by striving for the best.<sup>33</sup>

---

<sup>32</sup>Admittedly I have been assuming that the point of philosophy is to arrive at the truth; steadfastness would have a different role if point were rather to win arguments, something which often seems more credible in philosophy than elsewhere. On the alternative approach see Hugo Mercier and Dan Sperber 'Why do humans reason? Arguments for an argumentative theory', *Behavioral and Brain Sciences* 34 (2011), 57–111.

<sup>33</sup> A point which economists seem to have grasped somewhat before philosophers. The seminal article is R.G. Lipsey & Kelvin Lancaster, 'The General Theory of the Second Best', *The Review of Economic Studies*, 24 (1956–7), pp. 11–32.