

Hanging On To The Edges:

How my theory explains everything: And can make you happier, healthier, and wealthier

Version 1.0 of May 2017

Those with any imagination and understanding
are filled with doubt and indecision.
- Bertrand Russell, *New Hopes*

The quintessence of science is doubt. In other kinds of belief systems, the statements held to be true are hardened into dogma, declared as absolute and certain. In science, they are held provisionally, constantly questioned, continually refined and replaced. At any moment, new observations or reasoning will lead to them being qualified, revised or thrown out entirely. In dogmatic belief systems, all our epistemic effort is directed towards the confirmatory instances of what we believe; see, I went to sacred waterfall yesterday, and today my cold is better. In science, by contrast, all our epistemic effort is directed to the anomalies; the cases where the prediction is not met; where the theory breaks down; the puzzling inconsistencies that help reject incorrect claims or stimulate the development of a whole new paradigm.

This is fine, and so far, so uncontroversial. But I spend a lot of my time reading what scientists write, and I have to ask myself: why, then, do they seem so damned sure of themselves? Where is all this doubt that is supposed to be so great? Did these people not get the memo about The Enlightenment? I mean, doubt hardly gets a look in in your average scientific paper or book. Doubt, if it appears at all, is mainly levelled at *other people's* claims. At best it plays cameo parts: for example in the opening section or two, clearing the stage of clutter for the author's grand proposal to enter. And grand it is: here is my paradigm, here is what we have shown, here are all the things my theory is good at explaining. Come into my camp; what we have in here is good. In other words, in individual bits of scientific rhetoric, scientific discourse looks uncomfortably *unlike* how science is supposed to work, and awfully *like* how persuasion is done in non-scientific realms.

§

The overplaying of one's own correctness, and the underplaying of doubt, take a number of different flavours. In empirical papers published in the primary scientific literature, the principal flavour consists of finding a positive result, building the paper around it, and discussing it as if it definitely represented a real and important generalization about life. By 'positive' here, we usually mean some pattern that can have a statistician's p -value attached to it, preferably a rather small one. This is taken to mean that the said difference or association is unlikely to be due to chance. The problems with this way of reasoning are well documented. For a start, there is a good argument that 'significant' p -values do represent chance in most instances. Hang on, you say, the p -value is *defined* as the probability that the result is due to chance; so if p is small, then chance is an unlikely explanation. Here it pays to distinguish between the probability of p being small given that the finding is false, and the probability that the finding is true given that p is small. They are not the same thing.

To see why, let us say that the discovery of new and surprising things about the world is rare. By now that ought to be the case. Anyway, experiments tend to be small-scale (to be low in what is known as statistical power), and so often fail to detect real effects even when those effects are there. So if you do experiments, maybe only one in a hundred correctly detects a genuine new and surprising result. Fine. But, by definition, one in twenty will produce a p -value less than 0.05, the usual cut-off for declaring a result 'significant'. So, for every hundred experiments you perform, you will get a significant p -value about 5 times by chance, and about one time because you have made a genuine discovery. So, 5 out of every 6 significant p -values are in fact due to chance. As a very famous paper put it: most published research findings are false¹.

I have some quibbles with the details of this argument. For a start, most published *findings* are not false. The data are not false. I think almost all scientists describe the data honestly. Rather, it is the *inference* that general patterns about the world have been discovered from these small p -values that is false in many cases. Also, people do not choose the experiments they perform at random, but rather go where pilot data and other information suggests there may be something going on. So the rate of true discovery may be higher. And then the argument neglects that many of the significant p -values reported in the social and behavioural sciences are neither due to chance, nor discoveries about the world, but just reflect semantics². An example is the literature on the personality trait of neuroticism as a predictor of clinical depression. This is a perfectly serious topic of investigation. However, the measure of the predictor (personality trait neuroticism) is a questionnaire containing questions like 'Do you often feel blue?', whereas the outcome (depression) is measured with a questionnaire containing questions like 'In the last two weeks, have you felt blue?'. It would be an odd world indeed where you didn't get a significant p -value in a study like this, because the null hypothesis (people who say they often feel blue are no more likely than anyone else to say they have felt blue in the last two weeks) is somewhat nonsensical. Its failure doesn't tell you much about humans beyond, perhaps, a basic capacity for self-consistency. It tells you more about the semantics of the questionnaires. But certainly this 'finding' is not a falsehood.

These quibbles aside, the general burden of the criticism holds. A small p -value does not mean that anything has been demonstrated that is repeatable. The picture is in fact rather worse than I have suggested so far. This is because of 'researcher degrees of freedom'³. Basically, for any reasonably complex dataset, there are several different statistical analyses that could be performed. For example, there may be several slightly different choices for the outcome measure; several nuisance variables that either could or could not be adjusted for; and several subsets of the data that could be discussed. As a result, the researcher has quite a few different goes at getting that all-important small p -value, meaning that the probability of finding it by chance is typically rather a lot higher than 0.05.

An example has recently come to my attention, as I am reviewing the literature on whether stress shortens people's telomeres. Telomeres are the DNA caps on the end of our chromosomes. They tend to shorten with age, and there are some decent reasons for believing the shortening is speeded up by stress. The typical modus operandi is to measure 'stress' in at least half a dozen different ways (a perceived stress scale, a depression symptom checklist, whether you have to care for a sick relative or not, life events, a childhood trauma questionnaire) in a sample of people who also have their telomeres measured. Often, one or even two of these measures will be a 'significant' ($p < 0.05$) predictor of telomere length. The paper then gets into a high-profile journal as evidence for the claim 'Stress shortens telomeres'. The trouble is, sometimes the 'significant' predictor is childhood trauma, sometimes perceived stress, sometimes depressive symptoms, etc. And the paper is organized around

¹ Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine* 2, 696–0701.

² Arnulf, J. K., Larsen, K. R., Martinsen, O. L., & Bong, C. H. (2014). Predicting survey responses: How and why semantics shape survey statistics on Organizational Behaviour. *PLoS ONE*, 9: e106361.

³ Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22: 1359–1366.

the atypical finding of the ‘significant’ association, from its title, through its summary, to its discussion. Remember that if you have six stress measures and one of them is a ‘significant’ predictor, the other five are not; and also that if you test enough associations, the chance of at least one having a p -value less than 0.05 approaches certainty. If you summed up or averaged the stress measures into a single key variable, it probably wouldn’t be associated with telomere length. So really, a better title for a paper describing the very same results would be ‘Measures of stress are generally unrelated to telomere length’. I am not singling these particular papers out as egregious. I actually have some faith in their general hypothesis, and I would like it to be true. It does show us how business is done, though.

§

All in all, then, it should be no surprise that when studies in the behavioural and life sciences are replicated, we don’t consistently see the patterns reported in the originals; the patterns argued so confidently to be real and to be a vindication of this or that theory⁴. A great deal has been written about this recently. It makes uncomfortable reading. The scales are falling from our eyes⁵.

As consumers of science, we have to shift our focus from thinking about individual findings as instruments of truth, and instead thinking in terms of a slowly-evolving *population* of findings⁶. If one study finds that eating broccoli is associated with less depression, that’s not news. Once the *population* of scientific studies starts to contain finding after finding associating, in diverse ways, broccoli with less depression, that starts to be something interesting; maybe even heading for knowledge counter. Even once there it will continue to be dissected and its causal basis probed. It’s important for the media to understand that the development of knowledge is a gradual, population process, else it looks from the headlines like science is changing its mind all the time. This invites a kind of scepticism and hostility about the whole enterprise. Say I am trying to work out the proportion of cars on the roads that are red. Imagine if I issued a press release when I have observed a red car: ‘Study finds that all cars are red!’. Then a rival university could do a debunking study. They observe a car. It’s not red. ‘Study finds that no cars are red. Earlier study was flawed!’. Both parties increase their scientific visibility. But the truth is: neither study is decisive (that’s why science is hard); and yet every decent study very slightly increases the precision of our collective knowledge (that’s why science works).

§

What interests me the most is not the near-universal over-selling that goes on in scientific papers describing the results of primary empirical studies. It is the programmatic over-selling you find when academics write their ‘big idea piece’, or BIP. The BIP is often a book, or more rarely, a long-form discursive article. If a book, it may be aimed at a more general audience than just the few research specialists in the field; the so-called ‘popular’ or ‘trade’ book. It’s the author’s chance to cement their position in the minds of the community; to recruit future devotees; to make their mark. (By the way, the pronoun in the preceding sentence is not a grammatical error, but none other than ‘singular they’, the American Dialect Society’s word of the year for 2015. And by the other way, let me pre-empt your inevitable thought: I am as guilty as anyone else of the sins to be described below.)

The BIP has a recurrent four-act structure.

⁴ Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science* 349: aac4716.

⁵ In particular, I thoroughly recommend taking Professor Sanjay Srivastava’s class on scientific methods, *PSY607: Everything is fucked*. See: <https://hardsci.wordpress.com/2016/08/11/everything-is-fucked-the-syllabus/>

⁶ McElreath, R., & Smaldino, P. E. (2015). Replication, communication, and the population dynamics of scientific discovery. *PLoS ONE* 10: e0136088.

Act One: There are huge problems to be solved. The science is in disarray. Our existing theories are either rubbish or incomplete.

Act Two: Heroically, through alarms and excursions, I've come up with this theory. It overturns the rubbish theories and completes the incomplete ones. It solves the problems. Here are some arguments in its favour.

Act Three: Here is another thing my theory is good at. And another. And another. The good things about my theory go beyond the problem I set out to solve. Within a broad domain, it is quite possibly the long-awaited theory of everything.

Act Four: You can apply my theory to life. Unification of the sciences? My theory. Economic uncertainty? My theory. Environmental degradation? My theory. Problems with time management? My theory. Unlucky in love? You've guessed it.

We can know these acts by nick-names, each of which tells you what the problem is. They sound like the syllabus for an old-fashioned class on bad reasoning:

- *Act One:* The Straw Man
- *Act Two:* The Great Man View of History
- *Act Three:* Confirmation Bias
- *Act Four:* If All You Have Is A Hammer, Everything Looks Like A Nail

The BIP is, more or less, a systematic over-statement of the relative merits of the position the author is peddling. I say relative merits, because not only is the author's position inappropriately bigged up; the alternatives are inappropriately belittled too. It is not enough for my theory to succeed; first others must fail. Usually, the existing theories are flayed and mis-represented to the point of denying their daily successes and the valid insights on which they were based. There is a tendency here to go for the tall poppies. Presumably you get more points for flaying a big theory than a little one. Favourite targets are 'economic theory'; 'social science' (as if that were a thing); 'Evolutionary Psychology'; and, in top spot, 'Darwin's theory of evolution'. As John Welch has recently pointed out, there is a quite a cottage industry of setting out to declare Darwin's theory fundamentally inadequate/incomplete/incorrect, usually as Act One to the author's own BIP⁷.

You can see why. If you start out by saying: 'evolutionary theory is basically fine as far as it goes, but I want to talk about phenomenon X somewhere in the world of living things', then you are just a worker at the mill. You pay the rent but don't get your picture on the cover of *Nature* magazine. If you start out by saying: Darwin's theory fundamentally requires refinement/completion/replacement (by the thing I want to talk about), then you are an intellectual giant-killer, and people pay attention. But some of the claims that follow can be a little over-cooked. Often they present a phenomenon that is, directly or indirectly, the *outcome* of genetic evolution as if it challenged our fundamental understanding of the *process* of genetic evolution. You may have seen examples: Richard Dawkins says evolution is all about selfish genes, but what about mutualism/mirror neurons/mariachi music? They're really important! Evolutionary theory as we know it can't be right! It's not that mutualism/mirror neurons/mariachi music aren't interesting, of course. Nonetheless, I can't read these kind of texts without the image coming to my mind of a philosopher of science, somewhere across the world, weeping quietly in order not to wake their spouse.

Whereas in Act One, all other theories are caricatured and over-criticized, in Act Two, the author's own big idea is suspiciously free of problems. It accumulates credit for things that are indeed

⁷ Welch, J. J. (2016). What's wrong with evolutionary biology? *Biology & Philosophy* 32:263–279.

consistent with it, but are probably consistent with many other theories too. And then we only get to visit the cases that show it to its best advantage. The odd thing here is that the author ought rationally to believe that their own theory, too, is likely to turn out false. Philosophers even have a name for this reasonable inference: the pessimistic meta-induction⁸. The grounds for the pessimistic meta-induction go like this: the vast majority of explanatory beliefs about the world that humans, including scientists, have entertained through our history have turned out to be wrong. So faced with a new one, a reasonable being's assumption should be that it too will turn out to be wrong. We should thus entertain a certain detachment and vigilance toward it. Yes, it could be interesting, worth thinking about, but at least in the current form, it's probably wrong and it's not going to be around for very long. The authors of BIPs are good at applying the pessimistic meta-induction to all the other theories; they just can't take the extra step and apply it to their own.

In some cases, Acts Three and Four become comical. The second half as farce. In Act Three, The Theory, extended further and further beyond any basis it had in the technical literature, becomes more and more under-specified, under-evidenced and under-grounded. It's not even a promissory note. It's a vague promise to meet you with an important package at some unspecified time in the future in a pub near Chingford. But by this point the author is three months late with their manuscript and subsisting on a diet of their own rhetoric. And then Act Four. You can almost feel hear the literary agent: 'Well, they like it, but it's a bit academic. It has a better chance of breaking through if readers could see how to apply it to their daily lives. Could you put in a chapter about how to make practical use of your big idea in internet dating/choosing a pension plan/promoting world peace? After all, we've got a big advance to pay off'.

But this is not how science ought to work, is it? Surely the BIP should be a little bit about (a) the *common features* of different existing intellectual approaches to a problem, with a view to how they can be synthesised; and a lot about (b) the *failures* of our current understanding: the anomalies, the failed predictions, the problems, the things that don't yet fit. That's where the scientific action is.

§

The reasons that BIPs over-state their cases do not seem too difficult to understand. Do we need to appeal to anything more than ordinary human self-interest? People respond to the incentives. They want to get their papers published. They want their grants funded. They want status. They want their book to make a stir. And a few of the best 'popular' or 'trade' books about science have made their authors surprisingly large amounts of money. So we have a perfect incentive set for ambitious writers and thinkers to over-sell their wares. What else would we expect them to do?

There's been a lot written lately about changing the incentive structure of science, for example so that the acceptance of data for publication does not depend on the size of the *p*-value. This obviously makes sense. It will make it harder to ignore anomalies and negative replications, and that in turn will allow more airtime for appropriate doubt. But it will not eliminate the BIP-problem (if we decide that it is a problem, a question to which I will return). BIPs are typically statements of broad theories or paradigms. In the human sciences at least, broad theories or paradigms seldom make individual 'line-in-the-sand' predictions that can be decisively judged to have failed. They make meta-predictions: I predict that it will prove more useful in approaching a topic you wish to study if you use my framework to make your predictions, than if you don't use my framework. And it's obviously hard to defeat this meta-prediction: maybe the meta-prediction is false, maybe you just didn't use the framework right (yet). And who knows what it would have looked like if you had used a different framework anyway.

⁸ See Doppelt G. (2007). Reconstructing scientific realism to rebut the pessimistic meta-induction. *Philosophy of Science* 74: 96–118. Note that in the philosophy of science, the pessimistic meta-induction is discussed as an argument against the realist view of science (that is, the view that science can gradually come to approximate the objective truth about the world). I am using it in a more informal and uncontentious sense: that a lot of individual scientific ideas have turned out to be wrong.

So there's plenty of wiggle-room for writers of BIPs to construct cases for their pet theories, and scope for ailing big ideas to persist.

Anyway, there may be deep reasons BIPs so over involve over-selling. That this is so is suggested by a recent book, actually a BIP, Hugo Mercier and Dan Sperber's *The Enigma of Reason*⁹. To explain their central thesis, first let us grant that there is a human capacity for reasoning, which is, roughly speaking, the capacity to produce and evaluate reasons for beliefs and actions. Obviously, this capacity is central to the possibility of science. Mercier and Sperber review decades of psychological research showing, pretty unanimously, that human reasoning is systematically biased. Importantly, this is not just true in science, but across a broad range of contexts, every-day as well as arcane.

Specifically, human reasoning often shows evidence of 'me-sidedness'. Individuals find it easy to accumulate reasons *for*, and hard to find reasons *against*, things they are anyway disposed to believe intuitively. When it comes to things they want to believe anyway, they will accept relatively weak reasons for doing so; indeed, when asked to generate reasons for their actions and beliefs, their initial offerings are typically weak and superficial. Only when really pushed by other people will they come up with better ones. By contrast, when evaluating beliefs or actions that other people find intuitive, that they themselves no strong intuitions about, they take a more balanced view of the pros and cons. They evaluate other people's reasons in a much more demanding way than they generate their own.

Common experience tells us that me-sidedness in reasoning is very widespread, but it has taken some experimental deviousness to demonstrate it directly. In one ingenious set of experiments, participants selected answers to some logic problems, and gave reasons for their answers¹⁰. The problems were of a kind that has an undeniably correct answer, but not one so obvious that everyone sees it straight off. In a second phase, the participants were given the (different to their own) answers and reasons of another participant, to see if they accepted these and wished to change their own conclusion. Here's the deviousness, though: in one condition, the answer and reason of the 'other participant' was in fact their own answer and their own reason, whilst the one attributed to themselves was actually that of someone else. Many participants failed to detect this, because the critical switch was hidden in a number of non-switched problems. Anyway, these problems were unfamiliar, and participants were unlikely to have any settled views on them such that they immediately recognised the departure from their own position.

What were the results? First, people got the problems wrong quite a lot of the time (the problems had of course been chosen to produce this outcome). Second, when invited to reason about their own initial answers, they easily generated weak and superficial reasons in favour of them. In fact, giving the participants more time to reason about their answers did not lead to them switching their initial answers very often at all, even when their initial answers were wrong. Reason, applied to the responses they knew to be their own, just tended to confirm whatever they had intuited anyway, even when it was bad. Third, and most importantly, when they re-evaluated their own answers and reasons believing them to be someone else's, they thought quite critically about them. In fact, they rejected them as invalid slightly more often than not. Reassuringly, rejection was particularly likely if their answer had in fact been wrong. Nonetheless, the force of the result stands. These were answers and reasons that *they themselves* had in fact generated about five minutes earlier. When they thought they were justifications of their own intuitions, the participants thought the arguments were fine. When they thought they were justifications of someone else's intuitions, they were appropriately and effectively sceptical.

The experiments described above uncovered me-sidedness in reasoning by making the person's own arguments appear as if they were someone else's. The converse sleight of hand—making someone

⁹ Mercier, H. & D. Sperber (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

¹⁰ Trouche, E., Johansson, P., Hall, L. & Mercier, H. (2016). The selective laziness of reasoning. *Cognitive Science* 40: 2122-2136.

else's arguments appear as if they were one's own—has also been done¹¹. Me-sidedness suggests that people will like theories more just because they feel that their own intuitions have led to them, and this is exactly what the experiments showed. Participants were introduced to material about an alien planet, and to a theory about the behaviour of two species of fictional creature on this planet. They rated their degree of belief in the theory. The theory seemed initially plausible on the information given, but more and more facts were gradually uncovered until the theory started to seem unlikely. The experimental manipulation was chillingly simple: in one condition, the theory and discoveries were attributed to an 'Alex', who was presented as a researcher finding out about the planet. In the other condition, the same theory and discoveries were attributed to 'you' (the participant). And guess what: the participants rated the theory as more likely to be true just because it was suggested (pretty subtly) that they themselves had come up with it. This remained the case once the contrary reasons started to pile in; though reassuringly, the participants in both conditions showed a decline in belief as more counter-evidence accumulated. Between them, these results seem clear, and are corroborated by many other findings. We like our own intuitions and positions, and will accept pretty weak reasons for them; whereas we are sceptical and demanding about other people's intuitions and positions.

These facts about reasoning constitute something of a puzzle. The classical view of reasoning is that its function is to help the thinker find true beliefs and adopt right actions. It's a distinctively human adaptation for making each individual better at discovering the truth about their world. On this view, the ubiquity of 'me-sidedness' looks like a daft flaw. Here we are with this great telescope, but we usually use it with a rose-tinted lens on it. Mercier and Sperber argue, though, that the original and typical function of reasoning is not for the purpose of solitary, internal truth-finding. The function of reasoning is to facilitate social interaction.

We humans coordinate our actions with other individuals to a remarkable extent, often for mutual benefit. It's hard to do that effectively given that individuals typically have different preferences, needs, expectations and experiences. These preferences, needs, and expectations are, within each individual, largely generated intuitively. We give reasons to each other as a way of bringing about smooth coordination and effective collective action. But of course, the interests of different individuals in a social group are typically only partially aligned. We want the group to coordinate effectively, but we would prefer it to coordinate effectively in doing what we want, not what the others want. And so reasoning is an adaptive capacity in which we seek to move the intentions and plans of others towards where we already want them to be.

Viewed in this light, me-sidedness is not a design flaw of reasoning, but a design feature. Of course reasoning should be good at finding arguments in favour of our own positions—that's what it is for! Of course it should be satisfied with the minimal acceptable argument in favour of our own position—we are already persuaded of it! Our reasons only need to be good enough to get traction with others, so only if other people dispute or reject them do we need to generate better ones. But on the other side of the coin, the evaluation of arguments, Mercier and Sperber's position rightly suggests we ought to be quite sceptical and demanding. After all, it's really not in our vital interests to go along with other people's agendas in life just docilely. Our interests are typically different to theirs, even if both parties are going to gain from coordination. We don't want to be dupes. When other people have intentions and courses of action, then, we should evaluate them quite carefully, and we should demand decent reasons why that course of action rather than another is in fact an appropriate one.

§

Applying all this to science, *of course* BIP-authors are going to over-state their cases. They are not being cynical, knowingly doing so for financial or status reasons. They are giving us the honest output

¹¹ Gregg, A.P., Mahadevan, N. & Sedikides, C. (2017) The SPOT effect: People spontaneously prefer their own theories. *The Quarterly Journal of Experimental Psychology* 70:6, 996-1010.

of their—often impressive—reasoning processes. It's just that their reasoning processes, if Mercier and Sperber are right, are intrinsically prone to being me-sided. Thus, as a matter of course, you can expect to find all the reasons for the position the author lives with and few of the reasons against, alternative positions down-played or used as foils, and a failure to tackle difficult counter-examples. At first blush, this all seems rather depressing for science. We look to science as the paragon of objectivity, but now we end up concluding that scientists are no better, no less partisan than politicians or quacks, and moreover that this is because of fundamental design features of human reasoning. But actually, Mercier and Sperber's thesis is not bad news for science.

For a start, what makes science revolutionary is not that the individual scientists are necessarily any better at reasoning than politicians or quacks. Why should they be; they are no more and no less human. No, what makes science revolutionary is the way that knowledge-evaluating processes are socialized. Individuals may often over-claim for their positions, but the scientific community has particular norms and institutions for counter-acting this: peer review of papers, critical review articles, replications, meta-analyses. In other words, the objectivity of science is not contained within the heads of the individual scientists who come up with the ideas, but rather is distributed across the community of people who review, argue, test, critique, revise, evaluate, and teach. In evolutionary biology, we are taught that genetic mutation proposes, but natural selection decides. In science, intuition plus me-sidedness proposes, but the community (eventually) decides, and it decides at least to some extent on the basis of evidence and arguments. It's not the mental processes of scientists, but the social norms and institutions for systematic evaluation of claims, that make science different from dogmatic belief systems. This is why it is so disappointing when the media present 'the scientists don't agree' as a reason for dubiousness about some area of science. Of course the scientists don't agree—if they did, they wouldn't be scientists! But here's the paradox of science: by never agreeing, by always doubting, we gradually and collectively come up with beliefs we can all agree on and which we do not need to doubt.

And actually, the reasoning experiments show that people are generally quite good at evaluating arguments as long as those arguments are not their own. This means that processes like peer review and replication, when implemented and executed wisely, will tend to do some good. These processes can be frustrating and arbitrary at times. There's nothing worse, when you are just getting going with your own me-sidedness, than having to deal with someone else's me-sidedness! Nonetheless, these painful exchanges constitute a critical selective pressure that on average improves the level of correctness in the population of beliefs that the community currently holds. They drag us very slowly up a selection gradient towards knowledge—with all the usual caveats that selection is probabilistic, that selection gradients are not uniform, and that you can get stuck at local maxima in the landscape.

Even more than this, reasoning research shows people will eventually abandon and revise their own positions if their reasons for holding them are challenged in a compelling enough way. Me-sidedness means people have strong priors in favour of their own intuitions, but they are not completely immune to updating their intuitions given enough evidence and argumentation. This, Mercier and Sperber suggest, leads to a more positive view of the scientific process than that contained in Max Planck's famous claim that: 'A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die'. It suggests that as long as scientists don't live in silos, as long as they talk, argue, justify their claims to each other, then individuals with a stake in the research will at least sometimes change their minds in a way that goes from worse beliefs to better ones.

§

All this casts the BIP in a slightly different light. If the community is going to decide, it needs to know what the strongest case is for each of the options it is deciding amongst. Thus, BIP authors, ridiculous though they can be, are playing a useful role in a wider drama. There is a parallel here with the

adversarial legal system in countries like England. In an English court, an impartial authority, perhaps a jury, will ultimately decide one way or another. One advocate for the prosecution and one for the defence will each present the strongest possible case for their side. The advocates are expected—indeed required—to show me-sidedness for their position, to accumulate arguments for it, and to minimise the arguments against. In the proper exercise of their functions, at least one advocate will be wrong, must be wrong, and both must be partial. So too, perhaps, in science: by writing a BIP, an author makes the best possible case for the prosecution or for the defence, not because reality probably is that way, but because the best way for the impartial community to evaluate will be to have laid out before it the strongest version of the case that can be assembled.

I can see that it is useful for science to bold exemplars of positions that it is trying to assess. Imagine if every statement of a theoretical position were hedged around with caveats, complete in its weighing of pros and cons, exhaustive in its treatment of possible alternatives and other factors, extensive in its generous discussion of how similar it was to other positions and frank about the narrowness of its applicability. It would be very hard to get your teeth into exactly what was at stake. If the impartial community is to be judge and jury, it needs the strongest and most entertaining possible cases for the prosecution as well as the defence. As a young man ostensibly studying psychology and philosophy, I actually spent most of my time reading popular books on evolution. Why? At least partly because the evolutionists had a big idea whose universal scope and power they presented without qualification. These were ideas singing at the tops of their voices, not mumbles and apologies. Social science writing, with its frequent insistence on variegation and specificity, on multiplicities of factors, can really lose out here in the airtime of public discussion.

We seem to be in danger of completely exonerating the authors of BIPs from their absurdities, though, and I am not happy with that. We have seemed to have freed them from even a minimal obligation of balanced decency and good scholarship. Rather than striving to overcome their proneness to me-sidedness, BIP-authors can claim it serves the greater good of science ('just doing my job, guv'nor'). This seems to be going too far. And there is an alternative to adversarial legal systems: inquisitorial (or non-adversarial) systems. A substantial fraction of the world's population lives successfully under such arrangements. Here, the court itself is involved in the gathering of evidence, both for and against, and must come to a balanced determination. The generation of arguments and the evaluation of their merits are unified in a single office, rather than being divided across the advocates and court respectively. Should science operate more like this?

It's said that inquisitorial systems may be better at discovering the truth, whilst adversarial systems may be better at giving all individuals a hearing and hence protecting them from the inappropriate exercise of power. I don't know. I suspect science needs both modes. Really new ideas, paradigm-shifters like continental drift, are not going to get any traction without me-sided advocates. So you need the adversarial model at an early stage in the development of a paradigm. But cocky advocates strutting against one another does get a bit wearing—tribal, sterile, prone to self-congratulation and self-perpetuation. It is incapable of sorting out the details and typically does not produce synthesis. You then need wise magistrates, a lot of them.

The population needs both behaviours, and it will always get them, because some scientists take readily to the adversarial mode, while others gravitate more to the inquisitorial. It depends to some extent on one's personal balance of approach and avoidance motivations. What gets you out of bed in the morning, the possibility of glory and renown, or the terror of turning out to be wrong? For some BIP-writers, it seems to be the former; for natural magistrates, the latter (or at least, strong scruple about fairness and correctness) seems to loom larger. And of course, the same personality diversity when it comes to reasoning is observable outside science too: some people want to persuade and charm their social circle, while others place a big emphasis on listening to all sides and forming a reasonable consensus.

We need to value our bold advocates, our BIP-writers. In the end, most of them, like Icarus, will fail, but in so doing they might just open up new terrains, inject creativity, inspire others, before they burn out. But they need to have a decent modicum of balance and openness, and we need to soften their more blatant partisanship. Mercier and Sperber's thesis suggests that we can do that in science the way we do it in the rest of life—through conversation. It is through conversation that people's reasons are challenged, questioned, refined, balanced. But it has to be conversation undertaken in good faith with others whose perspectives are different from one's own, otherwise all that results is entrenchment and polarization. BIP-authors should not be rewarded for, or by, living in disciplinary or paradigmatic silos. Rather they must be engaged in friendly and quizzical conversation. We also need to make sure our institutions value and reward the quieter and wiser magistrates too. We don't currently do this enough. In promotion, funding, publication, and visibility, more thoughtful, perhaps more honest, souls lose often out. It is easy to see how this ends up happening. Indeed it is related to the broader societal pattern of more extroverted people being rewarded more in the world of work, without obviously adding more value¹². It is not alright though.

Anyway, I have to stop this now and work on the draft of my new book. It's called: *As Far As It Goes: A Decent Theory That Isn't Revolutionizing All of Biology, and Probably Won't Change Your Life*. Do you think it is going to sell?

¹² Pehkonen J. (2010). Personality and labour market income: Evidence from longitudinal data. *Labour* 24:201–220.