Identifying bias. A Bayesian analysis of suspicious agreement between beliefs and values.

Political decisions are based on values, on the one hand, and beliefs about empirical facts on the other. Ideally, your beliefs about facts should be based on empirical evidence. They should not be tainted by your political ideology. Nevertheless, this often happens. Most ideologies are uncomfortable with some empirical facts. For instance, Western communists were uncomfortable with the fact that Stalin had murdered millions of his compatriots, and therefore denied it for a long time. Similarly, climate change denial is usually ideologically motivated. Climate change requires – or so many people think – regulations of the free market that conservatives and libertarians oppose. Therefore, some of them deny climate change in the face of overwhelming empirical evidence.

There is a mounting psychological literature suggesting that this kind of political bias is very widespread. For instance, Dan Kahan and his collaborators has found that conservatives and liberals are much more likely to be right on empirical questions where they like the true answer than on those where they dislike it (Kahan et al. 2011). Also, Kahan and colleagures have found that people are more likely to make mathematical mistakes which support their political prejudices than mistakes which go against their political prejudices (Kahan et al. 2013).

Thus people regularly acquire beliefs about facts because they are support their political prejudices. This does not mean, however, that all cases of agreement between factual beliefs and political views are caused by political bias. Some false beliefs which accord with political opinions are caused by other factors, such as misleading evidence.

It is usually not possible to decide which of these explanations is the right one if we just look at an individual belief. By looking at the *structure* of larger sets of beliefs on politically controversial issues, we can, however, get a better grasp of which explanation is true.

To see that, consider the following example. Paul has been an adherent of one of the political parties in the US since childhood. He is asked *n* factual Yes-No questions on politically controversial issues (e.g. regarding the economy, crime, etc). On some of these questions, the answer that favors Paul's political views is more likely to be right, whereas on others, the opposite is true. The questions are probabilistically independent of each other, and the geometric mean of the objective probabilities of the answers that support Paul's political views is .5 (and likewise for those that disconfirm them), meaning that there is a $.5^n$ chance that all of the true answers support Paul's political views.

Now suppose that we prior to hearing Paul's answers entertain the following three hypotheses:

Randomization (R): Paul has arrived on his beliefs on these issues at random. On each individual question, there is therefore a 50 % chance that he will answer "Yes", and a 50 % chance that he will answer "No", regardless of which answer is more likely to be right and of which answer supports his political views.

Bias (B): Paul is politically biased in the sense that on all questions, he has arrived at the answer which supports his political views.

Truth-tracking (T): Paul is a perfect truth-tracker who is right on all issues. This means that if the objective probability that a certain answer is correct is 60 %, there is a 60 % chance Paul will give that answer.

Now suppose that Paul does give answers which support his political views on all issues. How will that affect the probability of the randomization, bias and truth-tracking hypotheses?

To answer that, let $A_1,\ldots,A_n$ denote the content of Paul's answers to the *n* questions (e.g. "concealed carry laws decrease/increase crime"), and $E_1,\ldots,E_n$ denote the fact that Paul did give those answers (e.g. "Paul said that concealed carry laws decrease/increase crime"). It should be obvious that the probability of $E_1,\ldots,E_n$ given the randomization and the bias hypothesis, respectively, are as follows:

$$P(E_1,...,E_n|R) = .5^n$$

$$P(E_1,...,E_n|B) = 1$$

To calculate the probability of $E_1,\ldots,E_n$ given the truth-tracking hypothesis, let us first note that:

$$P(E_1,...,E_n|T) = P(A_1,...,A_n)$$

That is, the probability that Paul will *say* that $A_1,...,A_n$ equals the chance that $A_1,...,A_n$ are indeed true, given that he is a perfect truth-tracker.

Next, note that $P(A_1,...,A_n) = P(A_1) \times ... \times P(A_n) \times C_{Sh}(A_1,...,A_n)$,

where C_{Sh} is the so-called Shogeji measure of coherence (Shogenji 1999):

$$C_{Sh}(A_1,...,A_n) = \frac{P(A_1...,A_n)}{P(A_1) \times ... \times P(A_n)}$$

C_{Sh} is 1 in case $A_1,...,A_n$ are, all-in-all, probabilistically independent of each other, whilst it is greater than 1 in case $A_1,...,A_n$ are, all-in-all, positively probabilistically dependent of each other, and below one if they are, all-in-all, negatively dependent on each other. Hence C_{Sh} could also be seen as a measure of probabilistic dependence.

Next, note that since the geometric mean of $P(A_1),...,P(A_n) = .5$, $P(A_1) \times ... \times P(A_n) = .5^n$. Hence:

$$P(A_1,...,A_n) = .5^n \times C_{Sh}(A_1,...,A_n)$$

This means that if $A_1,...,A_n$ are independent of each other, the likelihood of the evidence ($E_1,...,E_n$) is the same on the truth-tracking and the randomization hypotheses (i.e. $.5^n$). Since this number rapidly decreases as *n* grows larger, the upshot is that neither the truth-tracking or the randomization can compete with the bias hypotheses for high *n* unless the latter has an extremely low prior probability.

On the other hand, if $A_1,...,A_n$ were to be strongly coherent, then the situation becomes quite different. Suppose that $P(A_1)=.5$ but that $P(A_i|A_1)=1$, for *i*,…,*n* – i.e. that A_1 entails all of the other members of $\{A_1,...,A_n\}$. Then

$$P(E_1,...,E_n|T)=P(A_1,...,A_n)=.5$$

That is not so much less than $P(E_1,...,E_n|B)=1$, which means that we do not get as strong support for the bias support as we do when $\{A_1,...,A_n\}$ is an independent set.

Let us calculate the posterior probability of the bias hypothesis given E_1,…,E_n using Bayes' theorem:

Theorem 1:

$$P(B|E_1,...,E_n)=$$

$$\frac{P(E_1,...,E_n|B)P(B)}{P(E_1,...,E_n|R)P(R)+P(E_1,...,E_n|T)P(T)+P(E_1,...,E_n|B)P(B)}=$$

$$\frac{P(B)}{.5^n P(R)+.5^n C_{Sh}(A_1,...,A_n)P(T)+P(B)}$$

We can draw two conclusions from theorem 1.

1) *The less coherent the factual beliefs are, the more the agreement between political views and factual beliefs indicates bias.*
2) *The larger the number of factual beliefs is, the more the agreement between political views and factual beliefs indicates bias.*

Of course, normally you would not just entertain the three hypotheses I use in this simplified model, but also include intermediate hypotheses, saying, e.g. that Paul's beliefs are based on a combination of political bias and truth-tracking. While the precise numbers would be different if you included such hypotheses, it would not, however, change the basic logic of the inference. Agreement between factual beliefs and political views would still be strong evidence of bias, especially if those factual beliefs were Shogenji coherent and numerous.

Causes of agreement between political views and factual beliefs

In this case, Paul had had his political views since childhood. That means that it's likely that the reason why his political views and factual beliefs are in such agreement is that his factual beliefs are coloured by his political views. That is not the only logical possibility, however. There are three possible explanations for this agreement:

a) The one we already discussed: he arrived at his political views first, and then acquired factual beliefs which supported them.
b) He arrived at his factual beliefs first, and then acquired political views which accorded with them.
c) There is a common cause of both his political views and his factual beliefs.[1]

The first hypothesis needs little further introduction. There are many psychological biases which could make this happen such as wishful thinking,[2] confirmation bias[3] and the halo effect[4].

Turning to the second hypothesis, note that it is a bit surprising that Paul just happened to arrive at lots of factual beliefs which favour a particular political view. This is a striking coincidence which calls out for an explanation. The most likely explanation – given that it by hypothesis cannot be that he has been coloured by his own political views – is that he has been fed information which supported those views (e.g. by politically biased media).

The third hypothesis is, in turn, perhaps the least plausible one in most cases. It could be, though, that a personality trait such as disgust sensitivity could cause both factual beliefs and political views.[5]

The most plausible hypothesis is instead normally the first hypothesis, I would argue. One reason to believe that it is normally more plausible than the second is that one study showed that the correlation between answers to factual questions and political views was substantially reduced when people were paid for being right on the factual questions (Bullock et al. 2013). The second hypothesis does not adequately explain this fact. It assumes that your factual beliefs were acquired independently of your political views. But then there is no obvious reason to

[1] Whether it is appropriate to use the term "bias" under the two latter scenarios might be discussed.

[2] This common bias makes us align our beliefs with our preferences. For instance, it disposes us to believe that a team that we are supporting is likely to win even in the face of evidence to the contrary. There is a huge literature on wishful thinking; see, e.g. (Bastardi et al. 2011), (Kunda 1990) and (Lord et al.1979).

[3] This is the name for our tendency to disregard evidence that disconfirms our beliefs, while laying heavy emphasis on evidence that confirms or supports them. Confirmation bias has been observed in a great number of studies and its existence and significance is by now one of psychology's most entrenched findings. See (Nickerson 1998) for an overview.

[4] The halo effect makes people more inclined to think that people who have some positive qualities also have other positive qualities. There are lots of papers on this effect; see, e.g. (Dion et al. 1972) and (Nisbett and Wilson, 1977).

[5] Disgust sensitivity is positively correlated with conservative political views. See (Inbar et al. 2012).

believe that you would change your factual beliefs so that they correlate less with your political views simply because you are getting paid to get them right.

On the other hand, this is precisely what we would expect if people let their political views influence their factual beliefs directly. The payment makes people think a bit more carefully on the questions, instead of just going with their political gut-feeling.[6]

The argument from suspicious belief-value agreement

Let us now look at a slightly different example. Suppose that a company's board has shortlisted five candidates for a high-ranking position. After having interviewed the candidates, the board members are asked to rank the candidates regarding the following criteria (the "big five" personality traits):

a) Openness to experience
b) Conscientiousness
c) Extraversion
d) Agreeableness
e) Emotional stability (absence of neuroticism)

Each board member is thus asked to provide five rankings. The highest ranked candidate (on any particular ranking) gets five points, number two gets four points, and so on, and the candidate with the most points in total gets hired. The vote is not secret, but the board members are asked to sign their rankings.

The board members are asked to vote sincerely in the sense that their votes should reflect their beliefs of the relative merits of the different candidates on the different criteria. The rationale for this is purely epistemic—it is assumed that this gives the best basis for aggregating their votes to an accurate overall ranking of the candidates.

Since the big five personality traits are uncorrelated or very weakly correlated[7] it is very unlikely that one of the candidates is the best one on to all criteria. Nevertheless, one of the board members, Jane, ranks one of the candidates, Leonard, as being more open to experience, more conscientious, more agreeable, more extroverted and more emotionally stable than all the others, meaning that he gets in total 25 points from her. What is the most plausible explanation of this surprisingly uniform set of votes?

Presumably it has got to be that Jane has not cast her votes impartially, but that she in some way or other is favouring Leonard. She could of course suffer from some cognitive bias such as wishful thinking, confirmation bias or the halo effect. But it could also be that she voted strategically (rather than sincerely) in order to maximize Leonard's chance of winning.

[6] It should be added that under the third hypothesis, you would expect betting to reduce the correlation between factual beliefs and political views, since betting would reduce the correlation between factual beliefs and the third factor that causes them both.

[7] See (van der Linden et al. 2010).

This example is structurally analogous to the Paul example above, but differs from it in two respects. Firstly, it is Jane's *stated* beliefs, rather than her *actual* beliefs, which are in strong agreement with her values. She may be misrepresenting her actual beliefs in order to maximize Leonard's chance of winning (i.e. voting strategically). Second, what causes her bias or strategic voting is not any political views of hers, but another goal or value, namely that Leonard should get the job.

Thus in general, a surprisingly high degree between stated or actual beliefs and values – be they political or of any other kind – indicates either bias or strategic voting. Call this *the argument from suspicious belief-value agreement.*[8]

Applications of the argument from suspicious belief-value agreement

Numerous studies show that people are strongly susceptible to confirmation bias and motivated reasoning such as wishful thinking in political contexts (see, e.g. Redlawsk et al. 2010, Lodge and Taber 2005), something that should give rise to suspicious belief-value agreement. We also do have some direct evidence indicating that people's political beliefs tend to exhibit suspiciously high belief-value agreement. For instance, in a classic study by Alhakami and Slovic (1994), subjects' perceptions of the benefits of a number of items (such as nuclear power, radiation therapy, vaccination, etc.) were inversely correlated with their perceptions of risks.[9] In other words, for each item, the subjects either thought that each item was either wholly good or wholly bad. Naturally, Alhakami and Slovic discussed the halo effect as a plausible explanation of this phenomenon.

Hence it seems that suspicious belief-value agreement is very common in politics. Coupled with the obvious importance of politics, this makes it particularly interesting to make use of the argument from suspicious belief-value agreement in political contexts. However, it could also be used in a wide range of other contexts. Indeed, given the ubiquity of biases and the strategic behaviour, it is hard to imagine any field where argument from suspicious belief-value agreement could not be used.

To get a sense of how useful this argument is, consider a slightly modified version of the Jane case. Suppose that the board members are not explicitly voting on the different criteria but that they are instead having an open and unstructured discussion, and refer to their views of the candidates' personality traits when arguing for or against a particular candidate. Furthermore, suppose that in this situation, Jane systematically rejects any claim to the effect that Leonard is worse than any of the other candidates on one of the criteria. Just like in the previous version of the example, her view is that Leonard is the best candidate on all criteria. Hence the set of her statements, or arguments, is strongly coherent.

Jane is, in effect, doing the same thing in both cases: namely, saying that Leonard is the best candidate across the board. However, there is a crucial difference between the two cases, namely that in the former case the pattern of Jane's votes is easy to spot (since the vote was not

[8] I mostly focus on actual beliefs in this paper, which is why I chose this term.

[9] This means that those who perceived an item to be beneficial also thought that it carried low risks, whereas those who perceived it to be risky deemed it low in benefits.

secret), whereas in an unstructured discussion it is often hard to notice how one-sided her statements are someone's argument is. This makes it much safer for Jane to work for Leonard in a consistently one-sided way in the latter case. In the former case, someone might point out that the unusual pattern of Jane's votes suggests that she voted strategically (contrary to the instructions) or was unconsciously biased, something which might lower the other board members' trust in her. It is much less likely that something similar would happen in the latter case. Not the least for this reason, I would think that cases like that are much more common.

Analogy with coherence reasoning

The argument from suspicious belief-value agreement says that certain structures of beliefs or statements indicate that they were caused by certain processes, such as bias or strategic voting. It thus amounts to a kind of *reverse engineering* of belief or statements structures. Another example of such reverse engineering is coherence reasoning, which has been extensively studied by epistemologists.

The following passage in the *locus classicus* of coherence theory, C. I. Lewis's *An Analysis of Knowledge and Valuation* (1946), is an important source of this line of reasoning:

> For any of these reports [from relatively unreliable witnesses who independently tell the same circumstantial story] taken singly, the extent to which it confirms what is reported may be slight. And antecedently, the probability of what is reported may also be small. But congruence of reports establishes a high probability of what they agree upon, by principle of probability determination which are familiar: on any other hypothesis than that of truth-telling, this agreement is highly unlikely; the story any one false witness might tell being one out of so very large a number of equally possible choices. (Lewis 1946, 346)

What Lewis argues is the following. Say that a number of witnesses have given identical—hence highly coherent—reports. Furthermore, say that they are independent (which rules out the possibility that they are, e.g. conspiring to frame a certain suspect). Lastly, assume that there is at least some positive probability that they are truth-tellers. Under such circumstances, the only plausible explanation of this high degree of coherence (which is a property of structures of beliefs or statements) is that the witnesses are indeed reliably telling the truth (which is a kind of belief- and statement-forming process), since it is highly unlikely that independent witnesses who are not truth-tellers would converge on the same story.

In their book *Bayesian Epistemology*, Luc Bovens and Stephan Hartmann showed that Lewis's line of reasoning was indeed valid, using probability theory: a sufficiently high degree of coherence does guarantee truth-telling (Bovens and Hartmann 2003, 62-65). Also, in (Schubert 2011, 2012a) a stronger theorem was proved; namely that the more Shogenji coherent a set of beliefs or testimonies is, the more likely it is to have been caused by reliable, truth-tracking processes.[10]

[10] This holds in most circumstances, though not in all, as shown in (Schubert 2012b).

This was proved using a witness scenario model which has many similarities to the simple formal model employed in this paper. The model compared two hypotheses: perfect truth-tracking and randomizing. These are of course exactly the same hypotheses as I employ in the above model, which is strongly inspired by that model. What I have added is the bias hypothesis, which does not exist in Bayesian coherence theory.

Now a further similarity between coherence reasoning and the argument from suspicious belief-value agreement is that in both cases, the structural feature that we are looking for is "agreement", in a wide sense. In coherence reasoning, the fact that the beliefs agree *with each other* indicates that the randomization hypothesis is false, and that the truth-telling hypothesis must be true. In the argument from suspicious belief-value agreement, the fact that (stated) beliefs agree to a surprisingly high degree *with the agent's values* indicates that the randomization and the truth-tracking hypotheses are false, and that the bias (or strategic voting) hypothesis must be true.

Coherence and belief-value agreement as tools in genetic arguments

Now it is important to realize that both coherence reasoning and the argument from suspicious belief-value agreement can be used as *genetic arguments*. Unlike ordinary arguments, genetic arguments say that *A* is (not) to be believed because the person saying *A* is (not) trustworthy.

Coherence arguments are, of course, positive genetic arguments (*ad verecundiam*) – they say that they fact that some beliefs or statements are coherent makes them likely to have been caused by reliable processes, which in turn makes them likely to be true. The argument from suspicious belief-value agreement, on the other hand, is a negative genetic argument – an *ad hominem*-argument. It says, for instance, that there is no reason to believe that what Jane says about Leonard is true, because given the suspicious level of agreement between her votes and her (supposed) preference for Leonard, there is no reason to believe that her votes tracks the truth.

Both of these arguments are, however, a rather special kind of genetic arguments, where you make a lot of use of information of what is being said.[11] You make a careful analysis not only of the probabilities of the uttered propositions, but also of how they are related to each other; what the structure of the set of propositions is. Once you have done that, you reverse engineer them: you hypothesize that this structure is most likely (conditional on your knowledge of the subject and other background knowledge) to be the result of, e.g. wishful thinking, strategic voting or truth-tracking. Once this is done, you go back to the propositions themselves and adjust your beliefs in them accordingly. If you think that they were caused by psychological biases or strategic planning, you adjust our belief in them downwards, while if you think that they were caused by reliable processes, you adjust them upwards.

Coherence and belief-value agreement usually have common causes

[11] When it comes to standard deference to experts – arguably the most common genetic argument – you do not look at what being said, beyond classifying it as something that the expert is (or is not) competent in.

Another thing coherence and suspicious belief-value agreement have in common is that they both often are signs of a *common cause* (see Reichenbach 1956). In the case Lewis discusses, the common cause of the coherent set of testimonies is the fact that the witnesses are reliable together with the fact that what they say is true. In the case of Paul, the common cause of his suspicious belief-value agreement was a politically biased belief-forming process.

This is not to say that all coherent sets, or all cases of suspicious belief-value agreement have common causes. To see that, let us consider three events which are probabilistically dependent on each other but have not influenced each other directly: “Linda has lung cancer” (A_1), “Linda has yellow fingers” (A_2) and “Linda’s hair is green” (A_3). Now suppose that the reason why they are dependent on each other is that A_1 and A_2 have smoking as a common cause, A_1 and A_3 have some sort of radioactive process (that makes your hair green but not your fingers yellow) as a common cause whilst A_2 and A_3 have Linda painting with her daughter as a common cause. In this case, $\{A_1, A_2, A_3\}$ is coherent even though the three events do not have a common cause. (Similar examples showing that belief-value agreement does not always have common causes can easily be constructed.)

Thus, not all coherent sets, or all cases of suspicious belief-value agreement, have a single common cause. However, my hunch is that a large proportion of them do. Also, in many cases where not *every* member of the set in question has a common cause, a large part of them will.

In order to give some support to this hunch, suppose that a large set of beliefs $\{A_1,\ldots,A_n\}$ is coherent, and that this is not due to a single cause, but rather to several different factors $C_1,\ldots,C_m$ which make the subsets of $\{A_1,\ldots,A_n\}$ probabilistically dependent on each other. Thus the high degree of coherence is not due to one single factor or cause, but to many different factors. In that case, however, we would like an explanation of why it is that there are several different factors all of which make the members of $\{A_1,\ldots,A_n\}$ more probabilistically dependent on each other. These factors could, in turn, have a common cause.

This amounts, in effect, to a sort of second-order coherence reasoning. The idea underlying first-order coherence reasoning was that it was unlikely that we get lots of information pointing in the same direction by mere chance. Hence we need to postulate a common cause. Now assume that the subsets of $\{A_1,\ldots,A_n\}$ are probabilistically dependent on each other, but that some of these dependencies are down to factor C_1, some down to C_2, some down to C_3, etc. In many cases (though not, perhaps, in the Linda example), we would think that to be unreasonable. It would not be the simplest and most parsimonious explanation. Instead, the simplest, most Occamite explanation would say that $C_1,\ldots,C_m$ in turn have a common cause. (Again, the same argument holds for belief-value agreement, *mutatis mutandis*).

Applications: internet tests

The argument from suspicious belief-value agreement could be used to devise different bias tests. Together with the American critical thinking organization ClearerThinking.org, I have published such a test: the Political Bias Test, aimed for the American market (Schubert and Greenberg 2015). The test, which was published on Vox.com on September 10, 2015 (Whittlestone 2015), works like this. First, you are asked for your political views on a number of different issues: are you a social liberal or conservative, an economic liberal or conservative, in favour or environmental regulations or not, etc. Second, you are asked 18 factual questions, on which we know the true answer, on politically controversial topics such as climate change, inequality, GMOs, and so on. The idea is that, e.g., someone who is opposed to environmental regulations is not going to like the fact that climate change is caused by humans – since that might be a reason to introduce environmental regulations – and that they therefore will turn climate skeptics if they are politically biased.

For any individual test-taker, there are eight answers that support their political views, eight that disconfirm them, and two that are neutral. Now, test-takers are given a bias score on the basis of the *asymmetry* between their true and false answers. If they are systematically right where the true answer confirms their political views, and systematically wrong when it disconfirms them, they are defined as politically biased.

A key difference from the model used in this paper is that we use questions on which we know the true answer in our test. The reason for this is that it is very hard to estimate the objective probability of any empirical question where we are not certain about the true answer (i.e. where the probability is neither 0 or 1), and that we need objective probabilities to calculate the degree of support that a high degree of asymmetry gives to the bias hypothesis. Also, there was no attempt to estimate to what degree the question answers cohere with each other (though my hunch is that they are on average fairly independent).

Obviously, the test is not intended to give a scientific estimate of bias. We did pre-test it at Mechanical Turk and found that test-takers came out as somewhat biased on average – which is in line with Kahan's research (see, e.g. Kahan et al. 2011) – but we don't claim that the bias score is reliable. The idea was to raise awareness about the problem of political bias, something I think we managed to do to some extent, since the test got a fair amount of attention.

There is definitely room for more tests like this. For instance, you could measure the degree to which political bias explains prediction errors in competitions like Philip Tetlock's *Good Judgment Project*, where participants try to forecast geopolitical events (Tetlock et al., 2011). Tetlock himself suggests in his and Dan Gardner's book *Superforecasting* (Tetlock and Gardner, 2015) that political bias is a common source of forecasting error. It would be great if that hypothesis could be more extensively explored.

Reverse engineering arguments as a research program in formal epistemology

Of course, we could apply the reverse engineering reasoning applied in this paper to all sorts of other structures of information besides those discussed here. For instance, psychologists have discovered a wealth of cognitive biases besides those that give rise to suspicious agreement between beliefs and values. Presumably all of these heuristics give rise to distinctive patterns or structures of beliefs, which should be detectable by reverse engineering. Once we know what the process is, we can evaluate how likely it is that the beliefs are true, using our knowledge of how reliable the process tends to be.

In principle, a whole research program in formal epistemology could be built on this idea. Epistemologists could go through the by now rather long list of psychological heuristics and biases and show, using simple formal models, under what circumstances we have reason to believe that this or that mechanism was operative.[12] The resulting list would have similarities with the list of logical fallacies (such as the slippery slope, guilt by association, etc.).[13]

While we do not necessarily need to use formal tools to do that, I think that they normally do help. Many patterns of beliefs and other items of information are so complicated so as to make a non-mathematical analysis all but impossible. If we do not use mathematical methods, we will be forced to stick to the simplest kinds of suspicious patterns. Many of these are so obviously suspicious that reading analyses of them will not be very interesting to the intelligent reader. There is a quite thin line between arguments that are too complicated to be analysed by non-technical means, and arguments that are so simple so as to not be worth writing a paper about.

References

Alhakami, A.S., Slovic, P. (1994) "A psychological study of the inverse relationship between perceived risk and perceived benefit." *Risk Analysis* 14 (6), 1085-96.

Bastardi, A., Uhlmann, E. L. and Ross, L. (2011). "Wishful Thinking: Belief, Desire, and the Motivated Evaluation of Scientific Evidence." *Psychological Science* 22 (6), 731–732.

Bishop, M. A. and Trout, J. D. (2004). *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press.

Bovens, L. and Hartmann, S. (2003a). *Bayesian Epistemology*. New York and Oxford: Oxford University Press.

Bullock, J. G., Gerber, A. S., Hill, S. J., and Huber, G. A. (2013). "Partisan Bias in Factual Beliefs about Politics". *NBER Working Paper* No. 19080.

Dion, K., Berscheid, E., and Walster, E. (1972), "What is Beautiful is Good", *Journal of personality and social psychology* 24 (3), 285–90.

Douven, I. and Meijs, W. (2007). "Measuring Coherence." *Synthese* 156, 405–425.

Fitelson, B. (2001). *Studies in Bayesian confirmation theory*. Dissertation, University of Wisconsin at Madison. Available online at http://fitelson.org/thesis.pdf.

Inbar, Y., Pizarro, D., Iyer, R., and Haid, J. (2012). "Disgust Sensitivity, Political Conservatism, and Voting." Social Psychological and Personality Science, 3 (5), 537-544.

Kahan, D., Jenkins-Smith, H. and Braman, D. (2011). "Cultural cognition of scientific consensus". *Journal of Risk Research*, 14 (2), 147-174.

[12] This research program would thus follow Bishop and Trout's (2004) call for naturalistic philosophers to make use of cognitive psychology to improve reasoning.

[13] One important difference between the two lists is, though, that this list would consist of *valid* arguments whereas the logical fallacies are *invalid*. It strikes me as more natural to focus on valid arguments (e.g. logicians do not primarily study invalid inferences, but valid ones) even though it could of course be worth pointing out common mistakes.

Kahan, Dan M. and Peters, E., Dawson, E. C. and Slovic, P. (2013). “Motivated Numeracy and Enlightened Self-Government”. *Yale Law School, Public Law Working Paper* No. 307.

Kunda, Z. (1990). “The Case for Motivated Reasoning.” *Psychological Bulletin* 108 (3), 480–498.

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. LaSalle, Illinois: Open Court.

Lodge, M. and Taber, C. S. (2005). “The Primacy of Affect for Political Candidates, Groups, and Issues. An Experimental Test of the Hot Cognition Hypothesis.” *Political Psychology* 26, 455-482.

Lord, C., Ross, C. and Lepper, M. (1979). “Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence.” *Journal of Personality and Social Psychology* 37 (11), 2098-2109.

Nickerson, R. S. (1998). “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.” *Review of General Psychology* 2(2), 175–220.

Nisbett, R. E. and Wilson, T. D. (1977) “The Halo effect: Evidence for Unconscious Alteration of Judgments.” *Journal of Personality and Social Psychology* (American Psychological Association) 35 (4), 250–56.

Redlawsk, D. P., Civettini, A. J. W. and Emmerson, K. M. (2010) “The Affective Tipping Point: Do Motivated Reasoners Ever ‘Get It’”. *Political Psychology* 31 (4), 563-593.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of Los Angeles Press.

Shogenji, T. (1999). “Is Coherence Truth Conducive?” *Analysis* 59, 338–345.

Schubert, S. (2011). “Coherence and Reliability: The Case of Overlapping Testimonies.” *Erkenntnis*, 74, 263-275.

Schubert, S. (2012a). “Coherence Reasoning and Reliability: A Defense of the Shogenji Measure”. *Synthese*, 187, 305-319.

Schubert, S. (2012b). “Is Coherence Conducive to Reliability?”. *Synthese*, 187, 607-621.

Schubert, S. and Greenberg, S. (2015). “The Political Bias Test.” Test published on http://www.clearerthinking.org/.

Tetlock, P., Mellers, B. and Moore, D. (2011). *Good Judgment Project*. Research project. Website: https://www.gjopen.com/

Tetlock, P. and Gardner, D. (2015). *Superforecasting*. Cornerstone Publishing.

van der Linden, D. te Nijenhuis, J. and Bakker, A. B. (2010). ”The General Factor of Personality: A Meta-analysis of Big Five Intercorrelations and a Criterion-Related Validity Study.” *Journal of Research in Personality* 44, 315-327.

Whittlestone, J. (2015). “How politically biased are you? Take this quiz to find out.” Vox.com, 10 September 2015. Available at: http://www.vox.com/2015/9/10/9188517/political-bias.