



# Thou Shalt Not Take Sides: Cognition, Logic and the Need for Changing How We Believe

André C. R. Martins\*

*Interdisciplinary Nucleus in Complex Systems, NISC – Escola de Artes, Ciências e Humanidade, EACH – Universidade de São Paulo, São Paulo, Brazil*

We believe in many different ways. One very common one is by supporting ideas we like. We label them correct and we act to dismiss doubts about them. We take sides about ideas and theories as if that was the right thing to do. And yet, from a rational point of view, this type of support and belief is not justifiable. The best we can hope when describing the real world, as far as we know today, is to have probabilistic knowledge. In practice, estimating a real probability can be too hard to achieve but that just means we have more uncertainty, not less. There are ideas we defend that define, in our minds, our own identity. And recent experiments have been showing that we stop being able to analyze competently those propositions we hold so dearly. In this paper, I gather the evidence we have about taking sides and present the obvious but unseen conclusion that these facts combined mean that we should actually never believe in anything about the real world, except in a probabilistic way. We must actually never take sides since taking sides compromise our abilities to seek for the most correct description of the world. That means we need to start reformulating the way we debate ideas, from our teaching to our political debates. Here, I will show the logical and experimental basis of this conclusion. I will also show, by presenting new models for the evolution of opinions, that our desire to have something to believe is probably behind the emergence of extremism in debates. And we will see how this problem can even have an impact in the reliability of whole scientific fields. The crisis around  $p$ -values is discussed and much better understood under the light of this paper results. Finally, I will debate possible consequences and ideas on how to deal with this problem.

## OPEN ACCESS

### Edited by:

Katarzyna Sznajd-Weron,  
Wrocław University of Technology,  
Poland

### Reviewed by:

Janusz Szwabiński,  
University of Wrocław, Poland  
Piotr Nyczka,  
Jacobs University, Germany

### \*Correspondence:

André C. R. Martins  
amartins@usp.br

### Specialty section:

This article was submitted to  
Interdisciplinary Physics,  
a section of the journal  
Frontiers in Physics

**Received:** 22 September 2015

**Accepted:** 03 March 2016

**Published:** 30 March 2016

### Citation:

Martins ACR (2016) Thou Shalt Not  
Take Sides: Cognition, Logic and the  
Need for Changing How We Believe.  
*Front. Phys.* 4:7.  
doi: 10.3389/fphy.2016.00007

**Keywords:** belief system, cognition, individual and group reasoning, logic, opinion dynamics, epistemology

## 1. INTRODUCTION

We believe. We believe in the honesty of certain people and in the dishonesty of others. We might believe we are loved. Some of us believe in deities, some believe in political or economical ideas, some believe in scientific results. We have beliefs on how the world ought to be and, sometimes, at least for those beliefs, we are aware that reality does not correspond to them. But we also hold beliefs about the actual state of the world, about how things actually are. And sometimes those beliefs are so strong that we feel justified in saying that we know. We often choose to say we know the truth and we defend those beliefs as if they were indeed true. We can also hold beliefs about abstract entities that do not actually exist in the world outside our minds and languages, such as numbers. We make claims of knowledge, and we expect to have real knowledge. This expectation

is so strong that in logic, it is often considered that the phrase “Someone knows  $p$ ” means that  $p$  is actually true [1]. The beliefs we defend have a vast importance in all social aspects of human life. We have several organizations based on the spread of ideas, on defending the beliefs that the people in those organizations hold to be true, as if they were superior to beliefs of others. We join groups that share our beliefs and we fight (sometimes only verbally) those we disagree with. We even often define ourselves based on what we believe, as, for example, by our religious or political beliefs. But, as we will see, taking sides is not only logically wrong. It also makes us dumber than we could be, as experiments have been showing lately.

The actual question of what beliefs and knowledge are is an old and, maybe surprisingly, still open problem in Philosophy [2]. The concept of belief seems so natural that we have never really doubted its usefulness; so far, we only tried to understand and define it properly. But, when we have asked what knowing and believing mean, we have usually been forced to ask if knowledge is even possible. And one common logical answer is that Skepticism [3] could be unavoidable. And yet, whenever we have taken a position that did not include the hardest forms of Skepticism, the concept that we would accept some ideas and reject others have always come naturally. We do not question it. Accepting and rejecting ideas about the real world, seems natural, as a concept, it has survived for a long time without being properly challenged. It could be the case we have never challenged it just because it is too obviously right. But it is a scientist duty to ask if this is indeed the case. And we must acknowledge that the concept might just feel obvious as consequence of some very deep instincts we have.

Indeed, recent experiments in cognitive psychology point to serious problems with the way we naturally work with our beliefs and how we update them. Our many flaws in reasoning have led to the conclusion that the main function of our reasoning might not be making better decisions. Our reasoning about ideological issues does seem to come in packages better described as irrationally consistent [4]. Instead of looking for truth and correctness, we seem to use reasoning to win arguments [5, 6] and to protect our own cultural identity [7]. Looking for the correct or best answer might even be irrelevant in too many cases. At best, finding better answers could be just a beneficial side effect of the actual purpose of our reasoning skills. Their main function seems to be a social one, to allow us to identify and defend the ideas of those groups that define the way we see ourselves.

These results pose interesting questions. Since we do not use our reasoning to find the truth under many circumstances, a natural question we should ask but usually avoid is how we could counter this insidious bias. But even this question might not address the problem fully. A more meaningful and complete question could actually challenge the very idea that we should have identity-defining beliefs. That is the point of view I will defend in this paper. Evidence shows clearly that when we embrace an idea, be it a religion or a political ideology, we start wishing to defend that idea. And our brains start working to defend this interest, instead of looking for the correct answers. While it is not clear if this could pose a problem for moral issues, where no correct answer is known to exist, we do have a problem when confronted with any ideology that makes claims on how

the world actually is. Here, I will show why we should make very strong efforts to avoid those types of beliefs. In Section 2, I will present the problem as we know it now. I will review some of the evidence that shows our many reasoning errors and biases and proceed to investigate, from a normative point of view, if there is any reason to hold and defend beliefs. I will also briefly show that real life might demand us to speak as if we were defending or attacking an idea but that this does not necessarily mean holding beliefs about it.

In Section 3, I use a computational model to highlight the potential damage the desire to have a belief can cause, even if, at first, the agents are in doubt. I will investigate, by using the tools of opinion dynamics [8–13], how the way we treat beliefs can have a profoundly deep impact on the extremism of our thoughts. And how a subtle change in the way we think about problems can have an absurdly strong effect on extremism. I will investigate the effects of two possible causes: the differences that arise from communication methods where uncertainties are expressed, and the differences on how strong opinions can become based on how our own internal expectations about the information we obtain. The simulations presented will make a very strong case in favor of why we must change the way we deal with beliefs, not only to become better when looking for true or best answers (as if was not already reason enough) but also to avoid extremism. The simulations will show that extremism can be a result of the expectation to find a definitive single answer instead of being satisfied with a mixed result.

In Section 4, I will address the consequences of all those conclusions for scientific practice in general. As we will see, avoiding taking sides can help correct some problems we still have with the reliability of the scientific enterprise. The question of whether or not scientific knowledge can be considered far more reliable than any other kind of knowledge is one that many people consider settled. And yet many of those people who think they know the answer strongly disagree with each other. They take sides and defend those sides. Individual points of view vary widely. Some people consider that all scientific knowledge is no more valid than any other type of opinion, as in some wild interpretations that followed the proposals in the strong program in the sociology of knowledge [14]. Others hold very strong realistic beliefs that we are actually uncovering the true laws of Universe. Or even bolder positions that allow Stephen Hawking to claim, maybe as a joke, that he believes Roger Penrose to be a Platonist [15]. While it would make no sense to deny that actual scientists are human beings and, therefore, subjects to our failures and social pressure, it also makes no sense to deny the incredible achievements Science has achieved or attribute them to sheer luck. It makes sense to be quite wary of these identity-defining ideas so strongly defended by their proponents. It might indeed be wiser to learn which aspects of the best description of the problem each of the sides got right. In the terminology of Section 3, we should strive here to be mixers instead of wishers.

That we have sent probes to the outer reaches of the Solar System, that we have eliminated diseases and dangers and we live much longer than we used to just a couple of centuries ago are clear indications that we must be doing something right.

On the other hand, since we should not take sides, there is some reason to be concerned about scientific work (or any work that requires intellect). Indeed, instead of a cold and strictly rational relationship with the word of ideas, scientists can be quite passionate about what they defend. We now have access to data that points to the fact that stronger intellects might actually show stronger polarization of opinion effects [16]. More intelligent people can use their intelligence to protect the things they believe; the larger the brain power, the better defense one can prepare. At one side, we have our achievements to tell we are doing something right, at the other side, evidence that tells us to doubt the conclusions of our most intelligent individuals.

A deductive analysis of what Science can prove does not get us more confident. Popper idea of false-ability [17] was indeed based on the trivial logical fact that when a theory makes a prediction, just observing that the prediction holds provides no proof at all that the theory is right. On the other hand, if the prediction fails, we do know that, since the conclusions were wrong, something must be wrong with the premises. For Popper, that meant that, after proper checks for other possibilities, the theory had to be false. But that is actually only true if too many conditions are met. First, there must be absolutely no chance for experimental error. But the theory we can reject then is only the whole set of ideas used to make the prediction. That includes every detail about the Universe assumed during calculations. If one uses a more traditional meaning to the word theory, Popper idea that a theory can be proved false is just wrong from a strict deductive point of view [18, 19]. The discovery of Neptune is the classical example of a prediction (the movement of Uranus) that was wrong not because the theory of Newtonian gravity was failing. Uranus did not behave the way the theory predicted because there was another unknown planet we did not know at the time, Neptune. Neptune gravitational influence could not have been considered in the calculations before its existence was known. The structure of the Solar System is not considered part of Newtonian mechanics or any theory of gravity, it is just an auxiliary hypothesis. That is, actual failure of any predictions can be due to the theory we are testing being wrong, of course. But the failure can also be caused by many other hypothesis we had to make to obtain the predictions. From a strictly deductive point of view, we can not hope to prove any theories neither right nor wrong. The existence of Neptune was verifiable and, therefore, it might seem at first that we could actually have disproved Newtonian mechanics if Neptune did not exist [20]. But other hypothesis could have been made. Real calculations involve approximations and we never really have an exact prediction from a theory. There will always be the possibility that a new, unthought hypothesis could save any given theory.

The real problem is, of course, as we will see in Section 2, the fact that deduction can not actually tells us anything about the real world. This is true for scientific reasoning, just as it is true for any other kind of reasoning. Instead, we need inductive tools to estimate which theory is more likely to describe the world correctly. Indeed, current descriptions of scientific work have proposed that scientist, when doing honest work, would behave as if they were Bayesian agents, an idea known as

Confirmation Theory [21–23]. While this is a solid prescription for how scientists should behave, there is actually little evidence that scientists reason better than any other intelligent person. While a Bayesian description does capture the qualitative aspects of our reasoning nicely [24, 25], we know we actually fail at most probability estimates [26–28]. Indeed, since the first experiments on how we humans change already probabilistic opinions when presented new data, it has been observed that we have a strong tendency to keep our old opinions, a bias known as conservatism [29]. It is therefore reasonable to expect that we will have problems in Science. These problems might be minimized by some of the characteristics of the scientific work, by social effects, or not at all. If they are minimized, understanding which characteristics are important for that minimization could, in principle, help us improve the way we reason as scientists, as citizens, and in our general lives. If, or when, it is not, we must learn to improve and correct the ways we, scientists, work.

We already know that previous observations of scientific work have suggested that scientists break some expected norms all the time. Surprisingly, we will see that part of this non-compliance with the investigated norms is in complete accordance to normative Bayesian reasoning. It might actually be one reason why some areas are quite successful in explaining their objects of study. While those observations, when they were first made, were considered evidence that we should adopt a vision where all ideas are equally valid, this proposed (and quite damaging) relativism is not justified at all. The concept that we should accept the creation of new ideas, regardless of how much they contradict what we currently know is actually a good prescription. But Induction dictates that all these ideas must be ranked in a probabilistic way and that means they will not be equal at all. Some ideas will be far more probable than others. There are theories that we can claim, in a harmless abuse of language, that we have proved to be wrong. Ideally, we have to remember that we should keep even those theories as very remote possibilities. It might be justified to discard them in cases where our limitations dictate we can not consider all possibilities and must focus only on the more probable ideas, of course. But no idea can actually be discarded by data and they should survive in a limbo of basically useless concepts. And that means, among other things, that we actually need pure theorists in every mature area and those theorists should be capable of providing testable predictions.

Finally, not taking sides has another very important consequence on how we should analyze data. We have created and we have promoted the widespread use of statistical tools that have the objective of discarding hypotheses. Far worse, we have made those tools central to the whole process of acceptance of ideas in some areas of knowledge. This human desire to know the absolute truth and avoid dealing with the undesired uncertainty of the probabilities is at the very core of the whole problem caused by the widespread use of  $p$ -values and tests of hypothesis. Those tests are making whole fields of knowledge far less trustworthy than we can tolerate. We actually cause too much harm by using techniques aimed at excluding ideas, instead of ranking them. And we must change as fast as possible what is considered proper statistical techniques as they make our desire to defend an idea a much stronger problem.

## 2. WHY WE SHOULD NOT HOLD ANY BELIEFS AT ALL

### 2.1. Individual and Group Reasoning

That we humans fail at reasoning far more often than we would like to believe is now a well established result. The so called paradoxes of choice of Allais [30] and Ellsberg [31] have shown that our decision making is not rational, in the sense that it can not be described by a simple maximization of any utility function. We actually behave as if any probability value told to us were subject to the application of a weighting function before we use the value to make any decisions [32, 33]. Among other things, that means we overestimate the probability of unlike events. We see correlations in data when there is none simply because they seem to make sense [34] and fail to see them when they are surprising [35]. We interpret probabilistic information too badly, even in professional settings. To the point that we already have efforts to improve the understanding of basic probabilistic concepts among health professionals [36].

Since the first experiments with probabilistic choices, scientists have also explored other aspects of our reasoning with results that can be considered even more disturbing. We actually fail at very trivial logical problems [37, 38] and we provide different answers to the same question simply because it was framed differently [39]. We are not even aware of how badly we reason, we overestimate our chances to get correct answers [40], with the possible exception of the cases when we are actually really good at those specific questions, when underestimation of successes can happen [41]. These two effects combined can generate a dangerous combination of overconfidence among the incompetent and under-confidence among competent people and the consequences of this could be potentially disastrous. A curse of incompetence seems to plague us in some areas, and this can be particularly hard to notice for the cases where the competence needed to perform a task and the competence needed to know if we are competent are exactly the same ones [42]. And there is more. Sometimes, when presented with new information about a problem, we can become more certain of our answers even when we are making actual worse evaluations [43]. When reasoning in groups, we, at least, have mixed, good and bad, results. Our combined reasoning can indeed be more solid under some circumstances, a phenomenon known as the wisdom of the crowds [44]. But interaction and social pressures inside groups can actually make groups reason far more poorly than individuals [45, 46]. Social pressure can make us fail even in absurdly easy tasks almost nobody would get wrong, as shown, already in 1955, by the Ash experiments [47, 48].

The combination of all these experiments paints a grim scenario about our reasoning skills. When compared to all our accomplishments as a species, it should be clear that a part of the answer is missing. Of course, there is another side to this story. In particular, there are fundamental issues associated with the fact that information processing is a costly effort. More brain power requires more energy consumption. That means that theoretically optimal decision rules might not be ideal in the real world, where a compromise between effort and correctness might be the actual ideal [49]. So, if there are situations when we can

get a reasonably good answer most of the time with less effort, by using some simplified rule of thumb, that informational cheap answer might be the optimal solution in the evolutive problem, even if it is not the most accurate answer to the problem. By spending less energy thinking, it might have been in our best interest to not explore the problem further when a reasonable answer is available [50]. Indeed, there are quite simple heuristics that have been shown to be very effective [51, 52] at giving good, reliable answers. Our brains might just be looking for workable approximations, not the correct ones. Indeed, our reasoning seems to be clearly not well adjusted to logical problems at all, not without some amount of real previous training. But, while we are naturally incompetent at solving novel logical problems, when formally identical problems using situations we are familiar with are presented, we are actually quite good at solving them [53]. Even our probabilistic biases can actually be understood as reasoning that is similar to a Bayesian inference problem [25] where we correct the probability values we hear assuming they might be just an estimate [24]. It is in new problems, those far from our daily life, that we seem to fail even in the simplest questions. The worrisome thing to notice here is that all scientific advance happens in exactly new problems.

More recently, a very interesting answer to why we actually reason and, therefore, why we seem to have so many biases and do so many mistakes, have started to become clear. While reasoning can actually be used for looking for correct answers, this might not be its primary function. The ability to look for correct answers might even be just a luck accidental consequence of our actual need and use for reasoning. The more we learn, the more it seems we use reasoning mostly as a tool for winning arguments [5, 6]. And winning an argument does not necessarily means being correct. It just means we want to convince the listeners or be convinced by them. When there is already agreement, there would be no need for further exploration of the problem. That explains the very well-known confirmation bias [54] where people tend to look solely for information that confirms what they already think. Interestingly, this tendency to use reason mainly as a social tool to convince others is a result that appears to be universal, and not dependent on specific cultures [55]. That is, the real function of our reasoning abilities would be the social one. Our quest for truth, the function we would like to believe, might have nothing to do with our reasoning. Of course, winning arguments is not something we all can do. Someone will lose and, for the cohesion of the social group, the group should adopt the winning argument. It is about convincing or be convinced.

Sadly, our reasoning flaws are not restricted to everyday problems. We already have evidence that the public perception of scientists opinions can be quite wrong in several issues, such as global warming [56] or the disposal of nuclear waste [57]. Kahan et al have observed that even the way we perceive whether there is a scientific consensus on a specific question is influenced by which behaviors we find socially acceptable and which behaviors we believe are socially detrimental [7]. We know we value consistency in our beliefs to the point that the set of beliefs we adhere to can be more easily described as irrationally consistent [4]. We accept independent ideas as a package, put together for one sole purpose, that they support a conclusion we

wish to be true. We call these packages ideologies and we take sides and defend them. We often define who we are based on the ideologies we prefer. But, by doing so, we condemn ourselves, even the brightest ones among ourselves, to irrationality and stupidity. Indeed, there is now good evidence that our reasoning works in an identity-protective way [58]. Given the amount of already existent evidence on how our sides influence our reasoning, Kahan has proposed that information should always be presented in ways that are compatible with the values and positions of the listener, in order to allow the new information to be absorbed with less prejudice [59]. The cleverer we are, the better we can create arguments to defend our identities and the harder it may be to learn we are actually wrong! That suggests once more that, when facing problems that are associated with the ways we define ourselves, the function of our reasoning is not to find correct answers. Instead, its real function is to protect that identity, to protect our group and our side of the discussion. That includes, as we have discussed, forming arguments that support our views. Indeed, tests of how scientific literacy as well as numeracy correlate with beliefs about climate change among member of the public showed that those who scored higher on literacy and numeracy presented the strongest cases of cultural polarization [16]. Those increased skills actually allow people to defend their chosen positions better, instead of making them better at analyzing the literature and the evidence. The conclusion we can arrive from those observations is a scary one. Brilliant individuals who take sides might not be trustworthy about their opinions at all. Their arguments are probably worth listening to and analyzing. But their final conclusions must be seen as irrelevant, at least when they show certainty.

## 2.2. Normative Reasoning

Certainty (or the desire for certainty) is indeed a central part of the problem. We should want our reasoning to help us to arrive at the right answers, we assume that is what reasoning does, instead of simply being a tool to advance the goals of our own specific social group. That seems fundamental. But, as commitment to one idea seems to make all of us reason poorly, it makes sense to ask if we should not just get rid of such commitments. Therefore, we must now ask if, from a normative, logical point of view, such commitments even make sense. We are so used to choosing ideas and defending them that we do not even question if we should keep doing it. We do not ask if there are better alternatives, we have never questioned if the concept of taking sides might not be just plainly wrong. But that is exactly what we need to do.

Defending an idea would make sense if we knew it to be the truth. So, the first thing to consider is if, when we claim to believe in something, we are actually saying we have concluded that something is true. There is, of course, a weaker meaning to the word believe, as when we say “I believe it will rain tomorrow.” In this case, we are not really claiming knowledge, we are just making a statement that can be loosely described as probabilistic, despite the fact that no probability was mentioned. There is clear uncertainty in the phrase. As such, this is not the meaning we are interested about, since our social group will not suffer and our identity will not be threatened at all if it turns out we were wrong. We usually do not really take sides on those questions and we

easily accept those beliefs can be wrong. It is when believing is used as if it meant knowing that we can be in trouble. Therefore, that is the meaning we must investigate.

The problem of what knowing means is actually a very old philosophical question. Despite several attempts at defining the meaning of knowing, no clear answer has ever been reached. Apparently reasonable answers, such as the concept of justified true belief, are known to have flaws and exception cases that make them not a real correct answer [60]. Deductive logic exists at least since Aristotle, allowing us to make proofs where no doubt exist. But only as long as we do not doubt the premises. Newer, more complete, and better versions of deductive logic have appeared but they all depend on what you accept initially as true. There is always the need to already believe some initial concepts as true in order to use the logical tools, just as we need postulates to start any kind of mathematical thought. First truths seem to be unreachable by any deductive logic, as far as we know today. We can prove conclusions, we can not prove premises, not unless we make more primitive premises.

The form of reasoning we have that allows us to accept propositions about the real world is Induction. As far as we know, it is as old as Deduction, but it has never led to certainties. Indeed, it is based, among other things, on the assumption that the patterns we have observed in the past will be repeated in the future. And that assumption will not be true in every situation [61, 62]. While it is reasonable to expect Gravity to keep working as it has always done, there are actually areas of knowledge where expecting the future to strongly resemble the past sounds as an absurd, such as technology, the behavior of financial markets where completely new products are always been created, or the evolution of our own societies and economies as more and more types of work can be replaced by machines. Induction can still be useful in these cases, as we try to understand some underlying laws that might be more basic than the superficial data. But serious doubt will remain.

Indeed, from what we have learned so far about its uses, inductive reasoning can be a powerful and very useful tool, but it means that the desire for knowing the truth must be abandoned. In its place, a probabilistic approach, where we only estimate the likelihood that a given statement is true, is the best scenario we can expect. How to estimate those probabilities and change them as we learn more can be a very difficult or even an impossible task in real problems, but, at least in theory, we have the prescription for how it must be done. More than that, it is interesting to notice that different assumptions on the same problem of how induction should be performed lead to the same rule for changing probabilities as we learn more about the world [63–65]. We just need to apply a deceptively simple probability rule known as the Bayes theorem. However, despite Bayes Theorem apparent simplicity, using it in the context of real problems is almost never easy. Far from it, we still are plagued with a large number of unsolved problems, some of them related to the specification of initial knowledge, others about the models about the world that are required to use the theorem for an actual complete description of the possible theories about the world.

One approach equivalent to the bayesian method was proposed 50 years ago by Solomonoff [66]. This approach

sheds some very interesting light on the full requirements of the bayesian induction. For a full analysis of any problem, probabilistic induction would require both infinite information processing abilities to generate all possible explanations that are compatible with the data we have, as well infinite memory to deal with all those explanations and use them to calculate its predictions. While those requirements do not make Bayesian methods wrong nor useless, they do mean that even when estimating probabilities, the best we can hope for is an approximation to the actual probability values. Uncertainty remains even about the “real” probability values. Better methods to estimate uncertainty can always be created.

The state of the area and the technical difficulties of actually using those methods are interesting problems, but beyond the scope of this paper. What matters here is that, given our current state of knowledge in deductive and inductive logics, probabilistic beliefs about the world actually make normative sense but certainties do not. While we sometimes use the word belief in a probabilistic sense that fits very well with inductive considerations (I believe I will get that job), there are also times when we use the word belief as a weird claim of unjustified certainty. When we say someone believes in some set of religious claims or in an ideology, for example, we are often saying that person thinks she actually knows those sets of claims are actually true. Some people would agree that certainty is not really warranted. Instead, the set of beliefs is their best guess, a guess that person is willing to defend. But willingness to defend an idea can be far stronger than just thinking the idea is very probable. It is one thing to behave like that in moral issues. It is a completely different problem to do the same on descriptions of the world. The world is the way it is, regardless of what we think or feel about it. To make strong statements about the world, when those statements are simply not warranted should be considered a serious form of irrationality. And a damaging one, due to the way we reason.

### 2.3. Psychology Experiments and Logic Combined

Let's see what we have. Actively believing and defending an idea brings a series of negative consequences to our ability to pursue the truth. Experiments show that we tend to accept concepts simply because they support the conclusions we wish were true. At the same time, we submit the ideas that we disagree to an actual rational analysis, looking for possible errors. We value the ideas we believe in ways that resemble an endowment effect [67], where we value an object more if it is our possession. While the experiments conducted so far were focused mostly on political views and popular opinions about scientific issues, their results seem consistent enough.

The special case of moral beliefs could be treated apart. It is theoretically possible that a person could know her moral preferences. But we have a serious problem when we consider beliefs about the real world. The choice of one idea over another is never justified from a rational point of view, at least not to the point of certainty. In situations of limited resources, choosing one idea because we have no time to consider all other alternatives is

a perfectly rational choice. For example, choosing one treatment for a patient is necessary, but that does not mean that we decide all other possible illnesses must be wrong. We just decide the best current course of action, based on the chances we estimate for each disease and how serious their consequences might be. That does mean that the correct thing to do is not taking sides and defend one specific diagnosis, the decisions we make today must be easy to change, as we learn more. Any physician who would choose one initial diagnosis and then simply defend that choice and never change it should be considered incompetent, damaging and dealt with accordingly. Things are not different for any other area of knowledge. Decision making is fundamental and unavoidable. But we do have the tools to deal with decision problems involving uncertainty. On the other hand, there is nothing wrong with making probabilistic statements. Claiming one idea is far more probable than its alternative is logically correct.

This analysis assumes the debate is happening between knowledgeable, honest people. When debating people who refuses to even learn about the evidence that already exists, taking sides might be almost unavoidable. But that should not mean defending one idea as true, even though things might look that way. It can be just a very basic defense of the need for rational thinking as opposed to the very unreliable human thinking. In such cases, the decision process might even tell us the best thing to do can be to take the side of competence and clear logical thinking against cherry-picking tactics and circular reasoning. It can also be true that the best tactics might be simply to ignore the irrational debaters. Correct decision making includes estimating the outcomes of each possible action, taking into consideration the different probabilities for each idea. And that could mean the best momentary decision is, at a certain point, to temporarily act to defend or oppose one idea. As long as we do not believe in it as a certainty, we are logically correct.

We should, of course, never forget that our brains are limited machines. Some ideas are so improbable that we can claim them wrong in any practical terms based on the evidence we have collected so far (we will see an example in Section 4). In those cases, it will often make sense to disregard concepts that we have sufficiently strong evidence to discard as highly improbable. But this is different from claiming our ideas to be correct. Or even to claim certainty that the ideas we have discarded are wrong; it is always possible that new evidence might force us to change our opinions. We, scientists, claim to actually know that and that ideas might return and that we always should be open to changing our minds when faced with that new evidence. However, experiments show that by belonging to any group where an idea is defended we make ourselves vulnerable to our human shortcomings. Observations of real scientists working actually agree that we are not different from the picture that emerges from the experiments. We are far too stubborn and we cling to the ideas we defend despite evidence, as we will review in Section 4. We are not different and, as any humans, we stop being truth-chasers and become idea-defenders. That actually makes our individual opinions unreliable when we take sides. Anyone that is out there defending an idea should actually be considered a non-reliable source of information. That just means that we

should not believe what we say just because we say it, that the arguments and data are the only real thing that matters. We know that, but it is probably far more important than we might have thought.

### 3. HOW THE WAYS WE HANDLE BELIEFS AND HOW WE COMMUNICATE CAN INFLUENCE EXTREMISM

Unfortunately, our desire to have one side to defend does not only makes us unreliable. It can also be a fundamental factor in the appearance and spread of extremist views. Extreme opinions can be observed in many different issues and they are not always a problem. While a very strong opinion is sometimes warranted (for example, about not leaving the 20th floor by the window), there are several other circumstances where the same strength of opinion makes no sense. Even if we could ignore the appalling influence of extremist opinions on terrorism and wars, we would still be left with a series of problematic consequences associated with too strong beliefs. As examples, we know that some types of extreme opinions can cause serious problems in the democratic debates between the opposing political parties [68] or in how minorities are perceived [69]. One way that we have used to understand how extremism evolves is using the tools provided by opinion dynamics [8–13]. Opinion dynamics studies how opinions, extreme or not, spread through a society of artificial agents. As such, it can help us understand the factors that contribute to the decision of holding strong beliefs. In issues where there is actually no clear known answer, understanding the dynamics of how ideas spread and become stronger might, in principle, help us prevent problems and maybe even avoid the loss of human lives.

Previous works in the area have studied different aspects of the question, always based on one specific *ad-hoc* model chosen by the authors of each paper. In purely discrete models, although there is no strength of opinion, it is still possible to introduce inflexible agents [70]. It is also possible to include a larger than two number of opinions to better represent not only the choices, but how strong they are [71, 72]. However, strength of opinion is either not measured in those cases, or is limited to very few possible values. It is worthwhile to have a continuous variable that will represent the strength of the opinion to better define who is an extremist. While discrete, Ising-like models [73] can be very helpful to study how inflexible agents can influence the opinion of the rest of the population, they leave questions answered. Therefore, there is a large number of studies in the area that use as basis the bounded confidence (BC) models and their variations [12, 74]. One of the questions that have been investigated with the use of BC models is how agents with extreme opinions, represented by values that are close to the limits of the possible opinion values, can influence the rest of the society [75, 76]. These studies included studying the effects on the spread of extremism from factors such as the type of the network [77, 78], the uncertainty of each agent [79], the influence of mass media [80], and the number of contacts between individuals [81]. Consequences of extremism have also been studied in this

context, both negative, such as escalation in intergroup conflict [82] as well as possible positive effects such as the possibility that extremism might help maintain pluralism [83].

BC models have a central limitation. In most implementations, an interaction either brings the agents to an intermediary opinion or their opinions do not change at all. That is, unless a mechanism for the strengthening of opinions is added by hand, opinions just don't get stronger than the initially existing ones. Agents with already extreme opinions have to be included in the initial conditions. On the other hand, in the continuous opinions and discrete action (CODA) model [13, 84], it is possible to start only with agents who have moderate initial opinions and, due to the interactions and local reinforcement, those agents can and usually do end up with very strong views. This was accomplished by combining the notion of choice from discrete models with a continuous subjective probability each agent holds about the possibility that each given choice might be the best one. By assigning a fixed probability that each neighbor might have chosen the best possible choice, a trivial application of the Bayes Theorem yields a simple additive model where extremism arises naturally [85]. The general idea of using Bayesian rules as basis for Opinion Dynamics has been investigated and it was observed that the generated models are actually a good description of the real observed behavior [86]. And, while there is indeed strong evidence that people do not update their opinions as strongly as they should, a bias known as conservatism bias [87], several of our biases can actually be explained by more detailed Bayesian reasoning, if the possibility of error in the information source is included [24].

By separating the internal opinion from what other neighbors observe, CODA model allows a better description of the inference process that underlies the formation of opinions. Indeed, the way CODA was built can be used as a general framework to produce different models [84]. These new models can be obtained simply by changing the assumptions on how agents communicate, think or use the information available to them. Traditional discrete models can be obtained as a limit case of the situation where an agent considers its own influence on the opinion of its neighbors [88]. Here the separation between internal opinion and communication is a central aspect that can be included in many different models [89]. If the communication happens not in terms of choices but in terms of a continuous average estimate of a parameter, the BC model can also be obtained as an approximation to a Bayesian rule model [90]. Extensions of the CODA model were proposed to study the emergence of inflexibles [91], the effects of several agents debating in groups [92], as well as the effect that a lack of trust between the agents can have on social agreement [93] and the effects of the motivation of the agents [94]. Interestingly, for a completely different application, unrelated to social problems, the same ideas can even be used to obtain simple behaviors of purely physical systems, with properties such as inertia and the harmonic oscillator behavior arising from continuous time extension of Bayesian opinion models [95]. Here I will use the fact that the Bayesian framework behind CODA can be trivially applied to describe how agents think and behave to study differences in communication and mental models. This makes it possible to

explore the circumstances where extremism is more likely to appear and what makes opinions become too strong.

We know that extremism usually does not develop from the interactions in BC models, while it is prevalent in the CODA model. In BC models the agents observe a continuous value as the opinion of the other agent. Therefore, a possible explanation for the lack of increase in extremism could be associated with this different form of the communication. In the BC case, we only communicate continuous values, signaling doubt about the discrete choices where only the supported option is shown. Perhaps the reason why extremism is much stronger in CODA might be because of the availability of doubts. In order to check this possibility, a version of the CODA model where the actual probability assigned to the two options is observed instead of the preferred choice will be introduced.

On the other hand, a subtle but important difference between CODA and BC models exist in the assumptions agents make. In BC models, agents aim to find a continuous value. It is reasonable to think they would be satisfied with an intermediary value and most of them quite often do. In CODA, the question is changed to which of two options is the best one. That means that there is a strong underlying assumption that a best answer exists. However, the ideal quest for finding the best answer might not be well described this way. It is perfectly reasonable to assume, in many cases, maybe in most cases, that the optimal choice is not one or the other, but a mixture of both, with proper weights. Examples are many. When choosing the best food among two options, the optimal solution might be take 75% of the first one and 25% of the second. In Politics, the optimal policy, when there are two competing ideologies, could also be to use 50% of the ideas of each one of them. And so on. From a normative point of view, a mixture is also a more complete model. This happens because the case where one of the two options is actually the best one is just a special case, a possible answer, when the optimal solution goes to 0% or 100%. Such compromise is not included in the CODA model nor in the variation I just mentioned. But it is central in BC models. In order to investigate this effect, a second variation of CODA is proposed here, where the agents look for the correct proportion, instead of the correct option.

### 3.1. The Models

While the first obvious difference between the BC and CODA models is what agents observe from the other agents, the assumptions about the mental model agents have of the world are also subtly different. BC models assume the agents are looking for a correct value inside a range, typically between 0 and 1. In CODA, while the agents have a probabilistic estimate of the correctness of a choice (therefore, also a number between 0 and 1), they hold the belief that only one of the options can be true. In particular, by assigning a probability  $p$  to  $A$  being the best choice and  $1-p$  to the alternative choice  $B$ , one is ignoring the possibility that the ideal answer could be an ideal proportion  $f$  around 0.5. Instead,  $p$  only represents the chance that  $A$  is optimal. It makes sense to investigate what happens if agents consider the optimal could be obtained as a proportion  $f$  of  $A$  and a proportion  $1-f$  of  $B$ . While this might sound similar to the way CODA works, it is a complete different model. From a mathematical modeling

point of view, the difference could not be clearer. Instead of a probability  $p$  assigned to one specific result, the case where all mixtures are acceptable requires a continuous distribution  $g(f)$  over the space of proportions  $f$ , where  $0 \leq f \leq 1$ . That means that a new terminology is needed. From now on, I will refer to agents who look for a single isolated answer as certainty wishers or, more simply, wishers. Those who accept that the best current answer can be a mixture of both choices will be called mixers.

Of course, not only the mental model is important. If communication is continuous (and honest, as we will assume all the time here), wishers will provide the probability that  $A$  is the best choice, while mixers would give us their best estimate  $f_i$  of how much of  $A$  should be included in the optimal answer. But communication can also happens in a discrete way. In this case, while using the same rule (if  $p$  or  $f$  are larger or smaller than 0.5), wishers are picking the option they believe more likely to be the best one, while mixers would be simply indicating which option they expect would be present more strongly in the optimal answer. To explore the difference in communication and mental expectations, we must therefore investigate how extremism emerges in all combinations, represented by the four cases below:

	Certainty wishers	Mixers
Discrete observation	CODA	New model 2
Continuous observation	New model 1	Bounded Confidence (BC)

Notice that every model in this table will involve continuous variables. It is what is observed (observation) and what is expected (choice) that can be either discrete or continuous. The literature already shows very clearly the effects of CODA and BC models on extremism. In CODA, extremism appears from local interactions even when only moderate opinions existed at first. In BC models, the same is not true. With no extremists in the initial conditions, no extreme opinions emerge; the general tendency in the population is that agents become more moderate in time. In order to obtain a complete picture, we must learn how extremism evolves in the two new cases. To make the comparison natural, a generative framework for different models is necessary. As we have seen, extensions of the CODA model did generate such a framework [84] and the ideas in the framework have been used to obtain a version of the BC models that agree qualitatively with its results [90]. Therefore, we can simply refer to the known BC results from the literature as if they had been generated by the same framework and investigate here the consequences of the two new variations.

That framework can be summarized as the following recipe. First, represent the subject as a variable  $x$  that agents want to estimate. Here,  $x$  will have different meanings, since for wishers it will be the probability  $p$  that  $A$  is the best option while for mixers  $x$  will be the proportion that one should include  $A$  in the final answer. Each agent  $i$  has a subjective opinion about  $x$ , represented by a probability distribution  $f_i(x)$  over all possible values of  $x$ . Communication happens as agent  $i$  manifests its opinion, usually not as the distribution  $f_i$ , but as a functional  $C_i[f]$  of  $f$  (for

example, the expected value of  $x$  given by  $f$ ). Of course,  $C_i[f]$  can be a discrete or continuous variable or even a vector, if the communication involves several different pieces of information. In CODA,  $C_j$  is obtained by the option that is most likely to be true, that is  $C_j[p] = +1$ , if  $p > 0.5$  (representing the observation that  $j$  prefers  $A$ ) or  $C_j[p] = -1$  when  $p < 0.5$ . In the new models, wishers will simply state their probabilities,  $C_j[p] = p$ , while mixers, when using discrete communication, will use a rule similar to CODA. The only difference is that instead of  $C_j[p]$ , they will apply the condition to the expected value of the frequency, that is, the functional will be defined by  $C_j[E[f]]$ . Finally, as agent  $j$  observes the opinion of  $i$  from  $C_i[f]$ , it needs a likelihood for the value it observes. That is, for every possible value of  $x$ , we need a model that states how likely it would be to observe  $C_i[f]$ , that is,  $p(C_j|x)$ . With this, Bayes theorem tells us how agent  $j$  should update its subjective opinion  $f(x)$  to a new distribution  $f(x|C_j)$ .

Implementation of the models in this paper followed this steps:

- All agents start with moderate opinions. That means their observed or preferred choice can change with just one observation of a disagreeing neighbor.
- At each iteration, one agent is randomly drawn as well as one of its neighbors. The neighborhood is obtained from a 2-d lattice with periodic boundary conditions, except in the cases where a complete graph was used.
- The first agent observes what the second agent communicates; only the first agent updates its opinion following one of the update rules.
- Time is measured in number of average interactions per agent.

### 3.2. Wishers with Continuous Observations (New Model 1)

When we have wishers who make statements in the form of their continuous choice, we have the case where each agent  $i$  believes there is one true optimal choice,  $A$  or  $B$ . And, instead of observing only the favorite choice of agent  $j$ , agent  $j$  provides as information its own probability  $p_j$  that  $A$  is the best choice. Communication, therefore is a continuous value, just as in the BC models. The difference is that, while any probability value can be a representation of the agent point of view, there is the underlying assumption that only  $p = 0$  or  $p = 1$  (and no other value) can be the correct answers. The simplest model for this case is to assume a simple likelihood for the value  $p_j$  agent  $j$  communicates, if  $A$  is the best choice (as there are only two choices, we only need to worry about one). For the bounded continuous variable, a natural choice is to use a Beta distribution  $Be(p_j|\alpha, \beta, A)$  as the likelihood that agent  $j$  will express  $p_j$  as his opinion when  $A$  is the best choice, where

$$Be(p_j|\alpha, \beta, A) = \frac{1}{B(\alpha, \beta)} p_j^{\alpha-1} (1-p_j)^{\beta-1}$$

and where  $B(\alpha, \beta)$  is obtained from Gamma functions by

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Here,  $\alpha$  and  $\beta$  are the usual parameters that define the shape of the Beta function. Since we assume that agent  $j$  has some information about  $A$  and that each agent is more likely to be right than wrong, we need to choose a shape for  $Be(p_j|\alpha, \beta, A)$  that favors larger values of  $p_j$  (since, in this case, we are assuming that  $A$  is true). That corresponds to  $\alpha > \beta$ . The difference between  $\alpha$  and  $\beta$  dictates how much the opinion of agent  $j$  will influence the opinion of agent  $i$ . For symmetry purposes, we will assume that, if we assume  $B$  is the best choice, the likelihood will be given by  $Be(p_j|\beta, \alpha, B)$ .

This choice makes the application of Bayes Theorem quite simple. We have, for the odds-ratio (to avoid the normalization term),

$$\frac{p_i(t+1)}{1-p_i(t+1)} = \frac{p_i Be(p_j|\alpha, \beta, A)}{(1-p_i) Be(p_j|\beta, \alpha, B)} = \frac{p_i}{1-p_i} \frac{p_j^{\alpha-\beta}}{(1-p_j)^{\alpha-\beta}} \quad (1)$$

Applying Bayes Theorem and using the same transformation to log-odds  $v_i$  as in the CODA model

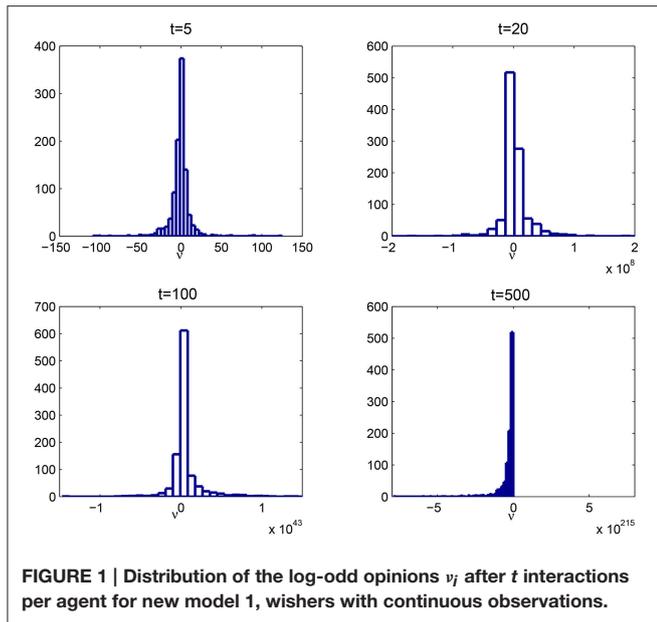
$$v_i = \ln\left(\frac{p_i}{1-p_i}\right),$$

we obtain the very simple dynamics

$$v_i(t+1) = v_i(t) + (\alpha - \beta)v_j. \quad (2)$$

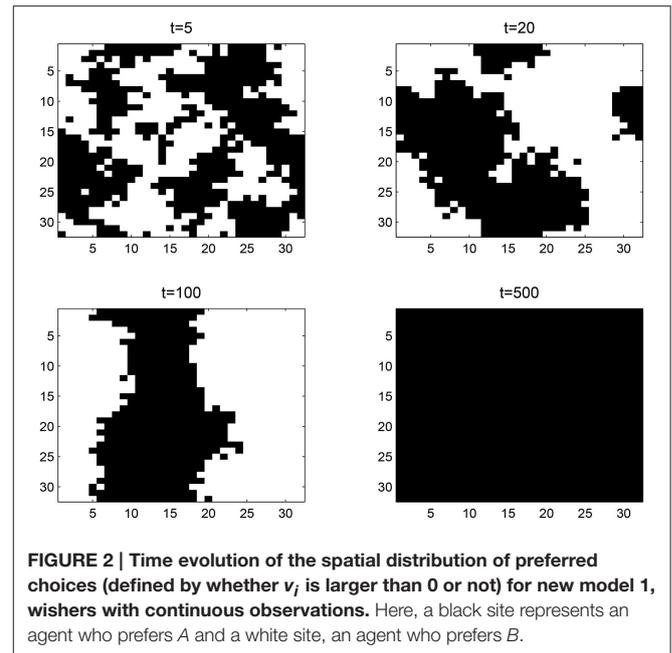
Equation (2) is similar to the CODA dynamics (where we would have  $v_i(t+1) = v_i(t) + \delta$ , with a fixed  $\delta$ ), except that now, at each step, instead of adding a term that is constant in size, we add a term that is proportional to the log-odds  $v_j$  of the opinion of the neighbor  $j$ . This means that the system should also tend to the extreme values, in opposition to the BC models where verbalization was also continuous but the agents were mixers. Here, as  $v_j$  increases, so will the change in each step, making the long term increase in  $v_i$  no longer linear. Instead, we have an exponential growth in  $v_i$ . Simulations were prepared to confirm the full effect of this change when each agent only interacts with a fixed neighborhood. Square lattices with periodic boundary conditions and von Neumann neighborhood were used and the state of the system observed after different average number of interactions  $t$  per agent. During the initial stages, we observe a behavior very similar to that of the CODA model, with a clear appearance of domains for both possible choices. However, instead of freezing, those domains keep changing and expanding and, eventually, one of the options emerge as victorious and the system arrives at a consensus.

**Figure 1** shows histograms with the observed distributions for  $v_i$  for systems that ran for different amounts of time. The important thing to notice there is how the scale of typical values for  $v_i$  changes with time. In the original CODA model, the extremist peaks, after the same number of average interactions per agent (500), corresponded to  $v_i$  around 430 (that is around 500 steps away from changing opinions, with the size of the step determined by  $a = 70\%$ ). That value was already quite extreme since it corresponded to probabilities of around 10–300. Now, however, instead of peaks of  $v_i$  around  $-430$ , we



can see at **Figure 1** we reach  $v_i$  values in the order of  $10^{215}$ , corresponding to the exponential increase in  $v_i$  predicted by Equation (2). Unlike CODA, there is no visible peak close to those much larger values. But one has to notice that, while the larger box close to 0 is the most frequent one, that does not mean moderate values are common. Indeed, since a limited number of boxes has to be used, for larger values of  $t$ , the range of the “moderate” box goes from 0 to very large values of  $v$ . If we take  $t = 20$  as an example, while the largest but rare values reach  $2 \times 10^8$ , the standard deviation of the  $v_i$  is still  $3.4 \times 10^7$ . Even the appearance that the mean might be close to 0 is an illusion as, in this case, the actual mean value was  $2.1 \times 10^6$ , a hundredth of the largest value, but still very large. The system as a whole observes a surprisingly fast appearance of extremists. After as little as five average interactions per agents, we can have opinions so extreme that it would have taken 100 CODA model interactions to obtain. The lack of freezing in the domains is actually due to these ever increasing opinions. As one side domain randomly gets a much larger value of  $v_i$ , this strength is enough to move the border freely and eventually, one of the options wins.

As an illustration, **Figure 2** shows the spatial configuration of the choices for the agents as time progresses. The choices are defined as in CODA, that is, each site is colored black if  $v_i > 0$  and white if it is not. While agents now observe the probabilities instead of the choices, we still have that values of  $v$  larger than 0 correspond to probabilities that  $A$  is the best choice larger than 0.5. Therefore, preference is still well defined. We can see the formation of domains after  $t = 5$ , just as it would be observed in the original CODA model. However, while in CODA those domains would freeze and only have their edges smoothed, keeping their basic shape after that time, the exponential growth of  $v_i$  for wishers cause the walls to move and some of the domains to grow and dominate the others, until only one remains.



### 3.3. Mixers with Discrete Observations (New Model 2)

The last remaining model to study is the opposite case. Now, each agent only states its best bet on which of  $A$  or  $B$  should make a larger part of the optimal answer. Instead of trying to determine which option is the best, the agent is trying to find the correct mixture of both choices. That is, we can no longer talk about a probability  $p_i$  agent  $i$  assigns to  $A$  being the optimal choice (and  $1 - p_i$  for  $B$ ). Instead, we need to have a continuous probability distribution  $g_i(f)$  over the values  $0 \leq f \leq 1$  for each agent  $i$  and a rule for the agent to make its choice based on this distribution. The simplest rule is that the agent communicates its best guess, based on whether the expected value of  $f$  is larger or smaller than 0.5.

The agent tries to determine the optimal frequency  $f_o$ . Assuming that other agents have information on  $f_o$ , a reasonable and simple model is that agent  $i$  might be trying to learn the proportion of choices in favor of  $A$  and  $B$ , assuming that the chance a neighbor will show choice  $A$  to be  $f_o$ . In this case, the inference on  $f_o$  becomes the simple problem of estimating the chances of success of a Binomial random draw. Once again, the  $Be(f_j|\alpha, \beta)$  distribution is useful, except this time, it is the natural choice for the prior opinion  $g_i(f)$ , since it is the conjugate distribution to the Binomial likelihood. The values of  $\alpha_i$  and  $\beta_i$  now represent the belief of the agent  $i$ ; the average estimate is very easily obtained from  $\frac{\alpha_i}{\alpha_i + \beta_i}$ . That basically means the agent  $j$  states it thinks a larger proportion of  $A$  is preferable if  $\alpha_j > \beta_j$ , and a larger proportion of  $B$  corresponds to  $\alpha_j < \beta_j$ .

The update rule one obtains is also trivially simple. If  $i$  observes  $j$  has a preference for  $A$ , it alters its distribution by making

$$\alpha_i(t+1) = \alpha_i(t) + 1 \quad (3)$$

and it keeps  $\beta_i$  constant. If it observes  $j$  chooses  $B$ , then  $\alpha_i$  is unchanged and we have

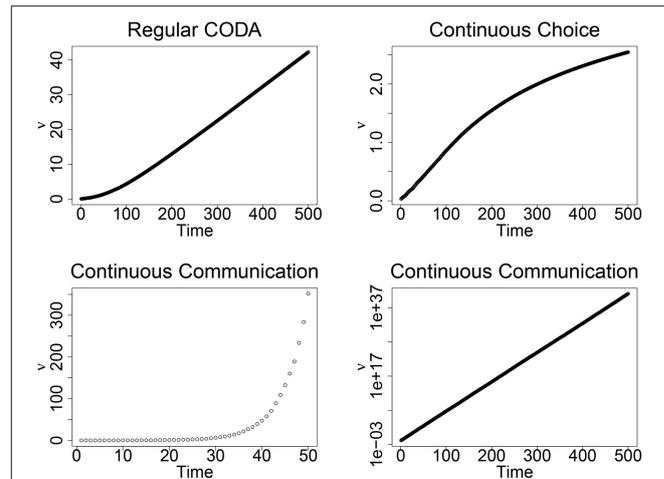
$$\beta_i(t+1) = \beta_i(t) + 1 \quad (4)$$

What is interesting and surprising here is that, while the values of the internal probability can be quite different, the qualitative rule for when the opinion of one agent changes from  $A$  to  $B$  is exactly the same as in the CODA model. Indeed, the above rule is equivalent to adding or subtracting one to  $\alpha - \beta$ , depending on whether  $A$  or  $B$  is observed. And the opinion changes when  $\alpha - \beta$  changes sign. That is the same rule of the CODA model. In CODA, the same dynamics was based on the log-odds  $\nu$ . As the actual dynamics of  $\nu$  in CODA can be renormalized to increments of size 1, both models will yield, given the same initial conditions and same random number generator, the exact same path in terms of configuration of choices on the lattice. Adoption of either choice will behave exactly the same way in both models. On the other hand, in the original CODA, probabilities change much faster. For example, if after 500 interactions, one specific agent had made 465 observations favoring  $A$  and 35 favoring  $B$ , we have seen that the effect of these 430 total steps for  $A$  meant the probability  $1 - p$  of  $B$  being the best option goes all the way down to 10–300. Here, if we start with  $\alpha \approx \beta \approx 1$  (a choice close to uncertainty), we will have, after the same 500 interactions, an average proportion for  $A$  of 466/502, something close to 93%. From this point, it would take the same number of influences in favor of  $B$  to change agent  $i$  decision toward  $B$  as it would have taken in the original CODA, that is 430 interactions favoring  $B$ . The opinion on which side should be more important is equally hard to change. But the very strong extremism we had observed there is gone, replaced by a reasonable estimate that the optimal proportion should still have some  $B$  (7%).

### 3.4. Communication or Assumptions: Comparing the Four Models

To compare the four cases easier, I have run a series of simulations where every agent can interact with all the others. At each interaction, agent  $i$  randomly draws an agent  $j$  to observe. With no local effects, all the systems will eventually reach consensus and we can focus only on how the strength of the opinions evolve. In all cases, the initial conditions included half of the agents showing some form of preference for each case and consensus emerges as one opinion randomly becomes prevalent. The magnitude of the difference in the evolution of the average opinions can be seen in the graphics shown in **Figure 3**. For the case where we have wishers and discrete communication, the strength of opinion increases so fast that it is simply impossible to show its evolution in the same graphic as the other two cases. If we get the other two cases in the same graphic in a way that their evolution can be observed, the wishers with discrete communication soon move out of range. Therefore, each case is shown in its own graphic. In order to understand the graphic better, note that odds of 1:10 translate to values of  $\nu = 2.3$ ; 1:100 corresponds to  $\nu = 4.6$ ; 1:1000 is  $\nu = 6.9$ , and so on.

We see that the difference between wishers and mixers is huge. Wishers opinions always become very strong, while the



**FIGURE 3 | Evolution of the average value of strength of the opinion, measured by  $\nu = \ln\left(\frac{p}{1-p}\right)$  as a function of time.** Time is measured in average interactions per agent and each point corresponds to the average value of 20 realizations of the problem. Due to the extreme difference in values, the evolution could not be visually shown in the same graphic. The upper left graphic corresponds to temporal evolution of a population of wishers with discrete communication, that is, the regular CODA model; the upper right, we have mixers with discrete communication (new model 2); while both graphics in the lower line show the case of wishers with continuous communication (new model 1), at left, just its evolution to 1/10 of the time the previous cases evolved. At lower right, the evolution of opinions for the new model 1 (wishers and continuous communication) is shown for the whole range of times with a logarithmic scale of the opinions.

same is not true for mixers. This suggests that our mental models are probably a crucial key to prevent extremism. Interestingly, for mixers, when they observe choices and not the internal probability, that still allows them to get opinions that become stronger with time. When they observe the continuous evaluations, we have the results of the Bounded Confidence models, even if we do a Bayesian version [90]. And, in this case, if no trust considerations are introduced, as it is the case of the other results here, the average opinion  $\nu$  will always stay around zero ( $p = 0.5$ ) for an initially random population, the exact opposite of extremism. Even more, each agent will move toward this value, becoming more moderate in time. That is, observing the amount of existing uncertainty allows mixers to estimate the actual initial proportion of support for each idea in the population. When they only state their preferences, random effects, even if small, do make one of those preferences obtain a little more support. As a consequence, the perceived optimal mixture shifts to favoring that one random choice. But we never get to real extreme values.

The exact opposite effect is observed for wishers. Just observing the choice of their neighbors makes wishers opinions already strong enough to be described as very extreme. While we might have expected that observing the uncertainty or doubt on the part of the other agents would lead them to some amount of doubt and, therefore, less extremism, that was not the case at all. The effect we observed for wishers can be described better in two steps. Observing doubt just once could mean a smaller change in

the probability the agent assigns to the best option (this can be easily adjusted by the model parameters). But in the long run, the opinions get stronger as constant reinforcement happens. Even if we start the system with smaller increases, as the agents observe their peers becoming more confident, their opinions start changing faster. It makes sense, as people with stronger opinions are expected to have a stronger influence. That is where the problem starts because there is no limit on the strength of these influences. At some point, the change in opinions will become as strong as it would have been if only the choice had been observed. From there, opinions will still keep increasing, making those changes larger and larger with each interaction. This influence actually grows exponentially and, given time, we can observe the absurd increase in certainty we see in **Figure 3**.

The message is clear. In these models, mixers opinions seem to never grow as strongly as the those of wishers. Indeed, either they will not grow at all, as in Bounded Confidence models, or they will approach certainty slowly (a mixture with frequencies of  $f = 1$  or  $0$ ) due to random effects causing one option to receive a greater amount of support. But even then, we only observe  $\nu$  growing logarithmically. On the other hand, wishers always observe a strong growth of the probability as shown by the linear or exponential growth of  $\nu$ , depending on how communication happens. Surprisingly, uncertainty in that communication can even make their opinions much stronger in the long run. And all of these results were obtained with models that do not include the fact that, as a person opinion becomes stronger, she will probably start disregarding opinions that disagree with her own.

Those results bring some remarkable evidence on how to deal with extreme opinions. They suggest that changing the way we communicate might not prevent extreme views at all. If we do not consider long term effects, it might even make things worse in the long run. We should notice that, in the cases studied here, the strengthening of the opinions happened only by social influence, no observations of the world were ever made. That means that there was no real reason for the system as a whole to go to such extremes as there was no real data confirming either possibility. The problem with unwarranted extreme opinions seems to be deeply connected with the way we think. It seems to be a consequence of our wish to find one truth, of our desire to believe and take the side of one option.

The message this exercise in Opinion Dynamics adds to the conclusions of this paper is clear. Not only our self-defining beliefs hinder our ability to think correctly and consider all evidence as well as not being supported by any kind of rational norm. This tendency to pick sides also seems to make our opinions absurdly stronger than they should be. And this tendency seems to be at the very heart of why extremism arises. The agents in these simulations were willing to change their opinions given enough influence. Even under these idealized conditions, the simple desire to have one option as the best answer already caused severe problems, making opinions absurdly and unwarrantedly strong. This reinforces the evidence we had, from cognitive psychology as well as both deductive and inductive logic, that we should not take sides. Indeed, the conclusion just became a little stronger: we should not even want to take sides. In the next Section, we leave the study of extremism

behind and I will inspect the consequences of our tendency to belief by discussing effects this bias causes on the ways we make Science.

## 4. SCIENTIFIC CHOICES, BELIEFS, AND PROBLEMS

### 4.1. Creating and Evaluating Theories

The naive description of the scientific enterprise says it is the job of Science to find the one truth about the world. Sometimes, scientists actually believe they have done just that. But this is the type of belief that would make them unreliable sources about the ideas they defend. Even when they recognize that we still are far from knowing it all, that our current ideas might be the best ones we have but are only tentative descriptions, scientists often expect that 1 day we will create that one winning theory, the Grand Unification that will explain everything. That might come to pass, of course. But given all we have seen so far, we should question how much of our human failures might be influencing the reliability of the Science we do today. And, if there are problems, what we can do to mitigate them.

The desire to have one winning theory, the one idea that will rule them all, as we have seen, ought to have consequences. When using inductive methods, it might actually be the case that a mixture of the available theories does a much better job than each one individual idea can do by itself. And, in every case, ideas are to be ranked by how probable they are, not discarded nor proved right. We have just seen in Section 3 that when we look for the one correct idea, our estimates can become much stronger than they should be. In some rare cases, when there is strong evidence that no idea does a perfect work, we actually have seen the survival and acceptance of a mixture of even incompatible theories. Such is the case of Quantum Mechanics and General Relativity, as each of them describes a different set of observations with what is considered remarkable accuracy and fail at explaining the phenomena the other theory describes. But this reasonably peaceful cohabitation of the minds of the researchers seems to be a rare case, caused by the utter failure of each theory to provide decent predictions for the set of experiments where the other one meets its greatest successes. It would be good to look at this example and think that the scientists are working as they should, but they actually never had a choice. Physicists still look for the one theory that will unify them both, as they should do. But the amazing success of physical theories is both a very lucky case that allows physicists to ignore most considerations about mixture of ideas and probabilistic induction and also a curse that has prevented them from learning those lessons.

We have always had idealized prescriptions about how the nature and demands of the scientific work and how scientists should behave. Standards are extremely important, of course. Given the enormous success of the scientific enterprise, it used to make sense to assume that individual scientists did not deviate much from the norms of the area. Some of them certainly did, but errors and deceptions ought to be the exception. However, since the first observations of how scientists actually work with their ideas, it has become very clear that we do not follow

the norms we believed we obeyed. That happens even among people working in some of our greatest technical feats, such as the Apollo missions [96]. The departure of the idealized norm seemed to be so prevalent that the personnel in those missions considered normal that, under the right circumstances, people they respected intellectually could hold dogmatic positions and not be convinced of their erroneous views even by data.

Feyerabend [97] observed, from historical cases as well as his own interactions with scientists, that our typical behavior did not match the ideas he had been told were the right way to do Science. His observations were quite influential and have even led people to think the whole reliability of the scientific enterprise could be seriously compromised. A shortened version of the rules he assumed we obeyed are the tentative norms that scientists should not use *ad-hoc* hypotheses, nor try ideas that contradict reliable experimental results, nor use hypotheses that diminished the content of other empirically adequate hypotheses, neither propose inconsistent hypothesis. And yet, he found evidence that all these norms were broken on a daily basis. More importantly, this breaking had been instrumental in many important advances. If those rules were indeed the source of Science credibility, it looked as if there was no reason to place more trust in scientific knowledge than in any other form of knowledge. Luckily that is not the case. One of his main conclusions, that all ideas are worth pursuing, is actually a prescription of a number of inductive methods. Instead of limiting the ideas that should be tested, inductive methods actually support the exact concept that every idea must be kept, tested as a possibility, and ranked. Of course, the fact that we are limited beings plays an important role. We can not know every possible idea and demonstration. For us, the need to test all ideas actually means that we must work to get more and more ideas created. Most of them will be utter failures, with probabilities so close to zero that they will be indeed useless. But we only know that after we come up with them and verify how well they describe the world.

If we want real probabilities and not only odd ratios between two ideas, Bayesian methods actually require, for the ideal situation, that we should consider all possible explanations [98]. Similarly, Solomonoff method for induction [66] requires a theoretical computer that would be able to generate every possible program, up to infinite length, and average over all programs that produce as output the data we already know. While both approaches are obviously impossible, they provide the ideal situation. As we work to get closer to the ideal, we have to realize that we must indeed catalog all possible ideas, including those very same hypothesis Feyerabend had thought to be forbidden. Obviously, hypotheses that contain logical inconsistencies are not acceptable *per se*. But even those can be useful as a middle step, if they inspire us to later obtain a consistent set of hypotheses, by correcting initial problems and inconsistencies. On the other hand, this prescription does not lead to any kind of relativism, contrary to what Feyerabend's followers would like to believe. We just need them all to verify which ideas actually match our observations better.

So, how close can we get to proving a theory “wrong”? The answer can be illustrated by the classical observation of

the advance of Mercury's perihelion and why this experiment was seen as a “proof” that Newtonian Mechanics had to be replaced by General Relativity. Here, we are talking about two theories that agree incredibly well with experiments. But they still disagree in some predictions and, as those cases started to be tested, there was no longer any doubts in the minds of the physicists about which theory was superior. The subtle differences in where Mercury was when it got closer to the Sun and how that varied over centuries was one of the first very strong “proofs.” When the two theories started competing, it was known that when we estimate every influence of the other planets on Mercury's orbit using Newtonian Mechanics, its perihelion should move  $5557.18 \pm 0.85$  s of arc per century. However, the actual observed displacement was  $5599.74 \pm 0.41$  [99]. On the other hand, General Relativity introduced a correction of 43.03 to the Newtonian prediction, bringing the actual estimate of the difference between observed measurement and prediction from  $42.56 \pm 0.94$  to  $0.47 \pm 0.94$  (the errors have been estimated by me simply adding the variances in Clemence article and are, therefore, not much more than an educated guess). We have the Newtonian prediction at a distance of 45.28 times the standard deviation from the observation and the General Relativity at a distance of 0.4997 standard deviations. Therefore, if we assume the observed variations have a Normal distribution, the likelihood ratio between the two theories would be

$$L = \frac{e^{-\frac{-0.4997^2}{2}}}{e^{-\frac{-45.28^2}{2}}} = \frac{0.8826}{7.19 \times 10^{-446}} = 1.2 \times 10^{445}. \quad (5)$$

In other words, if you started by believing General Relativity had just one in a billion chance to be a better description (1 in  $10^9$ ), you would end concluding it was actually better by a chance around 1 in  $10^{445-9} = 10^{436}$ .

It is very important to notice that this calculation is heavily dependent on the assumption that the errors would follow a Normal distribution, which is a rather thin-tailed distribution. If we do change the distribution of the experimental errors to a distribution with much larger tails, as for example, an arbitrary t-distribution with 10 degrees of freedom, things change dramatically. In this case, the likelihood ratio would be now

$$L = \frac{\left(1 + \frac{0.4997^2}{10}\right)^{-11/2}}{\left(1 + \frac{45.28^2}{10}\right)^{-11/2}} = \frac{0.873}{1.87 \times 10^{-13}} = 4.65 \times 10^{12}. \quad (6)$$

Now, an initial opinion of one in a billion against General Relativity would be transformed in something around 1 in 1000 in favor of it. However, the change in opinion is still quite impressive. And this was just one single experiment. Since then, many other experiments have confirmed the predictions of General Relativity. It is not difficult to see that today the probability of Newtonian Mechanics being correct is basically zero for all practical purposes. Results in Physics have been so discriminating between competing theories in most cases that physicists can work perfectly well if they behave as if the experiments were proofs.

The different likelihood ratios shows that the results can indeed depend heavily on the assumptions we make, even when the evidence is very clear. Those assumptions can complicate any probability estimates; dealing with this is beyond our objectives in this article. Some hypotheses about the errors will be more probable, others less so. We just need to observe that a theory can really become extremely unlikely given data. Technically, that does not mean Newtonian Mechanics probability is now a mathematical zero. But we can say that it is zero; saying it is not is a waste of effort, unless you are making a mathematical point. And yet, this exercise was not the complete problem, it was only a comparison between two theories. And we should include all possible alternatives, including unorthodox versions of the two theories we tested. The vast majority of other alternatives describe the data so much worse than the “wrong” Newtonian Mechanics that we do not even bother to enumerate them. Physicists have been quite lucky in the way data can discriminate between competing ideas. When that happens, concerns about our limited reasoning power are far more relevant than the prescription that all ideas should be kept.

On the other hand, Feyerabend correct conclusion that we must keep generating new ideas has some interesting consequences that people seldom consider. One of them is the actual strong usefulness of the existence of purely theoretical scientists. They are already a major part of the culture in the physical sciences and they also exist in some areas of Biology, such as Evolution. Having people dedicated solely to coming up with new better ideas means we can generate those needed ideas faster, meaning those fields are bound to benefit from their specialized knowledge. There are also incidental and very strong benefits from having a large number of people who do not make experiments or generate any data. And that is the simple fact that the people who will check their ideas are not the same as the proponents of the theories, allowing for a much less personally-biased analysis. The current tendency to demand that people produce the ideas and the data that supports those ideas is bound to result into biased choices of sets of data. And coupled with statistical techniques created to choose instead of ranking ideas, we already have whole areas where there is good reason to consider the conclusions of published papers to be unreliable. Someone proposing the idea and providing the tests for it is actually a very strong moral hazard and we should build our areas in a way that makes this hazard less probable. Having pure theorists is not only a good idea about dividing the work and generating needed new ideas faster. It is also a good idea because it avoids ethical problems in data analysis, as well as our tendency to defend our own ideas. Of course, theorists require that we know enough about a specific problem that we can start generating several theories and this might not be the case yet for some areas of knowledge.

There is another cognitive advantage of having theorists. A recent study on judicial rulings has shown that legal training and experience seems to provide some resistance to identity-protective cognition. Unfortunately, that resistance only applied to the situations about legal reasoning for which the judges had been trained [100]. This suggests that it is possible to train people to make decisions that are closer to neutrality, at least on matters

that are clear, logical applications of already known ideas. A central aspect of this training effect might come as a consequence that legal training is indeed based on a series of well defined rules. People learn to apply those rules, avoiding normal biases. This seems no different from the fact that, while we naturally fail miserably at simple logical problems, we are very good at solving them when they happen in an everyday situation and we can train ourselves to become good in normal logical problems. There are also several experiments that show that professionals actually fail at their own tasks and at evaluating their own rate of success when they do not have access to fact-correcting strategies. Simple training alone is not enough. We need training, we need to learn when we are making mistakes and, whenever possible, we need to making thinking tools to correct our human nature. Mathematics and logic play exactly that role, a similar role to the legal training that judges receive and allow them to overcome their biases. A mathematical or logical demonstration is not subject to our biases, anyone with the proper training will find mistakes that are not too subtle or hard. It is not mathematics that is unreasonably effective [101], it is the standard for comparison, our own reasoning, that is far poorer than we would like to admit. The importance of having logically sound reasoning in every field of knowledge makes the existence of classes of scientists specialized in applying mathematical tools to the exploration of each area another strong defense against human failings.

The need for the creation and acceptance of new ideas as part of the theoretical repertoire of an area has at least another very important consequence. While individuals might have difficulty accepting ideas that contradict their favorite theories, we must make it sure that this difficulty must not happen in any area as a whole. Cultural norms can help offset the bad behavior of the scientists [102]. In every field, new ideas should be always welcome or, at the very least, not forbidden, even though most of them might prove to be very improbable. Unfortunately, the needed openness to new theories might not be a common characteristic. Heavy reliance on the use of authority figures, a need to show that someone has said something similar before, these all too common practices could lead whole areas to stop advancing. If agreement with what has already been said is expected, advance might come to a halt. The refusal of many physicists to accept String Theory as a theoretical construct on the basis that it does not provide new predictions is epistemologically absurd, even more when we consider that its predictions actually work. The problem many people see with String Theory is that it just provides, for the experiments we can do, the same predictions of the older models. This goes directly against the belief physicists have that there will be only one true theory. And, as humans, they feel the urge to defend that random belief. Ironically, from an inductive point of view, the truth is that having more than one theory providing the same predictions is actually epistemologically useful. This is easier to see using Solomonoff framework. Identical predictions generated by different programs make those predictions more probable [66]! Finally, it is worth noticing that we already have some simulation evidence that, if social influence becomes too strong, an area might fail to replace an old theory for a better one [103, 104]. The habit of constantly referring and excessive citing

of old works is, indeed, something that should be considered worrisome. We need to allow new theories to be created and compared to reality. It does not matter if their predictions are not new or if we might not like them or if they may sound too weird at start nor if they go against what is the consensus. Competence is always mandatory, of course, but that is the only requirement of the scientific game. A good part of what makes an area trustworthy might be just its ability to allow the appearance of the new, the creation of ever new competitors for the position of the best description we currently have.

## 4.2. Tests and the Desire to Know

Of course, not only openness to new ideas matter. Simple competence is of fundamental importance, and one aspect of being competent is our competence in data analysis. But the history of the use of statistical methods in areas as Medicine and Psychology is one where competence was traded for ease of use allied to a desire to find the one true answer. It is not unreasonable to conclude that many statistical tools have been developed as statisticians were prey to our human tendencies; rejecting ideas is a recurrent theme in the development of the area. And that had to lead to a new set of problems. Things have become so out of control that there is now good reason to consider most results of those areas as unreliable [105, 106]. The causes are not only statistical methods but also the lack of deeper statistical knowledge, of course, coupled to absolutely damaging politics from the journals about when to accept an experimental paper. Still, it is clear that the use of methods that reject ideas is at the very heart of the observed problems. When using knowledge, full decision procedures are needed and that will often mean, in practical terms, to behave as if one of the possibilities were the true one. But this is true only for the decisions, such as which treatment to prescribe.

In particular, data analysis that wants to avoid generating biases in the scientists and readers minds should not include statements that could sound as if we were pointing at one idea as the correct one nor at one idea as the wrong one. We should avoid any kind of hypothesis testing. Ideally we should just point to which ideas are or seem more probable or, when that is not feasible just present and provide useful summaries of the data. An article that estimates rough probabilities is not a problem, if it is clear the estimates are rough. The same is valid for an article that only shows the characteristics of the data. Strong statements with no estimate of uncertainty are problematic. This prescription is something different researchers have been obtaining as they study the many errors associated with bad statistical usage and how it interacts with publishing practices. The absurdly disastrous practice of some areas where most journals would only publish results with “strong”  $p$ -values has led to nefarious consequences due to severe publication bias [105].

Recently, we have already observed many scientists also pointing the serious problems that arise from making tests, regardless of whether the journals would publish all experiments or just a severely biased collection of them [107]. In Medicine, cautionary articles about this misuse of statistical methods exist for more than two decades [108]. The widespread use of  $p$ -values as a way to avoid thinking about the statistical problem

has already been correctly labeled as surrogate science [109] and warnings about how  $p$ -values can be useless exist [110]. Correcting techniques have been proposed to allow us to use the existing reported  $p$ -values in the already existing literature in a non-testing manner [111] but, unfortunately, while they make sense from a statistical point of view, they can not compensate for the fact that the reported values we have in print are useless, as they were also the consequence of the publication biases. The need to completely abandon null-hypothesis significance testing has been acknowledged [112] but the suggestion to simply change it for confidence (or credibility) intervals is not necessarily an improvement from the point of view of the reader. Some readers might still read it as strong support to one idea [113]. On the other hand, acknowledgment of the dangers of null-hypothesis significance testing have finally reached the point where the method has simply been banned from at least one journal [114], so things are improving, albeit slowly. Misuse of statistics in the literature is not limited to problems with tests and  $p$ -values and represent a much larger problem with competence in general [115], but the area is already concluding that tests should not be used. This should be no surprise as tests are too similar to taking sides. Our desire for final answers is why hypothesis tests were created, despite the ever obvious fact that the tests never tell which hypothesis is true. But the terminology seemed to imply they did, with terms such as rejection or non-rejection. Estimating the actual probabilities might be too difficult for many problems. How to do it is often an open problem for researchers in statistics, but the fact remains that choosing sides about hypothesis is not warranted. Epistemologically, no idea should ever be rejected.

## 5. DISCUSSION

Changing from normal, committed beliefs to probabilistic, uncertain beliefs might not be easily done. Our societies and our institutions are currently built around the concept that ideas should be defended. We teach our kids to stand for what they believe, we teach them that debating means defending instead of analyzing ideas. But the fact that we have always done that does not mean this is the best way to do things. What we have seen here strongly supports the idea that we should start learning as soon as possible that ideas are to be challenged, not supported. They can be often safely discarded, as probabilities fly to zero. But openness to change is paramount. This is a concept that scientists have always claimed to follow. It might be useful to distinguish in our language the two types of beliefs, supporting unwarranted beliefs from probabilistic ones, to make it clear that they are not the same. Probabilistic beliefs (beliefs?) are correct while committed beliefs make us stupid. By claiming a probability we are not committing ourselves to an idea, at least not so strongly. It should be easier later to accept those ideas were wrong. Socrates sounds like he had grasped it, when he used to claim he had no knowledge about the world, but we have kept making this same mistake throughout the millennia since then. We should not teach anyone to debate and support an idea, we should teach people to look for mistakes in all the sides of the discussion. We

should not think that one political side will have the right ideas and should be supported against the wrong side. It is perfectly possible and sometimes very likely that one of the sides will be correct more often than the other side. But chances are that good and bad ideas will come from everywhere. My beliefs are my fallible opinion, at best they are only likely to be the current best option.

There will be cases when taking sides will be unavoidable. Moral issues might be an important exception to the rule of never taking sides. We do not know there is a right moral answer and, while the vast majority of us will agree on many basic ideas, it is not clear that there are right answers. Defending your own moral values might not be a problem, assuming you do know them perfectly well and that they do not suffer from logical inconsistencies. Unfortunately, there will also be times when taking sides about descriptions of the real world will be the best decision, given the circumstances. When people argue against vaccines, or that evolution is only a theory (the example on General Relativity should make it abundantly clear that some theories can become even more probable than any single experimental result), or any number of “crazy” ideas we see everyday, we often have too much data showing that they are wrong. Under some circumstances, we are actually in a situation where one of the probabilities has gone to basically zero. And then, accepting the death of little children or the spread of damaging world views, just for the sake of a trivial reminder that  $10^{-100}$  is technically (and only technically) not zero, would make no sense at all. When public discussion is a realm of incompetence case, simply stating the competent view will mean taking sides.

The need to change the way we debate should be clear by now. Taking sides can make us stupid. The more intelligent we are, the better equipped we are to defend our ideas and, paradoxically, the more stupid our beliefs can make us. Worse yet, the way we believe not only make it harder to get to the best answer. There is very good reason to think that the concept that there must be a correct answer is at the very core of the raise of extremist positions. Belief and belonging are very probably the main motors behind ideological violence. By changing what we consider to be acceptable ways of discourse, we might contribute to faster advances and technologies and also to a safer world. I believe or, more exactly, I strongly “believe” that, if we start paying attention to when we and others have group defining beliefs and if we start denouncing that behavior in ourselves and in others, there is a strong chance we will diminish the problems I have identified here.

The conclusion that we must not take sides also has consequences on how we researchers interact with our own scientific practice. We have been making mistakes in all fields of knowledge. While those mistakes have luckily been completely inconsequential in areas as Physics, they have been very costly in other fields of knowledge. It is not an exaggeration to claim our desire to take sides has cost many lives, since we have serious problems on the published results in Medicine. In the Humanities, the constant struggle between ideologies has certainly posed obstacles to our knowledge. It is close to impossible to figure out who might be right at many issues.

Taking sides and defending ideas has rendered the opinion of researchers completely useless in the cases we do not have clear data. Incompetence can always be detected, of course. But when arguments become more solid and sophisticated, the fact that there is no logical and mathematical foundations we can use to check them in a way that is independent of our severely biased human argumentation means that it might not be possible to even guess who is right. Work that simply describes situations is probably less prone to the problem.

We must also remember how the desire to obtain the one true answer has led the opinions of the agents to absurd extremes. The simulations are clearly not a representation of the whole process of human thinking. But the strength of the results does suggest that the difference between wishers, who want to find the one true idea, and mixers, who accept that they might need to live with a mixture of ideas coming from different theories, might be a fundamental problem. Checking the robustness of these results under different model assumptions is still needed. But the behavior we have observed in the models in this paper already poses significant evidence that this distinction is probably a very important one. That means that it is not only taking sides we should avoid. We must not even want to take sides. Add the psychological already observed effects to the quite probable problems arising from social interactions suggested by the simulations, we must learn that we should never use techniques to choose between ideas. Our statistical techniques must be developed and chosen to deal with the problem that arise from the fact that we are humans and quite fallible.

It is also interesting to notice that some of the ideas that gave birth to the most damaging forms of relativism are correct. They do not have the consequences people thought they have. But we do need to allow for the development of every type of theory and ideas. The crucial difference is that this actual need does not mean that all ideas are equal, some are more probable, others, as unlikely they can be safely dismissed. The fact that we need to test as many theories as possible means we should indeed have theoretical scientists in every field of knowledge. If the people who propose the ideas are not the same who obtain the data, observations and conclusions become far more trustworthy. Science must come to terms with our human shortcomings, with our desire to defend our own ideas and groups.

Before moving to the closing, some comments on the consistency of these conclusions is warranted. The conclusion that we should not hold beliefs is, of course, a belief and this might seem contradictory at first. It is actually not a contradiction, since we can have conclusions that are very likely to be true, instead of actually knowing them to be true. To evaluate its probability, we need to check the basis for the paper conclusion. It has two basis. One is the fact there is no real logical grounds for beliefs. This is pure logical reasoning; logic could change in time but guessing about future developments is simply beyond our abilities. The other source of doubt is the experimental results that show that we damage our reasoning when we pick sides. This conclusion seems quite robust and likely but, as with any experimental results, those results could be caused by sheer luck, which is very unlikely, or some effect we simply do not know about. Finally, the conclusion here is not self-defeating. While I, as the author,

am defending an idea, you, the reader, should really not trust me. But you can follow the reasoning and the logical proofs in the literature. And you can check the psychological data and arrive at the same conclusions yourself, without having to trust me.

We used to think we lived at the center of the Universe. We used to think we were not related to the animals. We used to think that we were governed by our reason and that our emotions did not have a significant impact on us. It is now time to realize that our reasoning is also deeply flawed. It does not work the way we always thought it did and we can not always trust our own arguments. We can only learn to use tools to correct for the severe problems in our very fallible reasoning. Those tools have brought us this far and they are what make Science deserve some trust, while other alternatives deserve none. Without those tools,

it seems that, every time we try to reason, we may lie to others and we may lie to ourselves. We are that incompetent when we do not correct our biases. For now, I leave the reader with the commandment of the title: “*Thou shalt not take sides.*”

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## ACKNOWLEDGMENTS

The author would like to thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for partial support to this research under grant 2014/00551-0.

## REFERENCES

- Fagin R, Halpern JY, Moses Y, Vardi MY. *Reasoning About Knowledge*. Cambridge: MIT Press (1995).
- Shope RK. Conditions and analyses of knowing. In: Moser PK, editor. *The Oxford Handbook of Epistemology*. Oxford: University Press (2002), pp. 25–70.
- Klein P. *Skepticism*. In: Zalta EN, editor. *The Stanford Encyclopedia of Philosophy*. Stanford: Summer (2014).
- Jervis R. *Perception and Misperception in International Politics*. Princeton: Princeton University Press (1976).
- Mercier H. Reasoning serves argumentation in children. *Cogn Dev*. (2011) **26**:177–91. doi: 10.1016/j.cogdev.2010.12.001
- Mercier H, Sperber D. Why do humans reason? Arguments for an argumentative theory. *Behav Brain Sci*. (2011) **34**:57–111. doi: 10.1017/S0140525X10000968
- Kahan DM, Jenkins-Smith H, Braman D. Cultural cognition of scientific consensus. *J Risk Res*. (2011) **14**:147–74. doi: 10.1080/13669877.2010.511246
- Castellano C, Fortunato S, Loretto V. Statistical physics of social dynamics. *Rev Mod Phys*. (2009) **81**:591–646. doi: 10.1103/RevModPhys.81.591
- Galam S, GEFEN Y, Shapir Y. Sociophysics: a new approach of sociological collective behavior: mean-behavior description of a strike. *J Math Sociol*. (1982) **9**:1–13. doi: 10.1080/0022250X.1982.9989929
- Galam S, Moscovici S. Towards a theory of collective phenomena: consensus and attitude changes in groups. *Eur J Soc Psychol*. (1991) **21**:49–74. doi: 10.1002/ejsp.2420210105
- Sznajd-Weron K, Sznajd J. Opinion evolution in a closed community. *Int J Mod Phys C* (2000) **11**:1157. doi: 10.1142/S0129183100000936
- Deffuant G, Neau D, Amblard F, Weisbuch G. Mixing beliefs among interacting agents. *Adv Compl Sys*. (2000) **3**:87–98. doi: 10.1142/S0219525900000078
- Martins ACR. Continuous opinions and discrete actions in opinion dynamics problems. *Int J of Mod Phys C* (2008) **19**:617–24. doi: 10.1142/S0129183108012339
- Bloor D. *Knowledge and Social Imagery*. Chicago: University of Chicago Press (1991).
- Hawking S, Penrose R. *The Nature of Space and Time*. Princeton: Princeton University Press (2010).
- Kahan DM, Peters E, Wittlin M, Slovic P, Ouellette LL, Braman D, et al. The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nat Clim Change* (2012) **2**:732–5. doi: 10.1038/nclimate1547
- Popper K. *The Logic of Scientific Discovery*. London: Hutchinson (1959).
- Quine WVO. Two dogmas of empiricism. *Philos Rev*. (1953) **60**:20–46. doi: 10.2307/2181906
- Duhem P. *La Théorie Physique: Son Objet et sa Structure*. Chevalier and Rivière, Paris (1906).
- Laudan L. Demystifying underdetermination. In: Savage CW, editor. *Scientific Theories*. Minneapolis: University of Minnesota Press (1990), pp. 267–97.
- Earman J. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge: MIT Press (1992).
- Maher P. *Betting on Theories*. Cambridge: Cambridge University Press (1993).
- Jeffrey RC. *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press (2004).
- Martins ACR. Probabilistic biases as bayesian inference. *Judg Decis Mak*. (2006) **1**:108–17. Available online at: <http://journal.sjdm.org/jdm06124.pdf>
- Tenenbaum JB, Kemp C, Shafto P. Theory-based Bayesian models of inductive reasoning. In: Feeney A, Heit E, editors. *Inductive Reasoning*. Cambridge: Cambridge University Press (2007), pp. 167–204.
- Plous S. *The Psychology of Judgment and Decision Making*. New York, NY: McGraw-Hill (1993).
- Baron J. *Thinking and Deciding*. Cambridge University Press (2007).
- Kahneman D. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux (2011).
- Phillips LD, Edwards W. Conservatism in a simple probability inference task. *J Exp Psychol*. (1966) **27**:346–54.
- Allais PM. The behavior of rational man in risky situations - A critique of the axioms and postulates of the American School. *Econometrica* (1953) **21**:503–46. doi: 10.2307/1907921
- Ellsberg D. Risk, ambiguity and the Savage axioms. *Q J Econ*. (1961) **75**:643–69.
- Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica* (1979) **47**:263–91. doi: 10.2307/1914185
- Birnbaum MH. New paradoxes of risky decision making. *Psychol Rev*. (2008) **115**:463–501. doi: 10.1037/0033-295X.115.2.463
- Chapman LJ, Chapman JP. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *J Abnorm Psychol*. (1969) **74**:271–80. doi: 10.1037/h0027592
- Hamilton DL, Rose TL. Illusory correlation and the maintenance of stereotypic beliefs. *J Pers Soc Psychol*. (1980) **39**:832–45. doi: 10.1037/0022-3514.39.5.832
- Gigerenzer G, Muir Gray JA, editors. *Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020*. Cambridge: MIT Press (2011).
- Watson PC, Johnson-Laird P. *Psychology of Reasoning: Structure and Content*. Massachusetts: Harvard University Press (1972).
- Tversky A, Kahneman D. Extension versus intuitive reasoning: the conjunction fallacy in probability judgement. *Psych Rev*. (1983) **90**:293–315.
- Tversky A, Kahneman D. The framing of decisions and psychology of choice. *Science* (1981) **211**:453–8.
- Oskamp S. Overconfidence in case-study judgments. *J Consult Psychol*. (1965) **29**:261–5.

41. Lichtenstein D, Fischhoff B. Do those who know more also know more about how much they know? The Calibration of Probability Judgments. *Organ Behav Hum Perform.* (1977) 3:552–64.
42. Dunning D, Johnson K, Ehrlinger J, Kruger J. Why people fail to recognize their own incompetence. *Curr Dir Psychol Sci.* (2003) 12:83–7. doi: 10.1111/1467-8721.01235
43. Hall CC, Ariss L, Todorov A. The illusion of knowledge: when more information reduces accuracy and increases confidence. *Organ Behav Hum Decis Process.* (2007) 103:277–90. doi: 10.1016/j.obhdp.2007.01.003
44. Surowiecki J. *The Wisdom of Crowds.* New York, NY: Anchor Books (2005).
45. Janis IL. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes.* Boston: Houghton Mifflin Company (1972).
46. Kerr NL, MacCoun RJ, Kramer GP. Bias in judgment: comparing individuals and groups. *Psychol Rev.* (1996) 103:687–719.
47. Asch S. Opinions and social pressure. *Sci Am.* (1955) 193:31–35.
48. Asch SE. Studies of independence and conformity: a minority of one against a unanimous majority. *Psychol Monogr.* (1956) 70:70.
49. Simon HA. Rational choice and the structure of environments. *Psychol Rev.* (1956) 63:129–38.
50. Gigerenzer G, Todd PM, Group TAR. *Simple Heuristics That Make Us Smart.* Oxford: Oxford University Press (2000).
51. Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. *Cogn Psychol.* (1973) 5:207–32.
52. Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. *Psych Rev.* (1996) 103:650–69.
53. Johnson-Laird PN, Legrenzi P, Legrenzi MS. Reasoning and a sense of reality. *Br J Psychol.* (1972) 6:395–400.
54. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol.* (1998) 2:175–220.
55. Mercier H. On the universality of argumentative reasoning. *J Cogn Cult.* (2011) 11:85–113. doi: 10.1163/156853711X568707
56. Goebbert K, Jenkins-Smith HC, Klockow K, Nowlin MC, Silva CL. Weather, climate, and worldviews: the sources and consequences of public perceptions of changes in local weather patterns. *Am Meteorol Soc.* (2012) 4:132–44. doi: 10.1175/WCAS-D-11-00044.1
57. Slovic P, Flynn JH, Layman M. Perceived risk, trust, and the politics of nuclear waste. *Science* (1991) 254:1603–7.
58. Kahan DM. Ideology, motivated reasoning, and cognitive reflection. *Judgm Decis Mak.* (2013) 8:407–24. Available online at: <http://journal.sjdm.org/13/13313/jdm13313.pdf>
59. Kahan D. Fixing the communication failure. *Nature* (2010) 463:296–7. doi: 10.1038/463296a
60. Gettier E. Is justified true belief knowledge? *Analysis* (1963) 23:121–3.
61. Hume D. *An Inquiry Concerning Human Understanding.* Oxford: Clarendon Press (1777).
62. Goodman N. *Fact, Fiction and Forecast.* Indianapolis: Bobbs-Merrill (1946).
63. Ramsey F. *Truth and probability.* In: Kyburg HE, Smokler HE editors. *Foundations of Mathematics and other Logical Essays.* 1931. Studies in Subjective Probability. New York, NY: Robert Krieger; K. Paul, Trench, Trubner and Co. (1980).
64. Jaynes ET. *Probability Theory: The Logic of Science.* Cambridge: Cambridge University Press (2003).
65. Caticha A. Relative entropy and inductive inference. In: Erickson G, Zhai Y, editors. *Bayesian Inference and Maximum Entropy Method.* No. 75 in AIP Conference Proceedings 707; ArXiv.org/abs/physics/0311093 (2004).
66. Solomonoff RJ. A formal theory of inductive inference. Part I. *Inf Control* (1964) 7:1–22.
67. Thaler R. Toward a positive theory of consumer choice. *J Econ Behav Organ.* (1980) 1:39–60.
68. Bafumi J, Herron MC. Leapfrog representation and extremism: a study of American voters and their members in Congress. *Am Polit Sci Rev.* (2010) 104:519–42. doi: 10.1017/S0003055410000316
69. Tileaga C. Representing the ‘Other’: a discursive analysis of prejudice and moral exclusion in talk about Romanians. *J Commun Appl Soc Psychol.* (2006) 16:19–41. doi: 10.1002/casp.846
70. Galam S. Local dynamics vs. social mechanisms: a unifying frame. *Europhys Lett.* (2005) 70:705–11. doi: 10.1209/epl/i2004-10526-5
71. Ben-Naim E, Krapivsky PL, Vazquez F, Redner S. Unity and discord in opinion dynamics. *Phys A Stat Mech Appl.* (2003) 330:99–106. doi: 10.1016/j.physa.2003.08.027
72. Nizamani S, Memon N, Galam S. From public outrage to the burst of public violence: an epidemic-like model. *Phys A Stat Mech Appl.* (2014) 416:620–30. doi: 10.1016/j.physa.2014.09.006
73. Galam S, Jacobs F. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Phys A Stat Mech Appl.* (2007) 381:366–76. doi: 10.1016/j.physa.2007.03.034
74. Hegselmann R, Krause U. Opinion dynamics and bounded confidence models, analysis and simulation. *J Artif Soc Soc Simul.* (2002) 5:3. Available online at: <http://jasss.soc.surrey.ac.uk/5/3/2/2.pdf>
75. Deffuant G, Amblard F, Weisbuch G, Faure T. How can extremism prevail? A study based on the relative agreement interaction model. *J Artif Soc Soc Simul.* (2002) 5:1. Available online at: <http://jasss.soc.surrey.ac.uk/5/4/1.html>
76. Weisbuch G, Deffuant G, Amblard F. Persuasion dynamics. *Phys A Stat Mech Appl.* (2005) 353:555–75. doi: 10.1016/j.physa.2005.01.054
77. Amblard F, Deffuant G. The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Phys A Stat Mech Appl.* (2004) 343:725–38. doi: 10.1016/j.physa.2004.06.102
78. Franks DW, Noble J, Kaufmann P, Stagl S. Extremism propagation in social networks with hubs. *Adapt Behav.* (2008) 16:264–74. doi: 10.1177/1059712308090536
79. Deffuant G. Comparing extremism propagation patterns in continuous opinion models. *J Artif Soc Soc Simul.* (2006) 9:8. Available online at: <http://jasss.soc.surrey.ac.uk/9/3/8.html>
80. McKeown G, Sheehy N. Mass media and polarization processes in the bounded confidence model of opinion dynamics. *J Artif Soc Soc Simul.* (2006) 9:11.
81. Boccaro N. Evolution of extremist opinions in a population of interacting agents. *Int J Mod Phys C.* (2010) 21:617–28. doi: 10.1142/S0129183110015373
82. Alizadeh M, Coman A, Lewis M, Cioffi-Revilla C. Intergroup conflict escalations lead to more extremism. *J Artif Soc Soc Simul.* (2014) 14:4. doi: 10.18564/jasss.2559
83. Gargiulo F, Mazzoni A. Can extremism guarantee pluralism? *J Artif Soc Soc Simul.* (2008) 11:9. Available online at: <http://jasss.soc.surrey.ac.uk/11/4/9.html>
84. Martins ACR. Bayesian updating as basis for opinion dynamics models. *AIP Conf Proc.* (2012) 1490:212–21. doi: 10.1063/1.4759605
85. Martins ACR. Mobility and social network effects on extremist opinions. *Phys Rev E* (2008) 78:036104. doi: 10.1103/PhysRevE.78.036104
86. Eguiluz VM, Masuda N, Fernández-Gracia J. Bayesian decision making in human collectives with binary choices. *PLoS ONE* (2015) 10:e0121332. doi: 10.1371/journal.pone.0121332
87. Edwards W. Conservatism in human information processing. In: Kleinmuntz B, editor. *Formal Representation of Human Judgment.* New York, NY: John Wiley and Sons (1968), pp. 17–52.
88. Martins ACR. Discrete Opinion models as a limit case of the CODA model. *Phys A Stat Mech Appl.* (2013) 395:352–7. doi: 10.1016/j.physa.2013.10.009
89. Fan K, Pedrycz W. Emergence and spread of extremist opinions. *Phys A Stat Mech Appl.* (2015) 436:87–97. doi: 10.1016/j.physa.2015.05.056
90. Martins ACR. Bayesian updating rules in continuous opinion dynamics models. *J Stat Mech Theory Exp.* (2009) P02017. ArXiv:0807.4972v1.
91. Martins ACR, Galam S. The building up of individual inflexibility in opinion dynamics. *Phys Rev E* (2013) 87:042807. ArXiv:1208.3290.
92. Diao SM, Liua Y, Zeng QA, Luo GX, Xiong F. A novel opinion dynamics model based on expanded observation ranges and individuals’ social influences in social networks. *Phys A Stat Mech Appl.* (2014) 415:220–8. doi: 10.1016/j.physa.2014.07.072
93. Martins ACR. Trust in the CODA model: opinion dynamics and the reliability of other agents. *Phys Lett A* (2013) 377:2333–9. doi:10.1016/j.physleta.2013.07.007
94. Si XM, Liua Y, Xiong F, Zhang YC, Ding F, Cheng H. Effects of selective attention on continuous opinions and discrete decisions. *Phys A Stat Mech Appl.* (2010) 389:3711–9. doi: 10.1016/j.physa.2010.05.010
95. Martins ACR. Opinion particles: classical physics and opinion dynamics. *Phys Lett A* ArXiv:1307.3304 (2014).

96. Mitroff II. Norms and counter-norms in a select group of the apollo moon scientists: a case study of the ambivalence of scientists. *Am Soc Rev.* (1974) **39**:579–95.
97. Feyerabend P. *Against Method*. London: Verso (1994).
98. Fitelson B, Thomason N. Bayesians sometimes cannot ignore even very implausible theories (even ones that have not yet been thought of). *Australas J Logic* (2008) **6**:25–36. Available online at: <https://ojs.victoria.ac.nz/ajl/article/view/1791/1642>
99. Clemence GM. The relativity effect in planetary motions. *Rev Mod Phys.* (1947) **19**:361–4. doi: 10.1103/RevModPhys.19.361
100. Kahan DM, Hoffman D, Evans D, Devins N, Lucci E, Cheng K. “Ideology” or “Situation Sense”? *An Experimental Investigation of Motivated Reasoning and Professional Judgment*. Philadelphia, PA: University of Pennsylvania Law Review (2015).
101. Wigner E. The unreasonable effectiveness of mathematics in the natural sciences. *Commun Pure Appl Math.* (1960) **13**:1–14. doi: 10.1002/cpa.3160130102.
102. Merton RK. 13: the normative structure of science. In: Storer NW, editor. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press (1979), pp. 267–78.
103. Martins ACR. Modeling scientific agents for a better science. *Adv Compl Sys.* (2010) **13**:519–33. doi: 10.1142/S0219525910002694
104. Martins ACR. Modelling epistemic systems. In: Dabbaghian V, Mago VK, editors. *Theories and Simulations of Complex Social Systems*. Berlin: Springer (2013), pp. 19–30.
105. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *J Am Med Assoc.* (2005) **294**:218–28. doi: 10.1001/jama.294.2.218
106. Ioannidis JPA. Why most published research findings are false. *PLoS ONE* (2005) **2**:e124. doi: 10.1371/journal.pmed.0020124
107. Nuzzo R. Statistical errors. *Nature* (2014) **506**:150–2. doi: 10.1038/506150a
108. Altman DG. The scandal of poor medical research. *Br Med J.* (1994) **308**:283–4.
109. Gigerenzer G, Marewski JN. Surrogate science: the idol of a universal method for scientific inference. *J Manage.* (2015) **41**:421–40. doi: 10.1177/0149206314547522
110. Cumming G. Replication and *p* intervals. *Perspect Psychol Sci.* (2008) **3**:286–300. doi: 10.1111/j.1745-6924.2008.00079.x
111. Held L. A nomogram for P values. *BMC Med Res Methodol.* (2010) **10**:21. doi: 10.1186/1471-2288-10-21
112. Cumming G. The new statistics: why and how. *Psychol Sci.* (2014) **25**:7–29. doi: 10.1177/0956797613504966
113. Savalei V, Dunn E. Is the call to abandon p-values the red herring of the replicability crisis? *Front Psychol.* (2015) **6**:245. doi: 10.3389/fpsyg.2015.00245
114. Trafimow D, Marks M. Editorial. *Basic Appl Soc Psychol.* (2015) **37**:1–2. doi: 10.1080/01973533.2015.1012991
115. Leek JT, Peng RD. P- values are just the tip of the iceberg. *Nature* (2015) **520**:612. doi: 10.1038/520612a

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JS and the handling Editor declared a current collaboration and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Martins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.